

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie chimique et de génie biotechnologique

Simplification de modèles mathématiques représentant des cultures cellulaires

Thèse de doctorat
Faculté de génie chimique et génie biotechnologique

Guillaume CARDIN-BERNIER

Jury : Jean-Sébastien DESCHÊNES
Ryan GOSSELIN
Bernard MARCOS
Philippe MICHEAU
Joël SIROIS (directeur)

RÉSUMÉ

L'utilisation de cellules vivantes dans un procédé industriel tire profit de la complexité inhérente au vivant pour accomplir des tâches complexes et dont la compréhension est parfois limitée. Que ce soit pour la production de biomasse, pour la production de molécules d'intérêt ou pour la décomposition de molécules indésirables, ces procédés font appel aux multiples réactions formant le métabolisme cellulaire. Afin de décrire l'évolution de ces systèmes, des modèles mathématiques composés d'un ensemble d'équations différentielles sont utilisés. Au fur et à mesure que les connaissances du métabolisme se sont développées, les modèles mathématiques le représentant se sont complexifiés. Le niveau de complexité requis pour expliquer les phénomènes en jeu lors d'un procédé spécifique est difficile à définir. Ainsi, lorsqu'on tente de modéliser un nouveau procédé, la sélection du modèle à utiliser peut être problématique. Une des options intéressantes est la sélection d'un modèle provenant de la littérature et adapté au procédé utilisé. L'information contenue dans le modèle doit alors être évaluée en fonction des phénomènes observables dans les conditions d'opération. Souvent, les modèles provenant de la littérature sont surparamétrés pour l'utilisation dans les conditions d'opération des procédés ciblées. Cela fait en sorte de causer des problèmes d'identifiabilité des paramètres. De plus, l'ensemble des variables d'état utilisées dans le modèle n'est pas nécessairement mesuré dans les conditions d'opération normales. L'objectif de ce projet est de cibler l'information utilisable contenue dans les modèles par la simplification méthodique de ceux-ci. En effet, la simplification des modèles permet une meilleure compréhension des dynamiques à l'œuvre dans le procédé. Ce projet a permis de définir et d'évaluer trois méthodes de simplification de modèles mathématiques servant à décrire un procédé de culture cellulaire. La première méthode est basée sur l'application de critères sur les différents éléments du modèle, la deuxième est basée sur l'utilisation d'un critère d'information du type d'Akaike et la troisième considère la réduction d'ordre du modèle par retrait de variables d'état. Les résultats de ces méthodes de simplification sont présentés à l'aide de quatre modèles cellulaires provenant de la littérature.

Mots-clés : Simplification, Réduction d'ordre, Modélisation, Culture cellulaire, Biologie des systèmes

REMERCIEMENTS

Je tiens tout d'abord à remercier les professeurs Joël Sirois et Bernard Marcos pour leur soutien pendant la durée de ce projet. Sans leurs apports, ce projet n'aurait pu être mené à terme. J'aimerais aussi remercier les membres du département de génie chimique et de génie biotechnologique de l'Université de Sherbrooke de m'avoir apporté de l'aide lorsque j'en ai eu besoin de même que pour m'avoir fourni un milieu de vie dynamique. Je remercie aussi le Conseil de recherche en sciences naturelles et en génie du Canada pour l'appui financier qui a été offert pour la réalisation de ce projet.

J'aimerais saluer Laurent Sloan, Renaud Lemaire-Dugré, Daniel Lesvesque, Alexandre Arsenault et Olivier Drevelle, sans eux, ces années de doctorat auraient été beaucoup moins agréables. Enfin, j'aimerais pouvoir trouver les mots pour exprimer ma reconnaissance à ma mère pour le support inconditionnel qu'elle m'a offert à travers les différentes épreuves que j'ai vécues.

TABLE DES MATIÈRES

RÉSUMÉ	i
REMERCIEMENTS	iii
TABLE DES MATIÈRES	iv
LISTE DES FIGURES	vii
LISTE DES TABLEAUX	ix
LISTE DES SYMBOLES	xii
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 OBJECTIFS ET MÉTHODE	5
2.1 Bioprocédés et modélisation	5
2.2 Optimisation d'un bioprocédé	8
2.3 Objectifs	9
2.4 Méthode	10
CHAPITRE 3 ÉTAT DE L'ART	12
3.1 Modèles mathématiques de cultures cellulaires	12
3.1.1 Éléments des modèles	12
3.1.2 Catégorisation des modèles	13
3.1.3 Cinétiques	16
3.1.4 Modèles mathématiques de culture cellulaire	18
3.2 Estimation de la valeur des paramètres	20
3.3 Identifiabilité	21
3.4 Simplification de modèles	22
3.5 Simplification des paramètres d'un modèle	23
3.5.1 Sensibilité du modèle à une variation de la valeur des paramètres	24
3.5.2 Analyse de sensibilité	25
3.5.3 Critère d'information d'Akaike	27
3.5.4 Critère d'information Bayésien	29
3.5.5 Analyse de niveau temporel	30
3.6 Réduction de l'ordre d'un modèle	32
3.6.1 Décomposition orthogonale propre	34
3.6.2 Techniques de troncature équilibrée	35
CHAPITRE 4 ÉLÉMENTS COMPOSANT LE MODÈLE	37
4.1 Données	37
4.1.1 Données expérimentales	38
4.1.2 Variables modélisées	38
4.1.3 Plage de données	38
4.2 Variables d'état	41
4.2.1 Variables d'intérêt et variables non essentielles	41
4.3 Modèles	42
4.3.1 Structure du modèle	42
4.3.2 Réactions	43
4.3.3 Impact réactionnel	44
4.3.4 Fonctions unitaires	44
4.3.5 Zone d'impact d'une fonction unitaire	45
CHAPITRE 5 ALGORITHMES DE SIMPLIFICATION	49

5.1	Simplification des paramètres par application de critères sur les structures du modèle	50
5.1.1	Critère A : Fonction unitaire ayant une valeur près de 1	50
5.1.2	Critère B : Fonction unitaire ayant une valeur près de 0	52
5.1.3	Critère C : Impact réactionnel ayant une valeur faible par rapport aux autres..	53
5.1.4	Critère D : Impact réactionnel ayant un impact faible sur l'estimation des variables.....	54
5.1.5	Critère E : Fonctions unitaires ayant une valeur constante	54
5.2	Simplification des paramètres du modèle par application du critère d'information d'Akaike	55
5.2.1	Utilisation d'un modèle bruité	56
5.2.2	Utilisation d'un RSS relatif.....	57
5.2.3	Utilisation d'un critère d'information d'Akaike modifié.....	58
5.2.4	Algorithme de simplification basé sur le critère d'information d'Akaike	58
5.3	Réduction d'ordre du modèle par retrait de variables non essentielles	59
5.3.1	Ordre de simplification des variables.....	59
5.3.2	Limite de simplification	60
CHAPITRE 6 ANALYSE DES MÉTHODES DE SIMPLIFICATION.....		62
6.1	Simplification des paramètres par application de critère sur les structures du modèle	62
6.1.1	Critère A : Fonction unitaire ayant une valeur près de 1	63
6.1.2	Critère B : Fonction unitaire ayant une valeur près de 0	64
6.1.3	Critère C : Impact réactionnel ayant une valeur faible par rapport aux autres..	67
6.1.4	Critère D : Impact réactionnel ayant un impact faible sur l'estimation des variables.....	70
6.1.5	Critère E : Fonctions unitaires ayant une valeur constante	72
6.1.6	Critères retenus.....	76
6.2	Simplification des paramètres du modèle par application du critère d'information d'Akaike	77
6.2.1	Critère d'information d'Akaike original	77
6.2.2	Utilisation d'un RSS relatif et Modèle bruité	78
6.2.3	Critère d'information d'Akaike modifié	83
6.3	Réduction d'ordre du modèle par retrait de variables non essentielles	88
6.3.1	Ordre de simplification des variables.....	88
6.3.2	Critères de simplification	89
CHAPITRE 7 ANALYSE DES MODÈLES SIMPLIFIÉS		94
7.1	Cloutier, 2007	94
7.1.1	Simplification par critères	94
7.1.2	Simplification par AICm	98
7.1.3	Simplification des variables non essentielles	100
7.2	Chassagnole, 2002	104
7.2.1	Simplification par critères	104
7.2.2	Simplification par AICm	106
7.2.3	Simplification des variables non essentielles	108
7.3	Iwamoto, 2007.....	111

