UNIVERSITÉ DE MONTRÉAL

MODÉLISATION DE LA PHASE DE CROISSANCE DE *Eschscholtzia californica*

FRANÇOIS RIOUX
DÉPARTEMENT DE GÉNIE CHIMIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU GRADE DE MAÎTRE ÈS SCIENCES APPLIQUÉES
(GÉNIE CHIMIQUE)
DÉCEMBRE 1996

© François Rioux, 1996.

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-26509-9

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

MODÉLISATION DE LA PHASE
DE CROISSANCE DE *Eschscholtzia californica*

présenté par : RIOUX, François
en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de:

M. RAMSAY Bruce, Ph. D., président
M. PERRIER Michel, Ph.D., membre et directeur de recherche
M. ARCHAMBAULT Jean, Ph.D., membre et codirecteur de recherche
M. GARNIER Alain, Ph. D., membre

Remerciements

Je tiens à remercier sincèrement les Professeurs Jean Archambault et Michel Perrier qui m'ont encadré et encouragé tout au long de ces deux années. Je les remercie de m'avoir communiqué leurs connaissances et leurs expériences. Merci au Professeur Claude Chavarie pour son appui durant les premières phases du projet.

Merci à Christophe Bazinet, Mario Jolicoeur, Alain Larose, et Marie-France Pépin pour nos discussions de corridors stimulantes et formatrices. Merci également aux autres membres du groupe de culture de cellules végétales : Dany Aubry, Martin Gaudette, Laurence Lamboursain, Luc Lavoie et Joël Sirois.

Un merci spécial à Robert Williams pour ses conseils et pour la préparation des inoculums et l'opération du HPLC.

Merci à mon épouse Alicia pour son encouragement constant et son appui sans faille qui m'a permis de mener ce navire à bon port.

Résumé

Ce travail s'insère parmi les travaux du groupe de culture de cellules de plantes de Biopro. L'objectif de ce mémoire est de modéliser la phase de croissance de cellules d'*Eschscholtzia californica* en vue d'une maximisation de la production de sanguinarine. La méthodologie en deux étapes, adaptée d'une approche développée pour des cellules animales et appliquée pour la première fois à des cellules végétales, consiste dans un premier temps à construire et à valider des schèmes réactionnels macroscopiques. Chaque schème est converti en un ou plusieurs modèles stoechiométriques qui sont résolus à l'aide de régressions multilinéaires. Les coefficients de ces schèmes réactionnels sont analogues aux coefficients de rendements de modèles simples.

La seconde étape est la modélisation des cinétiques de croissance. La méthodologie repose sur la construction de modèles polynômiaux. Cette étape permet de décrire la cinétique de croissance de la biomasse en fonction de nutriments choisis. Les paramètres utilisés pour la modélisation sont les sucres, l'ammoniac, le nitrate et le phosphate. De plus, deux indicateurs susceptibles d'influencer la cinétique de croissance ont été construits : un indicateur du phosphate intracellulaire, et un indicateur du taux d'accumulation d'ammoniac.

Pour les deux étapes, l'obtention des paramètres recherchés passe par une transformation de bilans de masse en systèmes matriciels sous la forme d'équations multilinéaires. La solution de ces équations est obtenue par la méthode des moindres carrés.

Les cultures sont réalisées en flacons agités et en bioréacteur de 6 l. Le bioréacteur consiste en un cylindre muni d'un agitateur à double ruban hélicoïdal. La concentration en oxygène dissous est contrôlée à 60 % de la saturation.

Seuls les modèles stoechiométriques à 2 ou 3 paramètres présentent une performance satisfaisante lors d'une validation croisée sur des cultures n'ayant pas servi à l'évaluation des paramètres. Par ailleurs, l'analyse des intervalles de confiances ne permet pas d'estimer avec satisfaction la valeur des coefficients pseudo-stoechiométriques ou des coefficients de rendements recherchés. Ces résultats découlent de deux facteurs : une variabilité entre les cultures reproduites dans des conditions similaires, et l'assimilation simultanée des nutriments mesurés qui entraîne une colinéarité des variables.

Des modèles cinétiques polynômiaux ont été évalués jusqu'à une complexité égale à un polynôme d'ordre 4, à 5 termes. Les résultats obtenus d'un modèle à l'autre sont cohérents pour le domaine expérimental où se retrouve la majorité des données. Pour l'extrapolation à l'extérieur de cette zone, les comportements varient d'un modèle à l'autre. Pour ce qui est des nutriments les plus importants pour la croissance de la biomasse, le phosphate intracellulaire est celui qui domine. Il est présent dans la majorité des modèles. Le sucre, substrat qu'on retrouve à la base de tous les modèles de la littérature, semble avoir une influence marginale sur les taux de croissance. Sa présence est nécessaire mais sa concentration ne semble pas influencer la croissance de façon significative.

Abstract

This work is part of the research done by the Plant Cell Culture Group of Biopro. The objective of this master's thesis is the modelling of the growth phase *of Eschscholtzia californica* cell cultures in order to maximise sanguinarine production. A two-phase methodology adapted from an approach used for animal cells was applied for the first time to plant cell culture. The first phase consist of building and validating reaction networks. These networks are then translated into stoichiometric models that are solved through multilinear regressions. The coefficients of reaction networks are similar to yield coefficients of simpler models.

The second phase consists of modelling the growth kinetics. This methodology relies on the construction of polynomial models. Through this step, the biomass growth kinetics are described as a function of the main nutrients. The parameters used in the modelling are sugars, ammonia, nitrate and phosphate concentrations. Furthermore, two indicators suspected of having an influence on growth are developed: an intracellular phosphate indicator and an ammonia absorption rate indicator.

For both phases, parameters are obtained through the transformation of mass balances into a matrix system of multilinear equations. The system is then solved with a least square method.

The cultures were grown in shake flasks and in a 6 liter bioreactor. The bioreactor was equipped with a helical ribbon impeller that minimises shear stress to the cells. Dis-

solved oxygen concentration is controlled was 60% of saturation.

The stoechiometric models containing 2 or 3 parameters present a satisfying performance after a cross validation with cultures not used for the parameter estimation. Large confidence intervals prevent estimation of stoechiometric-like coefficients or yield coefficients of the models. This originates from 2 possible causes: a large variability between cultures grown under similar conditions and a simultaneous absorption of various nutrients that results in the colinearity of the variables.

Polynomial kinetic models were tested up to a complexity equal to a four term fourth order polynomial. Results are coherent between models in the range where most data is concentrated. For extrapolation outside that zone, features differ from one model to the other. Intracellular phosphate is the dominant nutrient in growth kinetic modelling. It is found in a majority of models. However, sugar, the basis of all models found in the literature, seems to have little influence on growth rates. Sugar is required but its concentration does not seem to influence growth.

Table des matières

Remerciements iv

Résumé v

Abstract vii

Table des matières ix

Liste des figures xiv

Liste des tableaux xvi

Liste des annexes xviii

Nomenclature xix

1. Introduction 1

1.1 Culture de cellules de plantes 1

1.2 Cadre du projet 3

1.3 Objectifs 3

1.4 Approche de modélisation 4

1.5 Structure du mémoire 5

2. Revue de la littérature 7

2.1 Systèmes de culture de cellules végétales 7

2.2 Modes d'alimentation employés en culture de cellules végétales 9

2.3 Modélisation : Stoechiométrie et coefficients de rendement . 10

2.4 Modélisation de la cinétique de croissance de la biomasse .. 17

2.4.1 Structure et ségrégation d'un modèle 17

2.4.2 Régime permanent (croissance équilibrée) et état transitoire 18

2.4.3 Production de métabolites secondaires 19

2.4.4 Entretien cellulaire et mortalité 19

2.4.5 Modèles cinétiques recensés dans la littérature 20

2.4.6 Discussion sur les modèles cinétiques 26

3. Méthodologie 29

3.1 Schèmes réactionnels 29

3.2 Cinétique de croissance 31

3.3 Rendements, cinétique et contrôle 34

4. Matériel et méthodes 35

4.1 Lignée cellulaire 35

4.2 Description du bioréacteur 35

4.3 Préparation d'une culture en bioréacteur 36

4.4 Conditions d'opération 37

4.5 Stratégies d'alimentation 37

4.6 Acquisition de données et contrôle 38

4.7 Flacons témoins 39

4.8 Échantillonnage 39

4.9 Analyse des sucres et des ions 40

4.9.1 Hydrates de carbone 40

4.9.2 Chromatographie ionique 41

5. Schèmes réactionnels et bilans de masse 42

5.1 Schème le plus simple : croissance seule 43

5.2 Second schème réactionnel : croissance et maintenance 45

5.3 Troisième schème réactionnel 46

5.4 Quatrième schème réactionnel 50

5.5 Estimation par les moindres carrés 52

5.6 Notions sur la validation des modèles 52

5.6.1 Erreur résiduelle et coefficient de détermination 53

5.6.2 Coefficient de détermination multiple empirique ajusté 53

5.6.3 Validation croisée 54

5.6.4 Variance, intervalles de confiance et régresseurs significatifs 55

6. Application des schèmes réactionnels aux données expérimentales 57

6.1 Résultats expérimentaux 57

6.2 Présentation des résultats par catégories 59

6.2.1 Description des catégories 59

6.2.2 Résultats de régression pour un modèle et une catégorie de culture. 60

6.2.3 Discussions sur les erreurs 60

6.2.4 Validation des coefficients pseudo-stoechiométriques 62

7. Modèles polynômiaux de croissance de la biomasse 70

7.1 Modèles conventionnels 70

7.1.1 Comparaison entre les modèles cinétiques conventionnels et polynômiaux . 70

7.1.2 Modifications apportées à la méthodologie de Chotteau 74

7.2 Préparation des données pour la régression 75

7.3 Construction des familles de modèles 78

7.4 Sélection d'un modèle 81

7.4.1 Critères de validation d'un modèle 82

7.4.2 Intervalle de confiance et régresseurs significatifs 82

7.5 Construction d'indicateurs 82

7.5.1 Indicateur du phosphate intracellulaire 83

7.5.2 Indicateur de l'assimilation d'ammonium 85

8. Application des modèles cinétiques aux données expérimentales 87

8.1 Analyse préliminaire du taux de croissance 87

8.2 Évaluation des modèles cinétiques polynômiaux 87

8.3 Évaluation des modèles polynômiaux sans constante 98

9. Conclusions et recommandations 103

9.1 Conclusion sur les schèmes réactionnels 103

9.2 Conclusion sur la modélisation cinétique 104

9.3 Conclusions générales et recommandations 107

Bibliographie 110

Annexes 116

Liste des figures

Figure 3-1 Comparaison entre un modèle de type Monod et un modèle cinétique polynômial 33

Figure 6-1 Modèle 2d, Estimation de la biomasse à partir de paramètres mesurés..... 68

Figure 6-2 Coefficient de détermination multiple empirique ajusté (CDMEA) pour chacune des quatre catégories, en fonction des modèles 69

Figure 8-1 Erreur résiduelle de modélisation en fonction du nombre de termes du polynôme et de l'ordre des monômes pour modèles avec constante 93

Figure 8-2 Modèle cinétique de croissance avec OM=2,NR=3 avec constante 93

Figure 8-3 Dispersion des données de phosphate et de sucre pour l'ensemble des cultures 94

Figure 8-4 Résultat en ΔX (g) du modèle OM=2;NR=3 avec constante pour l'ensemble des cultures 95

Figure 8-5 Résultat en biomasse X (g/l) du modèle OM=2;NR=3 avec constante pour l'ensemble des cultures 96

Figure 8-6 Modèle OM=3, NR=4 avec constante, Ne=0 97

Figure 8-7 Erreur résiduelle de modélisation en fonction du nombre de termes du polynôme et de l'ordre des monômes pour les modèles sans constante..... 101

Figure 8-8 Modèle OM=3, NR=4 sans constante..... 102

Liste des tableaux

Tableau 2-1 Synthèse des différents modèles de culture de cellules végétales cités dans la littérature 15

Tableau 2-2 Schèmes découlant des différents modèles cinétiques de culture de cellules végétales recensés dans la littérature 16

Tableau 5-1 Définition des régresseurs linéaires pour le troisième schème réactionnel 49

Tableau 5-2 Définition des régresseurs linéaires pour les schèmes réactionnels 51

Tableau 6-1 Portrait de l'ensemble des cultures analysées 58

Tableau 6-2 Répartition des cultures parmi les catégories 65

Tableau 6-3 Paramètres de régression linéaire pour les différents modèles 65

Tableau 6-4 Valeur des régresseurs des modèles retenus pour chacune des 4 catégories 66

Tableau 6-5 Coefficients de corrélation "r" entre les paramètres de la culture RH7 66

Tableau 6-6 Paramètres de régression linéaire pour les différents modèles retenus 67

Tableau 7-1 Lexique des notations et définitions utilisées aux chapitres 7 et 8 77

Tableau 7-2 Construction des polynômes 80

Tableau 7-3 Concentration en phosphate de diverses cultures de cellules de plantes 86

Tableau 8-1 Paramètres utilisés pour modéliser la cinétique..... 91

Tableau 8-2 Erreur de modélisation et modèles avec constante 92

Tableau 8-3 Erreur de modélisation et modèles sans constante .. 100

Liste des annexes

Annexe A. Résultats expériementaux 116

A.1 Graphiques des données expérimentales 116

A.2 Description des cultures analysées 124

Annexe B. Transformations mathématiques 127

B.1 Transformations appliquées au premier schème 127

B.2 Transformations appliquées au second schème 130

B.3 Estimation par les moindres carrés 131

B.4 Préparation des données pour la régression 135

B.4.1 Construction de la forme intégrée matricielle 135

B.4.2 Discrétisation de la forme intégrée 139

B.4.3 Exemple de modélisation 141

Nomenclature

Générale

Symbole	Définition	unités
a_i	Coefficients des termes des polynômes	
k_X	Coefficient de régression	
k_S	Coefficient de régression	
k_N	Coefficient de régression	
k_M	Coefficient de régression	
k_{XV}	Coefficient de régression	
$\hat{k}_i$	Estimation du coefficient de régression i	
m_S	Taux d'entretien cellulaire	h^{-1}
p_{ij}	Coefficient des termes d'un polynôme	
$q_{produits}$	Taux de production spécifique de métabolites	h^{-1}
q_S	Taux de consommation spécifique de sucre	h^{-1}
q_O	Taux de consommation spécifique de l'oxygène	h^{-1}
q_X	Taux de production de biomasse spécifique	h^{-1}
r_i	Taux de réaction de production ou de consommation	$g\ h^{-1}$,$mmol\ h^{-1}$
r_i'	Taux de réaction pseudo stoechiométrique	$g\ l^{-1}h^{-1}$
p, n, m	Coefficients pseudo-stoechiométriques	$mmol\ g^{-1}$
s, s_1, s_2	Coefficients pseudo-stoechiométriques	$g\ g^{-1}$
C	Concentration d'un nutriment corrigée	$g\ l^{-1}$, $mmol\ l^{-1}$
C'	Concentration d'un nutriment mesurée	$g\ l^{-1}$, $mmol\ l^{-1}$

Symbole	Définition	unités
D	Taux de dilution	h^{-1}
E	Erreur quadratique résiduelle de modélisation	
F^{in}	Débit d'alimentation entrant au bioréacteur	$l\ h^{-1}$
F^{out}	Débit de milieu sortant du bioréacteur	$l\ h^{-1}$
K_S	Paramètre de l'équation de Monod	g
M^{in}	Concentration en nitrate de l'alimentation	$g\ l^{-1}$
N	Nombre de cultures utilisées pour la régression	
N^{in}	Concentration en ammoniac de l'alimentation	$g\ l^{-1}$
N^q	Nombre d'échantillons dans la q-ième culture	
P_{min}	Concentration intracellulaire en phosphate en début de culture	$mmol\ g^{-1}$
PCV	Volume des cellules compressées (Packed cell volume)	
R_i	Production ou consommation	g, mmol
S^{in}	Concentration en sucre dans l'alimentation	$g\ l^{-1}$
V	Volume de milieu	l
$V^{échant}$	Volume d'un échantillon	l
X	Concentration de la biomasse	$g\ l^{-1}$
$\grave{X}$	Estimé de la concentration de la biomasse	$g\ l^{-1}$
XV	Produit concentration de la biomasse*volume	g
$Y_{O/X}$	Rendement en biomasse par rapport à l'oxygène	h^{-1}
$Y_{X/S}$	Rendement en biomasse par rapport au sucre	$g\ g^{-1}$
γ_X	Degré de réduction généralisée de la biomasse	

Symbole	Définition	unités
γ_S	Degré de réduction généralisée du sucre	
ξ_i	Param. mesuré ou calculé influençant le taux de croissance	
η	Vecteur solution de la régression linéaire de la cinétique	
μ	Taux de croissance spécifique de la biomasse	h^{-1}
μ_{max}	Taux de croissance spécifique maximal de la biomasse	h^{-1}
ϑ_i	Coefficients des termes d'un polynômes	
ϕ_i	Monôme construit à partir des paramètres ξ	
ω	Nombre de régresseurs construits à partir de ξ	
ν	Nombre de paramètres pour modéliser le taux de croissance	
ρ	Nombre de termes d'un polynôme	
λ	Nombre de familles de modèles à évaluer	
σ	Nombre de termes d'un polynôme.	
ΔN	Ammoniac assimilé	mmol
ΔM	Nitrate assimilé	mmol
ΔS	Sucre consommé	g
ΔX	Biomasse produite	g
Θ	Matrice des coefficients de régression linéaire ϑ	
Θ_n	Matrice des coefficients correspondant au n ième polynôme	
Φ	Matrice des monômes	
Ψ	Matrice des régresseurs linéaires pour la cinétique	

Schèmes

Symbole	Définition	unités
p, n, m	Coefficients pseudo-stoechiométriques	mmol g^{-1}
s, s_1, s_2	Coefficients pseudo-stoechiométriques	g g^{-1}
M_e	Nitrate extracellulaire	mmol g^{-1}
M_i	Phosphate intracellulaire	mmol g^{-1}
N_e	Ammoniac extracellulaire	mmol g^{-1}
N_i	Phosphate intracellulaire	mmol g^{-1}
P_e	Phosphate extracellulaire	mmol l^{-1}
P_i	Phosphate intracellulaire	mmol g^{-1}
S, S_e	Sucre extracellulaire	g l^{-1}
S_i	Phosphate intracellulaire	mmol g^{-1}
X	Biomasse	g l^{-1}

Statistique

Symbole	Définition
b	Vecteur des éléments inconnus (k ou Θ)
$\hat{b}$	Vecteur des éléments inconnus (k ou Θ) estimés
b_j	Éléments inconnus (paramètres estimés, k_i ou ϑ)
c_{ij}	Élément de la diagonale de $[X'X]^{-1}$

Symbole	Définition
$s(b_j)$	Estimateur de l'écart-type de b_j
$s^2(b_j)$	Estimateur de la variance de b_j
$t_{\alpha 2}$	distribution de Student au niveau de signification α
e_i	Erreur de modélisation d'un échantillon
n	Nombre d'échantillons
p	Nombre de régresseurs
r	Nombre de cultures
y	Vecteur solution constitué des valeurs de biomasse (ΔX)
y_i	Paramètre mesuré ou calculé
$\bar{y}_i$	Valeur moyenne des paramètres mesurés ou calculés
$\hat{y}_i$	Paramètre estimé
R^2	Coefficient de détermination multiple empirique
$\bar{R}^2$	Coefficient détermination multiple empirique ajusté (CDMEA)
S_e^2	Estimateur de la variance de l'erreur
SCE	Somme des carrés des erreurs
SCT	Somme des carrés
X	Matrice des variables explicatives ($\Delta\xi$ ou ψ)

La nomenclature qui suit est celle utilisée spécifiquement pour le Tableau 2-2 et peut différer de celle utilisée dans le reste du document

Biomasse (Tableau 2-1)

Symbole	Définition
B	Biomasse structurelle
X	Biomasse totale
X_f	Biomasse humide
X_{vd}	Biomasse viable
X_{nd}	Biomasse non viable
X_l	Masse d'une vacuole
X_A	Cellules actives en division
X_R	Cellules au repos

Nutriments (Tableau 2-1)

Symbole	Définition
G_e	Glucose extracellulaire
H	Hormones
N	Nutriments non glucosés
P_e	Phosphate extracellulaire
P_f	Phosphate intracellulaire libre
P	Phosphate intracellulaire

Symbole	Définition
O_2	Oxygène
S, S_e	Sucre ou glucose extracellulaire

Métabolites, produits (Tableau 2-1)

Symbole	Définition
C	Amidon intracellulaire
CO_2	Dioxyde de carbonne
M	Métabolites secondaires
M_e	Métabolites secondaires extracellulaires
P_m	Produits de lyse de métabolites secondaires
P_l	Métabolites secondaires libérés dans le milieu
P_f	Produits de lyse de métabolites
P_X	Produits de lyse de biomasse
T	Métabolites intermédiaires
T*	Métabolites intermédiares structurels
T_e	Métabolites intermédiaires extracellulaires

Constantes et coefficients (Tableau 2-1)

Symbole	Définition
α	Coefficient de production de métabolites associés à la croissance (Luedeking-Piret)
β	Coefficient de production de métabolites non liés à la croissance (Luedeking-Piret)
k_a	Coefficient d'hydrolyse du sucrose
k_d	Coefficient de mortalité cellulaire
k_L	Coefficient de fuite de métabolites
m_S	Coefficient de maintenance cellulaire
K_d	Coefficient de dégradation des métabolites secondaires
V	Viabilité cellulaire
Z	Coefficient d'expansion cellulaire
$\kappa_{mortalité}$	Coefficient de mortalité cellulaire
χ	ratio poids humide/poids sec

1. Introduction

1.1 Culture de cellules de plantes

De tout temps, l'Homme a su utiliser les plantes pour améliorer sa qualité de vie. Que ce soit pour se nourrir, se vêtir, s'abriter, se chauffer ou se soigner, les plantes constituent encore de nos jours un apport inestimable au développement de l'être humain. Entre autres, les plantes sont une source importante de produits médicinaux pour tous les peuples. De nouvelles molécules actives sont régulièrement découvertes ou redécouvertes. Plusieurs substances ont un potentiel médical important. Ces substances phytochimiques sont souvent présentes en concentrations infimes chez des espèces exotiques croissant dans des environnements fragiles. La récolte à grande échelle de ces plantes pour en extraire quelques milligrammes de substances utiles est rarement réalisable au point de vue écologique ou économique.

Une voie de production est la synthèse chimique, complète ou partielle, des molécules actives recherchées. Malheureusement, cette solution est trop souvent irréaliste à cause des coûts reliés au grand nombre de réactions impliquées dans la synthèse de molécules complexes. L'utilisation de micro-organismes recombinants n'est pas actuellement envisageable à cause de la complexité des molécules phytochimiques recherchées.

Une alternative intéressante consiste à produire les substances recherchées au moyen de cellules de plantes cultivées en bioréacteurs. Ce procédé permet d'obtenir la substance recherchée dans un volume restreint et dans un environnement contrôlé.

Cette technique étudiée depuis des années présente d'importants défis technologiques qui doivent être relevés avant d'envisager une production industrielle. Il y a également place à l'amélioration des taux naturels de production par sélection des lignées et par contrôle des conditions de culture.

Les techniques de culture mises en oeuvre s'apparentent à la culture de micro-organismes. Cependant les taux de croissance faibles et, par conséquent, les longues durées de cultures (par rapport aux cultures bactériennes) exigent des précautions beaucoup plus importantes au niveau de la stérilité. La taille et la fragilité de ces cellules exigent également des conditions de mélange particulières.

Au niveau biologique, des problèmes importants demeurent au niveau de la sélection et du maintien de lignées cellulaires productrices stables. De nombreuses inconnues persistent concernant les mécanismes de croissance et de production. Il y a encore place pour l'amélioration de la croissance et de la production par le suivi et le contrôle des conditions de culture. La compréhension des divers mécanismes de croissance et de production en est encore à ses débuts. En ce sens, quelques efforts de modélisation ont été réalisés mais il n'existe pas encore de modèles biologiques complets et satisfaisants permettant de modéliser la croissance de cellules végétales et la production de métabolites secondaires.

1.2 Cadre du projet

Dans ce contexte de développement d'un procédé de production de substances phytochimiques par la culture de cellules végétales en bioréacteur, les travaux initiés en 1989

au laboratoire de culture de cellules de plantes du groupe Biopro de l'École Polytechnique de Montréal sont extrêmement pertinents. Entre autres sujets de recherche, on a étudié des cellules d'*Eschscholtzia californica* pour la production de sanguinarine, un alcaloïde benzophénanthridine, comme système modèle.

Aubry (1995) a établi un système de culture de cellules d'*Eschscholtzia californica* et de production de sanguinarine en bioréacteur. Celui-ci est constitué d'un bioréacteur à double ruban hélicoïdal de 6 l opérés à l'aide d'un système d'acquisition et de contrôle. Un autre bioréacteur du même type d'un volume de 2 l est également utilisé. La production de sanguinarine est réalisée en deux phases: une première phase de croissance de la biomasse, sans présence détectable de sanguinarine et une seconde phase de production, induite par élicitation. Les travaux présentés dans cette étude constituent une suite aux travaux de Aubry et s'appuient sur les systèmes de culture mis au point par ce dernier.

1.3 Objectifs

La revue de littérature montre qu'il y a encore place à l'originalité et à l'amélioration dans le domaine de la modélisation de la croissance de cellules végétales. Les modèles développés dans ce travail présentent une approche nouvelle qui permet de valider plusieurs modèles simultanément afin de sélectionner les meilleurs d'entre eux.

L'objectif principal de ce travail est de modéliser la phase de croissance de cellules végétales, en l'occurrence des cellules d'*Eschscholtzia californica*, lignée EC6 déve-loppée et maintenue par le Dr Robert D. Williams de l'École Polytechnique de Montréal.

Cet objectif sera réalisé en deux phases qui répondent aux objectifs suivants

◊ Établir et valider divers schèmes réactionnels décrivant l'évolution de la concentration en biomasse en fonction de divers nutriments dans le milieu.

◊ Proposer et valider divers modèles cinétiques de croissance reliant la croissance de la biomasse aux nutriments du milieu de croissance.

La définition des schèmes réactionnels est donnée à la section 3.1 du chapitre traitant de la méthodologie employée

1.4 Approche de modélisation

La modélisation consiste à estimer l'évolution de la concentration de la biomasse dans un bioréacteur en fonction de divers paramètres du milieu de croissance et des cellules. Ainsi les concentrations extracellulaires ou intracellulaires des principaux nutriments (sucre, phosphate, ammoniac, nitrate) sont utilisées dans les modèles. Plusieurs études (Rho et André, 1991; van Gulik et al., 1989) associent un ou plusieurs de ces paramètres à la cinétique de croissance. Ces paramètres ont l'énorme avantage d'être mesurables ou calculables par bilans de masse.

Le cadre de travail dans lequel s'insèrent ces travaux est celui d'une école d'ingénierie et d'un laboratoire axé sur la mise à l'échelle de bioprocédés. En ce sens, les aspects de contrôle sont très importants. Pour le contrôle d'un procédé, la précision et le détail des modèles sont, jusqu'à un certain point, sacrifiés en faveur d'une rationalisation permettant de définir l'état d'une culture par un nombre restreint de paramètres. Dans ce cadre,

les modèles développés dans le présent travail visent la simplicité plutôt qu'une description exhaustive des phénomènes et réactions impliqués.

Les schèmes réactionnels proposés ne tiennent pas compte de la phase de stagnation ou de mortalité qu'on retrouve à la fin de cultures en flacons car cette phase se produit après l'étape de production de biomasse du bioprocédé. De même, la phase de latence n'est pas explicitement modélisée quoique les modèles cinétiques s'y ajustent.

Ces objectifs sont réalisés par l'application à des cultures de cellules végétales de la méthodologie de modélisation développée pour des cellules animales par Chotteau (1995).

1.5 Structure du mémoire

Le mémoire est divisé en sections qui correspondent aux divers volets de la méthodologie. La revue de la littérature présentée au chapitre 2 est centrée sur la stoechiométrie et la cinétique des différents modèles publiés. Le chapitre 3 exprime de façon plus détaillée la méthodologie suivie pour l'obtention des coefficients de rendements et des modèles cinétiques polynômiaux. Le chapitre 4 introduit la méthodologie de culture ainsi que le matériel utilisé de la réalisation de la phase expérimentale de l'étude présentée ici.

Les deux chapitres suivants décrivent et appliquent la première phase de la méthodologie : l'étude de la stoechiométrie. Le chapitre 5 présente les schèmes réactionnels choisis ainsi que les équations de régression linéaire qui permettent d'obtenir les coefficients de rendements ou pseudo-stoechiométriques recherchés. Dans ce chapitre, les critères de sélection des différents schèmes sont décrits tout comme la construction des interval-

les de confiance autour des valeurs obtenues. Le chapitre 6 présente et discute des résultats obtenus suite au traitement des données expérimentales.

Les chapitres 7 et 8 traitent de la modélisation de la cinétique. Le chapitre 7 décrit les modèles polynômiaux utilisés pour obtenir des équations de cinétique ainsi que la méthode de construction des polynômes et les critères de sélection des meilleurs modèles. Le chapitre 8 est consacré à l'analyse des modèles obtenus à partir des données expérimentales. Le chapitre 9 présente la conclusion du travail et certaines recommandations.

2. Revue de la littérature

Cette revue porte d'une part sur la stoechiométrie et l'obtention de coefficients de rendements de systèmes biologiques et, d'autre part, sur les modèles cinétiques de croissance de cellules végétales.

Avant de présenter la revue comme telle, la littérature traitant des configurations de systèmes de culture ainsi que des modes de production de métabolites secondaires est recensée. Ces éléments sont utiles à la compréhension de la problématique de ce type de bioprocédé.

2.1 Systèmes de culture de cellules végétales

Les modes de cultures de cellules en suspension se regroupent en deux catégories principales: les cultures en flacons agités et les cultures en bioréacteurs. Ces deux catégories se distinguent principalement par les volumes de cultures, par la complexité des systèmes et par le suivi et le contrôle possible des paramètres de la culture.

Le système le plus simple, le flacon agité, est le plus utilisé pour la culture de cellules de plantes. Ce système consiste en un flacon de volume variant entre 50 ml et 1 l placé sur un agitateur horizontal. Outre la composition initiale du milieu, la concentration de l'inoculum, la température et l'illumination, peu de paramètres peuvent y être mesurés ou contrôlés en continu. Les interventions externes sur le système consistent à prélever un échantillon, à ajouter un volume correspondant de milieu, et parfois à ajouter certains nutriments au milieu de culture (Huang, 1993; Pielh, 1988; Schiel et al., 1984). On y

retrouve parfois une sonde à oxygène (Pépin, 1995; Pielh, 1988) permettant de suivre l'évolution de la concentration en oxygène dissous.

La seconde catégorie regroupe tous les autres types de bioréacteurs à agitation pneumatique (Payne et al., 1990; Smart et Fowler, 1984; Taticek et al., 1990) ou mécanique (Drapeau et Pareilleux, 1986; Ducos et al., 1986; Taticek et al., 1990; van Gulik et al., 1989, Kamen et al., 1992). Parmi ces derniers, on retrouve le bioréacteur à rubans hélicoïdaux (HRI). Ceux-ci possèdent des propriétés de mélange recherchées en cultures de cellules fragiles. Pour la même qualité de mélange et de transfert d'oxygène, le cisaillement est beaucoup plus faible (Kamen et al., 1992) que pour des bioréacteurs conventionnels à hélice ou à turbine. Pour cette raison un bioréacteur de type HRI a été utilisé pour réaliser les travaux présentés ici.

Par rapport aux flacons agités, les bioréacteurs offrent une plus grande flexibilité de suivi et d'intervention de la part de l'opérateur sur divers paramètres de croissance. On y mesure et contrôle généralement la concentration en oxygène dissous. On peut mesurer en continu le pH et la conductivité du milieu, la consommation d'oxygène et la production de CO_2 par l'analyse des gaz d'entrée et de sortie. On contrôle souvent la vitesse d'agitation, le débit et plus rarement la composition des gaz d'alimentation en surface ou en bullage. Grâce aux grands volumes disponibles, la fréquence d'échantillonnage pour analyse peut être plus élevée qu'en flacon agité.

2.2 Modes d'alimentation employés en culture de cellules végétales

Plusieurs types de manipulation peuvent influencer le déroulement d'une culture de cellules. Les méthodes et stratégies d'intervention dépendent du type de bioréacteur et des objectifs définis par l'opérateur.

La non-intervention est une possibilité. Les cultures en système fermé pour lesquelles la composition du milieu n'est pas modifiée sont dites en cuvée. C'est généralement le cas des flacons agités et des bioréacteurs. Le système n'a pas d'échanges de matière avec l'extérieur, sauf pour les gaz et pour les prélèvements lors d'échantillonnage, le cas échéant.

Une seconde stratégie, la cuvée alimentée (Huang et al., 1993; Pielh et al., 1988; Schiel et al., 1984), consiste à ajouter une solution de certains nutriments durant la croissance. Ces ajouts peuvent être faits pour alimenter un nutriment qui serait toxique s'il était présent en trop grande concentration ou, d'une façon plus générale, pour contrôler la concentration de certains nutriments dans le milieu. Cet ajout peut être fait de façon ponctuelle ou graduelle. La méthode habituellement utilisée consiste en l'ajout en une ou plusieurs étapes d'un volume connu d'une solution de nutriments. L'ajout graduel d'une solution de nutriments, selon un parcours programmé ou contrôlé en rétroaction par l'état de la culture, n'a pas encore été réalisé pour des cellules végétales. Aucun résultat présentant l'ajout en boucle fermée à partir de données acquises en continu n'a encore été publié. Seul Pielh (1988) présente le cas d'un ajout de sucre après échantillonnage et, analyse de la concentration en sucre dans le milieu.