7.3.1	Simplification par critères	112
7.3.2	Simplification par AICm	113
7.3.3	Simplification des variables non essentielles.....	115
7.4	Singh, 2006	117
7.4.1	Résultats de la simplification par critères	119
7.4.2	Résultats de la simplification par AICm	121
7.4.3	Résultats de la simplification des variables non essentielles.....	125
CHAPITRE 8 DISCUSSION ET APPLICATION COMBINÉE DES MÉTHODES ...		128
8.1	Simplification des paramètres par application de critères sur les structures du modèle	128
8.2	Simplification des paramètres du modèle par application du critère d'information d'Akaike modifié.....	130
8.3	Réduction d'ordre du modèle par retrait de variables non essentielles	131
8.4	Application combinée des méthodes de simplification.....	133
CHAPITRE 9 CONCLUSION ET PERSPECTIVES.....		137
9.1	Perspectives	138
ANNEXE A – Modèles Mathématiques		140
A.1	Modèle de Cloutier (2007).....	140
A.2	Modèle de Chassagnole (2002)	142
A.3	Modèle de Iwamoto (2007).....	147
A.4	Modèle de Singh (2006).....	149
LISTE DES RÉFÉRENCES.....		154

LISTE DES FIGURES

Figure 1.1 : Principales voies métaboliques du métabolisme (Berg, 2002)	2
Figure 4.1 : Données expérimentales associées à trois variables (Cardin-Bernier, 2010)	39
Figure 4.2 : Distribution de données expérimentales	40
Figure 4.3 : Plage de données expérimentales.....	40
Figure 4.4 : Exemples de la zone d'impact de fonctions unitaires (limitation et inhibition).....	46
Figure 4.5 : Comparaison entre la plage de données d'une variable et la zone d'impact de fonctions unitaires	47
Figure 4.6 : Exemple de plage de données ayant un recoupement faible avec la plage de données.....	48
Figure 5.1 : Déplacement de la zone d'impact de fonctions unitaires aux limites de la plage de données modélisées lorsque la fonction unitaire a une valeur près de 1.	51
Figure 5.2 : Déplacement de la zone d'impact de fonctions unitaires aux limites de la plage de données modélisées lorsque la fonction unitaire a une valeur près de 0.	52
Figure 6.1 : Valeurs dans le temps des fonctions unitaires liées à la croissance cellulaire du modèle de Cloutier (2007) A) limitation associée au phosphate intracellulaire (conservée), B) limitation associée au glucose intracellulaire (retirée), C) limitation associée à l'azote intracellulaire (conservée), D) inhibition associée au phosphate extracellulaire (retirée).....	64
Figure 6.2 : Valeurs dans le temps de la fonction unitaire ($\text{NH}_4/(\text{K}_{2,\text{NH}_4}+\text{NH}_4)$) du modèle de Cloutier (2007)	65
Figure 6.3 : Estimation du NH_4 selon le modèle de Cloutier (2007) avec et sans l'impact réactionnel ($-\text{V}_{\text{max},2,\text{NH}_4}*(\text{NH}_4/(\text{K}_{2,\text{NH}_4}+\text{NH}_4))*(\text{X}/\text{V})$) pour les deux expériences testées	66
Figure 6.4 : Impacts réactionnels liés à la cinétique du phosphate intracellulaire dans le modèle de Cloutier (2007).....	68
Figure 6.5 : Impact sur l'estimation du nitrate du retrait des 3 impacts réactionnels.....	72
Figure 6.6 : Valeur de la fonction $\text{C}_{6\text{pg}}/(\text{C}_{6\text{pg}}+\text{K}_{\text{PGDH},6\text{pg}})$	73
Figure 6.7 : Impact de l'application du critère E sur l'estimation de la concentration des sucres de réserve dans le modèle de Cloutier (2007)	76
Figure 6.8 : Impact de la simplification par la méthode de l'AICrel avec une erreur sur l'estimation de la biomasse.....	82
Figure 6.9 : Valeur de l'erreur relative au maximum moyenne en fonction de la valeur du paramètre	84
Figure 6.10 : Impact de la simplification par l'AICm sur l'estimation de la biomasse et du phosphate intracellulaire.....	87
Figure 6.11 : Résultat de l'algorithme de simplification grâce aux critères 1 et 2 sur les variables d'intérêt	92
Figure 6.12 : Résultat de l'algorithme de simplification grâce au critère 3 sur les variables d'intérêt.....	92
Figure 7.1 : Données modélisées par le modèle de Cloutier (2007) original et par le modèle simplifié.....	97
Figure 7.2 : Impact de la simplification par l'AICm sur l'estimation du phosphate intracellulaire et extracellulaire	99
Figure 7.3 : Impact de la simplification des variables non essentielles sur les variables d'intérêt	102
Figure 7.4 : Impact de la simplification par critère sur la variable C_{fdp}	106

Figure 7.5 : Impact de la simplification par l'AICm sur les variables Cfdp et Cpyr	108
Figure 7.6 : Schème réactionnel du métabolisme central du carbone (Chassagnole, 2002).....	109
Figure 7.7 : Impact de la simplification des variables non essentielles sur les variables Cg6p et C3pg.	111
Figure 7.8 : Impact de la simplification par l'AICm sur l'estimation des variables Y2 et Y9.	114
Figure 7.9 : Représentation du schème réactionnel utilisé dans le modèle de Singh (2006) ..	118
Figure 7.10 : Impact de la simplification par critères sur l'estimation des variables x4 (YAK) et x20 (STAT3N*-STST3N*).	121
Figure 7.11 : Impact de la simplification par l'AICm sur l'estimation des variables x16((IL6- gp80-gp130-JAK)* ₂ -SHP2) et x39 ((IL6-gp80-gp130-JAK)* ₂ -SHP2*).....	124
Figure 7.12 : Impact de la simplification par des variables non essentielles sur la variable x17	127

LISTE DES TABLEAUX

Tableau 3.1 : Classification des modèles selon Sirois (2000)	13
Tableau 3.2 : Caractéristiques des modèles étudiés	19
Tableau 3.3 : Caractéristiques des modèles étudiés	24
Tableau 6.1 : Valeur moyenne des fonctions unitaires liées à la croissance du modèle de Cloutier (2007) ainsi que leur évaluation selon le critère A.....	63
Tableau 6.2 : Valeur moyenne de la fonction unitaire ($\text{NH}_4/(\text{K}_{2,\text{NH}_4}+\text{NH}_4)$) du modèle de Cloutier (2007) ainsi que son évaluation selon le critère B.....	66
Tableau 6.3 : Analyse de la cinétique du phosphate intracellulaire du modèle de Cloutier (2007) selon le critère C	69
Tableau 6.4 : Impact du retrait de chacun des impacts réactionnels sur l'erreur relative au maximum de chaque variable.....	71
Tableau 6.5 : Analyse de la cinétique du nitrate du modèle de Cloutier (2007) selon le critère D	71
Tableau 6.6 : Analyse des fonctions unitaires de la cinétique des sucres de réserve selon le critère E.....	75
Tableau 6.7 : Liste des paramètres du modèle de Cloutier (2007) en ordre croissant selon leur AIC_m individuel.....	80
Tableau 6.8 : AIC_{rel} des modèles simplifiés après le retrait cumulatif des paramètres	81
Tableau 6.9 : Erreur relative au maximum de chacune des variables après la simplification du modèle par l'algorithme basé sur le critère d'information d'Akaike	82
Tableau 6.10 : Résultat de l'analyse de la valeur du multiplicateur du nombre de paramètres du critère d'information d'Akaike modifié	84
Tableau 6.11 : AIC_m des modèles simplifiés après le retrait cumulatif des paramètres.....	85
Tableau 6.12 : Erreur relative au maximum de chacune des variables du modèle original ainsi que du modèle simplifié par l'algorithme basé sur le critère d'information d'Akaike modifié	86
Tableau 6.13 : Classification des variables non essentielles en fonction des différents calculs de RSS	89
Tableau 6.14 : Résultat de l'application de l'algorithme de simplification des variables non essentielles selon les trois critères testés.	91
Tableau 7.1 : Nombre de fois où chacun des critères a suggéré une simplification dans le modèle de Cloutier (2007).....	95
Tableau 7.2 : Caractéristiques du modèle de Cloutier (2007) original et du modèle simplifié.	95
Tableau 7.3 : Erreur relative au max associée à la simplification du modèle de Cloutier (2007) pour les deux simulations (simulation 1 : Biomasse initiale :0.17 gDW/L, simulation 2 : Biomasse initiale :0.5 gDW/L).....	97
Tableau 7.4 : Erreur relative au maximum par rapport aux données et par rapport au modèle original du modèle de Cloutier (2007) simplifié	99
Tableau 7.5 : Caractéristiques du modèle de Cloutier (2007) original et du modèle simplifié par l'algorithme de simplification basé sur le critère d'information d'Akaike modifié.....	100
Tableau 7.6 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles.....	101
Tableau 7.7 : Erreurs relatives au maximum par rapport aux données simulées et par rapport au modèle original associées à la simplification du modèle de Cloutier (2007)	103
Tableau 7.8 : Caractéristiques de modèles simplifiés pour les deux scénarios de variables d'intérêt.....	103

Tableau 7.9 : Nombre de simplifications suggérées par chacun des critères pour le modèle de Chassagnole (2002)	104
Tableau 7.10 : Caractéristiques du modèle de Chassagnole (2002) original et du modèle simplifié selon l'algorithme de simplification par critère.	105
Tableau 7.11 : Erreur relative au max associée à la simplification par critère du modèle de Chassagnole (2002)	105
Tableau 7.12 : Erreur relative au max associée à la simplification AIC_m du modèle de Chassagnole (2002)	107
Tableau 7.13 : Caractéristiques du modèle de Chassagnole (2002) original et du modèle simplifié par l'algorithme de simplification basé sur le critère d'information d'Akaike modifié	108
Tableau 7.14 : Variables d'intérêt et résultat de la simplification du modèle de Chassagnole (2002) selon l'algorithme de simplification des variables non essentielles	110
Tableau 7.15 : Erreur relative au max des variables d'intérêt du modèle de Chassagnole (2002) associées à la simplification des variables non essentielles	110
Tableau 7.16 : Nombre simplifications suggérées par chacun des critères pour le modèle de Iwamoto (2007)	112
Tableau 7.17 : Caractéristiques du modèle de Iwamoto (2007) original et du modèle simplifié par l'algorithme de simplification basé sur les critères	112
Tableau 7.18 : Erreur relative au maximum du modèle de Iwamoto (2007) simplifié	113
Tableau 7.19 : Erreur relative au maximum par rapport aux données et par rapport au modèle original associée à la simplification par critères du modèle de Iwamoto (2007)	114
Tableau 7.20 : Caractéristiques du modèle de Iwamoto (2007) original et du modèle simplifié par l'algorithme de simplification basé l' AIC_m	115
Tableau 7.21 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles	116
Tableau 7.22 : Impact de la simplification des variables non essentielles sur l'estimation des variables d'intérêt	117
Tableau 7.23 : Nombre de simplifications suggérées par chacun des critères pour le modèle de Singh (2006)	119
Tableau 7.24 : Caractéristiques du modèle de Singh (2006) original et du modèle simplifié par l'algorithme de simplification basé sur les critères	119
Tableau 7.25 : Erreur relative au max associée à la simplification du modèle de Singh (2006)	120
Tableau 7.26 : Erreur relative au max associée à la simplification du modèle de Singh (2006) par l'algorithme basé sur l' AIC_m	122
Tableau 7.27 : Caractéristiques du modèle de Singh (2006) original et du modèle simplifié par l'algorithme de simplification basé sur l' AIC_m	125
Tableau 7.28 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles	126
Tableau 7.29 : Impact de la simplification des variables non essentielles sur l'estimation des variables d'intérêt	127
Tableau 8.1 : Simplifications réalisées par l'algorithme basé sur les critères	129
Tableau 8.2 : Simplifications réalisées par l'algorithme basé sur l' AIC_m	131
Tableau 8.3 : Simplifications des variables par l'algorithme basé sur la simplification des variables non essentielles.	132

Tableau 8.4 : Résultats de l'algorithme de simplification des variables non essentielles sur le modèle de Cloutier	134
Tableau 8.5 : Caractéristiques du modèle de Cloutier après application de l'algorithme de simplification des variables non essentielles	134
Tableau 8.6 : Résultats de l'algorithme de simplification combiné sur le modèle de Cloutier	135
Tableau 8.7 : Caractéristiques du modèle de Cloutier après application de l'algorithme de simplification des variables non essentielles	135

LISTE DES SYMBOLES

Symbole	Définition
[·]	Concentration
Σ	Modèle mathématique
β	Paramètre ou vecteur de paramètres
μ	Taux de croissance
m	Nombre de mesures
n	Nombre de paramètres du modèle
p	Nombre de sorties du système
t	Temps
u	Entrée du système
v	Nombre de variables d'état du modèle
v_i	Nombre de variables d'intérêt
w	Nombre d'entrées du système
x ou X	Variable d'état
y	Sortie du système
AIC	Critère d'information d'Akaike
AIC _m	Critère d'information d'Akaike modifié
B	Concentration cellulaire ou biomasse
BIC	Critère d'information Bayésien
I	Inhibiteur
RSS	Somme des carrés résiduels
RSS _{Rel}	Somme des carrés résiduels relatif
RSS _{Rel2max}	Somme des carrés résiduels relatif au maximum
S	Substrat

CHAPITRE 1

INTRODUCTION

La biotechnologie est l'utilisation d'organismes vivants ou de leurs dérivés pour fabriquer ou modifier un produit, pour améliorer des plantes ou des animaux ou pour développer un micro-organisme pour un usage spécifique (UN, 1992). Depuis le début du 20^e siècle, les applications de la biotechnologie se sont élargies grâce à une meilleure compréhension des mécanismes présents dans les différents êtres vivants.

Un procédé biotechnologique, appelé souvent bioprocédé, est un procédé dans lequel des cellules vivantes sont utilisées afin d'obtenir une transformation désirée. Par exemple, des micro-organismes (levures, bactéries, champignons, algues, cellules animales et végétales) sont utilisés pour la production de biomasse (ex. : production de *Saccharomyces cerevisiae* pour la boulangerie), pour la production de métabolites ou pour la bioconversion de molécules d'intérêt. Les domaines où sont utilisés les bioprocédés sont diversifiés. On retrouve l'utilisation de bioprocédés dans des industries tels l'agroalimentaire, l'industrie pharmaceutique, la production d'énergie ainsi qu'en environnement que ce soit pour le traitement de l'eau, des sols, etc. Selon le type de procédé, les raisons pour lesquelles des cellules sont utilisées peuvent varier. Par exemple, en pharmaceutique, des cellules sont utilisées à cause de leur capacité à produire des molécules très complexes et très spécifiques. En environnement et en énergie, c'est leur capacité à consommer des substrats parfois difficilement dégradables, comme source d'énergie pour leur métabolisme, qui est une propriété intéressante.

L'ensemble des réactions chimiques et biochimiques permettant à la cellule (composante fondamentale des organismes vivants) de vivre et de se reproduire est appelé métabolisme. Le métabolisme est composé de deux ensembles de réactions : l'anabolisme et le catabolisme. L'anabolisme est l'ensemble des réactions qui produisent des molécules à partir de plus petites unités. Ces réactions requièrent de l'énergie. Le catabolisme est l'ensemble de réactions où des molécules sont brisées en plus petites unités. Ces réactions produisent de l'énergie.

Ainsi, des milliers de réactions se produisent faisant interagir des millions de molécules différentes telles que l'ADN, l'ARN et les protéines. Afin d'assurer le fonctionnement de leur métabolisme, les cellules utilisent des voies de signalisation ainsi que des mécanismes de régulation (figure 1.1). De cette façon, le métabolisme cellulaire peut réagir à des perturbations environnementales par l'activation ou la désactivation de voies métaboliques (Albert, 2005). Plusieurs mécanismes cellulaires ont pour objectif de réguler la concentration des différentes molécules intervenant dans le métabolisme. Cela est nécessaire pour que les cellules puissent survivre aux perturbations environnementales. La biologie des systèmes est le domaine cherchant à décrire les processus qui sont à l'œuvre dans le métabolisme cellulaire.