Il est également possible d'opérer un bioréacteur sur une plus grande période. Le chemostat (de Gucht et van der Plas, 1995; van Gulik et al., 1989) en est un exemple. Un taux de dilution constant et une alimentation en milieu frais de composition constante sont les caractéristiques principales de ce mode d'opération qui permet d'obtenir un régime établi et un taux de croissance constant sur de longues périodes.

La culture en semi-continu (Guardiola et al., 1995; Quinlan, 1986; Schnapp et al., 1991) consiste en prélèvements réguliers de volumes connus et leur remplacement par la même quantité de milieu frais. Cela a pour avantage de répéter certaines conditions de croissance et d'obtenir des valeurs moyennes pour les conditions de croissance.

2.3 Modélisation : Stoechiométrie et coefficients de rendement

L'objectif de la modélisation est de formuler sous forme de relations ou d'équations, les interactions entre diverses composantes des cellules et de leur milieu. Dans le cas présent, seules les interactions macroscopiques entre la biomasse et les nutriments sont prises en compte. Bien que fondamentales pour la compréhension des mécanismes de croissance réels des organismes étudiés, les interactions biochimiques plus complexes ne sont pas abordées car elles sont peu utiles en regard des objectifs de modélisation en vue du contrôle d'un bioprocédé.

En biotechnologie, les coefficients de rendement indiquent les rapports entre les nutriments et la biomasse dans une réaction globale qui est la synthèse de cette biomasse.

Les coefficients de rendements décrivent la quantité d'un composé consommé ou produit

lors de la synthèse d'une quantité de biomasse. Ces coefficients peuvent être obtenus à partir de bilans massiques, de bilans énergétiques, de bilans de potentiel d'oxydoréduction. Ces paramètres sont très utiles pour suivre l'évolution d'une culture. En effet, si ces coefficients demeurent constants tout au long d'une culture, il suffit de connaître la variation d'un nombre restreint de nutriments pour estimer l'évolution de la culture.

Lorsque les systèmes de réaction sont complexes, on peut des utiliser coefficients pseudo-stoechiométriques qui sont similaires aux coefficients stoechiométriques des réactions chimiques (Chotteau, 1995, Dochain et al., 1990). Les coefficients pseudo-stoechiométriques permettent d'établir des relations plus complexes entre la biomasse, les nutriments, les produits, et les divers métabolites intermédiaires que nécessitent les modèles choisis. Dans tous les cas, cela demeure une simplification extrême de réactions biochimiques très complexes ayant lieu dans les cellules vivantes.

La revue de littérature a permis d'identifier deux équations décrivant les coefficients de rendement. Dans un premier temps, on peut définir un rendement global par :

$$Y_{X/S} = \frac{(\Delta X)}{(\Delta S)} \tag{2-1}$$

Cette définition de rendement (Bailey, 1986) correspond au modèle stoechiométrique le plus simple n'impliquant qu'un substrat et la biomasse.

$$sS \rightarrow X \tag{2-2}$$

où le coefficient pseudo-stoechiométrique s correspond au coefficient de rendement $Y_{X/S}$.

Ce rendement représente un rendement global qui repose sur l'hypothèse que tout le substrat S consommé est utilisé pour la production de biomasse. Cette définition est satisfaisante si la disparition du substrat du milieu est associé uniquement à la production de biomasse et son assimilation se fait au fur et à mesure des besoins c'est-à-dire qu'il n'y a pas d'accumulation. Pour les nutriments accumulés ou consommés comme source d'énergie, le rendement, selon cette définition, varie en fonction des conditions d'opération et particulièrement selon le taux de croissance.

Ainsi, les hydrates de carbone et l'oxygène ont une utilisation variée par les organismes. Pour en tenir compte, on définit le rendement d'une culture en séparant le terme de croissance $Y_{S/X}$ d'un terme de maintien m_S qui correspond à une consommation pour les besoins énergétiques de survie des cellules. Dans ce cas, la consommation de substrat est représentée par le taux de consommation spécifique "q_S" défini par l'équation de Pirt (1965):

$$q_S = \frac{\mu}{Y_{X/S}} + m_S \qquad (2\text{-}3)$$

Cette équation, développée pour des systèmes bactériens a été appliquée avec succès, avec ou sans les modifications présentées à l'équation 2-5, à des cultures de cellules de plantes. La plupart des études publiées ont été réalisées sur des cultures en chemostat ou en semi-continu avec divers taux de croissance car les cultures en cuvée ne permettent pas d'évaluer séparément Y et m.

Dans ce cas, le paramètre $Y_{X/S}$, est également un coefficient pseudo-stoechiométrique

d'un modèle de croissance à deux équations qui peut être représenté par le schème réactionnel suivant:

$$\begin{aligned} s_1S &\rightarrow X \\ s_2S + X &\rightarrow X \end{aligned} \tag{2-4}$$

où s_1 est le terme de rendement $Y_{X/S}$ et s_2 est associé au terme de maintenance m_S. Le Tableau 2-1 (p. 15) présente une synthèse de l'information relative au calcul des rendements des modèles recensés dans la littérature ainsi que de la méthode de calcul des rendements employée. Schnapp et al. (1991) présentent une étude sur *N. tabacum* en semi-continu. Le système de culture utilisé consistait en un flacon d'un litre contenant 200 ml de milieu. Toutes les 24 heures, un certain volume de milieu était remplacé par du milieu frais riche en glucose. Ils ont ainsi pu établir plusieurs régimes pseudo-permanents à divers taux de croissance (μ) et calculer les taux de consommation de sucre correspondant pour ainsi déterminer $Y_{S/X}$ et m_S pour le glucose à partir de l'équation de Pirt.

Pour estimer les rendements de cultures de cellules de *Catharanthus roseus* et *Nicotiana tabacum*, van Gulik et al. (1992) ainsi que de Gucht et van der Plas (1995) incluent à leur modèle la production de polysaccharides et autres métabolites qui comptent pour 6 à 8% de la biomasse selon van Gulik et al. (1992). Supposant que la composition élémentaire de ces métabolites est la même que celle de la biomasse, ils ont modifié l'équation de Pirt qui devient

$$q_s = \frac{\dot{q}_x}{Y_{x/s}} + m_s, \text{ ou } \dot{q}_x = D + q_{produits} \quad (2\text{-}5)$$

où $q_{produits}$ est le taux de production spécifique de métabolites et D le taux de dilution. Une équation similaire est utilisée pour les rendements en oxygène $Y_{O/X}$. Dans les deux cas, une correction statistique des données est effectuée en appliquant l'algorithme de Wang et Stephanopoulos (1983). Les bilans sont complétés par un bilan sur le potentiel d'oxydoréduction des composants impliqués:

$$\frac{\gamma_s}{4} q_s = \frac{\gamma_x}{4} \dot{q}_x + q_o \quad (2\text{-}6)$$

où γ_X et γ_S sont les degrés de réduction généralisés de la biomasse et du sucre respectivement, et q_O est le taux de consommation spécifique d'oxygène.

Le rendement en biomasse par rapport au sucre fait intervenir un ensemble d'équations pseudo-stoechiométriques qui sont présentées au Tableau 2-2 (p. 16).

Par ailleurs, quelques auteurs traitent des nutriments autres que le sucrose ou le glucose. Les nutriments modélisés sont le phosphate (Curtis et al., 1991; van Gulik et al., 1993) ou l'oxygène (Drapeau et al., 1986; Taticek et al., 1990; van Gulik et al., 1992). Le phosphate est utilisé pour la modélisation de la cinétique. L'oxygène introduit une redondance d'information sur les rendements et la cinétique qui permet de réduire les erreurs statistiques sur les paramètres.

La colonne intitulée Calcul des rendements du Tableau 2-1 présente les différentes définitions de rendement utilisées. Les auteurs semblent utiliser avec une égale satisfaction l'une ou l'autre des approches ou alors ils n'abordent pas l'aspect des rendements dans

leurs travaux (Bailey et Nicholson, 1989; Shibasaki et al., 1993; Hooker et Lee, 1992). Par ailleurs, personne n'a comparé les équations entre elles sur un même ensemble de données.

Tableau 2-1 Synthèse des différents modèles de culture de cellules végétales cités dans la littérature

Auteurs	Bioréacteur	Type de culture	Type de modèle	Calcul des rendements
Drapeau et al. (1986)	Chemap 14 l	cuvée	simple	-
Bailey, Nicholson (1989)	flacons	cuvée	structuré à 5 composantes	complexe
Frazier (1989)	flacons	cuvée	structuré	$\Delta X/\Delta S$
Taticek et al. (1990)	bioréacteur agité 1l, gazosiphon 1l	cuvée	simple	Pirt (O_2,S)
de Gunst et al. (1990)	bioréacteur à turbine 3 l	cuvée	avec ségrégation	Pirt
Schnapp et al. (1990)	flacons 1 l,	cuvée, semi-continu	simple	Pirt
Curtis et al. (1991)	flacons 500 ml	cuvée	avec ségrégation	$\Delta X/\Delta S$
Hooker et Lee (1992)	flacons	cuvée	structuré	-
van Gulik et al. (1992)	bioréacteur à turbine 2 l	chemostat	simple	Pirt corrigé pour produits
van Gulik et al. (1993)	bioréacteurs à turbine 2 l et 12 l	chemostat, cuvée	structuré	Pirt + lyse
Shibasaki et al. (1993)	flacons	cuvée	structuré	complexe
Guardiola et al. (1995)	flacons 250 ml	cuvée, semi continu	simple	Pirt, production de métabolite

Tableau 2-2 Schèmes découlant des différents modèles cinétiques de culture de cellules végétales recensés dans la littérature

Référence	Type de réaction	Schèmes	Paramètres cinétiques[a]
Drapeau (1986) et Taticek (1990)	croissance	$S + O_2 \rightarrow X$	S
	maintenance	$S + O_2 + X \rightarrow X$	m_S, m_{O2}
Bailey(1989)	croissance et production liée	$S \rightarrow X_{vd} + X_f + M$	S ,V, α
	production non liée	$X_{vd} \rightarrow X_{vd} + M$	β
	augmentation biomasse humide	$X_{vd} \rightarrow X_{vd} + X_f$	Z
	mortalité	$X_{vd} \rightarrow X_{nd}$	$\kappa_{mortalité}{}^{*}\chi$
	lyse	$X_{nd} \rightarrow P_X$	k_d
Frazier (1989)	synthèse métabolite intermétiaire	$S \rightarrow T$	S
	production de biomasse	$T \rightarrow X$	T
	relargage de métabolite	$T \rightarrow T_e$	T
de Gunst (1990)	croissance	$S \rightarrow X_A$	H, S
	changement d'état	$X_A \rightarrow X_R$	H
Curtis et al. (1991)	assimilation de phosphate	$P_e \rightarrow P_f$	instantanée
	croissance de la biomasse	$S + P_f \rightarrow X$	P_f
Hooker et Lee (1992)	hydrolyse du sucre	$S_e \rightarrow G_e$	k_a, S
	assimilations des nutriments	$G_e + N_e \rightarrow T$	G+N
	fabrication d'interméd. structuraux	$T \rightarrow T^*$	T, T*
	biomasse structurelle	$T^* \rightarrow B$	T*
	respiration endogène	$B \rightarrow R$	k_d
	maintenance	$T \rightarrow R$	G_e
	production de métabolite sec.	$T^* \rightarrow M$	T,T*
	relargage de métabolites	$M \rightarrow M_e$	k_L
van Gulik et al. (1992)	synthèse de biomasse	$S + O_2 \rightarrow X + CO_2$	S
	production de métabolite	$S \rightarrow M$	S,
	prod. endogène de métabolite	$X \rightarrow M$	S,
	maintenance	$S + X \rightarrow X$	S, m_S
	maintenance endogène	$X + X \rightarrow X$	S, m_S
van Gulik et al. (1993)	assimilation du phosphate	$P_e \rightarrow P$	P_e
	synthèse des précurseurs	$S + P \rightarrow E$	P, S
	synth. de la biomasse structurelle	$S + E \rightarrow B$	P, S
	accumulation de l'amidon	$S \rightarrow C$	P, S
	décroissance et lyse	$B \rightarrow P_X$	
	maintenance	$S + B \rightarrow B$	m_S, C
	maintenance	$C + B \rightarrow B$	m_S, C
Shibasaki et al. (1993)	croissance	$S \rightarrow X_{vb}$	S ,V
	accum. de produit dans vacuole	$X_{vb} \rightarrow X_{vb} + X_l + M$	Z
	mortalité, éclatement du vacuole	$X_{vd} + X_l + M \rightarrow X_{nd} + M_e$	$\kappa_{mortalité}{}^{*}\chi$
	lyse	$X_{nd} + M \rightarrow M_X$	k_d
Guardiola (1995)	croissance	$S \rightarrow X$	S
	production de métabolites	$S \rightarrow M$	S
	maintenance	$S+X \rightarrow X$	m_S, S
	maintenance (dégradation)	$S+X \rightarrow P_X$	S
	dégradation du métabolite	$M \rightarrow P_m$	K_d

[a] Toutes les réactions de croissance, de production et de lyse sont catalysées par la biomasse (sèche, viable ou nombre de cellules) et incluent implicitement ce terme dans l'équation cinétique.

2.4 Modélisation de la cinétique de croissance de la biomasse

Un modèle cinétique est une représentation mathématique d'un système réel. La structure d'un modèle est choisie de façon à faire ressortir certaines informations. La forme que prend un modèle dépend donc des objectifs poursuivis. Les modèles existants en biotechnologie sont basés sur une interprétation simplifiée de phénomènes observés. En effet, chaque cellule est une usine où des centaines de réactions interdépendantes et auto-régulantes ont lieu simultanément. Un modèle tente de saisir les tendances principales et de les résumer en quelques équations basées sur des bilans de masse macroscopiques.

Bailey et Ollis (1986) classifient les modèles biologiques selon divers critères proposés initialement par Tsuchiya et al. (1966) qui sont expliqués dans les sections suivantes. Ces regroupements englobent les modèles cinétiques de tous les micro-organismes et s'appliquent assez bien aux modèles de croissance de cellules de plantes qu'on retrouve dans la littérature.

2.4.1 Structure et ségrégation d'un modèle

Un modèle peut être structuré ou non structuré. Un modèle cinétique structuré représente une cellule comme un ensemble hétérogène d'entités distinctes. Ces entités peuvent être des structures physiques (cytoplasme, vacuole, paroi) ou des métabolites (phosphate, métabolites structuraux). Le modèle non structuré considère chaque cellule comme un tout uniforme.

Un modèle peut présenter ou non une ségrégation. Un modèle avec ségrégation représente une population comme un ensemble hétérogène d'individus. Un modèle sans ségrégation considère seulement une composition moyenne des cellules. Les modèles structurés avec ségrégation s'approchent le plus de la situation réellement observée. Dans la suite du document, le terme modèle simple est utilisé pour décrire les modèles non structurés et sans ségrégation.

2.4.2 Régime permanent (croissance équilibrée) et état transitoire

Différents modes de culture permettent différents types de croissance cellulaire. Dans le cas d'une culture en cuvée simple, un régime pseudo-permanent n'existe que pour de courtes périodes. On y observe typiquement quatre étapes soit une phase de latence, une phase de croissance exponentielle, une phase de stagnation et une phase de déclin. La composition cellulaire, la morphologie des cellules et le type de réactions observées changent significativement tout au long de la cuvée. Cette transition doit être prise en compte lors de la modélisation de tout le profil de croissance.

Pour un système de culture en continu, l'état du biosystème peut être constant : les facteurs physiques et chimiques (concentration en biomasse et en nutriments dans le milieu) s'ajustent en conséquence et demeurent stables. On peut s'attendre à un régime permanent et à une croissance équilibrée, c'est-à-dire à composition cellulaire constante. La cuvée alimentée est un cas particulier car les paramètres d'opération ne sont pas constants mais où l'hypothèse d'une croissance pseudo-équilibrée est parfois justifiée.

2.4.3 Production de métabolites secondaires

La présence de métabolites associés à la croissance influence également la structure d'un modèle. Lorsque la production de métabolite est associée à la croissance, un modèle cinétique de production de métabolite secondaire cherchera à reproduire le profil de production de ce composé.

Luedeking et Piret (1959) ont proposé un modèle générique où la production de métabolite procède de deux modes: le premier est associé à la croissance, le second en est indépendant. Le modèle proposé est de la forme

$$q_P = \alpha\mu + \beta X \tag{2-7}$$

ou q_P représente le taux de production d'un métabolite secondaire, α représente un coefficient indiquant la production liée à la croissance, et β le coefficient décrivant la production de métabolite non liée à la croissance mais dépendant de la concentration en biomasse.

Ce type de modèle est recensé dans la littérature de culture de cellules végétales (Bailey et Nicholson, 1989) mais n'est pas employé dans cette étude.

2.4.4 Entretien cellulaire et mortalité

Dans certains cas, l'utilisation du sucre ou de l'oxygène est modélisée en tenant compte de l'énergie requise pour les activités d'entretien cellulaire (maintenance). Généralement, l'entretien cellulaire est décrit par l'équation 2-3 (Pirt 1965) présentée précédemment. Parfois le taux de mortalité cellulaire est pris en compte, surtout pour la modélisation de cultures en cuvée qui se poursuivent au-delà de la phase de stagnation.

2.4.5 Modèles cinétiques recensés dans la littérature

Les caractéristiques des modèles recensés varient selon les buts poursuivis et les modes de culture des cellules étudiées. Tous ces modèles, ainsi que ceux qui sont développés dans cette étude, sont basés sur la croissance de la biomasse. Le Tableau 2-1 présente les principales caractéristiques des modèles recensés en culture de cellules végétales. Le Tableau 2-2 présente des schèmes réactionnels proposés pour exprimer les modèles recencés ainsi que les paramètres employés pour la modélisation des cinétiques de croissance et de production. Dans un effort d'unification, la nomenclature de certaines variables a été changée par rapport à la nomenclature originale.

Drapeau, Blanch et Wilke (1986) ont étudié la cinétique de croissance de cellules de *Dioscorea deltoidea* et de *Catharanthus roseus* en cultures en cuvée dans un bioréacteur Chemap de 14 l agité mécaniquement. Leur modèle simple relie le taux de croissance spécifique à la concentration en sucre du milieu. Ils constatent un lien entre la production de métabolites et les concentrations en sucre, en phosphate et en acide 2,4-dichloro-phénoxyacétique du milieu initial sans formuler de relation mathématique entre ces éléments.

Le modèle proposé par Taticek et al. (1990) est un modèle logistique appliqué à la croissance de cellules de *Eschscholtzia californica* en bioréacteurs agité mécaniquement et de type gazosiphon, tous deux d'un volume de 1 l. Ces deux premiers modèles se basent sur le schème réactionnel simple présenté au Tableau 2-2. Ce type de modèle est difficilement utilisable pour des procédés qui ne sont pas en cuvée simple. En effet, un

modèle logistique cache l'effet d'une limitation en un nutriment non identifié. L'ajout de nutriments au milieu modifiera donc l'allure sigmoïdale de la courbe de croissance.

Les travaux sur des cellules de *Vitis vinifera* cultivées en cuvée et semi-continu réalisés par Guardiola et al. (1995) ont permis de valider un autre modèle simple pour suivre le profil de croissance de la biomasse et de la production d'anthocyanes. Pour ces deux types de cultures, les cellules passent par un stade d'épuisement en sucre et par un métabolisme endogène. Les cinétiques de croissance de biomasse, de production de métabolite et la consommation de biomasse pour l'entretien cellulaire sont basées sur un modèle de type Monod. Les rendements de conversion du sucre en biomasse et en anthocyanes sont calculés simultanément avec les autres paramètres du modèle. Les constantes de Monod obtenues pour les différents taux de croissance et de production sont de 0,1 ou 0,4 g/l. D'une part, ces résultats indiquent qu'en absence de sucre la croissance cesse, ce qui est conforme aux observations. D'autre part, ces valeurs indiquent plutôt une dépendance d'ordre 0 face au sucre puisque, pour toute concentration en sucre supérieure à 1 g/l, la valeur maximale du taux de croissance ou de production est atteinte. Pour une culture en cuvée, la concentration est supérieure à 1 g/l pour presque toute la durée de la phase de croissance, jusqu'à peu avant son épuisement quand la concentration diminue rapidement; la zone de transition est très courte. Ces résultats indiqueraient que le sucre est essentiel à la croissance mais que, lorsqu'il est présent, sa concentration n'influence pas les taux de croissance ou de production.

Curtis et al. (1991) constatent l'absence de phase de croissance exponentielle de la bio-

masse pour leurs cultures de cellules de *Papaver somniferum* en cuvée. Diverses cultures croissant sur des milieux ayant subi des modifications de la concentration en phosphate par rapport au milieu MS standard les ont amenés à formuler un modèle du taux de croissance basé sur une fonction linéaire des réserves intracellulaires de phos-phate. Ces résultats sont comparés aux résultats de la présente étude à la section 7.5.1.

Les expériences de Bailey et Nicholson (1989) sur des cellules de *C. roseus* en flacons agités les ont amenés à développer un modèle structuré basé sur 5 composantes soit la masse sèche des cellules viables (X_{vd}), la masse sèche des cellules non viables (X_{nv}), la masse humide (X_f), la concentration en métabolites secondaires (M, ajmalicine et serpentine) et la concentration en substrat limitant (S, sucre, non mesuré). Dans ce modèle, la perte de viabilité est associée à l'augmentation du rapport de la biomasse humide sur la biomasse sèche qui est un indicateur de la taille des cellules et de la sensibilité au cisaillement. La relation entre le taux de croissance et la concentration en sucre (le substrat limitant) est supposée de premier ordre. Les auteurs observent une relation entre le rendement apparent en sucre et le taux de croissance en fonction de la température. Ce modèle structuré décrit assez bien les phases de croissance exponentielle, de stagnation et de déclin de la biomasse pour une culture en cuvée. La production de métabolites secondaires, prédite par un modèle non structuré de Luedeking-Piret (équation 2-7), ne représente pas de façon satisfaisante les résultats observés. Shibasaki et al. (1993) reprennent le modèle de Bailey (équation 2-1) auquel ils ajoutent une vacuole. Un métabolite secondaire est synthétisé par les cellules viables et, stocké à concentration constante dans la vacuole. Ces deux modèles sont basés sur des paramètres

physiques mesurables mais n'indiquent en rien les voies métaboliques utilisées comme c'est le cas pour les modèles qui suivent.

Hooker et Lee (1992) présentent un modèle structuré pour prédire la production de métabolites secondaires par des cultures de cellules de *Nicotiana tabacum* en flacons agités. Ce modèle structuré à trois compartiments tient compte de trois produits finaux: la biomasse structurale (B), la masse perdue par respiration (R, par métabolisme endogène), et les métabolites secondaires (M). Le modèle tient compte de l'hydrolyse du sucrose (Se) en glucose et fructose (Ge), des autres nutriments (Ne), des nutriments intracellulaires totaux (T) qui sont convertis en intermédiaires structuraux (T*) et ensuite transformés en biomasse structurale (B) ou en métabolite (M). Les cinétiques de chacune des réactions proposées ont la forme du modèle de Monod ou d'une cinétique chimique élémentaire. Les concentrations en nutriments intracellulaires sont exprimées en composition intrinsèque (par volume de milieu abiotique ou par gramme de biomasse sèche). La composition intrinsèque correspond mieux à la réalité en ce sens que les concentrations sont exprimées par rapport à la biomasse plutôt que par rapport au volume total du milieu.

Les travaux de Frazier (1989) portent sur des cultures en cuvée de cellules de *Dioscorea deltoidea*. La modélisation repose sur la présence d'un métabolite essentiel à la croissance qui est relargué vers le milieu. Le taux de croissance spécifique (μ) de la biomasse repose sur la concentration de ce seul métabolite selon un modèle de type Monod. Le taux de production de ce métabolite dépend de la concentration en substrat

dans le milieu, également selon un modèle de type Monod. Ce métabolite sert essentiellement à suivre la phase de latence d'une culture, ce qui correspond à l'idée que la mécanique biologique doit se mettre en marche en produisant des précurseurs et des enzymes avant de démarrer la croissance. À cela, Frazier ajoute le transfert vers le milieu de certains de ces composés essentiels qui peuvent ainsi prolonger la phase de latence.

Dans un premier temps, van Gulik et al. (1992) ont proposé un modèle simple pour la croissance de cellules de *C. roseus* et *N. tabacum* en cultures en cuvée et en chemostat. Ce modèle utilise une relation de type Monod entre le taux de croissance spécifique (μ) et la concentration en sucre du milieu. Les données provenant de l'analyse de la composition des gaz sont utilisées de la façon suivante : la consommation d'oxygène, la production de dioxyde de carbone et la consommation de sucre sont introduites dans le modèle à l'aide de bilans stoechiométriques. Les auteurs utilisent la redondance des données des bilans de masse pour effectuer une correction statistique sur les données (Wang et Stephanopoulos, 1983). Les bilans massiques sont bouclés par une analyse élé-mentaire de la biomasse qui donne la formule empirique de la biomasse et des produits.

Constatant les limites d'un modèle non structuré, ces mêmes auteurs (van Gulik et al., 1993) proposent un modèle structuré pour la croissance de cellules de *C. roseus*. Ce modèle comporte 4 composantes intracellulaires, soit le phosphate libre (P) rapidement absorbé du milieu, les précurseurs phosphorylés formés de phosphate libre (E) et de glucose (G) provenant du milieu. Ces derniers, (P et G), sont utilisés pour la formation

de biomasse structurale (B). Le glucose peut également être converti en amidon accumulé dans la cellule (C). Les cinétiques proposées sont de type Monod ou de premier ordre par rapport aux réactifs impliqués. La cinétique de production de biomasse est fonction des concentrations des précurseurs intracellulaires (E) et du glucose (G). De plus le phosphate (P) a un effet inhibiteur sur la production d'amidon (C). Il y a aussi une réaction de lyse de la biomasse (B). Les auteurs utilisent une composition intrinsèque pour exprimer les concentrations en nutriments intracellulaires.

Dans une approche originale sous plusieurs aspects, de Gunst et al. (1990) proposent un modèle stochastique avec ségrégation appliqué à une culture en cuvée de cellules de *N. tabacum*. Tous les calculs sont basés sur le nombre de cellules plutôt que sur la biomasse. L'hypothèse utilisée est que les cellules passent graduellement d'un état actif où elles se divisent à un état d'accumulation où elles grossissent, sans se diviser. Ce modèle indique que la probabilité qu'une cellule née à un temps donné sera fonction de la concentration en régulateur de croissance (l'acide 2,4 dichlorophénoxyacétique, ou 2,4-D) à cet instant. Le temps de division est fonction d'un temps de division minimum et d'une fonction stochastique qui dépend de la concentration en sucre du milieu. Ceci implique que le temps de doublement de la population augmente avec la progression de la culture puisque la concentration en sucre diminue et que la fraction de cellules en division diminue. Il semble que dans l'expérience présentée, la concentration initiale d'hormone (2,4-D) influence la production finale de biomasse et peut être considérée comme limitante. Les travaux présentés dans ce mémoire contredisent ces résultats. En effet, l'ajout de nutriments au milieu de croissance, sans ajout d'hormone, permet d'augmenter la con-

centration en biomasse. Donc, s'il y a un élément limitant ce n'est probablement pas les phytohormones, du moins pour des cultures de cellules de *E. californica*.

Cette revue présente des modèles de croissance de cellules végétales qu'on peut regrouper sous les catégories prédéfinies de structure et de ségrégation. Le seul modèle avec ségrégation (de Gunst, 1990) repose sur une approche stochastique, sur le compte cellulaire et sur l'observation microscopique de l'aspect des cellules. Des travaux récents du groupe de cultures de cellules végétales de Biopro mettent également en évidence l'importance du nombre de cellules par rapport à la seule mesure de la biomasse. Cependant, les modèles présentés dans le présent travail utiliseront l'approche plus classique de la mesure de la biomasse.

2.4.6 Discussion sur les modèles cinétiques

Les modèles recensés dans la littérature utilisent une cinétique de type Monod (équation 2-8) sur un ou plusieurs nutriments mesurés de façon à déterminer les taux de croissance de biomasse ainsi que le taux de consommation de nutriments ou alors la variation de paramètres physiologiques telle la viabilité cellulaire.

$$\mu(t) = \mu_{max} \frac{S(t)}{K_s + S(t)} \tag{2-8}$$

Les plus simples (Taticek et al., 1990; Guardiola et al., 1995; Frazier, 1989) et même d'autres beaucoup plus complexes (de Gunst, 1990) utilisent directement l'équation de Monod pour prédire le taux de croissance.

D'autres auteurs (Bailey et Nicholson, 1989; Shibasaki et al., 1993) ont basé leur cinétique sur des paramètres physiologiques comme la viabilité, le stress, le rapport biomasse sèche sur biomasse humide, etc. Les taux de croissance sont implicitement déterminés par le système d'équations différentielles reliant ces différents paramètres. Dans ces cas, les taux de croissance ne sont pas exprimés en fonction des concentrations en nutriments dans le milieu de culture.

Certains modèles récents (Hooker, 1992; van Gulik, 1993) utilisent une combinaison de cinétique de Monod et de réactions chimiques élémentaires pour exprimer les taux de synthèse et d'utilisation de métabolites intermédiaires.

Les modèles structurés font intervenir des métabolites intermédiaires et des chaînes de réaction. Ces modèles sont utiles pour conceptualiser la croissance et expliquer ses diverses phases. Cependant l'utilité de ces modèles est plutôt limitée pour le contrôle de la croissance à cause de la nécessité d'estimer la valeur des paramètres non mesurés. Un dernier constat porte sur l'utilisation généralisée du modèle de Monod pour déterminer la cinétique de la biomasse en fonction d'un substrat. Ce modèle a l'avantage d'être simple mais présente quelques inconvénients : il ne permet pas de détecter des effets plus complexes (inhibition, compétition) que la saturation par un composé et il doit être calculé par une méthode de résolution itérative qui requiert une certaine puissance de calcul.

Les modèles simples (sans structure ni ségrégation) présentent l'avantage de reposer sur des paramètres mesurables ou calculables par bilans de masse. La simplicité de ces mo-

dèles et la disponibilité des paramètres mesurables en font des candidats de choix pour des fonctions de contrôle.

Par ailleurs, seuls quelques modèles font intervenir des nutriments autres que le sucre. Parmi ceux-ci, on retrouve l'oxygène et le phosphate. Il est connu que le phosphate joue un rôle lors de la croissance et de la production de métabolites secondaires. D'autres auteurs (Rho et André, 1991) incluent l'azote parmi les nutriments principaux. Ces derniers nutriments sont également considérés lors de la formulation des modèles présentés dans cette étude.

De plus, l'approche utilisée par la plupart des auteurs cités consiste à proposer un modèle unique qui correspond à une conception macroscopique de la synthèse de la biomasse et de la production de métabolites, ce qui élimine d'autres modèles, peut-être plus simples, qui pourraient tout aussi bien suivre l'évolution de la biomasse. Sauf pour Shibasaki et al. (1993) qui proposent une amélioration au modèle de Bailey et Nicholson (1989), les modèles ne sont pas comparés entre eux. À l'exclusion de van Gulik et al. (1993), les sources citées n'utilisent pas d'outils statistiques pour valider les paramètres obtenus ou préciser leurs intervalles de confiance.

3. Méthodologie

La méthodologie de modélisation employée se divise en deux parties. Dans un premier temps, la stoechiométrie est étudiée à l'aide de schèmes réactionnels, indépendamment de la cinétique. Des modèles polynômiaux sont par la suite appliqués à la modélisation de la cinétique de la croissance de la biomasse. Cette division découle de la méthodologie employée par Chotteau (1995) pour la modélisation de la croissance de cellules animales.

3.1 Schèmes réactionnels

La première étape consiste à choisir divers schèmes réactionnels, ou systèmes de réactions, qui décrivent les relations entre les variables du milieu (biomasse, nutriments et métabolites). Ces schèmes, comme ceux présentés au Tableau 2-2, décrivent, sous forme de pseudo-réactions, un modèle simplifié de la réalité. Ces schèmes réactionnels s'apparentent à des réactions chimiques, avec les considérations suivantes (Dochain, 1990).

- Les schèmes réactionnels ne représentent pas une relation stoechiométrique entre les composants. Il s'agit plutôt d'une relation qualitative. Les coefficients des équations sont dits pseudo-stoechiométriques.

- Les composants d'un schème réactionnel sont généralement de 4 types: population (cellules), enzymes, substrats externes, produits ou substrats internes.

pour déterminer un modèle dynamique opérationnel. Seuls les composés d'intérêt sont inclus, au détriment de la loi de conservation de masse.

- Une réaction est un bilan d'un ensemble de réactions.

- Les cellules sont systématiquement les catalyseurs des réactions.

Ces schèmes ne s'opposent en rien aux réseaux de voies métaboliques décrivant le fonctionnement de la cellule. Il s'agit plutôt d'un complément donnant une vue macroscopique simplifiée d'un ensemble de réactions.

Plusieurs schèmes ont été développés et étudiés dans le présent travail; quelques-uns sont puisés dans la littérature, certains sont adaptés et d'autres sont originaux. Ces schèmes réactionnels peuvent être traduits en bilans de masse en régime transitoire. Ces bilans de masse, élargis aux prélèvements et aux ajouts faits aux cultures, peuvent s'exprimer sous la forme d'un système d'état. La résolution de ce système d'état à partir de données expérimentales permet de déterminer les coefficients pseudo-stoechiométriques qui sont, dans certains cas simples, identiques aux coefficients de rendements "Y" généralement utilisés dans la littérature. Des manipulations algébriques effectuées sur ces systèmes d'état permettent d'éliminer les composantes cinétiques des systèmes en ne conservant comme inconnues que les coefficients pseudo-stoechiométriques. Ces systèmes d'état peuvent être transformés en équations multilinéaires dont les coefficients sont reliés aux coefficients pseudo-stoechiométriques des schèmes correspondants. Ces coefficients sont déterminés à l'aide de simples régressions multilinéaires par la méthode

des moindres carrés.