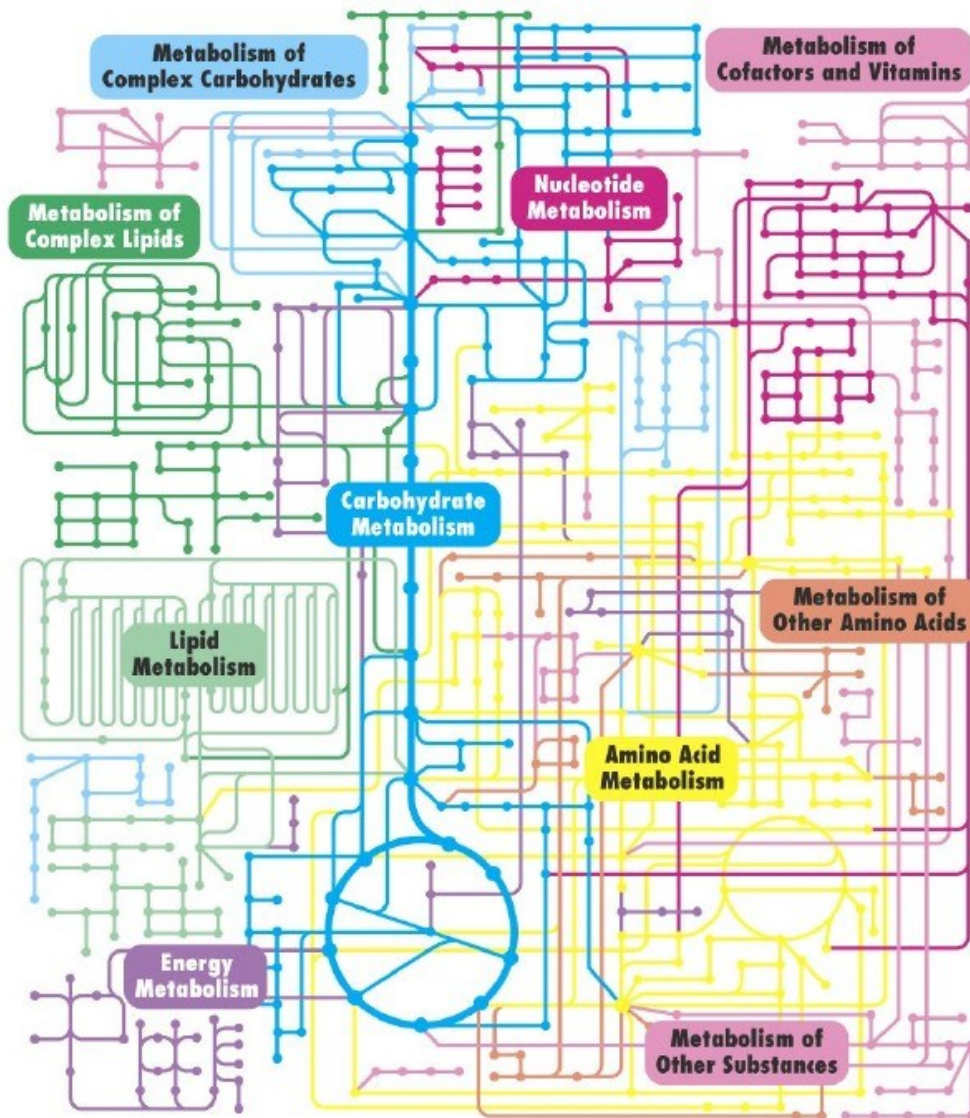


Figure 1.1 : Principales voies métaboliques du métabolisme (Berg, 2002)

Le métabolisme joue donc un rôle vital dans tout procédé utilisant des micro-organismes. C'est pourquoi une meilleure connaissance du métabolisme maximisera les chances de succès pour concevoir et opérer un procédé biotechnologique. Tenir compte de l'ensemble des connaissances actuelles sur le métabolisme cellulaire demanderait de considérer plusieurs milliers de variables (concentrations, température, etc.) et de réactions ce qui est en pratique impossible, notamment dans une perspective industrielle. Il est donc très difficile de modéliser un bioprocédé à cause de la complexité du métabolisme. Une deuxième difficulté vient du fait que l'on ne peut pas décrire chaque cellule séparément et que, par conséquent, on va travailler avec une représentation de la population cellulaire. Dans cette population, chaque cellule a des caractéristiques différentes et évolue dans des conditions locales différentes (concentrations, cisaillement, etc.). Tout modèle mathématique cherchant à représenter des concentrations se retrouve donc à être une approximation basée sur les valeurs moyennes de la population.

Une autre difficulté concerne l'obtention de données expérimentales. Lors d'une culture cellulaire, il est difficile d'obtenir des données expérimentales sur les conditions intracellulaires. En effet, la majorité des techniques permettant de mesurer des concentrations intracellulaires sont longues, complexes et destructrices. De plus, plusieurs métabolites ont des structures et des propriétés semblables rendant leur différenciation difficile. Ainsi, plusieurs techniques utilisées pour l'étude du métabolisme telles que la chromatographie, la diffraction de rayon X et le marquage par isotopes radioactifs sont peu appropriées pour le suivi de bioprocédés dans des conditions de productions industrielles, en raison du coût et de la complexité de ces techniques et des impacts possibles sur le produit fini.

Ainsi, plusieurs bioprocédés sont opérés au niveau industriel avec une connaissance limitée des phénomènes à l'œuvre. Bien que ce manque de connaissance n'affecte pas directement l'opération de bioprocédés, il limite considérablement leurs améliorations ou leur optimisation. En effet, au niveau industriel, l'objectif d'optimisation de procédés biologiques est d'obtenir la meilleure

productivité possible tout en conservant un procédé robuste (Waine, 2001). Or, si la dynamique du procédé est mal connue, il est difficile de proposer de meilleures conditions d'opération de façon à améliorer la productivité du bioprocédé. Il convient donc d'avoir une connaissance suffisante de la dynamique sans nécessairement modéliser de manière exhaustive le métabolisme cellulaire bien qu'il soit de plus en plus connu. S'il est possible d'obtenir des dynamiques représentatives du procédé en considérant un ensemble réduit du métabolisme cellulaire, il sera possible de concevoir et d'opérer des bioprocédés de manière plus efficace.

CHAPITRE 2 OBJECTIFS ET MÉTHODE

2.1 Bioprocédés et modélisation

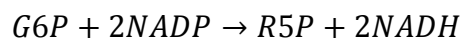
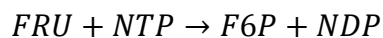
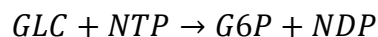
Les bioprocédés sont utilisés dans plusieurs industries. En 2004, les biotechnologies ont engendré une activité économique de près de 70 milliards de dollars au Canada. (Archambault, 2004). Les principaux secteurs faisant appel aux bioprocédés sont la pharmaceutique (43 milliards \$), l'agroalimentaire (8.8 milliards \$), l'énergie (3.4 milliards \$) et le traitement de la pollution (1.7 milliard \$). Les bioprocédés étudiés dans ce projet sont principalement les bioprocédés visant la production de molécules d'intérêt ainsi que les procédés visant la production de biomasse, mais n'y sont pas limités. Tous les procédés utilisant des modèles mathématiques respectant la forme définie peuvent être traités par les algorithmes de simplification développés dans ce projet.

De manière simplifiée, les bioprocédés comprennent principalement deux sections : la fabrication du produit d'intérêt réalisée dans un bioréacteur et la séparation du produit d'intérêt des autres sous-produits de la culture cellulaire. Le projet présenté ici ne porte que sur la première section. Un bioréacteur est un réservoir dans lequel est effectuée une culture cellulaire. Dans un bioréacteur, les conditions environnementales (température, pH, concentrations des différents nutriments) auxquelles sont exposées les cellules sont contrôlées. Selon les besoins du procédé, le bioréacteur comprend des entrées permettant d'ajouter différentes solutions en cours de culture, des ports pour différentes sondes permettant d'acquérir des données sur la culture en cours, de contrôler la température et le pH, etc.

Dans le cadre de ce projet, le système étudié est défini comme étant le milieu de culture ainsi que la population cellulaire présente dans celui-ci. Selon la complexité de la modélisation, les cellules peuvent être considérées comme un sous-système (milieu intracellulaire et milieu extracellulaire). Le volume contenu dans le bioréacteur peut varier en cours d'expérience (ex. : ajout d'une solution dans le bioréacteur). Les frontières du système sont donc les parois du bioréacteur ainsi que

l'interface entre la phase liquide et la phase gazeuse. Les différentes propriétés du système sont appelées variables d'état du système. Celles-ci sont influencées par la dynamique interne de même que par des transferts entre le système et son environnement. Deux types de transfert sont possibles, les entrées (ajout d'éléments de l'environnement vers le système) et les sorties (transfert d'éléments du système vers l'environnement). Dans le cadre de ce projet, les transferts (masse, énergie) sont supposés parfaitement connus. Des exemples possibles de transfert de l'environnement vers le système sont : l'ajout de substrat dans le bioréacteur ou le chauffage du bioréacteur par une source externe. Un exemple de sorties est la prise d'échantillons.

Afin de modéliser correctement le comportement de ce système, le métabolisme cellulaire doit être étudié et traduit sous forme de schème réactionnel. Le schème réactionnel est composé d'une ou plusieurs réactions. Chaque réaction représente une transformation qui se produit dans le procédé. Celle-ci peut être de nature chimique (ex. : oxydation), physique (ex. : absorption de nutriments) ou biologique (ex. : croissance cellulaire). À chaque réaction est associé un taux de réaction (aussi appelé vitesse de réaction) qui représente à quelle vitesse les substrats sont transformés en produit. Voici un exemple de schème réactionnel (Leduc, 2006) où chacun des symboles (GLC, NTP, G6P, NDP, ...) représente une espèce chimique :



...

Éq. 2.1

Les variables composant le schème réactionnel sont des variables d'état. Dans ce cas précis, chaque variable d'état utilisée dans le modèle de Leduc (2006) est une concentration de métabolites (GLC=glucose, NTP= nucléoside triphosphate, G6P= glucose-6-phosphate, ...). Les variables d'état représentent les éléments qui sont modifiés par le procédé et/ou par les transferts entre le procédé et l'environnement.

Des variables d'état courantes sont la concentration de différentes molécules, la concentration cellulaire ainsi que certaines propriétés physico-chimiques (température, pH, etc.).

La modélisation du schème réactionnel et des taux de réactions conduit à un ensemble d'équations différentielles appelé modèle. Le modèle décrit l'évolution des variables d'état en fonction des dynamiques internes du système dans les conditions d'opération du procédé. Les différents termes de ces équations représentent les cinétiques reliées au métabolisme cellulaire, à la consommation, à l'alimentation, à la dilution, etc. L'évolution des variables d'état du système telle que décrite par les équations dépend d'elles-mêmes, des transferts (entrée, sortie), ainsi que des dynamiques du système représenté par la forme de l'équation et la valeur des paramètres. L'équation 2.2 présente le modèle mathématique (Σ) :

$$\Sigma \begin{cases} \frac{dx(t)}{dt} = f(x(t), u(t), \beta) \\ y(t) = h(x(t), u(t), \beta) \end{cases} \quad \text{Éq. 2.2}$$

$$\text{Où } x(t) \in \mathbb{R}^v, y(t) \in \mathbb{R}^p, u(t) \in \mathbb{R}^w, \beta \in \mathbb{R}^n$$

où Σ est le modèle, x représente le vecteur de variables d'état, v est le nombre de variables d'état, y représente les sorties du système, p est le nombre de sorties, u représente les entrées du système, w est le nombre d'entrées, β représente les paramètres du modèle et n est le nombre de paramètres. Selon le type de procédé étudié, le niveau de complexité du modèle peut être très variable. Un modèle mathématique peut être très simple (une réaction et deux variables), ou il peut être extrêmement complexe. Un modèle représentant l'ensemble des connaissances actuelles sur le métabolisme cellulaire contiendrait plusieurs milliers de variables et de réactions. Ce qui entraîne la différence de complexité de différents modèles décrivant un procédé similaire est le niveau de détails accordé à la description du métabolisme cellulaire. Par exemple, les modèles les plus simples ne considèrent pas les concentrations intracellulaires. Les cellules sont alors considérées comme des

boîtes noires effectuant la transformation de composés présents dans le milieu de culture.

2.2 Optimisation d'un bioprocédé

La façon classique d'optimiser un bioprocédé se fait en deux étapes : la première étape est la sélection d'une souche productive. Ceci peut être réalisé par la sélection minutieuse d'une souche naturellement productive ou être réalisé par la manipulation génétique d'une souche soit pour augmenter sa productivité naturelle, soit pour induire une nouvelle productivité.

La seconde étape est la recherche des conditions optimales d'opération. Cette étape consiste à déterminer quelles sont les conditions nutritionnelles et physico-chimiques permettant d'obtenir le meilleur rendement possible. La façon classique d'optimiser les conditions d'opération du procédé est d'effectuer des essais à plus petite échelle en laboratoire de façon à couvrir la plage de valeurs la plus représentative possible (tamisage de la plage expérimentale, « screening »). Une fois ces conditions déterminées, le procédé est testé à l'échelle pilote où, si une condition d'opération pose problème, elle peut être réajustée. Puis, le procédé est amené à l'échelle industrielle où un contrôle serré est exercé afin de s'assurer que les conditions d'opérations respectent les critères établis. Les conditions opératoires du procédé sont donc liées aux tamisages effectués lors de l'étape initiale de laboratoire.

Une fois l'optimisation réalisée, le procédé est mis à l'échelle industrielle puis démarré. Lorsque le procédé est opéré au niveau industriel, de nombreux lots sont réalisés afin d'obtenir le produit désiré. Cela permet d'augmenter significativement le nombre de données expérimentales disponibles. Selon le procédé, les variables d'état peuvent être suivies de façon plus ou moins serrée. Dans un procédé avec très peu de suivi, les conditions initiales et les ajouts sont déterminés de façon fixe (le même volume d'une solution avec une concentration fixe est ajouté au même moment). Les variables d'état ne sont pas toutes mesurées et les transferts (entrées

et sorties) ne sont pas modifiés afin de s'adapter aux conditions réelles du système. Dans un procédé où le suivi est plus serré, plusieurs variables d'état sont mesurées et peuvent être utilisées pour optimiser le procédé en temps réel. Dans un cadre de bioprocédé pour la production industrielle, si le suivi est serré, il est possible de recueillir beaucoup de données sur plusieurs cultures semblables et d'utiliser ces données afin d'améliorer la compréhension du système dans le but d'optimiser le procédé.

2.3 Objectifs

L'objectif de ce projet est de générer un modèle simplifié pour un bioprocédé à partir du schéma réactionnel de modèles provenant de la littérature, tout en mettant à profit les données expérimentales recueillies à la fois lors de la phase d'optimisation et d'opération du bioprocédé. La simplification des modèles est la réduction du niveau de complexité d'un modèle mathématique. Cela peut être réalisé par la réduction du nombre de variables d'état ou la réduction du nombre de paramètres du modèle. En effet, les méthodes de simplification conduisent à quantifier l'impact de la simplification du modèle sur la qualité de l'estimation. L'avantage d'utiliser un modèle simplifié plutôt qu'un modèle complexe est qu'il est plus facile de comprendre les réponses du système face aux perturbations. Cela permet de cibler plus directement les actions à mettre en œuvre afin d'optimiser ce procédé.

La simplification est parfois négligée lors de la modélisation. Plutôt que de chercher à simplifier un modèle afin de comprendre les phénomènes en jeu et d'identifier les phénomènes les plus importants, on se fie plutôt au modèle ayant l'erreur d'estimation la plus faible afin de représenter les dynamiques du système. Cela s'exprime par des modèles surparamétrés, car le nombre de paramètres ayant un impact significatif est relativement faible par rapport au nombre total de paramètres du modèle.

La structure générale d'un schème réactionnel d'un modèle est soumise à différentes hypothèses sur les phénomènes affectant le procédé. Ces hypothèses sont associées à des phénomènes physiques, chimiques ou biologiques affectant les réactions du procédé telles les limitations, les inhibitions, l'induction, etc. Mathématiquement, ces hypothèses sont représentées par des modifications aux équations de cinétiques affectant le modèle. Lorsqu'un modèle est simplifié, le nombre de ces phénomènes considérés comme affectant le procédé est diminué. La difficulté de la simplification est d'identifier quels sont les phénomènes physiques ayant un impact réel observé et quels sont les phénomènes n'affectant pas les prédictions du modèle.

2.4 Méthode

Les méthodes de simplification présentées dans le cadre de ce projet se divisent en deux catégories.

La première catégorie comprend les simplifications par analyse comparative des données par rapport à la plage de concentrations où le phénomène est actif. Ainsi, différentes hypothèses possibles sont testées vérifiant l'impact des variables d'état sur l'évolution du système. Selon les phénomènes testés, des segments d'équations sont ajoutés ou retirés du modèle. La valeur des paramètres du modèle est recalculée par un algorithme de minimisation de l'erreur d'estimation. Ensuite, les segments d'équations associés aux différentes hypothèses sont testés afin de savoir s'ils affectent la prédiction du modèle. Si oui, on considère que le phénomène affecte le procédé. Sinon, le segment d'équation est retiré. Enfin, l'analyse des différents modèles (original et simplifiés) est réalisée afin d'identifier le modèle le plus approprié au procédé.

La seconde catégorie de simplifications étudiée est la réduction d'ordre du modèle, soit la réduction du nombre de variables d'état. Cela peut être attribuable au fait que la variable d'état a une valeur quasi constante (la cinétique globale associée

à la variable (dx/dt) a une valeur près de 0 pour l'ensemble des expériences). La seconde raison pour laquelle une variable peut être simplifiée est qu'elle n'influence pas de façon importante les autres variables du système. Elle n'est donc pas indispensable pour connaître la dynamique du système et peut être enlevée du modèle sans en changer significativement la dynamique. Cependant, dans certains cas, la variable doit être conservée, car celle-ci présente un intérêt au niveau de la productivité, de l'économie, etc. Cette situation est donc plus complexe, car elle demande d'identifier quelles sont les variables que l'on doit conserver (variable d'intérêt) et celles que l'on peut enlever (variable non essentielle).

La mise en œuvre de ce projet permettra plusieurs améliorations par rapport à ce qui se fait actuellement. Principalement, cela permettra de mettre à profit les données expérimentales recueillies lors de l'exploitation normale du bioprocédé. Ainsi, les données recueillies lors d'événements spécifiques, mais relativement fréquents en industrie (ex. : contaminations, modification aux conditions physico-chimiques, etc.) pourraient être utilisées afin d'améliorer la compréhension du procédé. Des phénomènes lents, comme l'évolution des souches microbiennes, pourraient être observés en comparant l'évolution des paramètres du modèle d'un lot à l'autre.