Les divers schèmes proposés sont donc comparés sur la base de l'erreur résiduelle de modélisation puis par validation croisée sur des cultures n'ayant pas servi à la régression. Par la suite, les paramètres de régression obtenus sont analysés à l'aide d'autres critères statistiques tels le coefficient de détermination multiple empirique ajusté et les intervalles de confiance.

3.2 Cinétique de croissance

L'outil de modélisation proposée par Chotteau est la validation d'une série de relations de forme polynômiale. Les composantes de ces polynômes peuvent être des substrats ou d'une autre nature : concentrations intracellulaires, métabolites intra ou extracellulaires, enzymes, etc. Les modèles peuvent être de complexités diverses selon le nombre de monômes et l'ordre des polynômes. Les modèles polynômiaux sont une combinaison linéaire de monômes qui sont fonction d'un ou plusieurs paramètres.

Chaque modèle créé exprime le taux de croissance spécifique de la biomasse (μ) par un polynôme dont les monômes sont un produit des paramètres cinétiques choisis. Les coefficients du polynôme sont obtenus par résolution par les moindres carrés du système multilinéaire obtenu à partir d'un bilan de masse sur la biomasse. Ainsi,

$$\mu(t) = a_0 + a_1 S(t) + a_2 S(t)^2 + a_2 S(t)^3 \qquad (3\text{-}1)$$

est un polynôme d'ordre 3 à trois termes basés sur un seul substrat S et

$$\mu(t) = a_0 + a_1S(t) + a_2P(t) + a_3S(t)P(t) + a_4S(t)^2P(t) \tag{3-2}$$

est un polynôme d'ordre 3 à 4 termes basés sur les substrats S et P. Ces modèles diffèrent du modèle de type Monod qui est généralement basé sur un seul substrat.

Les modèles ainsi créés n'ont pas de signification biologique mais permettent de déceler des relations cinétiques qui pourront par la suite servir à bâtir des modèles phénoménologiques. Par ailleurs, tous les modèles exprimés sous forme de quotient de polynômes (Monod, Haldane, etc.) peuvent être reproduits par un modèle polynômial. Ainsi, comme le montre la Figure 3-1, un modèle de Monod peut être approximé par un polynôme d'ordre trois avec une précision comparable à l'erreur de mesure des données expérimentales. Dans cette figure, les paramètres du modèles de type Monod sont μ_{max} = 100 et K_S = 30. L'équation du polynôme correspondant est:

$$\mu = 8{,}3*10^{-5}S^3 - 2{,}1*10^{-2}S^2 + 2{,}0S + 5{,}9 \tag{3-3}$$

Le plus grand avantage des modèles polynômiaux est leur simplicité de résolution à l'aide de régressions multilinéaires classiques. Des algorithmes de régression très performants permettent de résoudre ces systèmes d'équation avec peu de puissance de calcul. En corollaire, un ordinateur puissant permet de résoudre un grand nombre de modèles en un temps raisonnable. La méthode utilisée est basée sur le critère des moindres carrés. La simplicité de résolution ne s'applique pas à d'autres types de modèles (dont le modèle de Monod) qui requièrent une résolution par itération ce qui exige beaucoup de puissance de calcul. De plus, dans le cas de systèmes non linéaires, l'évaluation des intervalles de confiance est complexe et approximative, ce qui rend plus

difficile la détection d'une éventuelle surparamétrisation (Draper, 1981).

La préparation des données est semblable à celle réalisée pour la validation des schèmes réactionnels : la même séquence de manipulations mathématiques transforme les bilans de masse en systèmes d'état. Toutes les données relatives aux paramètres mesurés et calculés sont connues ainsi que le taux de croissance, et cela pour chaque point d'échantillonnage d'une culture. Chaque modèle polynômial est ensuite construit et ses coefficients de régression sont calculés par la méthode des moindres carrés.

Les critères de sélection sont, l'erreur résiduelle de modélisation pour comparer les modèles et un intervalle de confiance pour valider chacun des paramètres d'un modèle retenu.

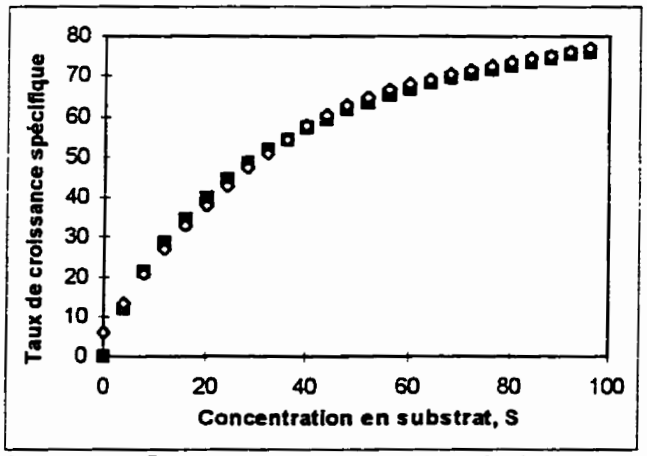


Figure 3-1 Comparaison entre un modèle de type Monod et un modèle cinétique polynômial (◊) : modèle polynômial, (■): modèle de Monod.

3.3 Rendements, cinétique et contrôle

La connaissance des coefficients pseudo-stoechiométriques obtenus de schèmes réactionnels validés permet d'inférer de plusieurs variables d'état en se basant sur la mesure d'un nombre limité de paramètres. En d'autres termes, il est possible d'estimer des états d'un modèle biologique représentant un système biologique à partir d'un petit nombre de paramètres. Ces informations sur les états du système combinées à la connaissance de la cinétique réactionnelle permettent d'envisager un contrôle prédictif du procédé avec un minimum d'échantillonnage.

En combinant les coefficients de rendements à la connaissance des cinétiques de croissance ou de production de métabolites par exemple, il est possible de prédire l'état d'une culture dans un temps ultérieur. Ceci ouvre la porte à un contrôle optimal permettant de maintenir des conditions favorisant la plus grande productivité possible par une alimentation en nutriments, par un contrôle de la température ou par un ajustement du pH.

Le présent travail se concentre sur les aspects nutritionnels de la modélisation. Les aspects directement associés au contrôle de procédé (observabilité, stabilité) ne seront pas traités.

4. Matériel et méthodes

4.1 Lignée cellulaire

Les cellules employées pour cette étude proviennent de la lignée cellulaire EC6 de l'espèce *Eschscholtzia californica* développée par le Dr Robert D. Williams du groupe de recherche en culture de cellules de plantes du centre BIOPRO. Le repiquage (30% V/V) s'effectue tous les 12 à 14 jours sur un milieu B5 (Wetter et al., 1982) auquel on ajoute 0,2 mg/L d'acide 2-4 dichlorophenoxyacétique (2,4-D), 0,1 mg/l de kinétine et 30 g/l de sucrose. Le pH est ajusté à 5,5 à l'aide d'une solution de KOH 0,1 N avant stérilisation (30 min à 121 °C, 1 bar). Les flacons (Erlenmeyer de 500 ml ou 1l, contenant 200 ml ou 400 ml de milieu) sont maintenus dans l'obscurité à 26 °C ± 1 °C dans un agitateur orbital opéré à 120 tours par minute.

Aux cultures d'inoculation, 3 ml d'une solution stérile contenant 2,2 g de $(NH_4)_2HPO_4$ et de 8,5 g de $(NH_4)_2SO_4$ par 100 ml sont ajoutés après 3 à 6 jours. Cet ajout a pour but de prolonger la période de croissance exponentielle et permet de réduire la phase de latence normalement observée suite à l'inoculation.

4.2 Description du bioréacteur

Le bioréacteur cylindrique d'un volume utile de 6 l a été réalisé par la firme Biogénie inc., Québec, Canada. Il s'agit d'un bioréacteur à rubans hélicoïdaux doubles spécialement conçu pour les cultures à faible résistance au cisaillement (Jolicoeur, 1992; Kamen, 1992).

L'axe des doubles rubans est entraîné par une courroie reliée à un moteur de 0,5 hp de vitesse réglable. Un refroidisseur à eau prévient l'évaporation. Le port d'échantillonnage facilite le prélèvement d'échantillons. Une chemise d'acier forme la base du bioréacteur à travers laquelle sont introduites deux sondes à oxygène dissous, le port d'échantillonnage et le diffuseur poreux. Cette chemise permet de contrôler la température du milieu par circulation d'eau chaude ou froide.

L'alimentation gazeuse (un mélange d'air et d'azote ou d'oxygène) est ajustée aux points de consigne par 3 contrôleurs de débit massique (0-200 cm^3/min, Tylan) calibrés pour chacun des gaz. Cette alimentation gazeuse est filtrée à l'aide de laine de verre et de filtres (0,45 µm, Gelman Science). Le bullage est introduit dans le bioréacteur au moyen d'un diffuseur en acier poreux (SS-316, 0,2 µm). Les gaz de sortie sont analysés au moyen d'un moniteur à CO_2 à l'infrarouge (0-3%, modèle 8700, Rosemount Instruments) relié au système d'acquisition de données. Deux sondes à oxygène dissous polarographiques stérilisables (Ingold), une sonde à pH (Cole Parmer) et plusieurs sondes à température RTD équipent le bioréacteur et sont reliées au système d'acquisition de données et de contrôle.

4.3 Préparation d'une culture en bioréacteur

Le bioréacteur HRI de 6 l est stérilisé à l'autoclave (60 min à 121 °C, 1 bar) puis refroidi sous hotte laminaire. L'inoculum consiste en 1,0 à 1,2 l de cultures en flacon agées de 10 à 12 jours. Le milieu de culture est B5 + 0,2 mg/l 2,4-D, + de 0,1 mg/l kinétine, le tout concentré 1,33 fois par rapport à la concentration originale. La concen-

tration en sucrose variait entre 20 et 40 g/l. Le milieu est stérilisé par filtration (Micro Culture Capsule, 0,2 µm, Gelman Sciences) sauf pour le sucre qui est préparé et stérilisé à la vapeur (30 min à 121 °C, 1 bar). De 4,5 à 4,8 l de milieu sont transférés dans le bioréacteur. Le volume initial de la culture variait entre 5,5 et 6,0 l.

4.4 Conditions d'opération

La température du bioréacteur est contrôlée à 26 °C ± 1 °C. La vitesse d'agitation est réglée à 60 tpm en début de culture et peut être augmentée à 75 ou 90 tpm afin d'augmenter le taux de transfert d'oxygène à la culture. Le débit d'aération de surface (air ou mélange air-oxygène) est fixé à 300 cm^3/min à l'aide d'un rotamètre. Le débit de bullage d'un mélange d'air et d'oxygène est fixé à une valeur comprise entre 50 et 250 cm^3/min. La concentration en oxygène du gaz varie entre 21 % (air) et 100 %.

4.5 Stratégies d'alimentation

Les ajouts de nutriments sont constitués de $(NH_4)_2HPO_4$ et de $(NH_4)_2SO_4$ à diverses concentrations et de sucrose. Les ajouts sont stérilisés (30 min à 121 °C, 1 bar) séparément. Ces nutriments sont ajoutés selon la stratégie choisie.

Deux types de cultures en cuvée alimentée ont été réalisées. La première consiste en un ou deux ajouts ponctuels de 200 à 500 ml, d'une solution de $(NH_4)_2SO_4$ et de $(NH_4)_2HPO_4$. Le moment des ajouts est déterminé par l'opérateur et réalisé lorsque les concentrations mesurées en ammoniac et en phosphate dans le milieu sont nulles. Les concentrations en ammoniac et en phosphate dans le milieu ne dépassent pas 15 mM et

2 mM respectivement, après l'ajout.

La méthode d'ajout graduel développée dans cette étude consiste en l'ajout au goutte-à-goutte d'une solution de $(NH4)_2SO_4$ et de $(NH_4)_2HPO_4$. Le critère d'ajout est une valeur plafond de pH. Lorsque le pH de consigne est atteint, le système de contrôle Bioexpert active une pompe péristaltique qui alimente environ 1 ml de la solution à la culture. Cet ajout entraîne une faible baisse de pH sous le point critique durant la consommation des nutriments. Suite à l'épuisement des nutriments ajoutés, le pH remonte vers le point de consigne pour un nouvel ajout. Cette forme d'ajout permet de maintenir le pH à une valeur cible tout en maintenant la concentration de nutriments à une valeur faible ou nulle et permet d'ajouter des quantités de nutriments supérieures à la méthode d'ajouts ponctuels.

Le sucrose est toujours ajouté de façon ponctuelle par volumes de 300 à 500 ml provoquant une augmentation de 15 à 25 g/l dans le bioréacteur.

4.6 Acquisition de données et contrôle

Tous les signaux (vitesse d'agitation, température, pH, oxygène dissous, débits gazeux) sont captés par le système d'acquisition Bioexpert (Biogénie, Québec, Canada) fonctionnant sur un compatible PC équipé d'un processeur 286-16MHz et de 2 mégabytes de mémoire vive.

Ce système capte les signaux d'entrée et met à jour les signaux de sortie toutes les 10 secondes. La température, les débits gazeux et la concentration en oxygène dissous sont

contrôlés à l'aide de contrôleurs PID digitaux programmés dans le Bioexpert. Le Bioexpert est programmé de façon à calculer le taux de consommation d'oxygène (OUR), le taux de production de CO_2 (CPR) et le taux de transfert d'oxygène (k_La) à partir de perturbations périodiques provoquées sur l'alimentation en gaz. L'acquisition des données pour analyse est faite aux 30 minutes.

4.7 Flacons témoins

Pour chaque culture en bioréacteur, de 3 à 6 flacons témoins (Erlenmeyer de 500 ml contenant 200 ml de milieu) sont inoculés simultanément avec un inoculum de même origine. Ces flacons sont placés sur un agitateur rotatif à 120 tpm et 26 °C

Pour certaines expériences, les flacons témoins reçoivent proportionnellement les mêmes ajouts ponctuels que le bioréacteur. Dans d'autres cas, les flacons témoins n'ont subi aucun ajout.

4.8 Échantillonnage

Des prélèvements sont effectués à des périodes variant de 12 à 48 h pour les cultures en bioréacteur et de 24 à 72 h pour les cultures en flacons. L'échantillonnage du bioréacteur consiste à recueillir en conditions stériles une purge d'environ 50 ml de culture dans un flacon laveur de 250 ml puis un échantillon d'environ 100 ml de la même façon. L'échantillonnage des flacons est fait en pipetant environ 13 ml de culture.

Le volume liquide recueilli est pesé. Des lectures externes de pH (Orion) et de conductivité (Orion) sont prises. Environ 12 ml de milieu sont pipetés dans un cylindre de 15

ml puis centrifugés (2000 tpm, 2 min) afin de déterminer le volume des cellules compactées (Packed Cell Volume, PCV). Parfois 1 ou 2 prélèvements de 2 ml sont faits pour le compte cellulaire. Le volume recueilli est pesé à nouveau. La biomasse est ensuite filtrée à 10µm (Miracloth, Mira), rincée à l'eau distillée, pesée, séchée durant au moins 24 heures à 80 °C puis pesée à nouveau. Ces manipulations permettent de déterminer les concentrations de biomasse humide et sèche. Un volume de 10 ml du filtrat recueilli avant le rinçage des cellules est congelé pour l'analyse ultérieure des ions et des sucres.

4.9 Analyse des sucres et des ions

Ces analyses sont pratiquées en série sur tous les filtrats congelés d'échantillons d'une même culture. Les échantillons sont décongelés à température ambiante puis filtrés à 0,2µm (Gelman Science) avant d'être analysés.

4.9.1 Hydrates de carbone

La concentration du milieu des cultures en sucrose, glucose et fructose est mesurée au moyen d'un chromatographe en phase liquide à haute performance (HPLC) comportant une pompe Waters modèle 6000A, un injecteur automatique Gilson modèle 231/401, un réfractomètre Gilson modèle 132, un élément chauffant Man-Tech modèle 7980 et un intégrateur Hewlet-Packard modèle HP3394A. La séparation des hydrates de carbone est assurée par une colonne Biorad HPX87C maintenue à 80 °C. Le volume d'injections est de 20µl. La phase mobile d'eau pure est alimentée à un débit de 1 ml/min.

4.9.2 Chromatographie ionique

La concentration des anions et cations est analysée par chromatographie ionique. Le système (Dionex) est composé d'une pompe, d'un injecteur automatique et d'un détecteur électrochimique opéré en mode conductivité. Ce système est contrôlé par le logiciel Dionex A1-450.

Les anions sont séparés par une colonne 4 x 250 mm IONPAC AS4A-SC (Dionex) et un suppresseur anionique (ASRS-1, Dionex). La phase mobile est une solution tampon aqueuse de bicarbonate (1,8 mM Na_2CO_3 et 1,7 mM $NaHCO_3$) alimentée à un débit de 2,0 ml/min. Les anions mesurés sont le chlorure, le nitrate, le phosphate et le sulfate.

Les cations sont séparés par une colonne 4x250 mm IONPAC CS-12 (Dionex) et un suppresseur cationique (CSRS-1, Dionex). La phase mobile est une solution aqueuse d'acide méthasulphonique (20 mM) alimenté à un débit de 1 ml/min. Les cations mesurés sont l'ammoniac, le calcium, le magnésium, le potassium et le sodium.

5. Schèmes réactionnels et bilans de masse

L'étude des mécanismes réactionnels macroscopiques est une étape préliminaire très utile au développement de modèles mathématiques (Dochain, 1990). Des schèmes réactionnels élaborés à partir de ces mécanismes permettent d'établir une relation stoechiométrique entre la variation de divers paramètres et permet ainsi de mieux suivre l'évolution d'une culture à partir d'un petit nombre de paramètres mesurés.

Dans ce chapitre, divers schèmes réactionnels sont présentés avec les systèmes d'état correspondants. Pour chaque schème, un ou plusieurs estimateurs de biomasse sont développés. La méthode utilisée est celle proposée par Chotteau (1995) appliquée au suivi de la biomasse plutôt qu'à tous les nutriments impliqués. Les schèmes présentés dans la présente étude ne tiennent compte que de la biomasse et des paramètres nutritionnels mesurés (ammoniac, nitrate, phosphate et sucre).

Tous les modèles rencontrés dans la littérature ne sont pas évalués. Certains sont trop complexes (Bailey, 1989; Hooker, 1992; van Gulik, 1993) par rapport aux données disponibles pour les raisons qui seront élaborées au chapitre 6. D'autres sont inapplicables, soit parce qu'ils sont basés sur des métabolites secondaires (absents de la phase de croissance de *E. californica*), soit parce qu'ils reposent sur des paramètres non mesurés dans ce travail (Guardiola, 1995; Shibasaki, 1993; de Gunst, 1990) ou pour ces deux raisons.

lier à l'utilisation du sucre (S) selon l'équation 5-1.

$$sS \xrightarrow{r_1} X \tag{5-1}$$

La biomasse peut s'exprimer en grammes par litre (g/l) ou en cellules par ml (10^6cell./ml). Les grammes par litres sont utilisés dans ce document. Cette unité correspond à des grammes de biomasse sèche par litre de suspension.

Ce schème peut s'exprimer sous forme d'un bilan de masse.

$$\frac{dX}{dt} = r'_1(t) \tag{5-2}$$

$$\frac{dS}{dt} = -sr'_1(t) \tag{5-3}$$

Afin d'augmenter la précision des calculs à haute concentration de biomasse, une correction est apportée aux concentrations en nutriments mesurées qui expriment une concentration par volume de milieu abiotique. Pour rapporter ces valeurs au volume total de milieu, il faut tenir compte du volume occupé par les cellules. La correction choisie est celle utilisée par van Gulik (1993) et exprimée par l'équation 5-4.

$$C = C' * (1 - 0{,}5 * PCV) \tag{5-4}$$

où C' est la concentration mesurée par les appareils d'analyse, et PCV est la fraction volumique des cellules compactées par rapport au volume total.

Cette correction, qui suppose que la fraction occupée par les cellules correspond à 50 % du volume des cellules compactées, est appliquée à toutes les valeurs de concentration présentées dans cet ouvrage. De plus, lors des calculs, tous les volumes de cultures sont

normalisés à 1 litre de façon à pouvoir comparer les cultures réalisées dans des volumes différents. Ainsi, les résultats exprimés en notation Δ (ex.: équation 5-7) sont calculés en fonction d'un volume initial de culture de 1l pour les flacons comme pour les réacteurs afin d'accorder à chacune des cultures une pondération similaire lors des régressions. Les volumes d'ajouts et de prélèvements sont ajustés en conséquence.

En tenant compte des ajouts de sucre et des prélèvements de biomasse et de culture lors des échantillonnages, le bilan de masse devient plus complexe.

$$\frac{dX(t)V(t)}{dt} = r_1(t) - F^{out}(t)X(t) \tag{5-5}$$

$$\frac{dS(t)V(t)}{dt} = -sr_1(t) + F^{in}S^{in}(t) - F^{out}S(t) \tag{5-6}$$

Ces bilans de masse sont traduits en systèmes matriciels, intégrés et se retrouvent après un changement de variable sous une forme simple et plus adaptée aux traitements numériques. Les étapes intermédiaires se retrouvent à l'annexe B.1. Le système matriciel correspondant au premier schème est donné par l'équation 5-7.

$$\begin{bmatrix} \Delta X(t) \\ \Delta S(t) \end{bmatrix} = \begin{bmatrix} 1 \\ -s \end{bmatrix} [R_1(t)] \tag{5-7}$$

Les variables précédées du symbole Δ signifient une consommation ou production totale au temps t par rapport au premier échantillonnage de la culture,

et $$R_1 = \int_0^t r_1(\iota)d\iota$$

Le système précédent peut être exprimé comme une relation linéaire entre la production

de biomasse et la consommation de sucre par l'équation 5-8.

$$\Delta X(t) = k_s \Delta S(t) \tag{5-8}$$

avec

$$k_s = -\frac{1}{s}. \tag{5-9}$$

5.2 Second schème réactionnel : croissance et maintenance

Le premier schème simple peut être augmenté d'un terme de maintenance et représenté par l'équation 5-10 :

$$\begin{aligned} s_1 S &\xrightarrow{r_1} X \\ s_2 S + X &\xrightarrow{r_2} X \end{aligned} \tag{5-10}$$

Ce schème représente les modèles utilisés par Drapeau et al. (1986) et Taticek et al. (1990) à l'exception de l'oxygène qui est absent du présent schème.

Après les transformations présentées à l'annexe B.2, on obtient le système linéaire donné par l'équation 5-11.

$$\Delta X(t) = k_s \Delta S(t) + k_{xv} XV(t) \tag{5-11}$$

avec k_S=-1/s_1 et k_{XV}=-s_2k_2/s_1

et où

$$XV(t) = k\int_0^t X(\tau)V(\tau)d\tau \tag{5-12}$$

5.3 Troisième schème réactionnel

Les deux premiers schèmes sont surtout utilisés pour des systèmes microbiens. Ils n'utilisent que l'information relative au sucre. D'autres schèmes utilisent une plus grande quantité d'information disponible telles les concentrations en nutriments importants. Les schèmes suivants utilisent également d'autres paramètres calculés telles les concentrations en nutriments intracellulaires. Ces paramètres ne sont pas disponibles par mesure mais peuvent être calculés par bilan de masse selon une méthode présenté à la section 7.5.

Le schème suivant inclut l'absorption des principaux nutriments (P=phosphate, N=ammoniac, M=nitrate) avant leur utilisation pour la formation de la biomasse:

$$\begin{aligned} P_e &\rightarrow P_i \\ N_e &\rightarrow N_i \\ M_e &\rightarrow M_i \\ S_e &\rightarrow S_i \\ s_2S_i + pP_i + nN_i &\rightarrow X \\ s_3S_i + pP_i + nM_i &\rightarrow X \end{aligned} \tag{5-13}$$

Les variables portant l'indice "e" sont les concentrations extracellulaires, exprimées en grammes ou millimoles par litre (g/l ou mM). Les variables portant l'indice "i" sont des concentrations intracellulaires exprimées en grammes ou millimoles par gramme de biomasse sèche. Ces dernières ne sont pas mesurées mais sont plutôt calculées par bilans de masse. Pour fins de simplifications, l'étape d'hydrolyse du sucrose en glucose et fructose est négligée, tous les sucres sont regroupés et traités comme une seule espèce chimique (S).

Dans ce schème, le phosphate est accumulé par les cellules en phosphate intracellulaire comme le propose le modèle de Curtis et al. (1991). La forme métabolique de ce phosphate n'est pas imposée. Le nitrate et l'ammoniac sont absorbés en nitrate et en ammoniac intracellulaires. Les sucres sont également absorbés en sucres intracellulaires.

Deux coefficients stoechiométriques peuvent être déterminés pour le sucre, un pour chacune des réactions de production de biomasse si les besoins énergétiques des deux réactions sont significativement différents.

Ce modèle à 6 équations peut être simplifié en supposant qu'il n'y a pas d'accumulation d'azote ou de sucre dans la cellule ou alors en supposant que ces réserves sont petites et constantes et peuvent être négligées, ce qui est équivalent pour la modélisation. Dans ce cas, on peut éliminer les équations d'absorption correspondantes et corriger l'équation de formation de biomasse.

$$\begin{aligned} P_e &\xrightarrow{r_1} P_i \\ s_2S_e + pP_i + nN_e &\xrightarrow{r_2} X \\ s_3S_e + pP_i + mM_e &\xrightarrow{r_3} X \end{aligned} \tag{5-14}$$

Ce schème peut être représenté sous forme de bilan. On doit tenir compte de l'élimination du phosphate intracellulaire avec la biomasse lors de l'échantillonnage.

$$\frac{d}{dt}\begin{bmatrix} XV \\ SV \\ NV \\ MV \\ P_eV \\ P_iXV \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -s_2 & -s_3 \\ 0 & -n & 0 \\ 0 & 0 & -m \\ -1 & 0 & 0 \\ 1 & -p & -p \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} + \begin{bmatrix} -F^{out}X \\ +F_2^{in}S^{in} - F^{out}S \\ +F_1^{in}N^{in} - F^{out}N \\ -F^{out}M \\ +F_1^{in}P^{in} - F^{out}P_e \\ -P_iF^{out}X \end{bmatrix} \tag{5-15}$$

L'équation 5-16 présente ce schème sous la forme d'un système d'état à l'aide des changements de variables utilisés pour le premier schème et étendus aux nouvelles variables:

$$\begin{bmatrix} \Delta X \\ \Delta S \\ \Delta N \\ \Delta M \\ \Delta P_e \\ \Delta P_iX \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -s_2 & -s_3 \\ 0 & -n & 0 \\ 0 & 0 & -m \\ -1 & 0 & 0 \\ 1 & -p & -p \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} \tag{5-16}$$

La surdétermination de ce système matriciel nous permet d'écrire plusieurs estimateurs de biomasse utilisant différentes combinaisons de variables mesurées.

$$\Delta X = k_N \Delta N + k_M \Delta M \tag{5-17}$$

avec k_N=-1/n, kM=-1/m,

$$\Delta X = k_N \Delta N + k_S \Delta S \tag{5-18}$$

avec k_N=-1/n+s_2/s_3n et k_S=-1/s_3,

$$\Delta X = k_M \Delta M + k_N \Delta N + k_S \Delta S \tag{5-19}$$

avec k_M=-1/m, k_N=s_2/s_3n et k_S=-1/s_3,

$$\Delta X = k_P(\Delta P_i X + \Delta P_e) \tag{5-20}$$

avec $k_P=-1/p$,

$$\Delta X = k_P(\Delta P_i X + \Delta P_e) + k_N \Delta N + k_S \Delta S \tag{5-21}$$

avec $k_P=-1/p$, $k_N=1/n+s_2/s_3n$ et $k_S=-1/s_3$.

Le Tableau 5-1 présente ces équations sous une forme plus compacte. D'autres combinaisons sont possibles. De fait, il se trouve que la plupart des combinaisons linéaires à 2, 3 ou 4 termes sont valables pour ce schème. Cependant, les coefficients des schèmes correspondants ne sont pas toujours idenfiables. Par exemple, dans le sys-tème linéaire présenté à l'équation 5-21, on peut établir explicitement les valeurs de p et s_3 mais les valeurs de "n" et "s_2" sont liées. Ainsi $p=-1/k_P$, $s_3=-1/k_S$ mais $s_2=(1-n*k_N)/k_S$.

Tableau 5-1 Définition des régresseurs linéaires pour le troisième schème réactionnel

Équations	**Interprétation des coefficients linéaires**			
	k_S	k_M	k_N	k_P
$\Delta X = k_N \Delta N + k_M \Delta M$		-1/m	-1/n	
$\Delta X = k_N \Delta N + k_S \Delta S$	$-1/s_3$		$-1/n+s_2/s_3n$	
$\Delta X = k_M \Delta M + k_N \Delta N + k_S \Delta S$	$-1/s_3$	-1/m	s_2/s_3n	
$\Delta X = k_P(\Delta P_i X + \Delta P_e)$				-1/p
$\Delta X = k_P(\Delta P_i X + \Delta P_e) + k_N \Delta N + k_S \Delta S$	$-1/s_3$		$1/n+s_2/s_3n$	-1/p

5.4 Quatrième schème réactionnel

Ce dernier schème reprends le schème précédent mais suppose une utilisation indifférenciée des deux sources d'azote en traitant ΔN et ΔM comme un seul paramètre. Ce bilan n'empêche aucunement une utilisation préférentielle de l'un ou de l'autre de ces nutriments. Ce nouveau schème présenté à l'équation 5-22 peut être traduit par le bilan de masse intégré de l'équation 5-23 à l'aide de la méthode développée précédemment.

$$\begin{array}{r} P_e \rightarrow P_i \\ sS_e + pP_i + n(N_e + M_e) \rightarrow X \end{array} \tag{5-22}$$

$$\begin{bmatrix} \Delta X \\ \Delta S \\ \Delta N \\ \Delta M \\ \Delta P_e \\ \Delta P_i X \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -s_2 \\ 0 & -n \\ 0 & -n \\ -1 & 0 \\ 1 & -p \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \tag{5-23}$$

Les solutions qu'on peut obtenir sont celles présentées au 3ieme schème en posant que le coefficient "m" est égal au coefficient "n". Par exemple le schème présenté à l'équation 5-17 devient l'équation 5-24.

$$\Delta X = k_N(\Delta N + \Delta M) \tag{5-24}$$

avec k_N=-1/n

Des schèmes réactionnels plus complexes tels ceux adaptés des modèles de croissance proposés par Hooker et al. (1992) et van Gulik et al. (1993) peuvent être évalués mais les résultats du chapitre 6 indiquent qu'il est impossible d'obtenir des valeurs significati-

ves pour les coefficients. Ainsi, seuls des modèles basés sur des paramètres mesurés ou calculés par bilan de masse sont retenus.

Le Tableau 5-2 présente la synthèse des régresseurs linéaires rattachés à leurs schèmes réactionnels respectifs.

Tableau 5-2 Définition des régresseurs linéaires pour les schèmes réactionnels

Schèmes	Équations	Interprétation des coefficients linéaires				
		k_S	k_M	k_N	k_P	k_{XV}
$sS \xrightarrow{r_1} X$	$\Delta X(t) = k_S \Delta S(t)$	-1/s				
$s_1 S \xrightarrow{r_1} X$	$\Delta X(t) = k_S \Delta S(t)$	-1/s				$-s_2k_2/s_1$
$s_2 S + X \xrightarrow{r_2} X$	$+k_{XV} XV(t)$					
$P_e \xrightarrow{r_1} P_i$	$\Delta X = k_N \Delta N + k_M \Delta M$		-1/m	-1/n		
	$\Delta X = k_N \Delta N + k_S \Delta S$	$-1/s_3$		$-1/n+s_2/s_3n$		
$s_2 S_e + pP_i + nN_e \xrightarrow{r_2} X$	$\Delta X = k_M \Delta M + k_N \Delta N$	$-1/s_3$	-1/m	s_2/s_3n		
	$+k_S \Delta S$				-1/p	
$s_3 S_e + pP_i + mM_e \xrightarrow{r_3} X$	$\Delta X = k_P(\Delta P_i X + \Delta P_e)$	$-1/s_3$		$1/n+s_2/s_3n$	-1/p	
	$\Delta X = k_P(\Delta P_i X + \Delta P_e)$					
	$+k_N \Delta N + k_S \Delta S$					
$P_e \rightarrow P_i$	$\Delta X = k_N(\Delta N + \Delta M)$			-1/n		
$sS_e + pP_i + n(N_e + M_e) \rightarrow X$						

5.5 Estimation par les moindres carrés

L'estimé $\hat{k}_s$ de k_s par la méthode des moindres carrés est obtenu en minimisant le carré de l'erreur entre $\Delta X(t)$ et l'équation du modèle. Dans le cas du premier schème,

$$\hat{k}_s = \arg\min_{k_s} \frac{1}{N}\sum_{k=1}^{N}\left[\Delta X(t_k) - (k_s \Delta S(t_k))\right]^2 \tag{5-25}$$

Dans ce cas, $\hat{k}_s$ est soumis à la contrainte $\hat{k}_s \leq 0$ puisque "s" est positif. $\Delta X(t_k)$ et $\Delta S(t_k)$ sont calculés à partir des valeurs mesurées de l'échantillon pris au temps t_k. N correspond au nombre d'échantillons traités.