La grande force de ce projet est que les coûts pour sa mise en application au niveau industriel sont minimes. Les algorithmes présentés dans ce projet peuvent être appliqués en continu s'ils sont reliés à une banque de données expérimentales. Chaque nouvelle donnée, qu'elle provienne d'une sonde ou d'un test manuel, peut être ajoutée à la base de données afin d'améliorer le modèle. Or, ce genre de base de données est courant dans l'industrie pharmaceutique où l'ensemble des données sur chaque cuvée doit être conservé. Ainsi, les risques du projet (coûts) sont très faibles par rapport aux gains possibles (meilleure compréhension du système, optimisation, etc.).

Afin de répondre aux objectifs de ce projet, plusieurs modèles ont été analysés pour comprendre l'impact des différents éléments des cinétiques sur les prédictions des modèles. Les étapes de modélisation et les méthodes de simplification seront analysées. Ainsi, l'analyse de la littérature scientifique présente la structure et la classification de modèles de culture cellulaire ainsi que quelques exemples de modèles. Différentes méthodes servant à simplifier les modèles mathématiques et à évaluer différents modèles en fonction de leur simplicité (nombre de paramètres du modèle) sont ensuite discutées. Différentes méthodes servant à effectuer l'estimation des paramètres et à caractériser l'identifiabilité de ceux-ci sont enfin présentées.

3.1 Modèles mathématiques de cultures cellulaires

3.1.1 Éléments des modèles

À partir du bilan de masse des différents composés, le modèle mathématique d'un bioprocédé est composé d'une ou plusieurs équations différentielles qui représentent l'évolution des différentes variables d'état du système.

$$\Sigma \begin{cases} \frac{dx(t)}{dt} = f(x(t), u(t), \beta) \\ y(t) = h(x(t), u(t), \beta) \end{cases} \quad \text{Éq. 3.1}$$

Où $x(t) \in \mathbb{R}^v, y(t) \in \mathbb{R}^p, u(t) \in \mathbb{R}^w, \beta \in \mathbb{R}^n$

Les différents types de variables sont (Sirois, 2000) :

- les entrées ($u(t)$), variables affectant le procédé. Elles sont supposées parfaitement connues;
- les sorties ($y(t)$), les résultats du procédé;
- les variables d'état (x), l'ensemble des conditions internes du système, principalement les concentrations. Les variables d'état évoluent en

fonction de la dynamique du système de même que des transferts entre le système et l'environnement.

Le terme (β) représente les paramètres du modèle. Les paramètres sont supposés constants au cours des simulations.

3.1.2 Catégorisation des modèles

En biologie des systèmes, les modèles mathématiques peuvent être classifiés selon leurs caractéristiques. Parmi les classifications intéressantes, notons celles décrites par Sirois (2000) telles que présentées dans le tableau 3.1 :

Stochastique Éléments probabilistes	Déterministe Structure précise fixée
Suspension distribuée La suspension est un tout	Suspension ségréguée La biomasse est distincte du milieu
Biomasse distribuée La biomasse est uniforme	Biomasse ségréguée Différentes biomasses
Macroscopique Modèles considérant les cellules comme une boîte noire	Métabolique Modèle décrivant le métabolisme cellulaire

Tableau 3.1 : Classification des modèles selon Sirois (2000)

Les modèles stochastiques possèdent des éléments statistiques de sorte que les prédictions du modèle ne sont pas basées sur des connaissances de fonctionnement du système tel que le métabolisme, mais sur une évaluation de l'évolution probabiliste des variables. Parmi les modèles stochastiques, on retrouve ceux issus d'algorithmes tels Monte-Carlo (Hammersley, 1964) ou d'algorithmes génétiques (Ashlock, 2006). Les modèles déterministes sont basés sur un schéma réactionnel décrivant le comportement du métabolisme. Dans ce projet, puisque

l'objectif est de tester les hypothèses formulées, les modèles étudiés sont des modèles déterministes.

Les modèles peuvent aussi être classifiés en fonction de leur ségrégation. La ségrégation peut avoir lieu par rapport à la suspension ou par rapport à la biomasse. La suspension est distribuée lorsque l'ensemble du contenu du procédé est considéré comme uniforme. Dans ce cas, il n'y a pas de distinction entre les phases liquides et solides de la suspension. Lorsqu'une distinction est faite entre les cellules et le milieu de culture, on dit que la suspension est ségréguée.

La ségrégation peut aussi être faite au niveau de la biomasse. Lorsqu'on considère que la population cellulaire est uniforme, la biomasse est dite distribuée. Lorsqu'il y a des distinctions entre différents types de biomasse, la biomasse est dite ségréguée. La ségrégation peut représenter différentes espèces cellulaires (ex. : acidogénèse et méthanogénèse lors de la biométhanisation), elle peut représenter l'état de la cellule dans le cycle cellulaire (phase G₀, G₁, G₂, S et M) (Csikasz-Nagy, 2009), ou elle peut représenter toute autre division dans la population cellulaire (cellules exprimant un gène, etc.). Les modèles utilisés dans ce projet peuvent être ségrégués ou distribués, à la fois au niveau de la biomasse et au niveau de la suspension.

Les modèles peuvent être aussi être classifiés de macroscopiques ou de métaboliques (Niu, 2013). Les modèles macroscopiques prennent en compte le métabolisme extracellulaire, mais accordent peu d'importance au métabolisme intracellulaire, considérant les cellules comme une boîte noire. Ils servent à décrire l'interaction entre les cellules et leur environnement (Junker 2006). Le plus connu est probablement le modèle à un substrat avec une cinétique de Monod (1949) :

Schème réactionnel : S → B

$$\frac{dB}{dt} = \mu * B$$

Éq. 3.2

$$\frac{dS}{dt} = \beta_1 * \mu * B$$
$$\mu = \mu_{max} * \frac{[S]}{\beta_2 + [S]}$$

où S est le substrat, B est la biomasse et μ est le taux de croissance. Les modèles macroscopiques ont tendance à être plus simples, surtout lorsque l'hypothèse d'un seul substrat limitant est posée. L'hypothèse d'un seul substrat limitant stipule que même si plusieurs variables peuvent affecter une cinétique, la valeur d'une variable est si limitante que l'impact des autres variables est négligeable. Le schéma réactionnel est généralement simple et contient peu de réactions. Les modèles macroscopiques ne contiennent pas de réactions d'absorption, de réactions intracellulaires, ni de variables intracellulaires. Les modèles macroscopiques sont des modèles à suspension distribuée (Haag, 2005). Les modèles métaboliques sont issus de la biologie des systèmes (Niu, 2013) et cherchent à décrire les réactions chimiques et biochimiques composant le métabolisme intracellulaire. Ces modèles sont plus complexes, car ils prennent en compte des variables d'état intracellulaires et leurs interactions. Les données expérimentales nécessaires à l'élaboration et la validation d'un modèle métabolique sont beaucoup plus difficiles à obtenir puisque ces modèles tentent de représenter des segments du métabolisme cellulaire, voir son entièreté (Niu, 2013). Ils sont donc le résultat de nombreuses et complexes expériences de laboratoire.

Les premiers modèles mathématiques du métabolisme cellulaire sont le résultat de l'analyse de données biochimiques ayant permis d'établir les cinétiques des différentes enzymes composant le métabolisme. Parmi ces modèles, notons celui du globule rouge humain par Joshi (Joshi, 1989(1,2) et 1990(1,2)) ainsi que le modèle de Rizzi représentant la culture de *Saccharomyces cerevisiae* (Rizzi, 1997(1,2)). Les paramètres cinétiques de ces modèles ont été établis en effectuant des mesures *in vitro* des taux de réactions enzymatiques avant d'être testés *in vivo* par réduction de l'erreur d'estimation du modèle. L'établissement de ces modèles a donc demandé un

nombre d'expériences considérable (plus de 50 réactions enzymatiques dont le taux de réaction doit être déterminé).

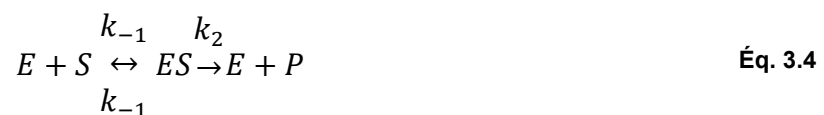
Les modèles métaboliques plus récents font appel à des techniques expérimentales permettant de suivre plus facilement l'évolution des molécules dans le métabolisme. Par exemple, la technique « Metabolic Flux Analysis » (MFA) peut être utilisée avec la cytométrie en flux et l'utilisation d'isotopes radioactifs (principalement du C13). Cela permet de mesurer les taux de réaction effectifs pour les réactions intracellulaires (Boghigian, 2010). Par exemple, Williams (2010) présente un modèle du métabolisme central du carbone (à l'aide de glucose contenant du C13) pour les cellules végétales *Arabidopsis thaliana*.