Une solution algébrique de l'équation 5-25 est possible car les équations algébriques permettant la résolution de systèmes linéaires sont simples et bien connues. La résolution de systèmes multivariables est généralement réalisée sous forme matricielle car il devient très complexe de traduire algébriquement le système d'équations linéaires lorsque le nombre d'équations est supérieur à 4.

La solution matricielle du premier schème est donnée par l'équation 5-26. La démonstration est présentée à l'annexe B.3.

$$\hat{k}_s = (\Delta S' \Delta S)^{-1} \Delta S' \Delta X \tag{5-26}$$

5.6 Notions sur la validation des modèles

Toutes les notions de probabilité et de statistiques utilisées dans le présent travail sont tirées de Drapeau et Smith (1981) et de Martel et Nadeau (1988).

5.6.1 Erreur résiduelle et coefficient de détermination

La seule valeur de l'erreur résiduelle (SCE, équation 5-28) calculée par la méthode des moindres carrés n'est pas suffisante pour comparer des modèles de complexité différente. En effet, l'augmentation du nombre de régresseurs entraîne automatiquement une diminution de l'erreur résiduelle et ne permet pas de détecter une surparamétrisation éventuelle. Une alternative est l'utilisation du coefficient de détermination multiple empirique qui mesure la proportion de la variation totale qui est expliquée par la régression.

$$R^2 = 1 - \frac{SCE}{SCT} \tag{5-27}$$

où SCE est la somme des carrés des erreurs telle que définie par

$$SCE = \sum e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{5-28}$$

et SCT est la somme des carrés totale

$$SCT = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{5-29}$$

À l'instar de l'erreur résiduelle qui diminue avec le nombre de régresseurs, ce coefficient R^2 présente le défaut d'augmenter avec le nombre de paramètres explicatifs (régresseurs).

5.6.2 Coefficient de détermination multiple empirique ajusté

D'autres critères permettent de compenser la diminution de l'erreur résiduelle en tenant compte du nombre de paramètres. Le coefficient de détermination multiple empirique

ajusté (l'acronyme CDMEA est employé par la suite pour alléger le texte) réalise cette fonction. Le CDMEA (également identifié par le symbole $\overline{R}^2$), dont la signification est semblable à celle du coefficient de détermination pour les régressions linéaires simples, est défini par l'équation 5-30

$$\text{CDMEA} = 1 - \frac{\text{SCE} / (n - p - r)}{\text{SCT} / (n - r)} \tag{5-30}$$

où n, p et r sont respectivement le nombre de points, le nombre de régresseurs et le nombre de réalisations indépendantes (nombre de cultures).

Ce critère impose donc une pénalité croissante avec le nombre p de régresseurs et passe par un maximum qui donne une indication sur le nombre optimal de variables explicatives à inclure dans le modèle de régression. Ce dernier critère sera employé comme critère de sélection.

5.6.3 Validation croisée

Une autre méthode pour détecter la surparamétrisation consiste à appliquer les modèles obtenus à des données n'ayant pas servi au calcul d'estimation des paramètres. Pour chaque catégorie de cultures qui est définie au chapitre suivant, certaines cultures qui ne servent pas à la régression sont réservées pour la validation croisée. Les CDMEA de ces dernières sont comparés aux CDMEA des cultures ayant servi au calcul des paramètres. Un CDMEA beaucoup plus faible pour la validation que pour la régression indique une surparamétrisation.

5.6.4 Variance, intervalles de confiance et régresseurs significatifs

Les deux critères présentés précédemment permettent de discriminer entre différents modèles de complexité variable.

L'objectif de cette étape de modélisation consiste non seulement à modéliser la croissance de la biomasse mais également à déterminer les coefficients pseudo-stoechiométriques des schèmes réactionnels qui sous-tendent les régressions. La précision des coefficients est fonction de l'incertitude associée à chacun. La meilleure régression peut ne pas fournir d'information valable sur un modèle auquel elle peut être assimilée si la précision des paramètres est mauvaise.

Une façon de présenter cette précision consiste à calculer un intervalle de confiance autour de chacun de ces régresseurs. Le calcul de l'intervalle de confiance pour un niveau de confiance (1-α) est basé sur une distribution de Student à (n-p-r) degrés de liberté et sur la variance estimée de chacun des régresseurs.

Cet intervalle, au niveau de confiance (1-α), prend la forme

$$\left[b_j - t_{\alpha/2}s(b_j), b_j + t_{\alpha/2}s(b_j)\right] \tag{5-31}$$

où $t_{\alpha/2}$ est la distribution de Student à (n-p-r) degrés de liberté et $s(b_j)$ est défini par l'estimateur de la variance du coefficient de régression empirique b_j

$$s^2(b_j) = S_e^2 c_{jj} \tag{5-32}$$

avec l'estimateur de la variance de l'erreur

$$S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{(n-p-r)} \tag{5-33}$$

et c_{jj} qui est un élément de la diagonale de la matrice définie par

$$[X'X]^{-1} \tag{5-34}$$

Si la valeur "zéro" est comprise dans l'intervalle, on peut conclure que, au niveau de signification α, le régresseur n'explique pas la variation de biomasse de façon significative. Par conséquent les coefficients stoechiométriques implicites derrière ce paramètre ne peuvent être calculés avec la moindre précision. Cela ne signifie pas que le modèle est faux mais seulement que l'analyse des données ne permet pas de calculer de façon satisfaisante les paramètres du modèle.

Ces critères de validation complètent ce chapitre. Dans le chapitre suivant, les régressions linéaires sont appliquées aux données expérimentales afin de déterminer les valeurs des coefficients pseudo-stoechiométriques des divers schèmes réactionnels.

6. Application des schèmes réactionnels aux données expérimentales

Dans ce chapitre, des régressions sont effectuées afin d'obtenir, à partir des données expérimentales, les paramètres des modèles stoechiométriques définis au chapitre précédent. Par la suite ces régressions sont évaluées par rapport à la qualité de la prédiction de biomasse qu'elles offrent.

6.1 Résultats expérimentaux

Les principales données expérimentales utiles pour ce travail sont présentées à l'annexe A. Le Tableau 6-1 résume les conditions expérimentales de chacune des cultures traitées. Les cultures qui portent le même numéro de séquence ont été réalisées simultanément. Souvent les cultures en flacons agités sont les témoins de cultures en bioréacteurs

On constate que la production de biomasse varie grandement d'une culture à l'autre. Cette variabilité n'est pas liée au mode de cultures sauf peut-être par le fait que les cultures en flacons agités cultivées en cuvée simple ont une production qui ne dépasse pas 8 g/l. On retrouve des cultures avec une grande production tant chez les cultures flacons agités (F43, F44) qu'en bioréacteur (RH7). Les résultats les plus faibles se retrouvent également en flacons (F33, F34) et en bioréacteur (H527, H529). Cependant certains résultats faibles proviennent de cultures en bioréacteur élicitées avant d'atteindre la phase de stagnation et une concentration en biomasse maximale. Les données post-

La section suivante utilise les données expérimentales de ces cultures pour évaluer les différents schèmes réactionnels proposés au chapitre précédent. Les résultats des régressions linéaires sont présentés par catégories. Ils sont ensuite discutés par rapport aux tests statistiques présentés au chapitre 5.

Tableau 6-1 Portrait de l'ensemble des cultures analysées

Nom de la culture	Type de bio-réacteur	Type d'alimentation	Biomasse maximale	No de séquence
F18	flacon 500 ml	cuvée simple	8 g/l	1
F19	flacon 500 ml	cuvée simple	8 g/l	1
F20	flacon 500 ml	cuvée simple	8 g/l	1
F22	flacon 500 ml	cuvée simple	8 g/l	2
F33	flacon 500 ml	cuvée avec ajouts ponctuels	5 g/l	3
F34	flacon 500 ml	cuvée avec ajouts ponctuels	6 g/l	3
F43	flacon 500 ml	cuvée avec ajouts ponctuels	13 g/l	4
F44	flacon 500 ml	cuvée avec ajouts ponctuels	14 g/l	4
F48	flacon 500 ml	cuvée simple	6 g/l	5
H527	réacteur HRI 6l	cuvée avec ajouts ponctuels	6 g/l	6
H530	réacteur HRI 6 l	cuvée avec ajouts ponctuels	10 g/l	5
H532	réacteur HRI 6 l	cuvée avec ajouts ponctuels	8 g/l	9
RH7	réacteur HRI 6 l	cuvée avec ajouts graduels	14 g/l	10
RH18	réacteur HRI 6 l	cuvée avec ajouts graduels	10 g/l	11

6.2 Présentation des résultats par catégories

6.2.1 Description des catégories

Quatre catégories ont été définies pour l'analyse des résultats en fonction du type de culture. Une ou deux cultures sont isolées et servent seulement à la validation (-v). Le Tableau 6-2 (p. 65) présente la distribution des cultures parmi les catégories utilisées. À l'intérieur de chaque catégorie, la plupart des cultures sont utilisées pour la régression (identifiées par l'extension -c)

La première catégorie (cs-f) regroupe les cultures en cuvée simple sans ajout. Deux autres catégories (ca-f et ca-r) représentent les cultures en cuvée avec ajout réalisées en flacon et en bioréacteur. La dernière catégorie (cag-r) comprend les cultures avec ajout graduel. Seules deux cultures sont présentes dans cette catégorie et elles sont utilisées pour les calculs. Il n'y a donc pas de culture disponible pour la validation croisée.

Divers schèmes réactionnels ont été présentés aux sections 5.1 à 5.4 pour déterminer les coefficients des équations linéaires obtenues et les coefficients pseudo-stoechiométriques associés lorsqu'ils sont identifiables. Cependant, pour faciliter l'analyse, les régressions sont classées en fonction du nombre de paramètres plutôt qu'en fonction de l'ordre de présentation des schèmes. De plus, certaines des régressions présentées ne sont associées à aucun schème en particulier. Le choix des paramètres pour les régressions linéaires des différents modèles est présenté au Tableau 6-3.

6.2.2 Résultats de régression pour un modèle et une catégorie de culture.

Des régressions linéaires multiples ont été réalisées sur les modèles présentés au Tableau 6-3. La Figure 6-1 présente un résultat typique. Il s'agit du modèle 2d (sucre et nitrate) appliqué aux cultures de la catégorie ca-r (H527, H530 et H532). Les points représentent les valeurs expérimentales de biomasse sèche. Les traits correspondent aux estimations de biomasse à partir des valeurs mesurées de sucre et de nitrate. On constate qu'on obtient une estimation correcte de la biomasse à partir du suivi des concentrations en sucre et en nitrate dans le milieu. La Figure 6-1 montre ces mêmes paramètres de régression appliqués à la culture de la même catégorie (H532) qui n'a pas été utilisée pour le calcul. On voit que dans ce cas la qualité de la régression est bonne. La validation croisée suggère que la régression choisie demeure valable sur d'autres cultures du même type et donc qu'il n'y a apparemment pas surparamétrisation des données.

La qualité de la régression est quantifiée par le coefficient de détermination multiple empirique ajusté, CDMEA, défini à la section 5.6.2. Les CDMEA du modèle 2d calculé par la régression et par la validation sur les cultures de la catégorie ca-r sont près de l'unité, soit respectivement 0,95 et 0,97.

6.2.3 Discussions sur les erreurs

Les mêmes régressions ont été appliquées aux diverses catégories et les CDMEA par modèle et par catégorie sont présentées à la Figure 6-2. Les coefficients de détermination multiples ajustés obtenus des cultures de validation (en gris) sont juxtaposés aux CDMEA des régressions ayant servi à leur calcul (en noir). Pour chaque modèle, la première colonne représente les coefficients ajustés pour les sous-blocs utilisés pour la

régression; la seconde colonne donne les valeurs des coefficients pour les sous-blocs de validation.

En observant ces résultats, on peut éliminer les modèles qui présentent de moins bons résultats afin de ne conserver que les plus performants. Cette analyse est réalisée en appliquant les critères développés précédemment.

Dans un premier temps, on constate que les modèles à un seul terme ont un CDMEA faible. Ceci est particulièrement vrai pour le groupe ca-f. Seul le modèle 1c y parvient mieux puisqu'il s'agit du paramètre définit pour calculer le coefficient de maintenance et il est construit à partir des données de biomasse. Il ne s'agit pas d'un modèle correspondant aux schèmes réactionnels présentés au chapitre 5.

Parmi les modèles à deux régresseurs, 2a, 2b, 2d donnent des coefficients supérieurs à 0,7 pour tous les groupes. Les modèles 2c, 2e, 2f sont faibles pour au moins un groupe. Les modèles à 3 paramètres donnent presque tous des coefficients de détermination ajustés supérieurs à 0,7 sur les cultures utilisées pour les calculs. Les modèles 4a et 4c sont également corrects pour les sous-blocs utilisés pour la régression.

Parmi les modèles retenus jusqu'à maintenant, les modèles 3b, 4a et 4c offrent une performance moindre sur les sous-blocs de validation et sont rejetés. Il apparaît donc que les modèles à 4 paramètres sont surparamétrés par rapport aux données disponibles.

A ce point de l'analyse, les modèles 1c, 2a, 2b, 2d, 3a et 3c sont retenus comme satisfaisant les critères de CDMEA et de validation croisée.

6.2.4 Validation des coefficients pseudo-stoechiométriques

Les équations de régression de chacun des modèles retenus pour les 4 catégories sont présentées au Tableau 6-4.

Les coefficients marqués en gris incluent la valeur "0" dans l'intervalle de confiance de 95 %. On peut donc conclure que les variables explicatives de ces coefficients n'affectent pas l'estimation de la biomasse de façon significative à ce niveau de confiance pour les données utilisées. Divers facteurs peuvent expliquer ces résultats. Il peut s'agir de la variabilité des différentes cultures, même sous des conditions d'opération semblable. Cette hypothèse est plausible puisqu'on peut voir à l'annexe A que les courbes de croissance et les rendements en biomasse sont très différentes d'une culture à l'autre.

Ceci peut également être dû à des mesures expérimentales très bruitées. Cependant, ce ne semble pas être le cas puisque les courbes de données préparées pour les régressions sont relativement lisses. La raison la plus plausible semble être le fort coefficient de corrélation reliant les paramètres. Le Tableau 6-5 présente les coefficients de corrélation pour les paramètres principaux. Les quatre premières lignes et colonnes montrent une forte corrélation (valeurs voisines de 1 ou -1) entre les diverses variables mesurées pour la culture RH7. Les coefficients négatifs indiquent une pente de corrélation négative. C'est le cas pour les mesures de sucre qui diminuent lorsque la biomasse augmente.

Cette très forte corrélation s'explique par le fait que les variables sont presque colinéaires. Cette dépendance entre les variables va à l'encontre d'une hypothèse de base des régressions multiples soit que les équations soient linéairement indépendantes et donc

qu'il n'y ait pas de colinéarité exacte entre les variables indépendantes (Martel, 1988). Cette presque colinéarité se traduit par une grande sensibilité au bruit, ce qui conduit à des coefficients de régression dont les valeurs ne se recoupent pas d'une catégorie à l'autre.

On peut conclure du Tableau 6-5 que chaque variable peut expliquer toutes les autres. Au niveau cellulaire, cela signifie que ces nutriments sont absorbés au même rythme et que, par conséquent, les données ne permettent pas de distinguer des modes de croissance plus élaborés. Cela justifie la décision de ne pas évaluer les schèmes complexes de van Gulik ou Hooker.

Le Tableau 6-6 résume les caractéristiques des six équations retenues. Les paramètres les plus utilisés pour l'estimation de la biomasse sont le sucre (Se) et le produit XD*V qui est associé à la maintenance. Viennent ensuite le nitrate (Me) et l'ammoniac total (Ne+Me). Le schème qui représente le mieux les six équations est le schème numéro quatre auquel on ajoute une équation de maintenance et on enlève l'équation associée à l'assimilation du phosphate. En effet, le phosphate n'est pas présent parmi les modèles stoechiométriques retenus et le paramètre XV qui correspond au terme de maintenance est présent dans quatre des six modèles.. Ce nouveau schème est présenté à l'équation 6-1.

$$\begin{aligned} s_1 S_e + n(N_e + M_e) &\rightarrow X \\ s_2 S_e + X &\rightarrow X \end{aligned} \qquad (6\text{-}1)$$

Les modèles pouvant être définis à partir de ce schème sont les modèles 1a, 1b, 2f et 2a,

2b et 3c respectivement. De la liste précédente, les modèles 2a, 2b et 3c se retrouvent parmi les modèles ayant réussi les étapes préliminaires de sélection.

Pour ce qui a trait à la valeur des coefficients pseudo stoechiométriques, comme on peut le constater au Tableau 6-4, pour ces modèles la variance des coefficients ne permet pas de déterminer avec précision la valeur des coefficients pseudo-stoechiométriques. Un schème plausible a été déterminé mais on ne peut en déterminer les coefficients pseudo-stoechiométriques. D'autres modèles sont probablement également valables mais les données et le comportement des cultures ne permettent pas de les valider.

Tableau 6-2 Répartition des cultures parmi les catégories

Catégorie	Type de culture	Calcul ou validation	Cultures associées
cs-f-c	cuvée sans ajout, flacons	calcul	F18, F19, F22
cs-f-v	cuvée sans ajout, flacons	validation	F20, F48
ca-f-c	cuvée alimentée, flacons	calcul	F33, F34, F43
ca-f-v	cuvée alimentée, bioréacteurs et flacons	validation	F44
ca-r-c	cuvée alimentée, bioréacteurs	calcul	H527, H530
ca-r-v	cuvée alimentée, bioréacteurs	validation	H532
cag-r-c	cuvée alimentée graduelle	calcul	RH7, RH18
cag-r-v	cuvée alimentée graduelle	validation	aucune

Tableau 6-3 Paramètres de régression linéaire pour les différents modèles

Modèle	Se	Me	Ne	(Ne+ Me)	Pe	Pi	XD*V	Schème de référence
1a	X							1
1b				X				4
1c							X	
2a	X						X	2
2b				X			X	4[1]
2c		X	X					3
2d	X	X						
2e	X		X					3
2f	X			X				4
3a	X	X	X					3
3b	X				X	X		
3c	X			X			X	4[1]
4a	X	X			X	X		3
4b	X			X	X	X		4
4c	X	X	X				X	3[2]

[1] *Quatrième schème auquel on a ajouté un terme de maintien*
[2] *Troisième schème auquel on a ajouté un terme de maintien*

Tableau 6-4 Valeur des régresseurs des modèles retenus pour chacune des 4 catégories

Modèle	Catégorie	Équation (incluant les intervalles de confiance)					
1c	cs-f	0.87 ● 0.08					
	ca-f	0.79 ± 0.14	*XDV				
	ca-r	1.17 ± 0.22					
	cag-r	1.55 ± 0.19					
2a	cs-f	-0.18 ● 0.12		0.60 ± 0.19			
	ca-f	-0.06 ± 0.09	*Se	0.68 ± 0.21	*XDV		
	ca-r	-0.24 ± 0.26		0.53 ± 0.71			
	cag-r	-0.40 ± 0.13.		0.31 ± 0.43			
2b	cs-f	-0.01 ± 0.05		0.84 ± 0.16			
	ca-f	-0.03 ± 0.03		0.61 ± 0.25			
	ca-r	-0.17 ± 0.14	* (Me+Ne)	0.29 ± 0.76	* XDV		
	cag-r	-0.24 ± 0.17		0.18 ± 0.99			
2d	cs-f	-0.55 ± 0.17		0.02 ± 0.10			
	ca-f	0.25 ± 0.13	* Se	-0.61 ± 0.14	* Me		
	ca-r	-0.11 ± 0.30		-0.36 ± 0.33			
	cag-r	-0.37 ± 0.14		0.14 ± 0.17			
3a	cs-f	-0.56 ● 0.17		0.04 ± 0.11		-0.17 ± 0.43	
	ca-f	0.27 ± 0.11	* Se	-0.56 ± 0.12	* Me	-0.03 ● 0.02	* Ne
	ca-r	-0.09 ± 0.30		-0.28 ± 0.37		-0.10 ± 0.18	
	cag-r	-0.34 ± 0.12		0.30 ± 0.32		-0.43 ± 0.28	
3c	cs-f	-0.22 ± 0.13		-0.22 ± 0.13		0.61 ± 0.19	
	ca-f	-0.03 ± 0.11	* Se	-0.02 ± 0.04	* (Me+Ne)	0.59 ± 0.26	* XDV
	ca-r	-0.14 ± 0.27		-0.13 ± 0.16		0.14 ± 0.81	
	cag-r	-0.34 ± 0.14		-0.11 ± 0.12		-0.14 ± 0.64	

Tableau 6-5 Coefficients de corrélation "r" entre les paramètres de la culture RH7

	Se	Me	Ne	XDV	XD
Se	1	0.9575	0.979	-0.9798	-0.9953
Me	0.9575	1	0.9928	-0.9473	-0.9514
Ne	0.979	0.9928	1	-0.9599	-0.9786
XDV	-0.9798	-0.9473	-0.9599	1	0.971
XD	-0.9953	-0.9514	-0.9786	0.971	1

Tableau 6-6 Paramètres de régression linéaire pour les différents modèles retenus

Modèle	Se	Me	Ne	(Ne+ Me)	Pe	Pi	XD*V	Schème de référence
1c							X	
2a	X						X	2
2b				X			X	4[1]
2d	X	X						
3a	X	X	X					3
3c	X			X			X	4[1]

[1] *Quatrième schème auquel on a ajouté un terme de maintien*
[2] *Troisième schème auquel on a ajouté un terme de maintien*

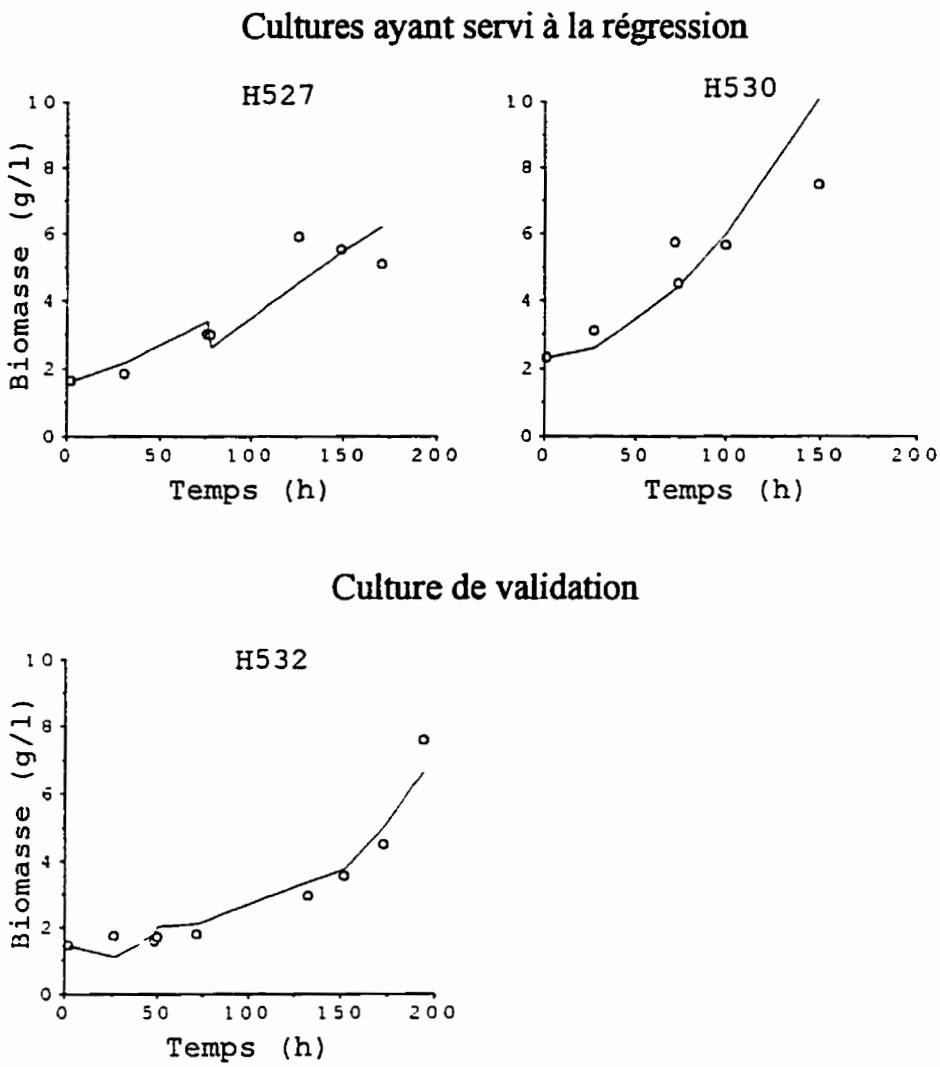


Figure 6-1 Modèle 2d, Estimation de la biomasse à partir de paramètres mesurés (0): Données expérimentales; (-): valeurs calculées.

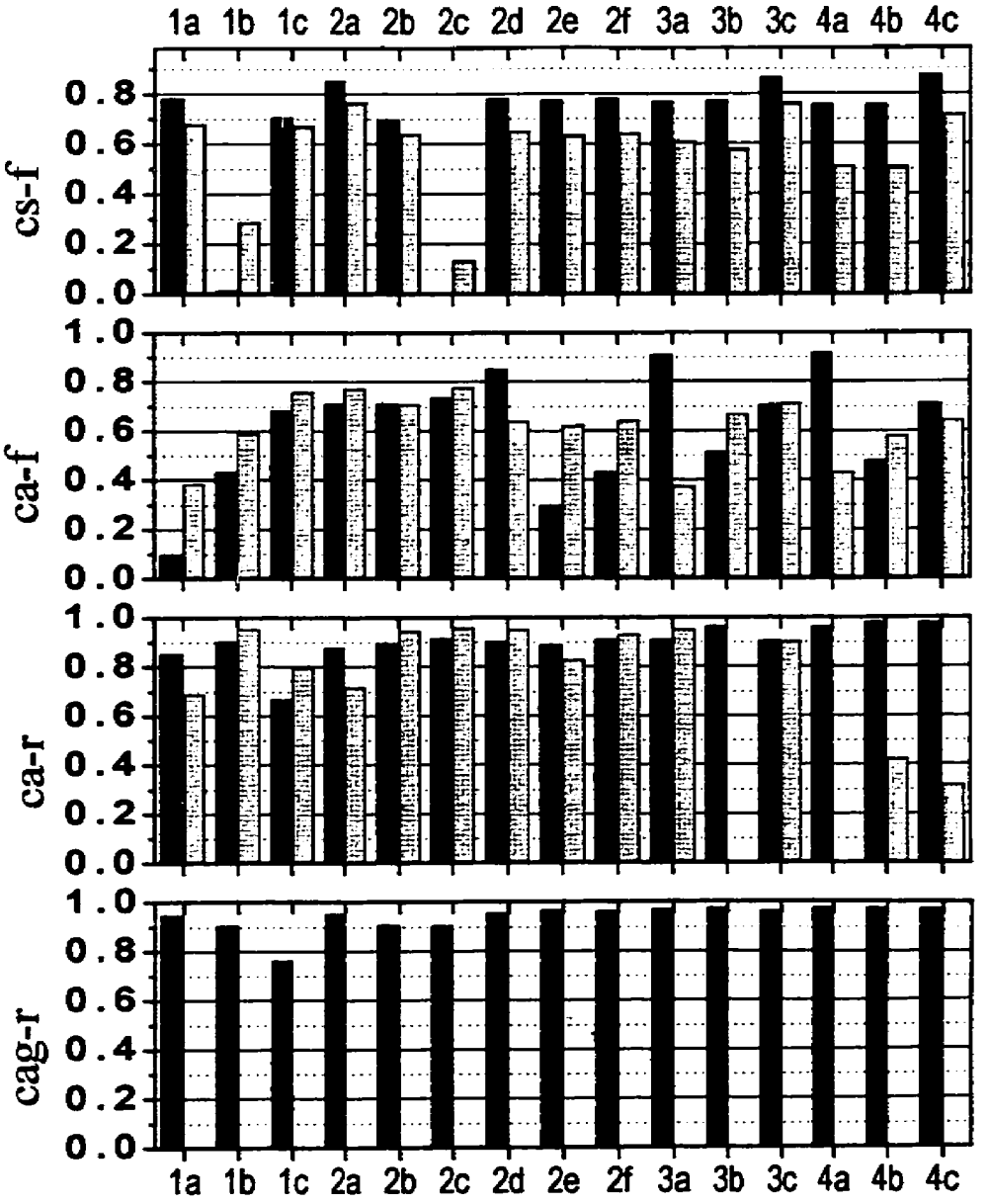


Figure 6-2 Coefficient de détermination multiple empirique ajusté (CDMEA) pour chacune des quatre catégories, en fonction des modèles
(■): Cultures utilisées pour la régression, (■): Cultures de validation croisée

7. Modèles polynômiaux de croissance de la biomasse

L'originalité de la démarche proposée par Chotteau et employée ici est l'utilisation de modèles cinétiques polynômiaux plutôt que de modèles plus conventionnels souvent basés sur le modèle présenté par Monod.

La structure générale d'un modèle polynômial décrivant le taux de croissance est

$$r_x(t) = f(\xi_1, \xi_2) = p_o + p_{10}\xi_1(t) + p_{01}\xi_2(t) + \ldots + p_{20}\xi_1^2(t) + p_{11}\xi_1\xi_2 + \ldots \qquad (7\text{-}1)$$

Une équation similaire peut décrire le taux de croissance spécifique (μ) :

$$\mu(t) = g(\xi_1, \xi_2) = p_o + p_{10}\xi_1(t) + p_{01}\xi_2(t) + \ldots + p_{20}\xi_1^2(t) + p_{11}\xi_1\xi_2 + \ldots \qquad (7\text{-}2)$$

où p_{ij} est une constante ou un coefficient des termes du polynôme, et ξ_i représente un paramètre mesuré ou calculé pouvant influencer le taux de croissance.

Contrairement à ce qui est proposé par Chotteau, la modélisation s'appliquera ici au taux de croissance spécifique "μ" pour des raisons qui seront détaillées à la section 7.1.2.

7.1 Modèles conventionnels

7.1.1 Comparaison entre les modèles cinétiques conventionnels et polynômiaux

La littérature recensée présente généralement un seul modèle pour lequel les paramètres sont calculés. Lorsqu'ils traitent de cinétique, les auteurs posent une structure rationnelle pour laquelle ils déterminent les coefficients. La forme de loin la plus populaire est la structure de Monod (équation 2-8) définie par une variable (un substrat limitant) et par deux paramètres à déterminer : le taux de croissance spécifique maximal (μ_{max}) et la constante de saturation (K_s) qui indique à quelle concentration la cinétique

de la valeur maximale.

Cette forme à deux paramètres fut initialement proposée par Michaelis et Menten en 1913 pour exprimer le taux de réaction de réactions enzymatiques, avec un seul substrat. Monod l'a appliqué à la croissance de micro-organismes sans aucune justification biologique. En fait, comme le mentionnait Monod lui-même en 1942, "toute courbe d'allure sigmoïde pourrait être ajustée aux données expérimentales".

D'autres formes ont été proposées dont certaines, comme le modèle de Haldane, permettent de tenir compte d'une inhibition de la croissance en présence d'une forte concentration d'un nutriment.

En cultures de micro-organismes, l'effet sur la cinétique de plusieurs paramètres est généralement exprimé par un produit d'équations de Monod ou d'équations similaires. On obtient alors plusieurs paramètres à déterminer simultanément. Plusieurs problèmes se posent alors quant au calcul de la valeur des paramètres. En admettant que l'équation choisie représente effectivement bien la situation réelle, il faut expérimenter sur une plage de concentrations de nutriments qui couvre toutes les zones importantes. Dans le cas du modèle de Monod, les concentrations observées doivent couvrir tant la zone linéaire que la zone d'ordre 0. Ces plages d'opération ne peuvent généralement être étudiées que lors d'expériences spécifiques.

Par exemple, il est très difficile de caractériser l'effet de la concentration du sucre sur le taux de croissance pour des cultures en cuvée. En effet, la concentration en sucre est

élevée durant presque toute la durée de l'expérimentation et diminue très rapidement en fin de culture lorsque la biomasse est maximale: cela ne permet d'obtenir qu'un petit nombre de points de mesure dans cette zone à faible concentration et l'évaluation des paramètres est d'autant plus incertaine.

D'autre part, si la relation réelle entre la concentration d'un nutriment et le taux de croissance n'est pas d'allure sigmoïdale, des relations permettant une meilleure prédiction du taux de croissance sont ignorées au détriment de la précision du modèle.

La complexité des algorithmes de résolution itératifs de modèles non linéaires est telle qu'il est coûteux en puissance de calcul d'évaluer un grand nombre de modèles pour chacun des paramètres étudiés. De plus, l'étude de la variance et des intervalles de confiance est complexe et approximative dû à la nature non-linéaire de ces modèles (Draper, 1981).

L'utilisation de modèles polynômiaux empiriques permet, grâce à la simplicité de résolution de systèmes d'équations multilinéaires, d'évaluer un très grand nombre de possibilités. Cette méthode sera présentée un peu plus loin dans ce chapitre.

Cette méthode possède cependant les inconvénients des méthodes de lissage utilisant les polynômes de régression. C'est-à-dire que la régression est valable seulement sur la plage étudiée si l'ordre des polynômes est petit par rapport au nombre de points sur lesquels s'applique la régression et si les points étudiés sont bien répartis dans la plage. On peut parfois obtenir des modèles prédisant une croissance négative sous certaines condi-

tions. Ces abérations mathématiques se produisent dans une zone qui ne fait pas partie de la plage expérimentale où se concentre la majorité des données et elles peuvent être considérées commes des extrapolations abusives des données.

Il faut noter que le modèle de Monod fut développé à partir de cultures de micro-organismes en cuvée. Pour ce type de culture, l'arrêt de la croissance correspond à l'épuisement du substrat limitant. De cela découle qu'une concentration en substrat nulle implique une croissance nulle. Des conditions de cuvée alimentée par rétroaction ou de profils exponentiels d'alimentation permettent d'étudier avec plus de finesse les plages de faible concentration et d'observer des comportements différents. D'ailleurs, dans un chemostat stabilisé où tout le substrat est consommé au fur et à mesure de son alimentation, la croissance de la biomasse est forte malgré une concentration en substrat non détectable.

Les modèles conventionnels ne détectent que les tendances présentes dans la structure proposée. Les modèles polynômiaux permettent de détecter des relations complexes entre divers paramètres. L'analyse de ces tendances peut éventuellement servir à construire *a posteriori* des modèles phénoménologiques plus complexes.

Les modèles polynômiaux ont donc des avantages et des inconvénients différents des modèles conventionnels. Il s'agit d'un outil de modélisation qui est appliqué pour la première fois à la culture de cellules de plantes.

7.1.2 Modifications apportées à la méthodologie de Chotteau

Le modèle polynômial proposé par Chotteau est basé sur l'estimation des taux de réaction volumiques r exprimés en g ou moles par l par h ($g(l^{-1})(h^{-1})$ ou $mmol(l^{-1})(h^{-1})$). Dans le présent travail, le taux de croissance spécifique μ sera modélisé (h^{-1}). Ce choix s'explique d'une part par l'utilisation généralisée du taux de croissance spécifique en culture de cellules de plantes et d'autre part par l'objectif sous-jacent de ce travail qui est l'optimisation des conditions de croissance afin d'augmenter la productivité en métabolites secondaires. Cette optimisation vise à maintenir pour une grande période la phase de croissance exponentielle caractérisée par un taux de croissance spécifique (μ) constant maximal. Ce choix entraîne une légère augmentation de la complexité de l'algorithme de résolution par rapport à la forme proposée par Chotteau.

La seconde modification majeure est une simplification. Chotteau suggère de modéliser le taux de croissance de la biomasse ainsi que les taux de consommation des nutriments et de production de métabolites indépendamment l'un de l'autre. Il a été démontré dans le chapitre précédent qu'il existe une très forte corrélation entre la production de biomasse et la consommation de nutriments. Cette information permet de réduire la modélisation à un seul terme, soit le taux de croissance (spécifique) de la biomasse. L'évolution de la consommation des nutriments peut être calculée subséquemment à partir de la variation de la biomasse et des rendements pseudo-stoechiométriques.

Le fait de limiter la modélisation au seul taux de croissance spécifique de la biomasse a pour conséquence d'éliminer le couplage des taux de réaction par rapport à un schème

réactionnel. Ainsi, la modélisation de la croissance n'est plus basée sur des schèmes réactionnels mais sur les seuls paramètres mesurés.

Le cheminement mathématique présenté à la section suivante pour en arriver aux équations polynômiales tient compte de ces modifications.

7.2 Préparation des données pour la régression

Cette section présente la démarche mathématique permettant la construction des combinaisons de régresseurs à partir des données expérimentales. On cherche à déterminer le taux de croissance spécifique (μ) à partir des données expérimentales et d'un bilan de masse tel que présenté à l'équation 7-3. La démonstration complète est présentée à l'annexe B.4. Le Tableau 7-1 (p. 77) présente un lexique des notations et concepts utilisés dans ce chapitre et dans le suivant.

$$\frac{dX(t)V(t)}{dt} = \mu(t)X(t)V(t) - F^{OUT}(t)X(t) \tag{7-3}$$

Les modèles qui sont construits ont une structure polynômiale semblable à l'équation 7-2. En utilisant une notation matricielle, un modèle cinétique polynômial peut s'écrire:

$$\mu(t) = \Theta\Phi(t) \tag{7-4}$$

où la matrice des coefficients Θ est définie par $\Theta \equiv [\vartheta_1 \quad \cdots \quad \vartheta_v \quad \cdots \quad \vartheta_\omega]$ avec ϑ_i qui est un coefficient d'un terme du polynôme et où $\Phi(t)$, le vecteur contenant tous les monômes, est défini par $\Phi(t) \equiv [\phi_1(t) \quad \cdots \quad \phi_v(t) \quad \cdots \quad \phi_\omega(t)]'$ avec ϕ_v qui représente une combinaison de paramètres formant les termes d'un polynôme.

Les différentes familles de modèles sont obtenues en fixant à 0 la valeur de certains

coefficients de Θ. Les ϑ_i non nuls sont estimés par régression multilinéaire.

Suite à l'intégration du bilan et à l'introduction de la notation ΔX définie au chapitre 5, le bilan de masse peut s'exprimer par l'équation 7-5.

$$\Delta X = \Theta \int_0^{t_k} \Phi(\tau) X(\tau) V(\tau) d\tau \tag{7-5}$$

Après un nouveau changement de variable présenté à l'annexe B.4.2, on obtient une forme matricielle compacte du bilan de masse présentée à l'équation 7-6.

$$\Delta X(t_k) = \Theta \Psi(t_k) \tag{7-6}$$

où

$$\Psi(t_k) = \begin{bmatrix} \psi_1(t_k) \\ \vdots \\ \psi_v(t_k) \\ \vdots \\ \psi_\omega(t_k) \end{bmatrix} \tag{7-7}$$

pour lequel

$$\psi_v(t_k) = \psi_v(t_{k-1}) + \phi_v(t_k) X(t_k) V(t_k)(t_k - t_{k-1}) \tag{7-8}$$

L'équation 7-6 est une représentation matricielle d'une équation multilinéaire dont une méthode de résolution est présentée à l'annexe B.3. Le vecteur Θ qui résoud le système matriciel de l'équation 7-6 est également la solution de l'équation 7-4 et donne un modèle polynômial. Un exemple de modélisation est présenté à l'annexe B.4.3.

Tableau 7-1 Lexique des notations et définitions utilisées aux chapitres 7 et 8

Concept ou notation	Définition	Commentaire ou exemple
X	Biomasse (g/l)	Mesurée expérimentalement
ΔX	Production nette de biomasse (g)	Obtenu par intégration d'un bilan de masse et discrétisation
μ	Taux de croissance spécifique (h^{-1})	Forme polynômiale: $\mu(t) = p_o + p_{10}\xi_1(t)+...+p_{11}\xi_1\xi_2+..$ $\mu(t) = \Theta\Phi(t)$
ξ_i	Paramètre cinétique mesuré ou construit	S_e, N_e, M_e, P_e, P_i ou dN
ϕ_v	Régresseur (monôme). Produit de paramètres ξ_i	Ex.: $\phi_1=1$, $\phi_i=S_e$, $\phi_j=P_i^2$, $\phi_k=P_i*M_e^2$
Φ	Vecteur de tous les régresseurs pour les modèles d'une complexité donnée	$\Phi = \{\phi_1,...,\phi_v,...,\phi_\omega\}$
ϑ_v	Coefficient de régression à déterminer	$\Theta=[\vartheta_1\ \vartheta_2\ \vartheta_\upsilon\ \vartheta_\omega]$
Θ	Vecteur des coefficients ϑ des régresseurs (à déterminer)	La famille de modèles "n"est définie en fixant à 0 tous les coefficients sauf ceux définissant cette famille $\Theta_n=[\vartheta_1\ \vartheta_2\ 0\ 0\ \vartheta_5\ 0]$
η	Coefficients du meilleur modèle d'une famille. Vecteur solution de la régression linéaire : $\Delta X(t_k) = \Theta\Psi(t_k)$	$\hat{\Delta X}(t_k) = \eta\Psi(t_k)$ $\eta_n=[0{,}05\ \ 1{,}12\ \ 0\ \ 0\ \ 0{,}47\ \ 0]$
ω	Nombre de combinaisons de régresseurs pour une complexité donnée (nombre de ϕ différents)	
λ	Nombre de familles de modèles à construire et évaluer	
Famille de modèles	L'ensemble des combinaisons linéaires pour un Θ_n donné.	Famille de modèles "n" : $\mu(t) = \Theta_n\Phi(t)$
Modèle	Pour une famille de modèle, la combinaison linéaire qui minimise l'erreur de $\Delta X(t_k) = \Theta_n\Psi(t_k)$	Modèle "n" $\mu(t) = \eta_n\Phi(t)$
Complexité	Fonction de : • nbre de paramètres cinétiques ξ, • ordre des polynômes (OM), • nbre de régresseurs (monômes) par polynôme (NR)	Les familles de modèles d'une complexité donnée incluent les familles de complexité inférieure. Ex.: Les familles d'ordre 4 incluent les familles d'ordre 1, 2 et 3.

7.3 Construction des familles de modèles

La terminologie choisie pour décrire les modèles est la suivante. On appelle famille de modèles un polynôme constitué d'une constante et d'un certain nombre de régresseurs construits à partir des paramètres mesurés ou dérivés. Chaque polynôme constitue une famille de modèles.

La solution de la régression linéaire d'une famille de modèles constitue un modèle optimal pour cette famille. Afin de simplifier l'écriture, on appelle modèle la solution par régression multilinéaire dans le sens des moindres carrés d'une famille de modèles.

Selon la notation utilisée, on appelle modèle l'équation $\mu=\eta_n\Phi$ qui est la solution obtenue par régression de l'équation 7-6 qui représente une famille de solutions où Θ_n est un vecteur ligne dont tous les termes autres que ceux correspondant aux régresseurs $\Phi_i,\ldots$ Φ_k qui constituent la famille du modèle choisi sont nuls.

Le nombre des familles de modèles pouvant être construites dépend du nombre de paramètres choisis pour décrire la cinétique, de l'ordre maximal des monômes pouvant faire partie de la solution et du nombre de régresseurs constituant le polynôme. D'une façon générale, le nombre ω de régresseurs ϕ est calculé à partir de l'analyse combinatoire selon la formule d'analyse combinatoire

$$\omega = \frac{(\nu+\rho)!}{\nu!\rho!} \tag{7-9}$$

où ν est le nombre de paramètres servant à modéliser le taux de croissance et ρ est

l'ordre du polynôme.

Prenons un problème où le nombre de paramètres mesurés (ξ) est de 3 ($\nu = 3$) pour lequel on souhaite évaluer tous les polynômes d'ordre 3 ($\rho = 3$) à 4 termes. Le nombre ω de régresseurs dans ce cas est de 20. Ces 20 régresseurs sont présentés au Tableau 7-2.

Le nombre de familles de modèles pouvant être construites à partir de ces régresseurs dépend du nombre de régresseurs et du nombre de monômes choisis pour le polynôme. L'analyse combinatoire permet également de déterminer le nombre λ de familles de modèles qui se calcule grâce à l'équation 7-10

$$\lambda = \binom{\omega - 1}{\sigma} = \frac{(\omega - 1)!}{\sigma!(\omega - 1 - \sigma)!} \tag{7-10}$$

où σ correspond au nombre de termes du polynôme (nombre de régresseurs, + 1 pour la constante).

Pour construire un polynôme composé de 3 monômes et d'une constante (σ=4) à partir de 3 paramètres mesurés, le nombre de combinaisons λ est de 969, ce qui correspond au nombre de familles de modèles à calculer au moyen de régressions linéaires.

Le nombre de combinaisons, et donc le temps de résolution, augmente très rapidement avec l'augmentation de chacun des paramètres ν, ρ et σ. Ainsi, le calcul des meilleurs régresseurs de toutes les familles composées de polynômes d'au maximum 5 monômes d'ordre 5 construits autour de 5 paramètres ξ exige la résolution de 9 691 375 équations construites autour de 252 régresseurs. Il y a quatre ordres de grandeur entre le nombre

de solutions à calculer dans les deux cas; le second exemple exige plus de 10 000 fois le temps de calcul du premier car la complexité de la régression augmente également avec le nombre de régresseurs (σ).

Tableau 7-2 Construction des polynômes

no.	OM[1]	$\phi_v(t)$	no.	OM[1]	$\phi_v(t)$	no.	OM[1]	$\phi_v(t)$
1	0	1	5	2	$\xi_1\xi_1$	11	3	$\xi_1\xi_1\xi_1$
			6	2	$\xi_1\xi_2$	12	3	$\xi_1\xi_1\xi_2$
2	1	ξ_1	7	2	$\xi_1\xi_3$	13	3	$\xi_1\xi_1\xi_3$
3	1	ξ_2	8	2	$\xi_2\xi_2$	14	3	$\xi_1\xi_2\xi_2$
4	1	ξ_3	9	2	$\xi_2\xi_3$	15	3	$\xi_1\xi_2\xi_3$
			10	2	$\xi_3\xi_3$	16	3	$\xi_1\xi_3\xi_3$
						17	3	$\xi_2\xi_2\xi_2$
						18	3	$\xi_2\xi_2\xi_3$
						19	3	$\xi_2\xi_3\xi_3$
						20	3	$\xi_3\xi_3\xi_3$

[1]OM : ordre des monômes

7.4 Sélection d'un modèle

La détermination de la valeur des paramètres ϑ associés à cette structure ou famille de modèles s'effectue par régression multilinéaire. Chaque famille de modèles constitue l'ensemble des combinaisons linéaires possibles pour un choix de régresseurs $\{\phi_i,\ldots,\phi_k\}$ de Φ.

Soit $\Theta_n\Phi$ une famille de modèles où Θ_n est un vecteur dont tous les éléments autres que $\vartheta_i,\ldots,\vartheta_k$ sont nuls. En d'autres termes, pour le produit $\Theta_n\Phi$, seuls les coefficients $\vartheta_i,\ldots,\vartheta_k$ sont inconnus et à déterminer. La solution de ce problème est le vecteur η qui peut être déterminé par la résolution du système multilinéaire 7-6 par la méthode décrite à l'annexe B-3. L'erreur quadratique résiduelle de modélisation peut être exprimée par l'équation 7-11.

$$E = \sum_{q=1}^{N}\left(\frac{1}{N^q}\sum_{k=1}^{N^q}\left(\Delta\mathbf{X}^q(t_k) - \eta\mathbf{\Psi}^q(t_k)\right)^2\right)^{1/2} \tag{7-11}$$

où N est le nombre de cultures sur lesquelles s'applique la régression, N^q est le nombre d'échantillons dans la q-ième culture et où ψ^q représente les régresseurs calculés pour la q-ième culture et ΔX^q les données de la q-ième culture.

Le vecteur η est également la solution de l'équation différentielle 7-4 qui représente l'équation 7-2 et le modèle cinétique correspondant est

$$\mu(t) = \eta\mathbf{\Phi} \tag{7-12}$$

et la biomasse peut être calculée à partir du bilan de masse discrétisé par l'équation 7-13.

$$\hat{X}(t_k) = \eta\Psi(t_k) + X(t_0)V(t_0) + \sum_{i=1}^{k-1} V^{\text{échant}}(t_i)\hat{X}(t_i) \tag{7-13}$$

7.4.1 Critères de validation d'un modèle

Dans le cas de modélisations polynômiales, le très grand nombre de modèles impliqués rend inefficace la comparaison des modèles obtenus sur différentes catégories de données (comme ce fut le cas pour les modèles stoechiométriques) ou de complexité différente. En effet, les meilleurs modèles de chaque catégorie peuvent avoir une structure passablement différente, ce qui rend difficile leur comparaison. L'alternative retenue est de modéliser l'ensemble des cultures en un seul bloc.

Le problème de la surparamétrisation ne se pose pas dans le cas présent car les résultats du chapitre suivant démontrent que pour un nombre de régresseurs supérieurs à 3 ou 4 la réduction de l'erreur par l'ajout d'un nouveau terme est marginale. Cela limite l'utilisation de modèles plus complexes. Pour chaque famille de modèles, le modèle retenu est celui qui présente la plus faible erreur résiduelle telle que définit par l'équation 7-11.

7.4.2 Intervalle de confiance et régresseurs significatifs

La méthode de calcul de la variance autour des régresseurs (coefficients) est celle décrite à la section 5.6.4.

7.5 Construction d'indicateurs

Cette section décrit deux paramètres construits qui peuvent servir à l'estimation de la cinétique. Des paramètres cinétiques évidents sont les conditions mesurées du milieu soit les concentrations en nutriments jugés importants. Ces paramètres sont les concen-

trations extracellulaires de sucre, sous forme de sucre total, de phosphate, d'ammoniac et de nitrate.

Deux des nutriments mesurés (le phosphate et l'ammoniac) sont cependant rapidement épuisés dans le milieu en conditions de cultures en cuvée et en cuvée alimentée. En cuvée à alimentation graduelle, ces concentrations sont également faibles ou nulles lorsque l'ajout est bien contrôlé puisque tel est l'objectif de ces ajouts graduels. Ces paramètres sont donc difficilement utilisables pour la modélisation et d'autres paramètres doivent être considérés.

Deux indicateurs pouvant servir de paramètres cinétiques sont décrits dans les sous sections qui suivent. Le premier indique la concentration en phosphate intracellulaire et le second estime le taux d'accumulation d'ammoniac.

7.5.1 Indicateur du phosphate intracellulaire

Il est reconnu que, pour des cultures en cuvée, le phosphate s'accumule au début de la culture dans les cellules de plantes pour usage ultérieur. La disparition rapide du phosphate extracellulaire et la croissance soutenue par la suite dans le cas de cultures en cuvée viennent confirmer ce fait. La concentration de phosphate qui varie alors est la concentration en phosphate intracellulaire. Ce phosphate comprend du phosphate métabolisé et du phosphate libre. Le phosphate métabolisé se retrouve, par exemple, dans les phospholipides formant la paroi cellulaire et dans l'ADN. Le phosphate libre est accumulé sous forme vacuolaire ou est présent dans le cytoplasme et participe à la chaîne respiratoire en tant qu'ADM-ADP-ATP (Horton, 1993). Pour les expériences décrites

dans le présent travail, aucune de ces formes n'a été mesurée, ni même le phosphate intracellulaire total. La méthode de calcul du phosphate intracellulaire repose donc sur un simple bilan de masse sur le phosphate extracellulaire et une composition initiale moyenne en phosphate.

Cette composition initiale moyenne en phosphate intracellulaire a été calculée à partir des données recueillies de 12 cultures en flacons agités sans ajouts. Cette valeur a été calculée à partir des données initiales et finales, au maximum de la biomasse, en phosphate et en biomasse des cultures choisies. Ces valeurs correspondent à l'inverse du rendement en phosphate $Y_{X/P}$. Cette valeur s'approche de la composition minimale en phosphate.

Le calcul de la concentration en phosphate intracellulaire peut être représenté par l'équation

$$P_{min} = \frac{\Delta P(t_{final})}{\Delta X(t_{final})} = \frac{P(t_{final})V(t_{final}) - P(t_0)V(t_0)}{X(t_{final})V(t_{final}) - X(t_0)V(t_0)} \tag{7-14}$$

où, comme définit précédemment, ΔP est la consommation nette de phosphate exprimée en µmoles, ΔX en grammes, P est la concentration intracellulaire en µmol/g, X est en g/l et V en l.

La valeur obtenue est de 67 µmol PO_4 par g de biomasse sèche. Cela correspond à une composition massique de 0,0053g PO_4/g.

Cette concentration minimale est plus faible que celle calculée par Curtis et al. (Curtis, 1991) pour différentes cultures de plantes en phase stationnaire (réf. Tableau 7-3). Les concentrations maximales en phosphate intracellulaire atteintes par culture durant la

phase exponentielle (entre 400 et 800 µmol PO_4/g biomasse sèche, voir annexe A) sont supérieures aux valeurs présentées par Curtis et al. pour les concentrations en phosphate intracellulaire durant la phase de croissance exponentielle.

Cette valeur P_{min} est utilisée comme concentration intracellulaire en phosphate au temps initial ($P_i(t_0)$) dans le calcul du phosphate intracellulaire donné par l'équation 7-15.

$$P_i(t_k) = \frac{P_i(t_0) * X(t_0) * V(t_0) + \Delta P_e(t_k)}{X(t_k)V(t_k)} \tag{7-15}$$

7.5.2 Indicateur de l'assimilation d'ammonium

Contrairement au phosphate, l'ammoniac est rapidement métabolisé plutôt qu'accumulé. Comme source d'azote, l'ammoniac est métabolisé plus rapidement que le nitrate. La voie métabolique de l'ammoniac est beaucoup plus courte que celle du nitrate qui implique la réduction de l'azote d'une valence de -5 à +3.

Cette métabolisation rapide peut se traduire par une croissance plus rapide durant l'utilisation de l'ammoniac comme source d'azote. Par ailleurs, certains auteurs (Beck et Renner, 1989) citent également une accumulation facilitée du nitrate en présence de faibles concentrations d'ammoniac dans le milieu. Dans les deux cas, l'accumulation d'ammoniac fait augmenter le taux de croissance de la biomasse.

Un estimateur a donc été développé pour mesurer le taux d'accumulation d'ammoniac entre deux échantillons.

Cet estimateur (dN) s'exprime par l'équation 7-16.

$$dN(t_k) = \frac{\Delta N(t_{k+1}) - \Delta N(t_k)}{X(t_k)V(t_k)(t_{k+1} - t_k)} \tag{7-16}$$

La valeur du paramètre en un point est calculée à partir des valeurs d'ammoniac mesurées à ce point et au point suivant pour tenir compte de l'assimilation qui a lieu durant la période suivant l'échantillonnage. Ce paramètre n'ajoute pas d'information aux cultures en cuvée mais peut être pertinent pour les cultures avec ajout et surtout pour les cultures avec ajout graduel où la concentration en ammoniac demeure faible tant que l'assimilation est rapide par rapport au taux d'ajout. Il est à noter que cet estimateur est différentiel et donc sujet à des variations importantes entre 2 points. De plus, ce paramètre calculé n'est pas lissé, le nombre de points d'échantillonnage étant trop faible.

Tableau 7-3 Concentration en phosphate de diverses cultures de cellules de plantes (réf. Curtis, 1991)

	Phase de croissance	
Type de cellules	**exponentielle (µmol P/g DW)**	**stationnaire (µmol P/g DW)**
Poppy	333	108
Tobacco	742	154
Bindweed	237	96
Potato	447	136
Carrot	400	104
Millet	309	135
Banana	352	106

8. Application des modèles cinétiques aux données expérimentales

Dans ce chapitre, la modélisation est appliquée aux données expérimentales récoltées dans le cadre de ce travail. Tel que mentionné précédemment, les structures de modèles construits sont appliquées à l'ensemble des données simultanément.

Diverses catégories de familles de modèles sont présentées par complexité croissante. Le nombre de régresseurs passe de 2 à 5 pendant que l'ordre des polynômes varie de 1 à 4. Un deuxième type de modèles cinétiques polynômiaux est développé pour des motifs qui sont expliqués à la section 8.4 où ces modèles sont présentés : ce sont les modèles sans constante. Les meilleurs modèles de chaque catégorie sont présentés. Certains sont analysés en détail et présentés de façon graphique.

8.1 Analyse préliminaire du taux de croissance

Une première analyse des taux de croissance montre que ceux-ci se situent en général entre 0,005 et 0,015 h^{-1} (0,12 et 0,36 j^{-1}), ce qui est comparable avec les données obtenues par Aubry (1995) pour la même souche, en flacons et en bioréacteurs. L'annexe A présente les résultats expérimentaux sous forme de graphiques de même qu'une analyse des cultures utilisées pour la modélisation.

8.2 Évaluation des modèles cinétiques polynômiaux

Une régression multilinéaire est appliquée à toutes les familles de modèles comprenant les polynômes à deux, trois ou quatre termes (incluant une constante) construits autour

de six paramètres présentés au Tableau 8-1 (page 91). Les polynômes de régression obtenus suite à la régression multilinéaire par la méthode des moindres carrés sont présentés au Tableau 8-2. Ces modèles sont ceux présentant la plus faible erreur résiduelle de chaque famille de modèles. Pour en arriver à ce tableau, 77 238 876 modèles différents ont été évalués.

La Figure 8-1 présente les erreurs résiduelles de ces modèles par ordre maximal du polynôme et par nombre croissant de régresseurs. La première donnée illustrée représente la régression avec la seule constante comme paramètre. C'est ce modèle qui présente la plus grande erreur résiduelle. Les autres séries représentent des polynômes constitués de 2, 3 et 4 termes.

Un premier constat intéressant provient de la deuxième série: toutes les erreurs des modèles à 2 paramètres sont identiques. Le Tableau 8-2 montre qu'effectivement, dans les quatre cas, le modèle est le même. Il convient de se rappeler qu'une catégorie contient toutes les structures de modèles des catégories inférieures. On conclue donc que de tous les modèles à 2 termes, c'est le modèle d'ordre 1 qui permet le mieux de représenter les données. Dans ce cas, la concentration en sucre est le régresseur qui permet le mieux de prédire le taux de croissance.

Le même phénomène se produit pour les modèles à trois paramètres. Dans ce cas la meilleure combinaison de régresseurs est la concentration de phosphate intracellulaire à la première et à la seconde puissance (P_i et P_i^2). La Figure 8-2 présente la fonction décrite par ce polynôme par rapport à la concentration calculée en phosphate intracellu-

laire Pi. La courbe de modélisation en forme de parabole inversée indique une augmentation du taux de croissance avec la concentration en Pi jusqu'à un maximum puis une inhibition de croissance à des concentrations en Pi supérieures à 0,52 mmol/g. La Figure 8-3 présente la dispersion des données de phosphate et de sucre pour l'ensemble des cultures analysées. On constate que la majorité des données de phosphate est située dans la plage inférieure à 0,5 mmol/g. La plus faible concentration de points d'échantillonnage au-delà de cette borne augmente l'incertitude sur la partie de droite de la courbe de la Figure 8-2. L'effet d'inhibition du phosphate intracellulaire est donc probable mais reste à confirmer ou à préciser avec de nouvelles données expérimentales.

Les équations développées au chapitre précédent minimisent l'erreur entre les valeurs estimées et calculées de ΔX. Les résultats indiquent que les modèles présentant la plus faible erreur résiduelle pour les ΔX présente également la plus faible erreur sur la biomasse mesurée (X, échantillonnée). Les Figures 8-4 et 8-5 comparent les résultats de simulation pour le modèle OM=2, NR=3 (représenté par la Figure 8-2) avec les données expérimentales transformées (ΔX) et directement mesurées (X) respectivement.

Cette série de modèles à 3 régresseurs se démarque également par le terme constant qui est presque nul. D'autres modèles à constante nulle sont également présents parmi les bons modèles d'autres catégories à un niveau d'erreur légèrement plus élevé que les modèles retenus qui ont par ailleurs tous un terme constant autour de 0,0155. Les modèles à constante nulle représentent une avenue à explorer car ils permettent d'éliminer un effet des modèles polynômiaux complets qui, à cause de la constante, peuvent prédire

une croissance non nulle à concentration nulle des nutriments comme c'est le cas pour les modèles à quatre termes présentés à la Figure 8-6. Ces nouvelles structures de modèles, qui présentent une forme plus près du modèle de Monod, sont évaluées à la section suivante.

Les modèles à 3 ou à 4 régresseurs ne peuvent être entièrement représentés graphiquement car ils impliquent tous plus de deux paramètres. Cependant une projection tridimensionnelle (en fixant la valeur d'une des variables) peut être réalisée en imposant arbitrairement une valeur de 0 à l'ammoniac extracellulaire. Suite à l'assimilation rapide de l'ammoniac, c'est la situation qui prévaut pour la majeure partie du temps pour les cultures en cuvée simple et en cuvée avec ajout. La Figure 8-6 illustre une partie l'équation cinétique lorsque Ne=0 pour le modèle où OM=3 et NR=4 avec constante. On constate un taux de croissance égal à une constante lorsque tous les paramètres sont nuls(point identifié A sur la Figure 8-6). Dans ce cas, la structure du modèle présente une croissance identique à concentration de sucre nulle. Les cellules d'*Eschscholtzia californica* n'accumulent pas de sucre et l'expérience démontre que la croissance cesse lorsqu'il y a épuisement du sucre,. La Figure 8-3 indique cependant que très peu de points se situent dans la zone où la concentration en sucre est inférieure à 10 g/l et, par conséquent, cette plage se retrouve à l'extérieur de la zone de validité du modèle.

Par ailleurs, on constate au Tableau 6-2 que tous les paramètres estimés ont un intervalle de confiance qui exclue la valeur "0". Ce qui indique que les paramètres retenus contribuent de façon significative au modèle.

Tableau 8-1 Paramètres utilisés pour modéliser la cinétique

Paramètre cinétique	Symbole
Sucre (sucrose, glucose, fructose) extracellaire	Se
Ammoniac extracellulaire	Ne
Taux d'accumulation d'ammoniac	dN
Nitrate extracellulaire	Me
Phosphate extracellulaire	Pe
Phosphate intracellulaire	Pi

Tableau 8-2 Erreur de modélisation et modèles avec constante (avec l'intervalle de confiance de 95%)

OM[1]	NR[2]	Err	Equation * 10^{-3}						
1	1	19.42	+(8.22 ± 0.68)						
1	2	12.10	+(15.70 ± 1.33)	-(0.37 ± 0.06)	Se				
2	2	12.10	+(15.70 ± 1.33)	-(0.37 ± 0.06)	Se				
3	2	12.10	+(15.70 ± 1.33)	-(0.37 ± 0.06)	Se				
4	2	12.10	+(15.70 ± 1.33)	-(0.37 ± 0.06)	Se				
1	3	11.81	+(15.80 ± 1.31)	+(0.74 ± 0.70)	Pe	-(0.39 ± 0.06)	Se		
2	3	10.42	+(0.73 ± 1.10)	+(41.20 ± 5.64)	Pi	-(39.35 ± 6.01)	Pi^2		
3	3	10.42	+(0.73 ± 1.10)	+(41.20 ± 5.64)	Pi	-(39.35 ± 6.01)	Pi^2		
4	3	10.42	+(0.73 ± 1.10)	+(41.20 ± 5.64)	Pi	-(39.35 ± 6.01)	Pi^2		
1	4	11.39	+(15.84 ± 1.27)	-(0.52 ± 0.41)	Ne	+(2.76 ± 1.72)	Pe	-(0.39 ± 0.06)	Se
2	4	9.60	+(16.44 ± 1.09)	-(0.42 ± 0.05)	Se	-(5.64 ± 1.75)	Pi*Pe	+(0.17 ± 0.05)	Pe*Se
3	4	9.22	+(14.67 ± 1.06)	-(0.43 ± 0.05)	Se	-(1.29 ± 0.34)	$Ne*Pi^2$	+(0.02 ± 0.01)	$Pi*Se^2$
4	4	9.06	+(16.12 ± 1.02)	-(0.48 ± 0.06)	Se	-(0.05 ± 0.01)	$Ne*Pi^2*Se$	+(0.00 ± 0.00)	$Pi*Se^3$

[1] OM: ordre ou degré des monômes

[2] NR: nombre de régresseurs (incluant la constante)

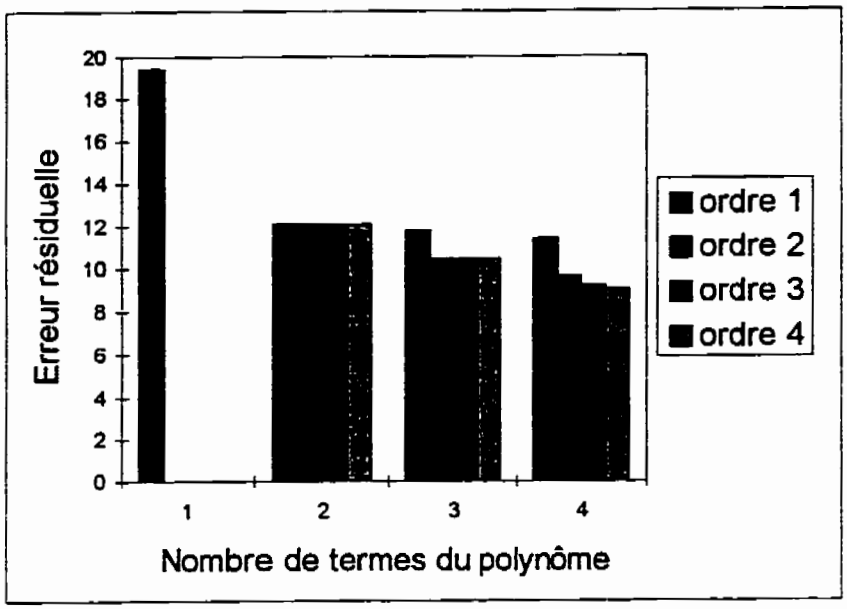


Figure 8-1 Erreur résiduelle de modélisation en fonction du nombre de termes du polynôme et de l'ordre des monômes pour modèles avec constante