3.1.3 Cinétiques

Les premières tentatives visant à représenter mathématiquement le comportement des éléments composant le métabolisme portaient sur le développement de l'équation de Michaelis-Menten (Michaelis, 1913) servant à décrire la cinétique d'une réaction biochimique catalysée par une enzyme.

$$\mu = \mu_{max} * \frac{[S]}{K + [S]} \quad \text{Éq. 3.3}$$

Cette équation décrit la vitesse d'une réaction catalysée par des enzymes en se basant sur l'hypothèse de formation d'un complexe enzyme-substrat réversible suivi d'une réaction irréversible de formation du produit. Suite à cette hypothèse, le schème réactionnel est donc :



Briggs et Haldane (1925) ont développé d'autres équations décrivant le même type de phénomènes en se basant sur l'hypothèse de quasi-équilibre selon laquelle la

concentration du complexe enzyme-substrat est constante. D'autres équations ont été développées afin d'intégrer l'impact d'autres variables tels les inhibiteurs et les activateurs sur la vitesse de réaction enzymatiques (Walsh, 2007).

a.	Inhibition compétitive	$\mu = \mu_{max} * \frac{[S]}{K \left(1 + \frac{[I]}{K_I}\right) + [S]}$	Éq. 3.5
b.	Inhibition non compétitive	$\mu = \mu_{max} * \frac{[S]}{(K + [S]) \left(1 + \frac{[I]}{K_I}\right)}$	
c.	Inhibition mixte	$\mu = \mu_{max} * \frac{[S]}{K \left(1 + \frac{[I]}{K_I}\right) + [S] \left(1 + \frac{[I]}{\beta * K_I}\right)}$	
d.	Réaction réversible	$\mu = \frac{\mu_{max,S} * \frac{[S]}{K_S} - u_{max,P} * \frac{[P]}{K_P}}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}}$	

Plusieurs cinétiques de croissance cellulaire utilisent des équations ayant des caractéristiques similaires à celles décrivant les cinétiques enzymatiques. Par exemple, l'équation 3.6 présente des cinétiques utilisées pour décrire la croissance cellulaire (Nelson, 2009 et Khavarpour, 2011).

a.	Tessier (1936)	$\mu = \mu_{max} * \left(1 - \exp\left(\frac{-[S]}{K}\right)\right)$	Éq. 3.6
b.	Monod (1949)	$\mu = \mu_{max} * \frac{[S]}{K + [S]}$	
c.	Moser (1958)	$\mu = \mu_{max} * \frac{[S]^n}{K + [S]^n}$	
d.	Contois (1959)	$\mu = \mu_{max} * \frac{[S]}{[S] + K * [B]}$	
e.	Andrews (1968)	$\mu = \mu_{max} * \frac{[S]}{K + [S] + [S]^2/K_i}$	

Les modèles macroscopiques plus complexes intègrent l'impact de plusieurs substrats sur la croissance cellulaire et la génération de produit. Ils utilisent cependant des cinétiques utilisant des fonctions mathématiques similaires (Tessier,

Monod, etc.) (Cloutier, 2007). Par exemple, la cinétique de croissance cellulaire du modèle Cloutier (2007) est décrite par l'équation 3.7.

$$\mu = \mu_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} \quad \text{Éq. 3.7}$$

Comme on peut le constater, cette cinétique est affectée par quatre variables : le phosphate (Pc), les sucres (Sc) et l'azote (Nc) intracellulaires ainsi que le phosphate extracellulaire (Pi). Tous les autres termes de l'équation sont des paramètres.

Une caractéristique de ce type d'éléments de cinétiques $\left(\frac{[S]}{K+[S]}, 1 - \exp\left(\frac{-[S]}{k}\right), \frac{[S]}{[S]+k*[B]}, \text{ etc.} \right)$ est que leur valeur se situe entre 0 et 1 pour l'ensemble des valeurs possibles de la variable associée (S). Pour cette raison, dans le cadre de ce projet, ces éléments sont appelés fonctions unitaires. Les fonctions unitaires servent à décrire l'impact des variables du modèle sur les cinétiques des réactions. Deux types de fonctions unitaires ont été observés dans les modèles étudiés : les limitations et les inhibitions. Les limitations sont des fonctions dont la valeur augmente lorsque la valeur de la variable associée augmente. Les inhibitions sont des fonctions unitaires dont la valeur diminue lorsque la valeur de la variable associée augmente.

3.1.4 Modèles mathématiques de culture cellulaire

Différents modèles décrivant des systèmes biologiques ont été identifiés. Certains modèles décrivent des procédés de culture de cellules végétales, d'autres décrivent des voies métaboliques et leurs régulations dans des cellules humaines. Ces modèles ont été choisis afin de représenter une bonne diversité au niveau de leur complexité (nombre de variables, de paramètres et de fonctions unitaires) et de leurs caractéristiques. Ce sont des modèles déterministes ayant une suspension

distribuée ou ségréguée et une biomasse distribuée ou ségréguée. Ce sont des modèles métaboliques plus ou moins complexes, comprenant entre 10 et 70 variables d'état, entre 25 et 125 paramètres et entre 3 et 75 fonctions unitaires. Parmi les différents modèles présentés au tableau 3.2, certaines cinétiques contenant des fonctions unitaires affectent plusieurs variables. Cela fait en sorte que des fonctions unitaires présentes dans ces cinétiques se retrouvent dans plusieurs équations différentielles du modèle. Ainsi, on retrouve dans le tableau 3.2 le nombre de fonctions unitaires différentes (uniques) ainsi que le nombre de fonctions unitaires totales. Il est à noter que le modèle de Cloutier est classé comme un modèle macroscopique même s'il contient des variables intracellulaires. La raison pour cela est que ces variables sont uniquement considérées dans la production de biomasse. Le modèle ne contient pas de réactions intracellulaires. Ce cas est un peu à la frontière entre les deux types de modèles.

Caractéristiques	Cloutier 2007	Chassagnole 2002	Iwamoto 2007	Singh 2006
Type de Modèle	Déterministe Suspension ségréguée Biomasse distribuée Macroscopique	Déterministe Suspension distribuée Biomasse distribuée Métabolique	Déterministe Suspension distribuée Biomasse ségréguée Métabolique	Déterministe Suspension distribuée Biomasse distribuée Métabolique
Nombre de variables d'état	10	26	30	68
Nombres de paramètres	31	122	76	107
Nombre de réactions	30	86	126	321
Nombre de fonctions unitaires (différentes)	15	28	6	2
Nombre de fonctions unitaires (totales)	73	39	6	3

Tableau 3.2 : Caractéristiques des modèles étudiés

Le modèle de Cloutier (2007) est un modèle macroscopique décrivant la croissance cellulaire de trois souches de cellules : *Catharanthus roseus*, *Daucus carota* et *Eschscholzia californica* en suspension. Le même modèle est utilisé pour

les trois souches, avec quelques différences. Le modèle pour *Eschscholzia californica* utilise du glucose plutôt que du sucrose comme source de carbone, les modèles pour *Catharanthus roseus* et *Daucus carota* utilisent du sucrose et nécessitent donc une cinétique supplémentaire associée à la dégradation de sucrose en glucose et fructose. De plus, le modèle pour *Eschscholzia californica* utilise de l'ammonium et du nitrate comme source d'azote alors que les autres modèles utilisent uniquement du nitrate. Les valeurs des paramètres du modèle sont calculées indépendamment pour chacune des souches cellulaires. Le modèle pour *Eschscholzia californica* est sélectionné pour être étudié plus en profondeur.

Le modèle de Chassagnole (2002) décrit la dynamique du métabolisme central du carbone chez *Escherichia coli* en réaction à une impulsion de glucose (c'est-à-dire un ajout soudain et connu). Le modèle produit par Iwamoto (2007) décrit l'impact de dommages à l'ADN sur le cycle cellulaire. Le cycle cellulaire est l'ensemble des états dans lesquelles les cellules peuvent se retrouver au niveau de la reproduction. Ce cycle est composé de 5 étapes soit la mitose (S), la méiose (M), deux interphases (G1 et G2) et un stade de dormance (G0). L'état du cycle cellulaire a un impact important sur le procédé, puisque ces phases influencent les voies métaboliques exprimées. Le modèle produit par Singh (2006) décrit le mécanisme de réponse et de contrôle du métabolisme de cellules hépatocytes à la présence d'interleukine-6. Ce mécanisme fait partie de la réponse du corps à une blessure ou une inflammation.

Il existe plusieurs dizaines d'autres modèles représentant la culture cellulaire en bioréacteurs (Leduc (2006), Kiran (2009), Cintolesi (2012), etc.). Ces modèles décrivent le comportement de cultures cellulaires de différentes espèces. Le niveau de complexité de la représentation du métabolisme diffère, mais les types de phénomènes représentés (limitations et inhibitions) sont les mêmes.