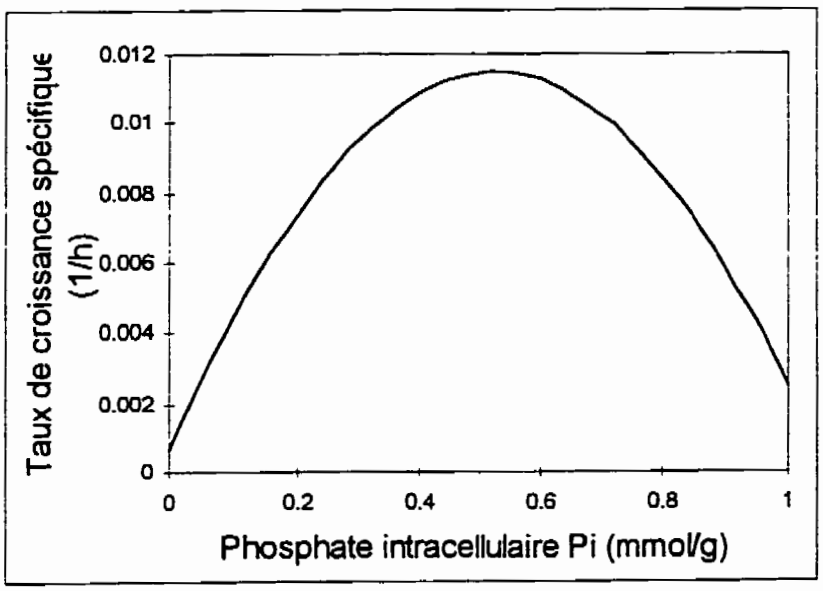


Figure 8-2 Modèle cinétique de croissance avec OM=2,NR=3 avec constante

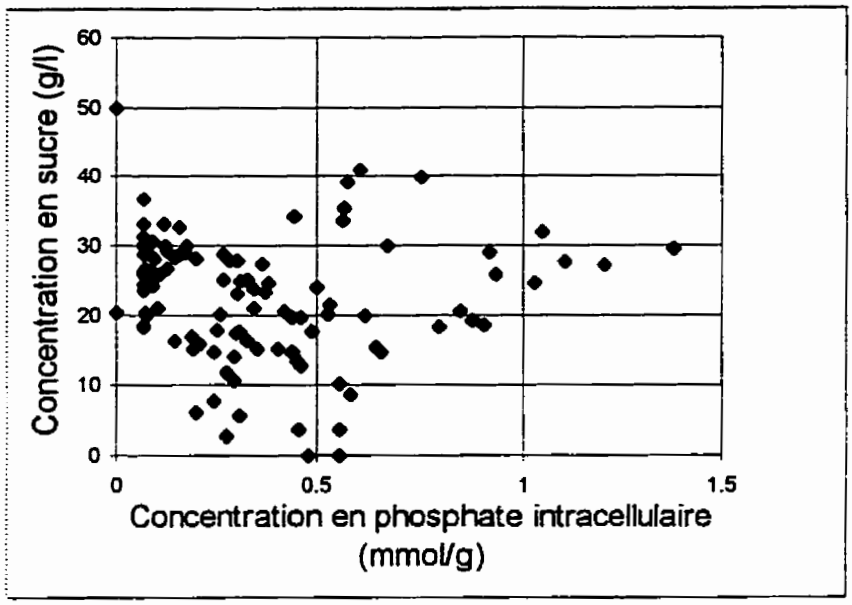


Figure 8-3 Dispersion des données de phosphate et de sucre pour l'ensemble des cultures

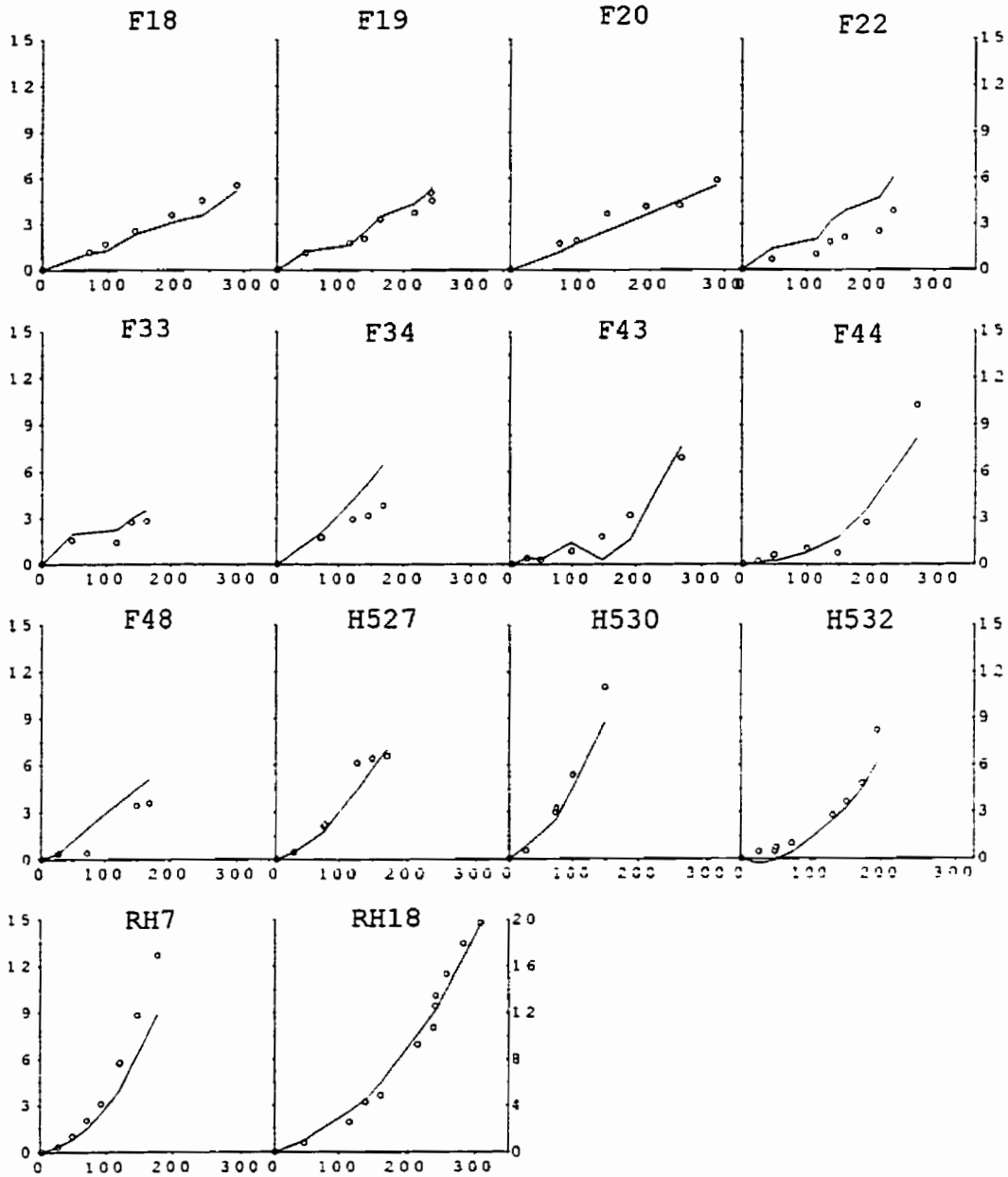


Figure 8-4 Résultat en ΔX (g) du modèle OM=2;NR=3 avec constante pour l'ensemble des cultures
(0)*: Données expérimentales,* (-)*: Simulation*

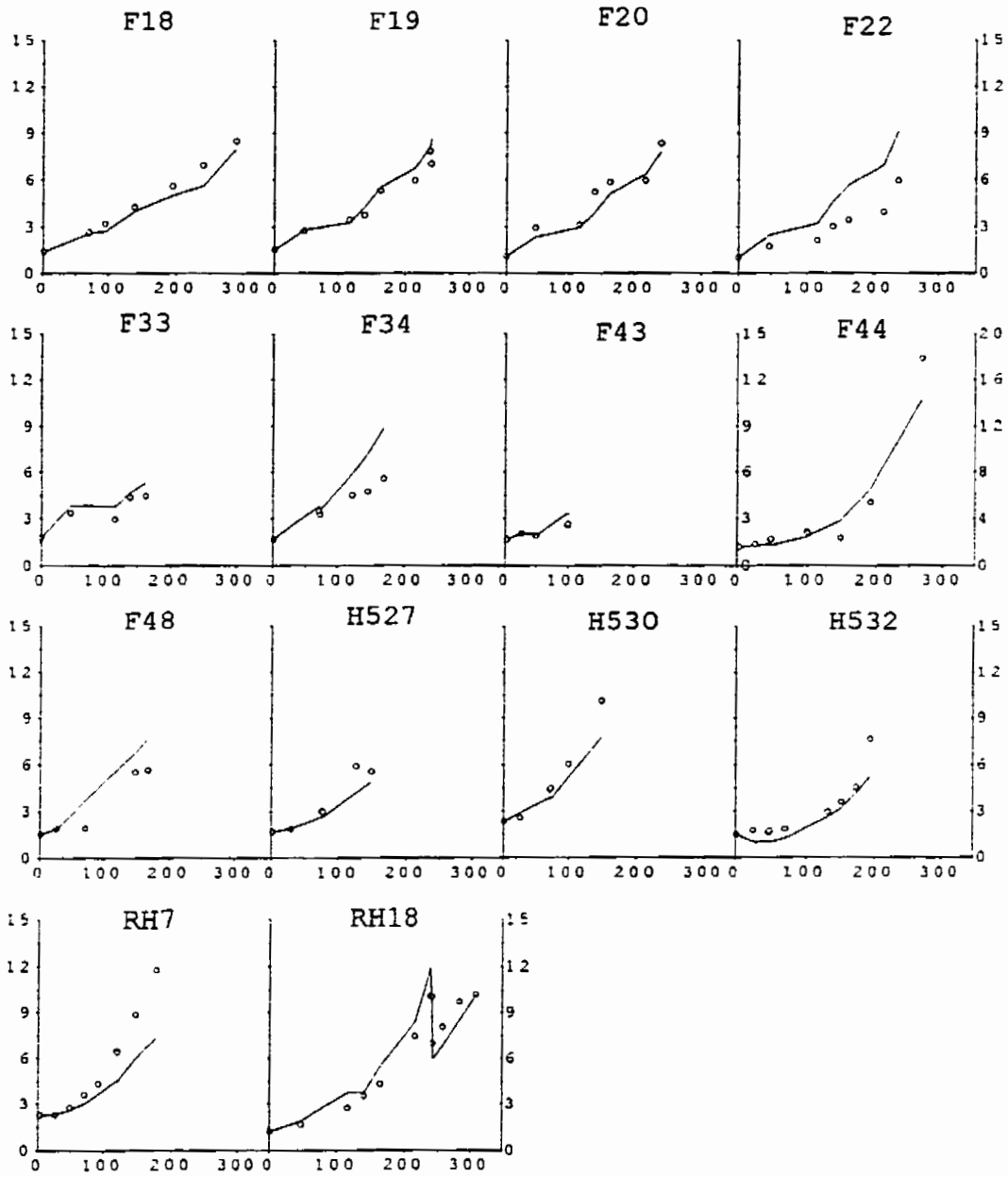


Figure 8-5 Résultat en biomasse X (g/l) du modèle OM=2;NR=3 avec constante pour l'ensemble des cultures
(0)*: Données expérimentales,* (-)*: Simulation*

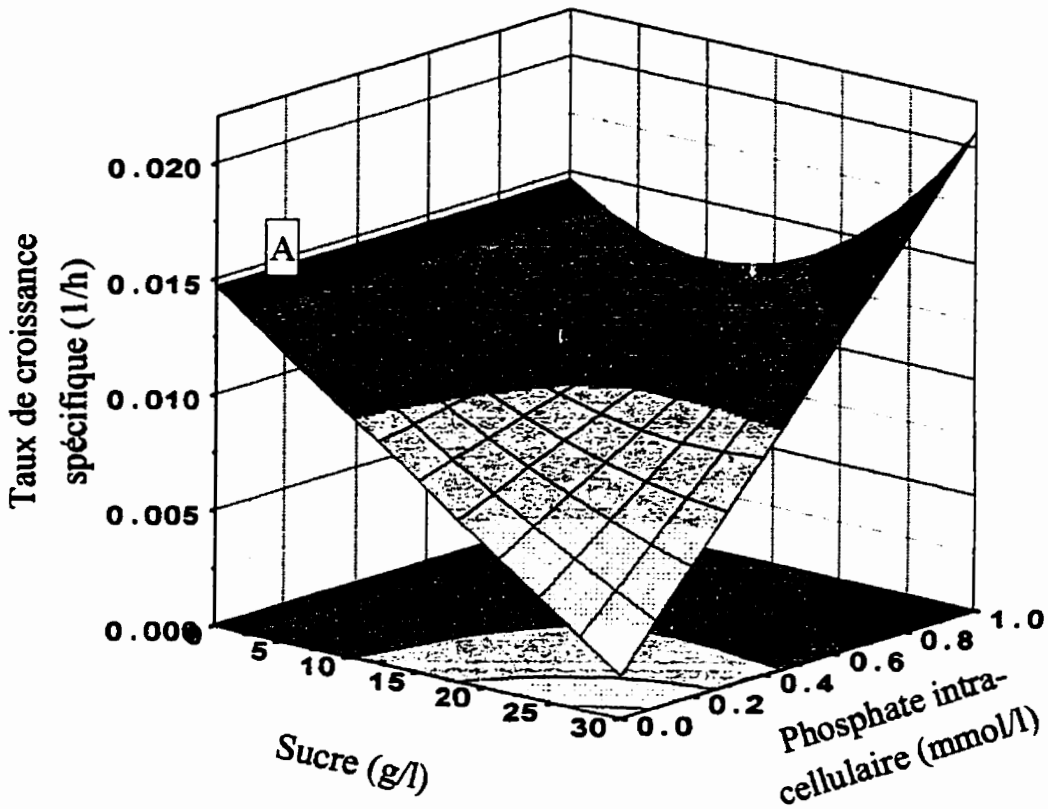


Figure 8-6 Modèle OM=3, NR=4 avec constante, Ne=0

8.3 Évaluation des modèles polynômiaux sans constante

Les modèles présentés dans cette section se révèlent plus conformes aux modèles conventionnels de type Monod qui présentent la caractéristique de prédire une croissance nulle en l'absence des nutriments essentiels. Le Tableau 8-3 présente donc les meilleurs régresseurs en fonction de l'ordre des régresseurs et du nombre de monômes. La Figure 8-7 présente les erreurs de ces modèles sans constante regroupées par ordre de polynômes et par nombre de régresseurs.

On constate encore une fois qu'avec un nombre minimum de régresseurs, le meilleur régresseur est d'ordre 1. Dans le cas présent, le Pi est le nutriment qui suit le mieux le taux de croissance. Cependant l'erreur est plus élevée que pour les modèles à deux paramètres présentés à la section précédente (24,76 contre 19,42).

Pour les polynômes à deux régresseurs d'ordre 2 ou supérieur, on retrouve, sans la constante, les mêmes monômes que ceux obtenus par les catégories à trois régresseurs de la section précédente. L'erreur de modélisation est légèrement plus élevée : ceci est cohérent avec la logique de réduction de l'erreur avec l'augmentation de la complexité du modèle.

Tous les modèles sans constante présentés reposent sur la concentration en phosphate intracellulaire. Tous ces modèles reposent sur les paramètres Pi et Pi^2. La valeur des coefficients varie cependant en fonction des autres termes du modèle.

Dans cette catégorie de modèles comme dans la précédente, les paramètres Me et dN ne sont pas utilisés. Bien qu'on les retrouve dans certains modèles moins performants de chaque famille de modèles, ils sont absents des 32 meilleurs modèles présentés dans ce chapitre.

La Figure 8-8 présente le modèle à 4 régresseurs d'ordre 3. Le modèle est fonction des concentrations en phosphate intracellulaire et en sucre seulement. On constate un optimum de croissance pour une concentration en Pi variant entre 0,40 et 0,60 pour toute concentration en sucre extracellulaire (point A). Ce modèle démontre également un léger effet positif de la concentration en sucre à forte concentration de phosphate intracellulaire (point B). Le modèle indique une croissance négative lorsque la concentration en Pi est élevée pendant que la concentration en Se est faible (point C). Il faut cependant noter que ces conditions n'ont été explorées que de façon incomplète, même durant les cultures avec ajout graduel, et sont donc situées à l'extérieur des plages expérimentales et ne sont pas statistiquement validés.

Les autres modèles à plusieurs termes présentent également des taux de croissance négatifs, mais encore une fois pour des conditions qui ne font pas partie des plages expérimentales observées.

Tableau 8-3 Erreur de modélisation et modèles sans constante
(avec l'intervalle de confiance de 95%)

OM	NR	Err	Equation * 10^{-3}							
1	1	24.76	+(17.80 ± 2.00)	Pi						
2	1	24.76	+(17.80 ± 2.00)	Pi						
3	1	24.76	+(17.80 ± 2.00)	Pi						
4	1	24.76	+(17.80 ± 2.00)	Pi						
1	2	21.46	+(22.40 ± 2.40)	Pi	-(4.00 ± 1.50)	Pe				
2	2	10.53	+(44.50 ± 2.80)	Pi	-(42.20 ± 4.20)	Pi^2				
3	2	10.53	+(44.50 ± 2.80)	Pi	-(42.20 ± 4.20)	Pi^2				
4	2	10.53	+(44.50 ± 2.80)	Pi	-(42.20 ± 4.20)	Pi^2				
1	3	18.96	+(18.10 ± 2.80)	Pi	-(4.20 ± 1.30)	Pe	+(0.10 ± 0.00)	Se		
2	3	10.28	+(46.00 ± 3.10)	Pi	-(46.00 ± 5.50)	Pi^2	+(0.40 ± 0.30)	Pe^2		
3	3	10.24	+(46.10 ± 3.10)	Pi	-(46.10 ± 5.40)	Pi^2	+(0.10 ± 0.10)	Pe^3		
4	3	10.12	+(41.60 ± 3.60)	Pi	-(105.20 ± 24.60)	Pi^3	+(64.60 ± 20.80)	Pi^4		
1	4	18.72	+(18.20 ± 2.70)	Pi	-(0.50 ± 0.70)	Ne	-(2.30 ± 2.90)	Pe	+(0.10 ± 0.00)	Se
2	4	9.89	+(47.60 ± 3.30)	Pi	-(0.60 ± 0.40)	Ne	+(0.30 ± 0.20)	Ne*Pe	-(48.20 ± 5.90)	Pi^2
3	4	9.46	+(72.40 ± 12.60)	Pi	-(97.40 ± 24.50)	Pi^2	-(1.10 ± 0.50)	Pi*Se	+(2.20 ± 1.00)	Pi^2*Se
4	4	9.28	+(65.60 ± 8.90)	Pi	-(69.70 ± 11.70)	Pi^2	-(0.90 ± 0.40)	Pi*Se	+(0.00 ± 0.00)	Pi^2*Se^2

[1] OM: ordre ou degré des monômes
[2] NR: nombre de régresseurs

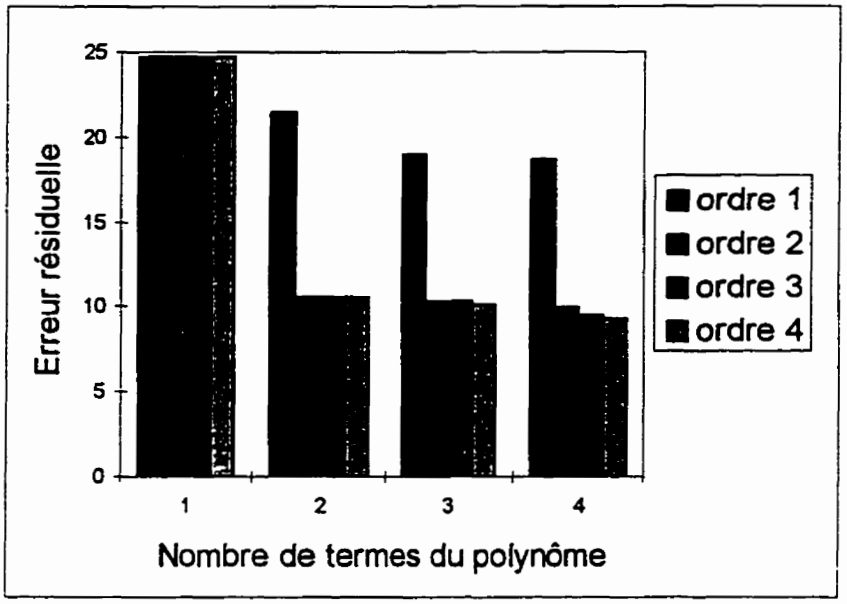


Figure 8-7 Erreur résiduelle de modélisation en fonction du nombre de termes du polynôme et de l'ordre des monômes pour les modèles sans constante

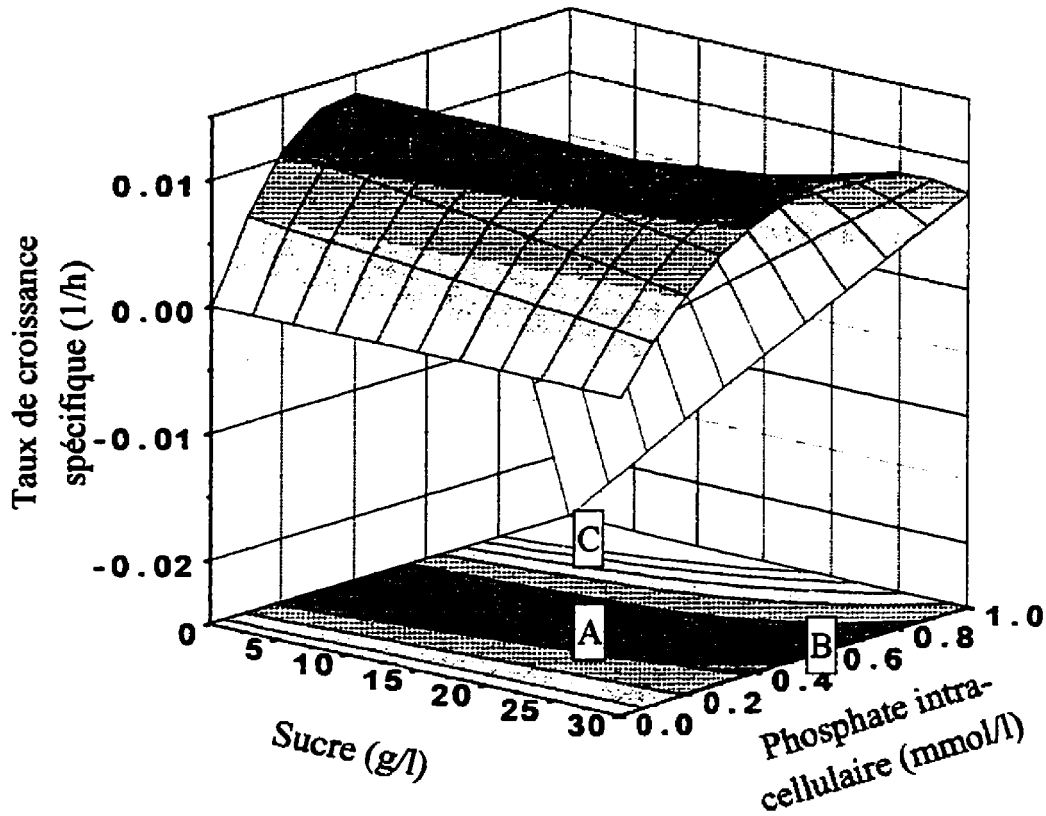


Figure 8-8 Modèle OM=3, NR=4 sans constante

9. Conclusions et recommandations

9.1 Conclusion sur les schèmes réactionnels

Pour la phase modélisation de la stoechiométrie, quatre schèmes réactionnels ont été développés pour décrire la production de biomasse de cultures cellulaires d'*Eschscholtzia californica* en fonction des taux d'assimilation des nutriments. De ces schèmes, quinze équations linéaires furent construites. Les coefficients de ces équations linéaires ont été évalués par régression multilinéaire par la méthode des moindres carrés sur 14 cultures réalisées en flacons agités et en bioréacteurs. La validation de ces équations a permis d'en éliminer certaines présentant un coefficient de détermination (CDMEA) faible. D'autres furent retirées à cause d'une dégradation des résultats d'un test de validation croisée.

Cinq des six équations décrivant le mieux l'évolution de la biomasse comptent 2 ou 3 paramètres. L'équation à un seul paramètre est dérivée d'une équation de maintenance. Trois autres équations reposent également sur ce paramètre de maintenance. Un autre paramètre est utilisé par 4 équations : le sucre. Le nitrate et la combinaison nitrate-ammoniac combinés sont aussi présents dans 2 modèles chacuns. Le modèle qui exprime le mieux l'ensemble des équations linéaires retenues est défini par l'équation 6-1.

$$\begin{aligned} s_1 S_e + n(N_e + M_e) &\rightarrow X \\ s_2 S_e + X &\rightarrow X \end{aligned} \tag{6-1}$$

L'étude de la variance des coefficients ne permet pas de déterminer avec précision la

plupart des cas, les intervalles de confiance de 95% autour des coefficients incluent la valeur 0. Ceci indique que le paramètre en question ne contribue pas de façon significative à l'explication de la variable dépendante, soit la biomasse. Cette variance importante semble avoir deux causes qui sont la variabilité des cultures et l'assimilation conjointe des nutriments qui entraine une presque colinéarité d'équations indépendantes.

9.2 Conclusion sur la modélisation cinétique

La méthode développée par Chotteau et appliquée au chapitre précédent permet de déterminer des modèles cinétiques polynômiaux pour prédire le taux de croissance spécifique de cultures de cellules d'*E. Californica*.

Les modèles polynômiaux retenus pour chaque niveau de complexité sont généralement différents d'une famille de modèles à l'autre. Une étude plus détaillée de plusieurs des meilleurs modèles permettrait de trouver plus de similitudes entre des polynômes de complexités différentes. Dans le cas présent, seul le polynôme présentant l'erreur de modélisation la plus faible de chacun des 32 groupes de complexité a été retenu parmi les 154 millions de modèles évalués, négligeant de ce fait d'autres modèles aussi valables dont l'erreur est très semblable. Les différentes tendances des modèles pour les mêmes données s'expliquent probablement par la concentration des données sur une plage relativement petite. L'élargissement de cette plage expériementale permettrait, d'une part, de tester de nouvelles conditions d'opération et, d'autre part, de restreindre les variations entre les divers modèles.

On a vu que les modèles polynômiaux peuvent prédire des comportements non observables : la plupart des modèles avec constante prédisent une croissance forte en l'absence d'un ou plusieurs nutriments et certains modèles sans constante (par exemple la Figure 8-8) indiquent une croissance négative sous certaines conditions. Il a été démontré que la plage pour laquelle la croissance est négative ne fait pas partie de la plage expérimentale et doit être considérée comme une extrapolation abusive du modèle. Ce problème est inhérent à la méthode utilisée : les modèles polynômiaux, et les régressions multilinéaires en général, ne sont strictement valables que pour la plage de données pour laquelle ces paramètres ont été calculés. Cette contrainte des modèles obtenus ne limite pas l'emploi du modèle pour des fins d'analyse dans la région valide quoique cette zone de validité ne soit pas calculée avec précision. Pour des fins de contrôle normal, l'inconvénient peut être assez faible si on s'en tient à des conditions d'opération faisant partie du domaine de validité du modèle.Si le modèle est utilisé en contrôle ou à des fins de simulation et que celle-ci diffère significativement de la réalité pour certaines conditions d'opération ou d'alimentation, il peut être utile de recalculer les modèles en incluant les nouvelles données dans les calculs, élargissant ainsi le domaine d'application du modèle.

Contrairement à la méthode employée pour choisir les schèmes réactionnels, aucun critère de sélection n'a été développé pour déterminer le meilleur modèle parmi les modèles retenus. Ce choix est évidemment plus complexe car il ne s'agit pas de déterminer des valeurs numériques aux coefficients de rendements. Le critère de choix dépendra de l'utilisation souhaitée pour le modèle. Pour des fins de contrôle, on privilégie un modèle simple qui offre une erreur de prédiction raisonnable : dans ce cas le modèle sans cons-

tante à deux termes (équation 9-1) est le meilleur choix. La Figure 8-2 présente une courbe très semblable au modèle de l'équation 9-1. On voit un effet d'inhibition du phosphate. Cette inhibition est plausible mais la précision de la courbe d'inhibition est plus faible dû à un manque de points expérimentaux dans cette plage de concentration de phosphate intracellulaire. Des modèles plus complexes ont été évalués mais l'augmentation de la complexité ne réduit que de peu l'erreur résiduelle de modélisation.

$$\mu = [+(44.50 \pm 2.80)*P_i - (42.20 \pm 4.20)*P_i^2]*10^{-3} \qquad (9\text{-}1)$$

Pour des fins d'optimisation des conditions de croissance et de l'alimentation, les modèles plus complexes sont préférables. Ils permettent de déceler des interactions plus complexes entre les nutriments. La connaissance de ces interactions peut permettre de suggérer des modèles à base biologique plus complexe que les modèles simples de type Monod généralemenent utilisés.

En ce qui à trait aux paramètres ayant la plus grande influence sur la cinétique, pour les modèles sans constante, la concentration en sucres apparait rarement dans les meilleurs modèles. Cela étonne puisque la plupart des modèles cinétiques qu'on retrouve dans la littérature sont basés sur ce substrat. Pour cette lignée, l'épuisement en sucre entraîne l'arrêt de la croissance (Aubry, 1995). Cependant, lorsque le sucre est présent dans le milieu, on constate que sa concentration a peu d'influence sur le taux de croissance.

Parmi les paramètres présents dans les modèles retenus, la concentration en phosphate intracellulaire (P_i) est le paramètre le plus utilisé, soit seul dans un régresseur, soit sous sa forme au carré (P_i^2) ou en combinaison avec d'autres paramètres. Ce paramètre est

souvent utilisé avec le régresseur Pi^2 ou en combinaison avec d'autres nutriments Beaucoup de modèles utilisent également ce paramètre au carré (Pi^2). Le phosphate extracellulaire (Pe), l'ammoniac extracellulaire (Ne) et l'indicateur d'accumulation d'ammoniac (dN) n'ont que très rarement été retenus parmi les régresseurs. Cela suggère de porter une attention plus importante aux nutriments secondaires et intracellulaires. Les substrats secondaires jouent un rôle important dans la cinétique de croissance de cultures de cellules de *E. californica* et probablement pour la plupart de cellules végétales.

Des gains significatifs de productivité peuvent être atteints en ajoutant en cours de culture ces nutriments par ajouts ponctuels. La méthode d'ajout graduel utilisée dans ce travail va également dans ce sens.

9.3 Conclusions générales et recommandations

L'approche de modélisation utilisée dans ce mémoire est basée sur les travaux de Véronique Chotteau en culture de cellules animales. La méthodologie a été transposée à la croissance de cellules végétales avec quelques modifications, la plus importante étant la modélisation des taux de croissance spécifiques.

Différents schèmes réactionnels ont été élaborés sur la base des connaissances actuelles du comportement des cellules végétales cultivées en bioréacteurs et en flacons agités. Une analyse statistique poussée a révélé que les intervalles de confiance sont généralement plus grands que la valeur des coefficients obtenus ce qui ne permet pas de valider les modèles avec satisfaction. La colinéarité des variables des modèles et la variabilité des cultures expliquent ces résultats. En conclusion, seuls les modèles les plus simples

peuvent être validés et même dans ce cas, les coefficients de rendement ne sont pas significatifs, les intervalles de confiance étant trop grands. Aucun jugement certain ne peut être porté sur les autres modèles étant donné que les données ne permettent pas de définir avec satisfaction les coefficients de rendement de ces modèles.

Les modèles cinétiques de production de biomasse ne sont généralement pas fonction de la concentration en sucre. Cela va à l'encontre de la plupart des modèles de la littérature. Les modèles développés dans le présent travail révèlent une dépendance beaucoup plus importante face au phosphate intracellulaire.

L'auteur est d'avis que des percées importantes en culture de cellules de plantes, au niveau de l'amélioration de la productivité et de la compréhension des mécanismes de croissance ou de production, passent par un contrôle précis des conditions d'alimentation des différents nutriments. Le contrôle des concentrations de phosphate intracellulaire exige une modélisation plus poussée que le simple calcul de concentration de sucre ou d'autres nutriments dans le milieu.

Ces ajouts sont complexes et requièrent un algorithme de contrôle poussé. Des modèles relativement simples, comme le modèle sans constante à deux paramètres (équation 9-1), pourront éventuellement réaliser ces fonctions. Il offre une faible erreur résiduelle et demeure relativement simple.

En outre, toutes les données disponibles pouvant indiquer l'état des cultures sont utiles. La mesure de la production de CO_2 et de la consommation en oxygène fournissent des

données relatives aux taux de croissance. Ces données fournissent des informations en continu très utiles au contrôle de la culture. L'inclusion de ces données aux schèmes réactionnels, de façon directe ou par corrections statistiques basées sur les bilans de masse, peut permettre de diminuer l'incertitude sur les coefficients de rendements.

La modélisation du nombre de cellules, conjointement à la biomasse, semble également prometteuse et devrait être poursuivie. Les travaux de Pépin (1995) et des données partielles obtenues des travaux réalisés dans le cadre de ce mémoire mais non présentées indiquent une différence significative entre la courbe de croissance du nombre de cellules et de celle de la biomasse, surtout en fin de culture lorsque des cellules accumulent de la biomasse sans se diviser.

La recherche en culture de cellules de plantes est essentielle et doit être poursuivie. Grâce à elle, on guérira peut-être un jour le cancer, on cultivera des déserts et les plantes continueront à fournir à l'Homme les éléments essentiels à sa survie et à son développement.

Bibliographie

ARCHAMBAULT, J. (1991). Large scale (20L) culture of surface-immobilized *Catharanthus roseus* cellules. Enzyme Microb. Technol., 13: 882-892.

AUBRY, D. (1995). Développement et commande d'un bioprocédé de production de sanguinarine à partir de la culture en suspension de cellules de plante. Mémoire de maîtrise, École Polytechnique de Montréal.

BAILEY, C.M., H. NICHOLSON (1989). A new structured model for plant cell culture. Biotechnol. Bioeng., 34: 1331-1336.

BAILEY, J.E., D.F. OLLIS (1986). Biochemicals Engineering Fundamentals. New York, McGraw Hill.

BECK, E., U. RENNER (1989). Ammonium triggers uptake of NO3- by *Chenopodium rubrum* suspension culture cells and remobilization of the vacuolar nitrate pool. Plant Cell Physiol, 30(4): 487-495.

CHOTTEAU, V. (1995). A general modeling methodology for animal cell cultures. Thèse de doctorat, Université Catholique de Louvain.

CHOTTEAU, V., G. BASTIN (1993). Identification of mathematical models for biotechnological processes : Recipe and application to a culture of animal cells, Centre for systems engineering and applied mechanics, Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

CURTIS, W.R., P.M. HASEGAWA, A.H. EMERY (1991). Modelling linear and variable growth in phosphate limited suspension cultures of opium poppy. Biotechnol. Bioeng., 38: 371-379.

DE GUCHT, L.P.E., L.H.W. VAN DER PLAS (1995). Growth Kinetics of Glucose-Limited *Petunia hybrida* Cells in Chemostat Cultures: Determination of Experimental Values for Growth and Maintenance Parameters. Biotechnol. Bioeng., 47: 42-52.

DE GUNST, M.C.M., P.A.A. HARKES, J. VAL, W.R. VAN WET (1990). Modelling the growth of a batch culture of plant cells: a corpuscular approach. Enzyme Microb. Technol. 12: 61-71.

DOCHAIN, D., G. BASTIN (1990). Adaptative Control of Fedbatch Bioreactor. Chem. Eng. Comm., 87: 67-85.

DRAPEAU, D.H., W. BLANCH, C.R. WILKE (1986). Growth kinetics of *Dioscorea deltoidea* and *Catharanthus roseus* in batch culture. Biotechnol. Bioeng., 28: 1555-1563.

DRAPER, N. R., H. SMITH (1981), Applied regression analysis, 2ième édition., New York, Wiley

DUCOS, J.-P., A. PAREILLEUX (1986). Effect of aeration rate and influence of pCO2 in large-scale cultures of *Catharanthus roseus* cells. Appl. Microbiol Biotechnol. 25: 101-105.

FRAZIER, G.C. (1989). A simple, leaky cell growth model for plant cell aggregates Biotechnol. Bioeng., 33: 313-320.

GUARDIOLA, J.IBARRA., J.L.M. CANOVAS (1995). A Model That Links Growth and Secondary Metabolite Production in Plant Cell Suspension Cultures. Biotechnol. Bioeng., 46: 291-297.

HOOKER, B.S., J.M. LEE (1992). Application of a new structured model to tobacco cell cultures. Biotechnol. Bioeng., 39: 765-774.

HORTON, H, L. MORAN, R. OCHS, J. RAWN, K. SCRIMGEOUR (1994), Principes de biochimie, Bruxelles, De Boeck-Wesmael S.A.

HUANG, L.-C., C.-M. CHI, H. VITS, E.J. STABA (1993). Population and biomass kinetics in fed-batch culture of *Daucus carota L.* somatic embryos. Biotechnol. Bioeng., 41: 811-818.

JOLICOEUR, M.C. CHAVARIE, P.J. CARREAU, J. ARCHAMBAULT (1992). Development of a helical-ribbon impeller bioreactor for high-density plant cell suspension culture. Biotechnol. Bioeng., 39: 511-521.

KAMEN, A. A.,C. CHAVARIE, G. ANDRÉ, J. ARCHAMBAULT, (1992) Design parameters and performance of a surface baffled helical ribbon impeller bioreactor for the culture of shear sensitive cells. Chem. Eng. Sci., 47: 9-11

LUEDEKING, R., E.L. PIRET (1959). J. Biochem. Microbiol. Technol. Eng., 1: 343-442.

MARTEL, J.-M, R. NADEAU (1988), Statistique en gestion et en économie, Boucherville, Gaëtan Morin Éditeur

PAREILLEUX, A., R. VINAS (1983). Influence of the aeration rate on the growth yield in suspension cultures of *Catharanthus roseus* (L.) G. Don. J. Ferment Technol., 61: 429-433.

PAYNE, G., F.S.W. DAVIDSON, J.L. TATE (1990). Experimentals constraints to studying the effects of dissolved oxygen and dissolved carbon dioxide on plant cell growth. Developments in industrial Microbiology, 31 (Journal Article of Industrial Microbiology, suppl. no 5): 293-301.

PÉPIN, M.-F., J. ARCHAMBAULT, C. CHAVARIE, F. CORMIER (1995). Growth kinetics of Vitis vinifera cell suspension cultures: I. Shake flask cultures. Biotechnol. Bioeng., 47: 131-138.

PIELH, G.-W., J. BERLIN, C. MOLLENSCHOTT, J. LEHMANN (1988). Growth and alkaloid production of a cell suspension culture of *Thalictrum rugosum* in shake flasks and membrane-stirrer reactors with bubble free aeration. Appl. Microbiol Biotechnol, 29: 456-461.

PIRT, S.J. (1965). The maintenance energy of bacteria in growing cultures. Proc. Roy. Soc. Lon. (Biol.), 163: 224-231.

QUINLAN, A.V. (1986). A semicontinuous culture model that links cell growth of extracellular nutrient concentration. Biotechnol. Bioeng., 28: 1455-1461.

RHO, D., G. ANDRÉ (1991). Growth and stoichiometry of a *Catharanthus roseus* cell suspension culture grown under nitrogen limiting conditions. Biotechnol. Bioeng., 38: 579-587.

ROKEM, J.S., I. GOLDBERG (1985). Secondary metabolites from plant cell suspension cultures: Methods for yield improvement. Advances in biotechnological processes. I. Alan R. Liss. 4: 241-274.

ROSE, D. ET S.M. MARTIN (1973). Parameters for growth measurement in suspension cultures of plant cells. Can. J. Bot., 52: 903-912.

ROSE, D., S.M. MARTIN, P.P.F. CLAY (1972). Metabolic rates for major nutrients in suspension culutres of plant cells. Can. J. Bot., 50: 1301-1308.

SCHIEL, O.K., J.-REDECKER, G.W. PIEHL, J. LEHMANN (1984). Increased formation of cinnamoyl-putresciness by fedbatch fermentation of cell suspension cultures of *Nicotiana tabacum*. Plant Cell Rep., 3: 18-20.

SCHNAPP, S.R., W.R. CURTIS, R.A. BRESSAN, P.M. HASEGAWA (1991). Estimation of growth yield and maintenance coefficient of plant cell suspensions. Biotechnol. Bioeng., 38: 1131-1136.

SHIBASAKI, N.T., YONEMOTO, T. TADAKI (1993). Application of five-state structured model for growth and secondary metabolite production in culture of *Nicotiana tabacum*. J. Chem. Tech. Biotechnol, 0268-2575/93.

SMART, N.J., M.W. FOWLER (1984). An Airlift column bioreactor suitable for large-scale cultivation of plant cell suspension. Journal of Experimental Botany, 34(154): 531-537.

SPIELER, H., A.W. ALFERMANN, E. REINHARD (1985). Biotransformation of Beta-methyldigitoxin by cell cultures of *Digitalis lanata* in airlift and stirred tank reactors. Appl. Microbiol. Biotechnol., 23: 1-4.

SUCKER, H, .G.M. JORDAN, H.M. EPPENBERGER, F. WIDMER (1994). Bubble bed reactor: a reactor design to minimise the damage of bubble aeration on animal cells. Biotechnol. Bioeng., 44: 1246-1254.

TATICEK, R., A.M. MOO-YOUNG, R.L. LEGGE (1990). Effect of bioreactor configuration on substrate uptake by cell suspension cultures of the plant *Eschscholtzia californica*. Appl. Microbiol. Biotechnol., 33: 280-286.

TATE, J.L., G.F. PAYNE (1991). Plant cell growth under differents levels of oxygen and carbon dioxide. Plant Cell Reports, 10: 22-25.

TSUCHIYA, H.M, .A.G. FREDRICKSON, R. ARIS (1966). Dynamics of microbial cell populations. Adv. Chem. Eng., 6: 125.

VAN GULIK, W.M., J.J. MEIJER, H.J.G. TEN HOOPEN, K.C.A.M. LUYBEN (1989). Growth of a Catharanthus roseus cell suspension culture in a modified chemostat under glucose limiting conditions. Appl. Microbiol. Biotechnol., 30: 270.

VAN GULIK, W.M.H., J.G. TEN HOOPEN, J.J. HEIJNEN (1992). Kinetics and stoichiometry of growth of plant cell culture of *Catharanthus roseus* and *Nicotiana tabacum* in batch and continous fermentors. Biotechnol. Bioeng., 40: 863-874.

VAN GULIK, W.M.H., J.G. TEN HOOPEN, J.J. HEIJNEN (1993). A structured model describing carbon and phosphate limited growth of *Catharanthus roseus* plant cell suspension in batch and chemostat culture. Biotechnol. Bioeng., 41: 771-780.

WANG, N.G., G. STEPHANOPOULOS (1983). Application of macroscopic balances to the identification of growth measurement errors. Biotechnol. Bioeng., 20: 2177-2208.

WETTER, L. R., CONATABEL, F., (1982). Plant tissue culture method, 2nd ed., Saskatoon, Conseil national de la recherche du Canada.

WILLIAMS, R.D., C. BÉDARD, C. CHAVARIE, J. ARCHAMBAULT (1996). Production of sanguinarine by plant cell culture II. Further nutritional aspects., Biotechnol. Bioeng., 46: 107-120.

WILLIAMS, R.D., N. CHAURET, C. BÉDARD, J. ARCHAMBAULT (1992). Effect of polymeric adsorbents on the production of sanguinarine by *Papaver somniferum* cell culture. Biotechnol. Bioeng. 40: 971-977.

Annexe A. Résultats expériementaux

A.1 Graphiques des données expérimentales

Pour les figures A-1 à A-14, les graphiques présentent les données expérimentales telles que recueillies lors des échantillonnage pour les cultures présentées au Tableau 5-1.