3.2 Estimation de la valeur des paramètres

Afin d'estimer les paramètres associés aux différents modèles, on cherche à minimiser la somme des carrés résiduels (RSS) entre les variables mesurées (X_{exp}) et les variables estimées ($X_{\text{modèle}}$) selon l'équation 3.8.

$$RSS(\beta) = \sum_{i=1}^m [X_{\text{exp},i} - X_{\text{modèle},i}(\beta)]^2 \quad \text{Éq. 3.8}$$

Il existe plusieurs algorithmes servant à estimer la valeur des paramètres d'un modèle à partir de données expérimentales. Parmi les plus importants, notons les algorithmes de type Gauss-Newton (Madsen, 2004) tel que l'algorithme Levenberg-Marquardt (Cloutier, 2006), l'algorithme de Nelder-Mead (Conolly, 2003) et les algorithmes génétiques (Shi, 2012). Ces techniques sont itératives, cherchant des minimums locaux. À partir d'une estimation initiale de la valeur des paramètres, celles-ci tentent de trouver un nouvel ensemble de paramètres dont l'erreur d'estimation est plus faible. Pour ce faire, le processus continue jusqu'à ce qu'aucun nouvel ensemble de paramètres ne donne d'erreur plus faible (minimum local).

3.3 Identifiabilité

L'estimation des paramètres est une étape importante de la modélisation. Les modèles étudiés ont comme propriété de comprendre beaucoup de paramètres. Dans certains cas, il est possible que plusieurs ensembles de paramètres donnent des dynamiques très similaires. Ces modèles sont dits non identifiables.

Il existe plusieurs définitions de l'identifiabilité.

- Selon Little (2010), un ensemble de paramètres β est identifiable s'il n'existe pas d'ensemble de paramètres $\hat{\beta} \in \Omega$ (où Ω est l'espace des paramètres possibles) permettant d'obtenir la même distribution des variables que β .

- Pour Raue (2009), un paramètre est identifiable si l'intervalle de confiance de son estimation est fini.

La définition de l'identifiabilité utilisée dans ce projet est celle proposée par Little (2010) selon laquelle un modèle est identifiable s'il n'existe qu'un ensemble de paramètres permettant de minimiser l'erreur entre les variables mesurées et prédites pour des conditions expérimentales données. S'il existe plusieurs ensembles de paramètres permettant de minimiser l'erreur d'estimation du modèle, les paramètres du modèle sont dits non identifiables.

Raue (2009) fait la distinction entre identifiabilité structurelle et identifiabilité pratique :

1. Non-identifiabilité structurelle : la non-identifiabilité structurelle arrive lorsque la structure même du modèle fait en sorte que deux ensembles distincts de paramètres donnent la même prédiction pour les variables d'état. Cela peut être dû à une paramétrisation redondante, c'est-à-dire que l'impact d'un paramètre peut être compensé complètement par une modification à la valeur d'un autre paramètre. Ainsi, même avec un nombre infini de mesures exactes, il est impossible de déterminer une valeur unique pour ces paramètres.
2. Non-identifiabilité pratique : la non-identifiabilité pratique survient lorsque les mesures disponibles (nombre et incertitudes) ne permettent pas de déterminer une valeur unique pour les paramètres. Dans ce cas, deux ou plusieurs ensembles de paramètres conduisent à des prédictions équivalentes à l'intérieur de l'incertitude de mesure.

3.4 Simplification de modèles

De façon générale, la simplification d'un modèle est la réduction de la complexité de ce modèle que ce soit par la réduction du nombre de paramètres ou la réduction du nombre de variables définissant ce modèle. La simplification de modèle par réduction du nombre de paramètres est appelée simplification des paramètres du

modèle. La simplification de modèle par réduction du nombre de variables est appelée réduction de l'ordre du modèle.

L'objectif des méthodes de simplification est de réduire la complexité d'un modèle mathématique décrivant un système tout en conservant la description de la relation entre les entrées et les sorties de ce système (Penzl, 2006). Pour ce faire, il faut quantifier l'impact de la simplification du modèle en fonction de la qualité de l'estimation et décider si l'amélioration de la qualité de l'estimation justifie l'utilisation d'un modèle plus complexe. Cette simplification peut permettre d'améliorer la compréhension du procédé en mettant en valeur les dynamiques fondamentales du système et en éliminant les éléments ayant peu d'impact sur les prédictions du modèle (Eberlein, 2006). De plus, travailler avec des modèles plus simples permet d'obtenir l'estimation des variables d'état de façon plus rapide et plus fiable.

Les modèles mathématiques représentant le métabolisme cellulaire ont souvent de hauts niveaux de complexité. Ainsi, ces modèles sont souvent sujets à des problèmes de non-identifiabilité structurelle (Okino, 1998). De plus, même s'ils sont structurellement identifiables, les données disponibles limitées pour les variables intracellulaires font en sorte qu'ils sont aussi sujets à des problèmes de non-identifiabilité pratique. Une solution à ces problèmes est de réduire la complexité des modèles utilisés grâce à des techniques de simplification.

3.5 Simplification des paramètres d'un modèle

La simplification des paramètres d'un modèle consiste à retirer les paramètres qui n'affectent pas de façon importante l'erreur d'estimation en conservant une forme de cinétique permettant de ne pas modifier les prédictions du modèle. Les techniques de simplification des paramètres présentées dans le cadre de ce projet sont basées sur la mesure de l'erreur d'estimation du modèle due à une variation des paramètres, l'analyse de sensibilité, le critère d'information d'Akaike, le critère d'information Bayesian et l'analyse de niveau temporel.

3.5.1 Sensibilité du modèle à une variation de la valeur des paramètres

Il est possible d'évaluer l'impact des paramètres de façon simple et rapide en modifiant une par une les valeurs des paramètres et en vérifiant si cette modification entraîne une variation significative de l'erreur d'estimation des modèles. Cette technique permet d'obtenir une estimation grossière de la sensibilité du modèle $\left(\frac{\partial X}{\partial \beta}\right)$. Dû à sa simplicité de mise en place, cette technique est une des premières utilisées pour étudier l'impact de la valeur des paramètres sur les prédictions du modèle. Le tableau 3.3 présente les résultats de cette analyse pour quatre modèles tels que présentés dans les articles dont ils sont tirés. Chacun de ces articles présente une analyse tentant de déterminer quels sont les paramètres qui ont un impact significatif sur l'erreur d'estimation du modèle. Pour ce faire, les méthodes utilisées dans ces articles appliquent une variation différente de la valeur des paramètres et utilisent un seuil différent pour définir ce qui est considéré comme une variation significative de l'erreur d'estimation du modèle.

Modèle	Nb. Paramètres	Caractéristiques de l'analyse		Résultat de l'analyse	
		Variation des paramètres	Variation significative de l'erreur	Nb. Paramètres significatifs	% de paramètres significatifs
Jolicoeur, 2003	51	±10% et ±50%	≥30%	9	17.6
Leduc, 2006	89	±50%	≥5%	19	21.3
Cloutier, 2007	39	±25%	>25%	5	12.8
Iwamoto, 2008	75	Plusieurs	≥10%	6	8.0

Tableau 3.3 : Caractéristiques des modèles étudiés

Il est possible de constater qu'il n'y a pas de seuil absolu pour la variation à appliquer à la valeur des paramètres de même qu'il n'y a pas de valeur unique pour représenter une modification significative de l'erreur d'estimation. Le choix de ces valeurs dépend de la nature du modèle. Dans le tableau 3.3, il est intéressant de constater que pour tous ces modèles, le nombre de paramètres ayant une influence

considérée comme significative est faible (<25%). Cette information peut être utilisée afin de cibler les phénomènes ayant un impact sur la culture (Iwamoto, 2008).

Cependant, cette méthode ne permet pas de conclure sur la présence ou l'absence de phénomènes, puisque, dans certains cas, même si la valeur d'un paramètre peut être doublée sans changer significativement l'erreur du modèle, cela ne veut pas dire qu'il est possible de le retirer sans affecter les prédictions du modèle. Par exemple, dans le cas d'une cinétique de Monod dont la valeur du paramètre est beaucoup plus petite que les concentrations de la variable d'état qui y est associé, la présence de la cinétique fera en sorte que le taux de réaction soit nul lorsque la variable associée tend vers zéro. Cependant, la valeur effective du paramètre a peu d'impact sur la prédiction du modèle tant que l'ordre de grandeur de la concentration de la variable associée est plus élevé. Les paramètres dont la valeur n'influence pas la prédiction du modèle peuvent être associés à des phénomènes d'un ordre de grandeur différent de celui des variables associées, mais situé à l'intérieur de la plage de données expérimentales (exemple précédent). Dans ce cas, on ne peut pas enlever ce paramètre (ou la fonction unitaire associée), car ceci changerait notablement les prédictions du modèle.

3.5.2 Analyse de sensibilité

La sensibilité d'une variable par rapport à la valeur d'un paramètre est définie comme la dérivée partielle de la variable par rapport à dérivée partielle de la