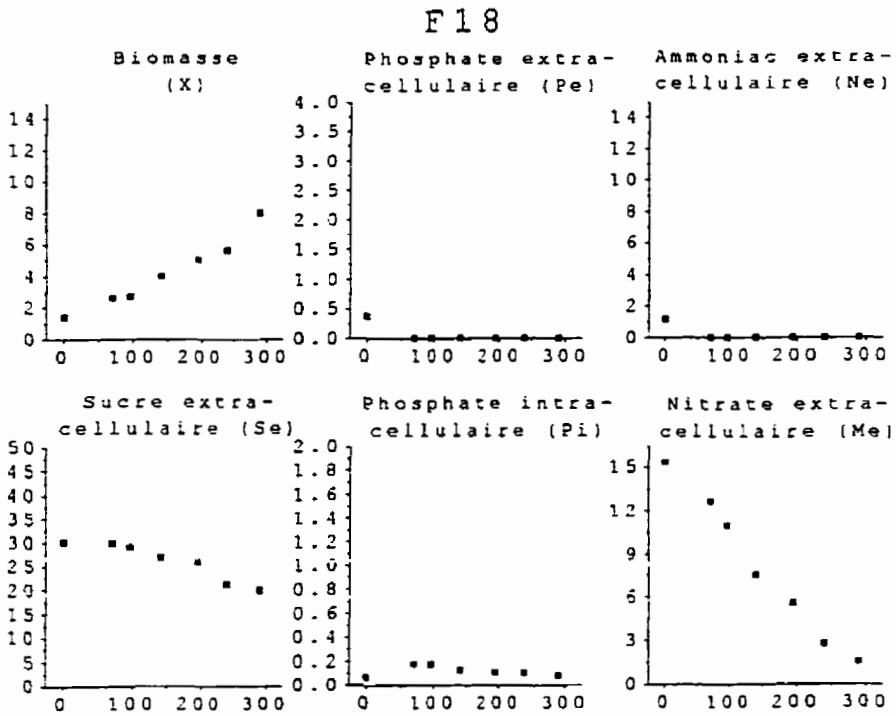


Figure A-1 Résultats bruts de la culture F18

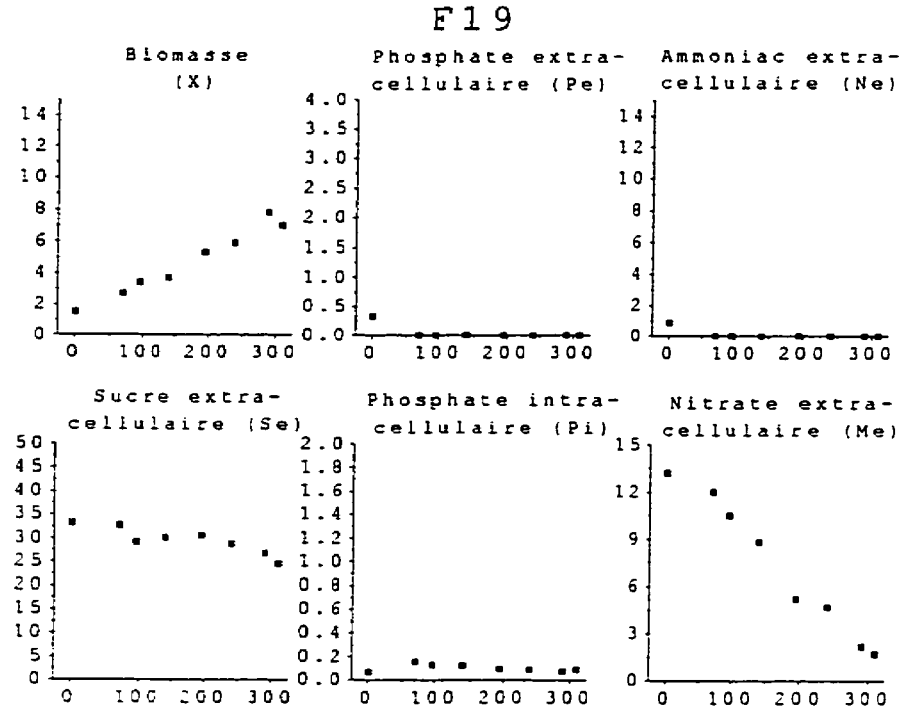


Figure A-2 Résultats bruts de la culture F19

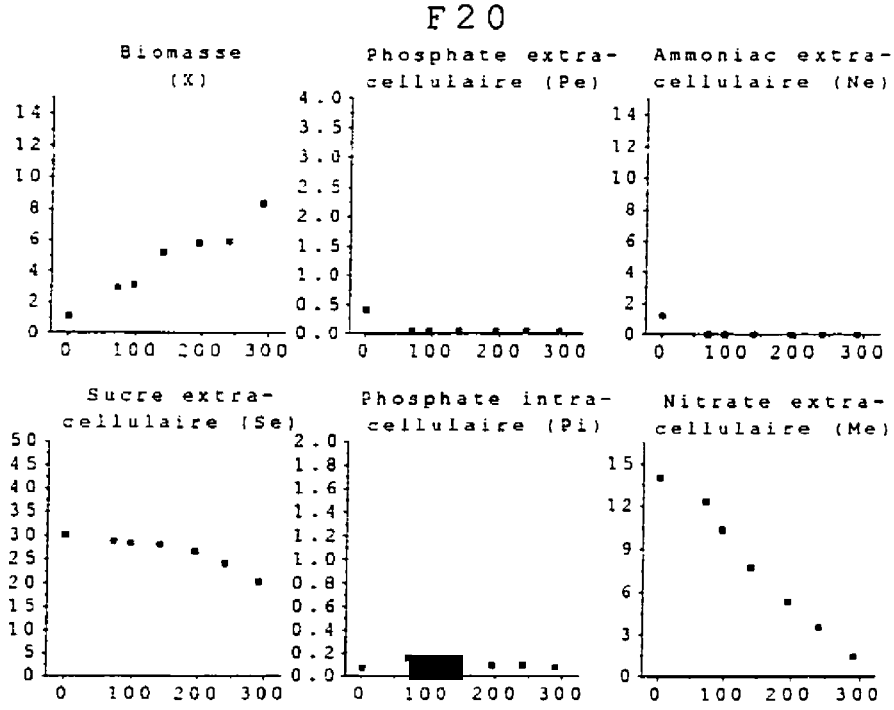


Figure A-3 Résultats bruts de la culture F20

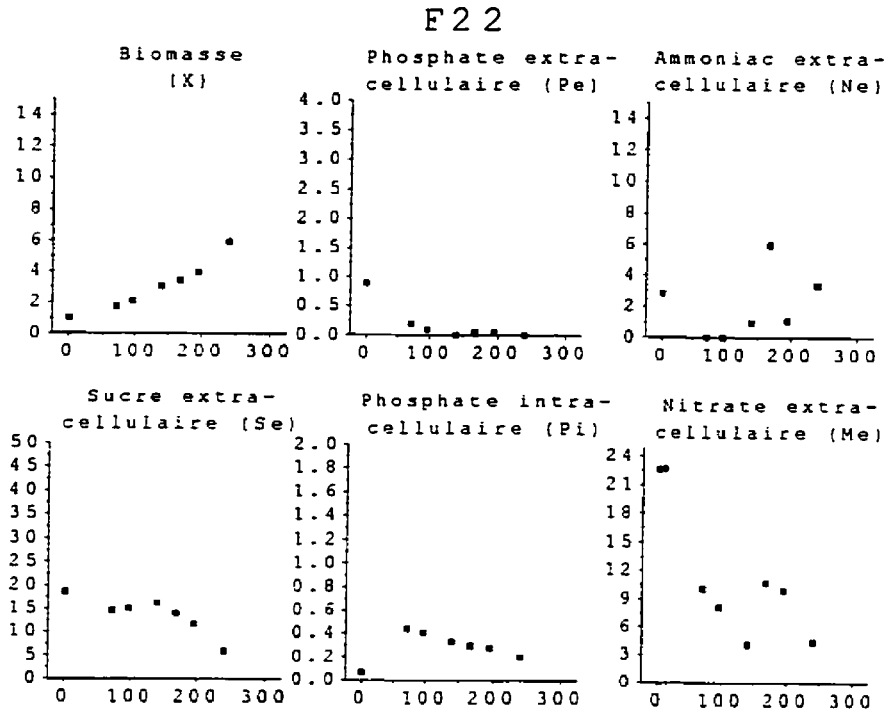


Figure A-4 Résultats bruts de la culture F22

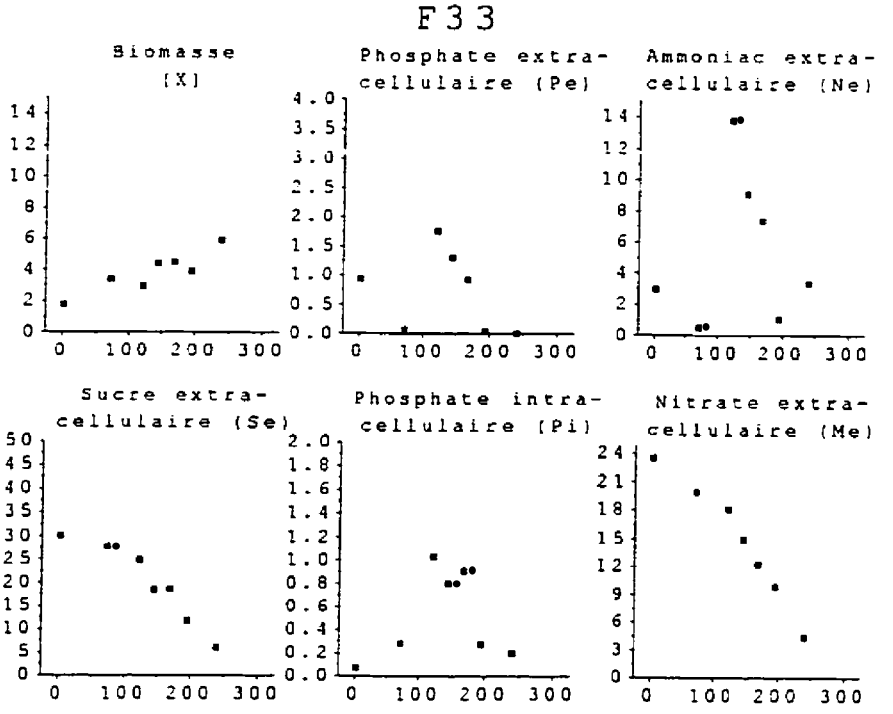


Figure A-5 Résultats bruts de la culture F33

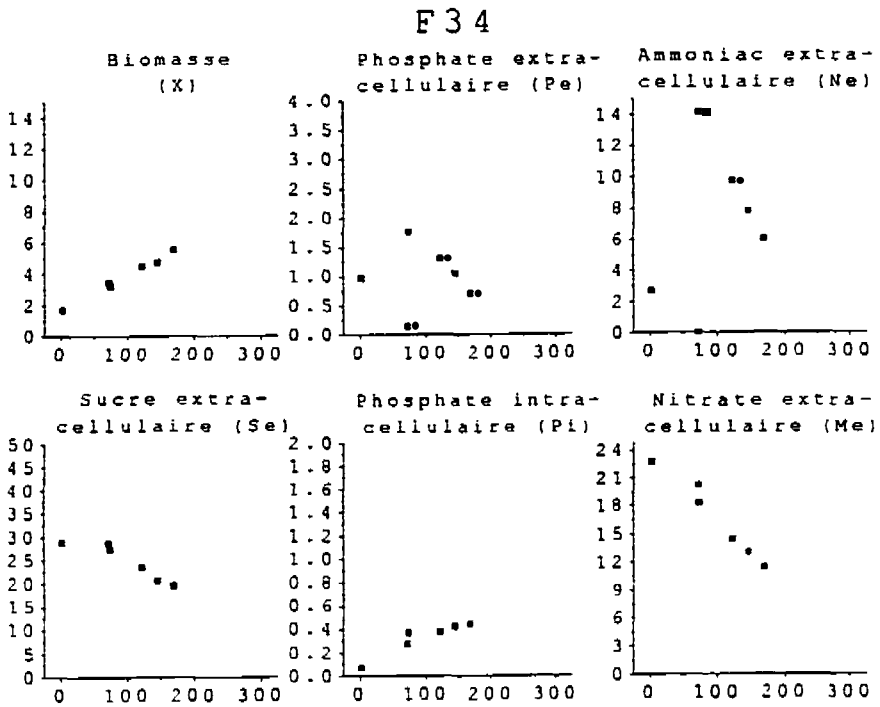


Figure A-6 Résultats bruts de la culture F34

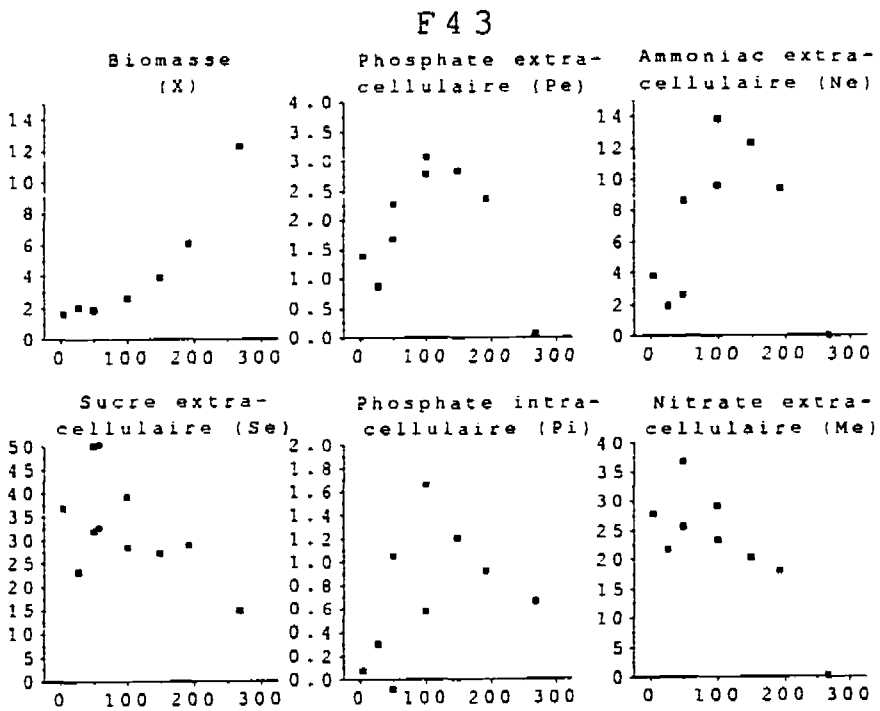


Figure A-7 Résultats bruts de la culture F43

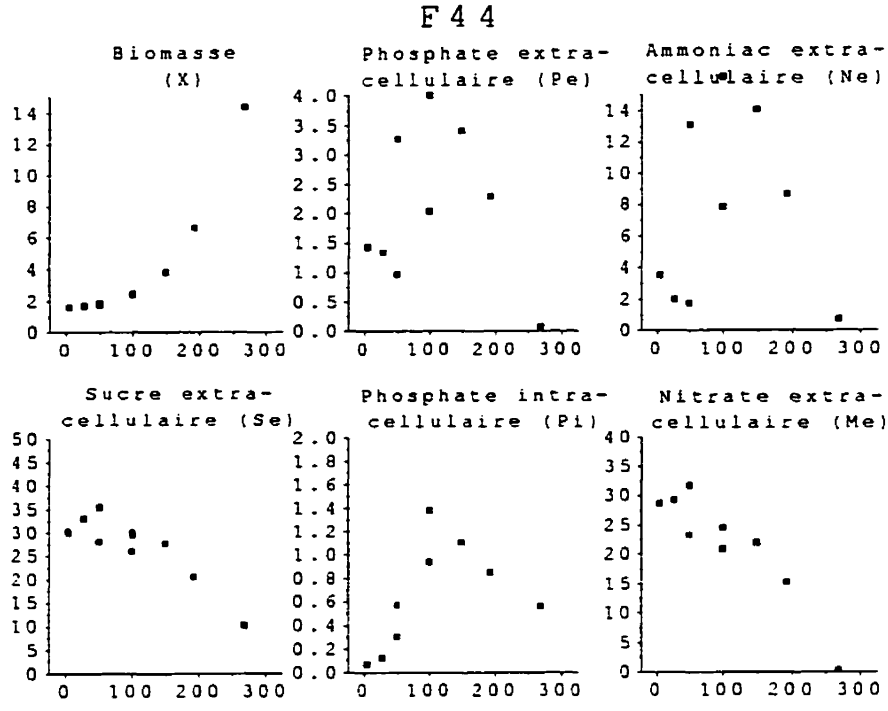


Figure A-8 Résultats bruts de la culture F44

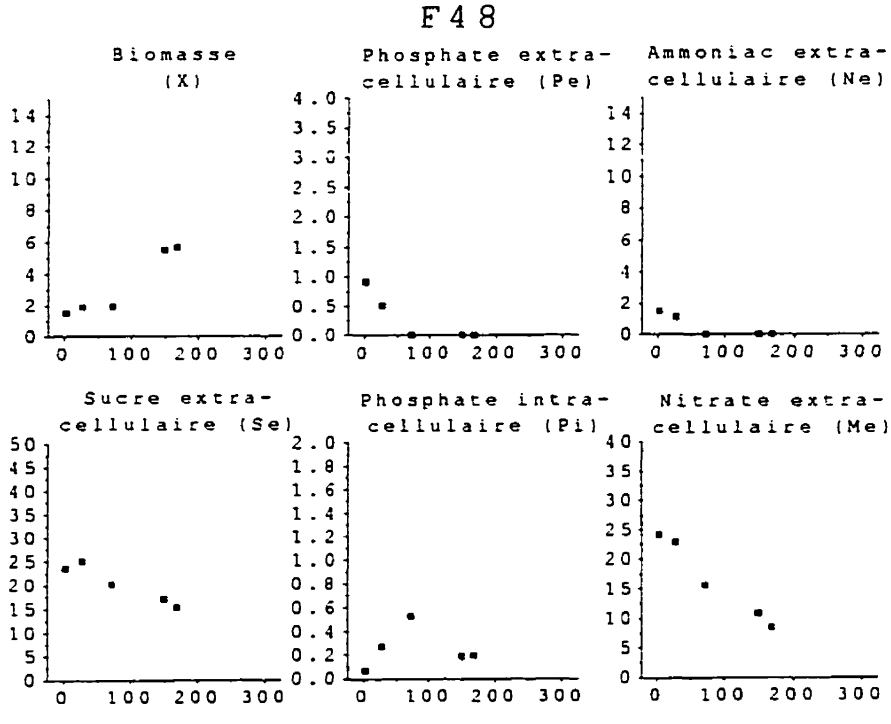


Figure A-9 Résultats bruts de la culture F48

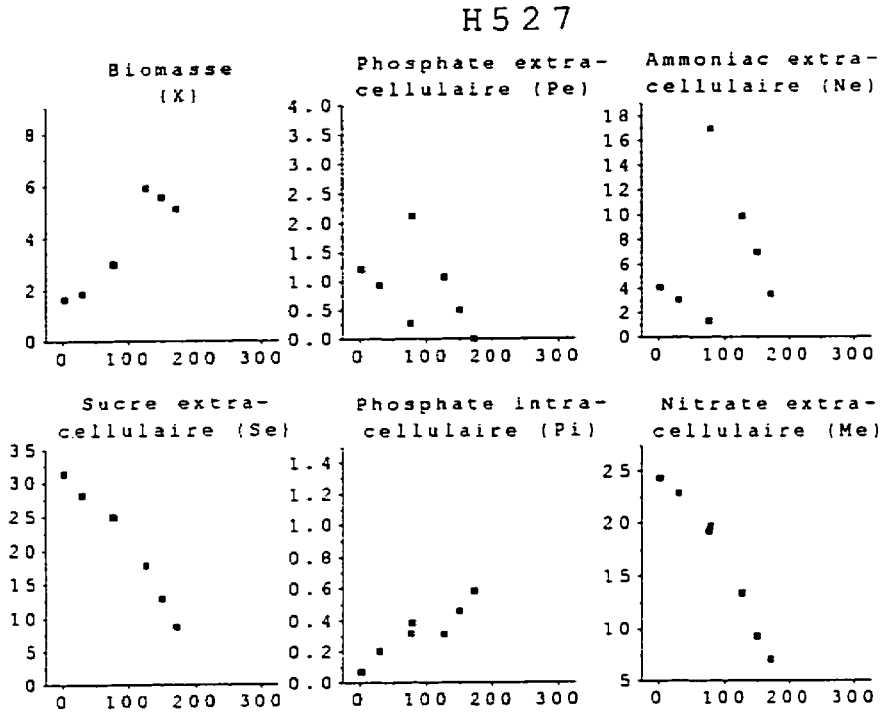


Figure A-10 Résultats bruts de la culture H527

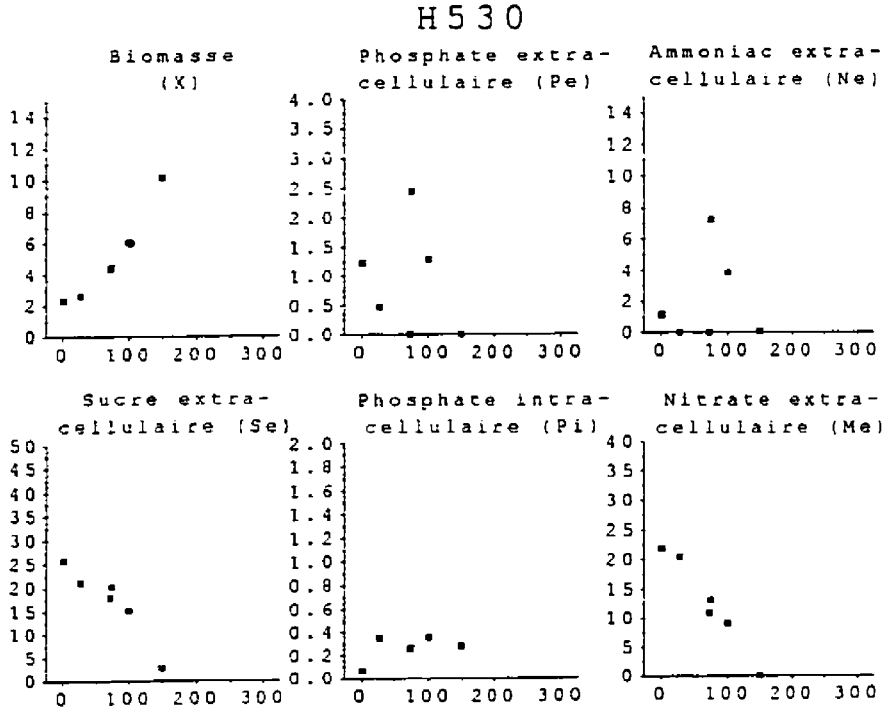


Figure A-11 Résultats bruts de la culture H530

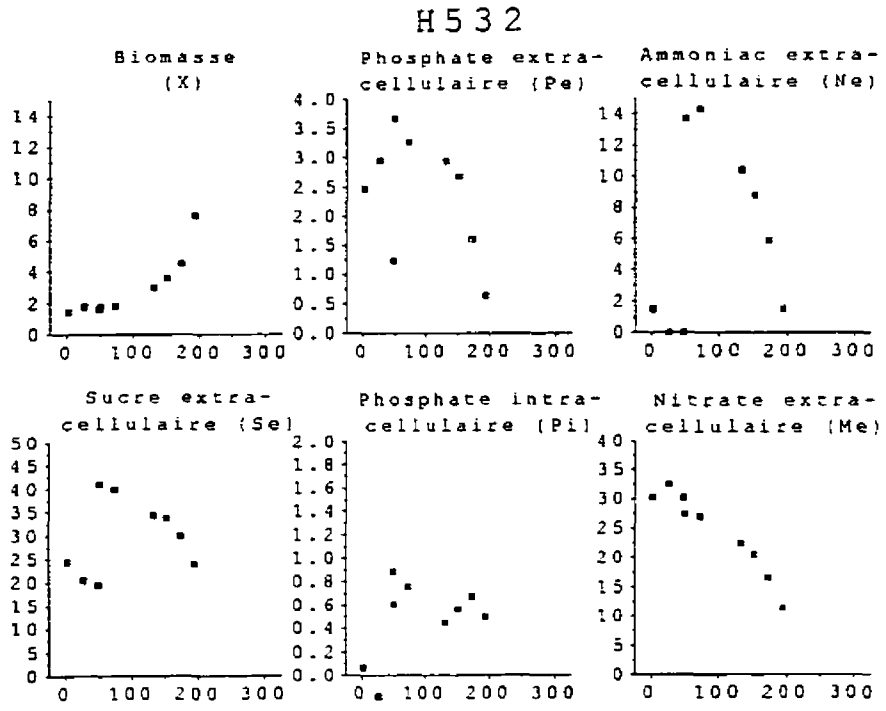


Figure A-12 Résultats bruts de la culture H532

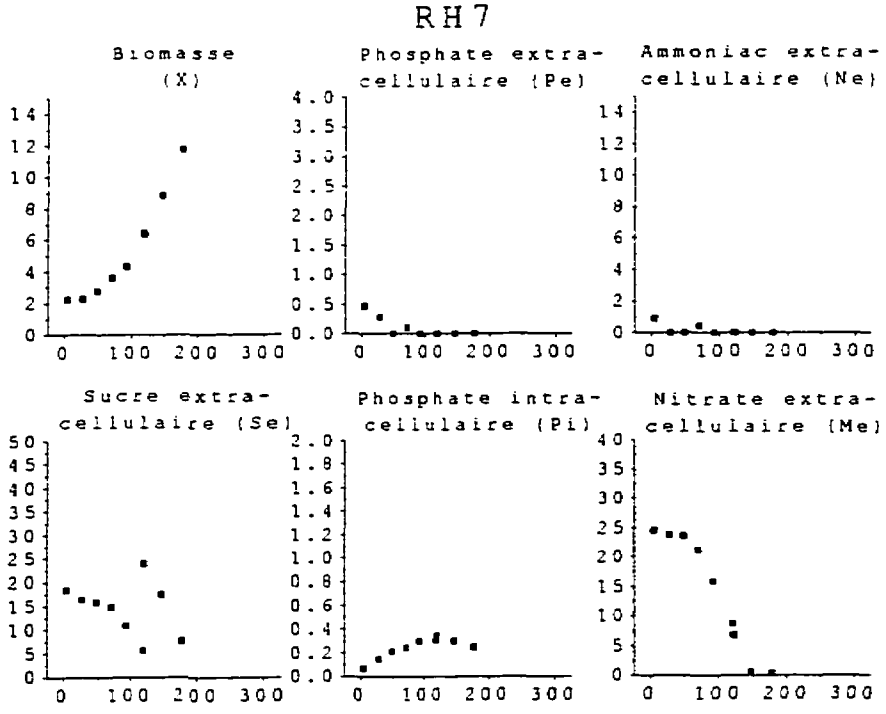


Figure A-13 Résultats bruts de la culture RH7

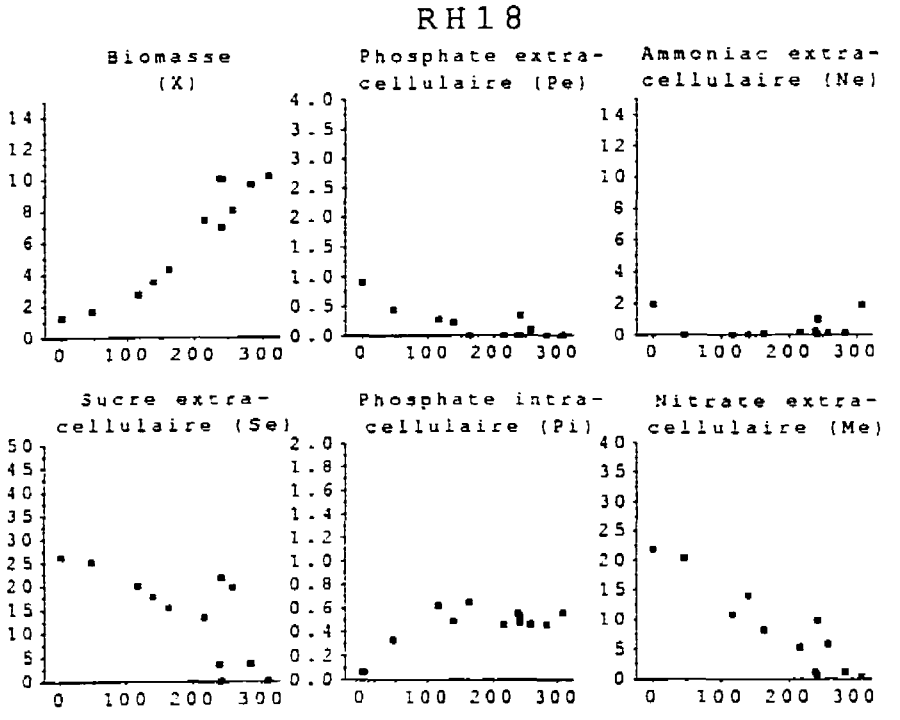


Figure A-14 Résultats bruts de la culture RH18

A.2 Description des cultures analysées

Certaines des cultures présentées à l'annexe A.1 semblent interrompues avant d'avoir atteint un plafond dans la biomasse. Ceci ce produit losrque ces cultures ont été élicitées. Les données au delà de l'élicitation des flacons ne sont par montrées et ne participent pas à l'étude de la modélisation.

Les cultures F18, F19 et F20 proviennent du même inoculum. Les trois cultures présentent une croissance plutôt linéaire. Sans ajout, il s'agit d'un échantillon typique de croissance des inocula utilisés pour la sous-culture. On constate que le phosphate et l'ammoniac sont absorbés en moins de 75h. Les valeurs non nulles de phosphate après ce temps sont dues à un biais dans la mesure de l'analyseur ionique. Le sucre et le nitrate sont consommés régulièrement jusqu'à la fin de la croissance.

Les cultures F32, F33 et F34 proviennent d'un même inoculum. La culture 32 a été réalisée sans ajouts. Les cultures F33 et F34 ont reçu un ajout de 40 mmol de $(NH_4)_2HPO_4$ et de 40 mmol de $(NH_4)_2SO_4$ à la 74ième heure par ajouts de 10 ml chacuns. La présence d'une forte concentration d'ammoniac (14 mM) dans le milieu semble avoir nuit à la croissance des cellules. En effet, le maximum atteint après 174 heures est de 5 et 6 g/l pour les cultures avec ajout, comparativement à 7 g/l pour la culture sans ajout.

Les cultures F43 et F44, réalisées simultanément à la culture en bioréacteur H529, ont connu une très bonne croissance, atteignant 13 et 14 g/l respectivement en 275 heures.

Le comportement erratique des courbes de nutriments reste à expliquer. Le dernier échantillon de H529 a été effecuté avant l'élicitation. À ce moment (à la 192ième heure) la biomasse avait atteint près de 6 g/l soit sensiblement les mêmes valeurs (6 à 7 g/l) que celles obtenues après cette période par les cultures témoins en flacons.

Les deux cultures sans ajout F47 et F48 furent réalisées conjointement et simultanément à la culture en bioréacteur H530 et présentent une croissance plutôt faible. La biomasse n'a pas atteint plus de 7 g/l en 168 heures. La phase de latence a duré plus de 75 heures. L'ammoniac et le phosphate étaient pratiquement entièrement consommés après 75 heures. La culture H530 avait atteint 10 g/l en 150 heures. L'élicitation fut éffectuée à ce moment. Les cultures témoins sans ajouts (F47 et F48) n'ont pas atteint 7 g/l en 175 heures. La croissance fut donc beaucoup plus rapide en réacteur dans ce cas.

La culture H527 présente une croissance de 6 g/l avant un plafonnement pour des raisons autres qu'un manque de sucre ou d'azote. La présence de plus de 15 mM d'ammoniac et de 2 mM de phosphate dans le milieu suite à l'unique ajout ne semble pas avoir ralenti la croissance qui a doublé durant les 48 heures suivant la 78ième heure.

La culture H532 présente une phase de latence de plus de 75 heures. Cette phase a peut être été prolongée à cause de la présence d'ammoniac et de phosphate a concentration élevée dès la 50ième heure à cause d'un ajout hâtif.

La culture RH7 présente un ajout proportionnel réussi. On constate une phase de latence relativement courte et une croisance soutenue entre les heures 100 et 200. La

biomasse diminue par la suite, vraisemblablement dû à un épuisement du sucre ou du nitrate dans le milieu. Tout le phosphate et l'ammoniac ajouté a été consommé au fur et à mesure de l'alimentation par la biomasse, d'où des concentrations non détectables dans le milieu après la 50ième heure, à l'exception de la 75ième heure ou de faibles concentrations sont détectées. L'ajout d'ammoniac et de phosphate semble avoir eu un effet très positif sur la croissance de la biomasse qui a atteint près de 14 g/l en moins de 200 heures.

La culture RH18 présente encore le profil d'ajout graduel. L'ammoniac et le phosphate sont demeurés à des niveaux faibles ou nuls durant la période d'ajout, sauf pour le dernier échantillonage où l'ion ammoniac a été détecté à une concentration de 2 mM. Un ajout de milieu B5 a été fait au même moment que l'ajout de sucre de façon à augmenter le volume de culture à un niveau raisonnable pour poursuivre la culture et les échantillonnages. Sans la dilution, la biomasse aurait atteint un niveau semblable à celui de la culture RH7.

Annexe B. Transformations mathématiques

Les manipulations algébriques et matricielles qui suivent sont données à titre d'exemple à l'aide de ce premier schéme réactionnel simple. Les mêmes transformations sont appliquées aux schèmes plus complexes mais seuls les résultats seront présentés.

B.1 Transformations appliquées au premier schème

Le schème le plus simple permettant de décrire la croissance de la biomasse (X) consiste à la relier à l'utilisation du sucre (S) selon l'équation 5-1 présentée à la section 5-1.

$$sS \xrightarrow{r_1} X \tag{5-1}$$

Ce schème peut s'exprimer sous forme d'un bilan de masse.

$$\frac{dX}{dt} = r'_1(t) \tag{B-1}$$

$$\frac{dS}{dt} = -sr'_1(t) \tag{B-2}$$

On peut développer un estimateur linéaire de biomasse à partir des deux équations différentielles précédentes. Certaines variables mesurées tels les débits d'alimentation et d'échantillonnage (F^{in} et F^{out}), les concentrations de l'alimentation et du milieu (S^{in}, S) et la biomasse (X). Le taux de réaction r_1 peut être éliminée de ces bilans par de simples manipulations algébriques, laissant une relation linéaire entre la production de biomasse et la consommation de sucre.

Présenté sous la forme matricielle d'un système d'état, les bilans précédents peuvent s'exprimer sous la forme suivante:

$$\frac{d}{dt}\begin{bmatrix} X(t)V(t) \\ S(t)V(t) \end{bmatrix} = \begin{bmatrix} 1 \\ -s \end{bmatrix}\begin{bmatrix} r_1 \end{bmatrix} + \begin{bmatrix} -F^{out}(t)X(t) \\ +F^{in}(t)S^{in} - F^{out}(t)S(t) \end{bmatrix} \tag{B-3}$$

La résolution numérique du terme de gauche implique une différentiation du produit X(t)*V(t). En traitement de signal, cette différentiation correspond à un filtre passe-haut qui amplifie le bruit de mesure et détériore la précision de l'estimation. Une façon de circonvenir ce problème consiste à intégrer l'équation précédente par rapport au temps.

L'intégration de 0 à t donne:

$$\begin{bmatrix} X(t)V(t) \\ S(t)V(t) \end{bmatrix} - \begin{bmatrix} X(0)V(0) \\ S(0)V(0) \end{bmatrix} = \int_0^t \begin{bmatrix} 1 \\ -s \end{bmatrix}\begin{bmatrix} r_1 \end{bmatrix} d\tau + \int_0^t \begin{bmatrix} -F^{out}(\tau)X(\tau) \\ +F^{in}(\tau)S^{in} - F^{out}(\tau)S(\tau) \end{bmatrix} d\tau \tag{B-4}$$

Les notations suivantes sont introduites:

$$\Delta X(t) = X(t)V(t) - X(0)V(0) + \int_0^t F^{out}(\tau)X(\tau)d\tau \tag{B-5}$$

$$\Delta S(t) = S(t)V(t) - S(0)V(0) + \int_0^t (F^{out}(\tau)S(\tau) - F^{in}(\tau)S^{in}(\tau))d\tau \tag{B-6}$$

$$R_1 = \int_0^t r_1(\tau)d\tau \tag{B-7}$$

Ici les variables précédées du symbole Δ signifient une consommation ou production totale au temps t par rapport au premier échantillonnage de la culture exprimée en

grammes ou en mmoles. L'utilisation des nouvelles variables construites qui incluent les termes d'échantillonnage et d'ajout permettent de simplifier l'écriture du système en mettant en évidence les variations de biomasse et de nutriments causées par la seule croissance de la biomasse.

Le bilan peut être réécrit:

$$\begin{bmatrix} \Delta X(t) \\ \Delta S(t) \end{bmatrix} = \begin{bmatrix} 1 \\ -s \end{bmatrix} \left[R_1(t) \right] \tag{B-8}$$

Et de façon explicite:

$$\Delta X = R_1(t) \tag{B-9}$$

$$\Delta S = -sR_1(t) \tag{B-10}$$

On peut maintenant éliminer R_1 et obtenir une relation linéaire entre ΔX et ΔS.

$$\Delta X(t) = -\frac{1}{s} \Delta S(t) \tag{B-11}$$

Ce schème simple est donc similaire à la première définition de rendement $Y_{X/S}$ présentée dans Bailey et Ollis (Bailey 1986, voir équation 2-1).

La connaissance de "s" permettrait d'estimer la biomasse. Cependant comme ce coefficient est inconnu, il devra être déterminé à partir des données expérimentales. En effectuant une simple paramétrisation sur les coefficients des variables mesurées ou calculées, dans ce cas

$$k_s = -\frac{1}{s} \tag{B-12}$$

on obtient une équation qui pourra être résolue par une simple régression linéaire

$$\Delta X(t) = k_s \Delta S(t) \tag{B-13}$$

On pourra parfois, mais ce n'est pas toujours le cas pour les schèmes plus complexes, déterminer les coefficients pseudo-stoechiométriques du schème réactionnel qui sous-tend le bilan de masse. Dans ce cas-ci

$$s = -\frac{1}{k_s} \tag{B-14}$$

Ces opérations sont triviales pour ce premier schème réactionnel et sont effectuées à titre d'exemple. La même démarche est appliquée aux autres schèmes.

B.2 Transformations appliquées au second schème

Le second schème est le suivant:

$$\begin{aligned} s_1 S &\xrightarrow{r_1} X \\ s_2 S + X &\xrightarrow{r_2} X \end{aligned} \tag{B-15}$$

Sous forme de système d'état, le bilan de masse peut s'écrire:

$$\frac{d}{dt}\begin{bmatrix} X(t)V(t) \\ S(t)V(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -s_1 & -s_2 \end{bmatrix}\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} + \begin{bmatrix} -F^{out}(t)X(t) \\ +F^{in}(t)S^{in} - F^{out}(t)S(t) \end{bmatrix} \tag{B-16}$$

Après intégration et changement de variable, tel que présenté à la section précédente, on obtient l'équation de bilan

$$\begin{bmatrix} \Delta X \\ \Delta S \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -s_1 & -s_2 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \tag{B-17}$$

et quelques manipulations, on peut obtenir l'estimateur de biomasse

$$\Delta X(t) = -\frac{1}{s_1}\Delta S(t) - \frac{s_2}{s_1}R_2(t) \tag{B-18}$$

A priori, R_2 ne peut être éliminé algébriquement. Cependant, le taux de maintenance r_2 est proportionnel à la quantité de biomasse dans le milieu et peut être remplacé par

$$r_2(t) = k_2 X(t)V(t) \tag{B-19}$$

En effectuant le changement de variable

$$XV(t) = \int_0^t X(\tau)V(\tau)d\tau \tag{B-20}$$

L'intégration donne

$$R_2(t) = k_2 XV(t) \tag{B-21}$$

Cette équation peut être incluse au bilan présenté à l'équation B-17 qui devient

$$\begin{bmatrix} \Delta X \\ \Delta S \\ XV \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -s_1 & -s_2 \\ 0 & 1/k_2 \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \tag{B-22}$$

et de laquelle on peut obtenir le système linéaire

$$\Delta X(t) = k_S \Delta S(t) + k_{XV} XV(t) \tag{B-23}$$

avec $k_S = -1/s_1$ et $k_{XV} = -s_2 k_2 / s_1$

B.3 Estimation par les moindres carrés

Les équations algébriques permettant la résolution de systèmes linéaires sont simples et bien connues. La résolution de systèmes multivariables est généralement réalisée sous forme matricielle car il est pratiquement impossible de traduire algébriquement le système d'équations linéaires lorsque le nombre d'équation est supérieur à 4.

Dans cette section des changements de variable sont proposés de façon à présenter l'équation d'estimation des moindres carrés sous une forme classique. Les variables y, X et b sont définies à partir des variables obtenues aux sections précédentes et correspondent à la nomenclature statistique standard (y=Xb) telle qu'on la retrouvre dans le livre de Martel et Nadeau (1988) qui sert de référence à cette section.

Un indice doit être ajouté aux équations définies précédemment de façon à marquer les différentes cultures. Ainsi, ΔX_i correspond aux données de biomasse obtenues de la i-ième culture choisie pour la régression et pour la cinétique.

Pour la suite des démontrations, on définit $\Delta\xi_i$ qui peut prendre les valeurs d'un des paramètres mesurés ou calculés (ΔS, ΔP_i, ΔP_c, ΔN ou ΔM). La matrice Ψ_i correspond au Ψ de la i-ième culture tel que défini par l'équation 7-9.

Donc, soit "y" le vecteur solution constitué des valeurs de biomasse pour chacune des cultures sur lesquelles s'applique la régression

$$y = \begin{bmatrix} \Delta X_1 \\ \Delta X_2 \\ \vdots \\ \Delta X_r \end{bmatrix} \tag{B-24}$$

où $\Delta X_i = [\Delta X_i(t_0)\ \Delta X_i(t_1)\ \ldots\ \Delta X_i(t_{ki})]'$,

Soit X la matrice des variables explicatives, c'est-à-dire les paramètres mesurés choisis pour la modélisation. La composition de cette matrice est différente selon qu'on modélise les rendements où la cinétique. L'équation B-25-a présente la forme utilisée pour calculer les rendements et l'équation B-25-b donne la forme utilisée pour calculer les cinétiques.

$$X = \begin{bmatrix} \Delta\xi_{1-1} & \Delta\xi_{1-2} & \cdots & \Delta\xi_{1-p} \\ \Delta\xi_{2-1} & \Delta\xi_{2-2} & \cdots & \Delta\xi_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta\xi_{r-1} & \Delta\xi_{r-2} & \cdots & \Delta\xi_{r-p} \end{bmatrix} \tag{B-25-a}$$

$$X = \begin{bmatrix} \Psi_{1-1} & \Psi_{1-2} & \cdots & \Psi_{1-p} \\ \Psi_{2-1} & \Psi_{2-2} & \cdots & \Psi_{2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{r-1} & \Psi_{r-2} & \cdots & \Psi_{r-p} \end{bmatrix} \tag{B-25-a}$$

Soit b le vecteur des coefficients inconnus défini par les équation B-26-a et B-26-b

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} \tag{B-26-a}$$

ou chaque élément b_i correspond aux coefficients k à déterminer (k_S, k_N, k_M, ...)

$$b = \Theta' \tag{B-26-b}$$

où Θ est le vecteur des coefficients du modèle polynômial.

Le système multilinéaire à résoudre est donc décrit par la forme matricielle compacte standard définit à l'équation B-27.

$$y = Xb \tag{B-27}$$

La minimisation de la somme des carrés des erreurs empiriques e_i revient à trouver le minimum de la fonction Q définie par

$$Q(b_0, b_1, ..., b_p) = \arg \min_{(b_0, b_1, ..., b_p)} \sum e^2 \tag{B-28}$$

Étant donné "e" le vecteur des résidus calculés pour chacun des points

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = [y - Xb] \tag{B-29}$$

alors la fonction "Q" s'écrit sous la forme matricielle suivante

$$\begin{aligned} Q(b_0, b_1, ..., b_p) &= \arg \min_{(b_0, b_1, ..., b_p)} e'e \\ &= (y - Xb)'(y - Xb) \\ &= y'y - 2b'X'y + b'X'Xb \end{aligned} \tag{B-30}$$

Si l'on dérive successivement cette fonction par rapport à b_0, b_1, ..., b_p, on obtient le système d'équations normales suivant:

$$X'Xb = X'y \tag{B-31}$$

Et en conséquence les estimateurs de moindres carrés sont donnés par l'équation B-32. Le vecteur solution $\hat{b}$ correspond donc soit aux coefficients k ou aux coeffcients des polynômes recherchés.

$$\hat{b} = (X'X)^{-1}X'y \tag{B-32}$$

À titre d'exemple, pour résoudre le premier schème présenté à l'équation B-13 à partir des données d'une seule culture de façon matricielle, on obtient

$$\begin{aligned}\hat{k}_S &= \hat{b} = (X'X)^{-1}X'y \\ &= (\Delta\xi_1'\Delta\xi_1)^{-1}\Delta\xi_1'\Delta X_1 \\ &= (\Delta S_1'\Delta S_1)^{-1}\Delta S_1'\Delta X_1\end{aligned} \tag{B-33}$$

B.4 Préparation des données pour la régression

Cette section présente la démarche mathématique permettant la construction des combinaisons de régresseurs à partir des données expérimentales.

B.4.1 Construction de la forme intégrée matricielle

Le taux de croissance spécifique est au coeur d'un bilan de masse de réaction autocatalytiques comme c'est le cas dans les systèmes biologiques.

À partir d'un bilan de masse sur la biomasse

$$\frac{dX(t)V(t)}{dt} = \mu(t)X(t)V(t) - F^{OUT}(t)X(t) \tag{B-34}$$

qu'on peut réécrire sous la forme

$$\frac{dX(t)V(t)}{dt} + F^{OUT}(t)X(t) = \mu(t)X(t)V(t) \tag{B-35}$$

on constate que le terme de gauche est complètement déterminé par les mesures et que le terme de droite contient le terme cinétique à modéliser (μ).

Deux alternatives se présentent pour solutionner ce système. On peut isoler $\mu(t)$ de l'équation précédente avant de résoudre. Il est également possible de travailler avec l'équation ainsi posée. La seconde avenue est choisie afin de maintenir une plus grande cohésion avec les opérations effectuées au chapitre précédent.

Pour des raisons expliquées précédemment, le bilan est intégré par rapport au temps.

$$\int_0^t \left(\frac{dX(\tau)V(\tau)}{dt} + F^{OUT}(\tau)X(\tau) \right) d\tau = \int_0^t \mu(\tau)X(\tau)V(\tau)d\tau \tag{B-36}$$

La résolution du terme de gauche donne lieu au changement de variable utilisé aux chapitres précédents

$$\Delta X(t) = X(t)V(t) - X(0)V(0) + \int_0^t F^{out}(\iota)X(\iota)d\iota \tag{B-37}$$

Avant de présenter les manipulations du terme de droite, il est utile de rappeler que la forme proposée pour le modèle cinétique est une forme polynomiale.

$$\mu(t) = f(\xi_1, \xi_2) = p_o + p_{10}\xi_1(t) + p_{01}\xi_2(t) + p_{00_1}\xi_n \ldots + p_{20}\xi_1^2(t) + p_{11}\xi_1\xi_2 + \ldots \tag{B-38}$$

où ξ_1, ξ_2, ξ_n représentent les paramètres utilisés dans le modèle cinétique (concentration en sucre, en phosphate, en nitrate, ...). La structure linéaire de l'expression est conservée à travers l'intégration, comme il est démontré plus loin.

Le terme de droite de l'équation B-36 peut s'exprimer

$$\int_0^t \mu(\tau)X(\tau)V(\tau)d\tau = p_o\int_0^t X(\tau)V(\tau)d\tau + p_{10}\int_0^t \xi_1(t)X(\tau)V(\tau)d\tau + p_{01}\int_0^t \xi_2(t)X(\tau)V(\tau)d\tau + \ldots$$
$$+p_{00\ldots1}\int_0^t \xi_n X(\tau)V(\tau)d\tau + \ldots + p_{20}\int_0^t \xi_1^2(t)X(\tau)V(\tau)d\tau + p_{11}\int_0^t \xi_1\xi_2 X(\tau)V(\tau)d\tau + \ldots \tag{B-39}$$

Une écriture compacte permet de simplifier l'écriture et les manipulations sur les paramètres.

Par exemple, le polynôme

$$p_0 + p_1\xi_1(t) + p_2\xi_1(t)\xi_2(t) \tag{B-40}$$

peut s'écrire sous la forme matricielle compacte $\Theta\Phi(t)$ où $\Phi(t)=[1\ \xi_1(t)\ \xi_1(t)\xi_2(t)]^T$, les variables du polynôme et où $\Theta=[p_0\ p_1\ p_2]$ est le vecteur des coefficients correspondants.

Afin de poursuivre le formalisme, quelques définitions sont nécessaires. Soit

- ρ le nombre de paramètre pouvant être utilisés dans la régression ([Se], [PO_4e], etc.);
- ν le degré maximal du polynome, c'est-à-dire la somme maximale des exposants de chacun des monomes;
- m le nombre de régresseurs par polynome avec ρ paramètre et d'ordre ν;
- ω le nombre total de monomes possibles;
- $\phi_i(t)$ un régresseurs, c'est-à-dire un produit de paramètres (ξ). Chaque ϕ_i correspond donc à la variable d'un monome; et

- J_i un ensemble comprenant au plus ν entiers correspondant à chacun des monômes.

Mathématiquement, tout régresseur ϕ_v est défini par

$$\begin{cases} J_i = \{l_1, \ldots, l_v\} \\ \phi_v(t) = \prod_{i \in J_v} \xi_i(t) \end{cases} \tag{B-41}$$

$$\begin{aligned} & l_i \in \{0, \ldots, \rho\} \\ & v = 1, \ldots, \omega \end{aligned} \tag{B-42}$$

avec

$$\xi_0(t) = 1 \quad \forall t \tag{B-43}$$

Ainsi, par exemple pour $J_v=\{1,2,4\}$, le régresseur correspondant $\phi_v(t)$ est $\xi_1(t)\xi_2(t)\xi_4(t)$.

Le modèle polynomial de l'équation B-38 s'exprime

$$\hat{\mu}_v(t) = \sum_{v \in K} \vartheta_v \phi_v(t) \quad \text{avec} \quad \begin{aligned} & K = \{l_1, \ldots, l_m\} \\ & l_i, l_j \in \{1, \ldots, \omega\} \\ & l_i \neq l_j \end{aligned} \tag{B-44}$$

En utilisant une notation matricielle, l'équation des cinétiques peut s'écrire

$$\hat{\mu}(t) = \Theta\Phi(t) \tag{B-45}$$

où la matrice des coefficients Θ est données par

$$\Theta \equiv \begin{bmatrix} \vartheta_1 & \cdots & \vartheta_v & \cdots & \vartheta_\omega \end{bmatrix} \tag{B-46}$$

et où $\Phi(t)$ est définie par

$$\mathbf{\Phi}(t) \equiv \begin{bmatrix} \phi_1(t) \\ \vdots \\ \phi_v(t) \\ \vdots \\ \phi_\omega(t) \end{bmatrix} \tag{B-47}$$

Les différentes structures de modèles sont obtenues en fixant à 0 la valeur de certains coefficients de $\mathbf{\Theta}$.

B.4.2 Discrétisation de la forme intégrée

L'intégrale doit être remplacée par une sommation de façon à pouvoir appliquer les données numériques calculées à partir des données mesurées.

L'intégration numérique de l'équation B-37 donne

$$\mathbf{\Delta X}(t_k) \cong X(t_k)V(t_k) - X(t_0)V(t_0) + \sum_{i=1}^{k-1} V^{OUT}(t_i)X(t_i) \tag{B-48}$$

La discrétisation du membre de droite de l'équation B-36 donne

$$\int_0^t \mu(\tau)X(\tau)V(\tau)d\tau = \mathbf{\Theta}\int_0^{t_k} \mathbf{\Phi}(\tau)X(\tau)V(\tau)d\tau \tag{B-49}$$

En posant

$$\psi_v(t) = \int_0^t \phi_v(\tau)X(\tau)V(\tau)d\tau \tag{B-50}$$

et en discrétisant par

$$\psi_v(t_k) = \psi_v(t_{k-1}) + \phi_v(t_k)X(t_k)V(t_k)(t_k - t_{k-1}) \tag{B-51}$$

on obtient le membre de droite de l'équation B-49

$$\Theta\int_0^{t_k}\Phi(\tau)X(\tau)V(\tau)d\tau \cong \Theta\Psi(t_k) \tag{B-52}$$

où

$$\Psi(t_k)=\begin{bmatrix}\psi_1(t_k)\\ \vdots \\ \psi_v(t_k)\\ \vdots \\ \psi_\omega(t_k)\end{bmatrix} \tag{B-53}$$

Ainsi le système à résoudre devient

$$\Delta\mathbf{X}(t_k)=\Theta\Psi(t_k) \tag{B-54}$$

Où les termes $\Delta\mathbf{X}(t_k)$ et $\Psi(t_k)$ sont entièrement définis par les données expérimentales.

Les manipulations précédentes ont permis de linéariser les résultats expérimentaux et d'éliminer l'effet négatif de la différentiation sur l'estimation des paramètres.

On cherche donc les éléments ϑ de Θ qui permettent de minimiser l'erreur entre $\Delta\mathbf{X}$, défini par l'équation B-37, et $\hat{\Delta\mathbf{X}}$, calculé par l'équation

$$\hat{\Delta\mathbf{X}}(t_k)=\Theta\Psi(t_k) \tag{B-55}$$

La résolution de ce système correspond à la détermination d'un modèle pour le taux de croissance spécifique μ présenté dans le bilan de masse initial.

L'équation précédente représente une équation linéaire dont la solution peut être obtenue par régression multilinéaire.

B.4.3 Exemple de modélisation

Un exemple peut permettre d'appliquer les définitions présentées jusqu'à maintenant. On désire calculer les coefficients d'un modèle cinétique polynômial d'ordre 2 constitués de 3 termes. On suppose que le taux de croissance est déterminé par les seules concentrations S et P, mesurées simultanément à la concentration en biomasse X. aux temps t_1, t_2, t_3, t_4 lors d'échantillonnage de volume $V^{échant}$.

Parmi les modèles disponibles, on choisit d'évaluer la famille de modèles décrite par

$$\mu(t) = \vartheta_1 + \vartheta_2 S(t) + \vartheta_5 S(t)P(t) \tag{B-56}$$

parmi tous les modèles d'ordre 2 pour lesquels l'ensemble des monômes est représenté par $\{S, P, S^2, SP, P^2\}$. Sous représentation matricielle, on écrit

$$\mu(t) = \Theta_n \Phi \tag{B-57}$$

où

$$\Theta_n = [\vartheta_1\ \vartheta_2\ 0\ 0\ \vartheta_5\ 0] \tag{B-58}$$

est un sous-ensemble de Θ, et où

$$\Phi(t) = \begin{bmatrix} \phi_1(t) \\ \vdots \\ \phi_v(t) \\ \vdots \\ \phi_\omega(t) \end{bmatrix} = \begin{bmatrix} 1 \\ S(t) \\ P(t) \\ S^2(t) \\ S(t)P(t) \\ P^2(t) \end{bmatrix} \tag{B-59}$$

Les manipulations précédentes ont démontré que la résolution du système

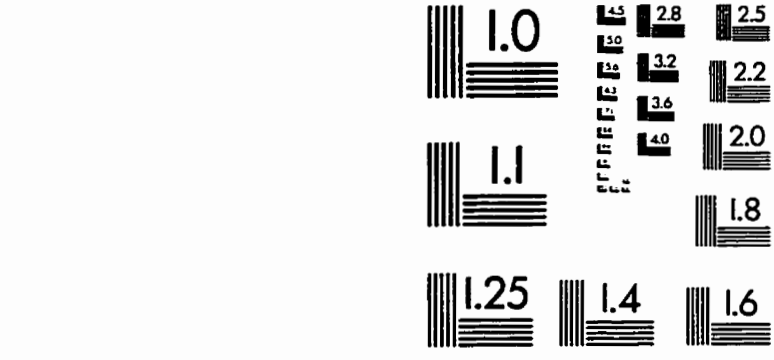
$$\frac{dX(t)V(t)}{dt} = \Theta_n \Phi(t)X(t)V(t) - F^{out}(t)X(t) \tag{B-60}$$

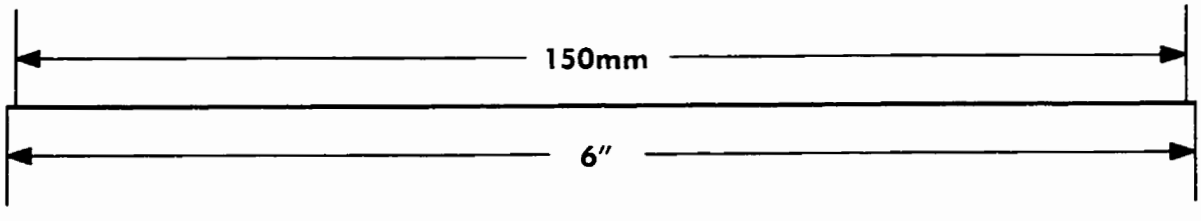
est équivalente à la résolution de

$$\Delta X(t_k) = \Theta_n \Psi(t_k) \tag{B-61}$$

Le choix des termes non nuls de la matrice des coefficients Θ détermine une famille de modèles pour laquelle la régression est effectuée et l'erreur est calculée.

IMAGE EVALUATION TEST TARGET (QA-3)





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved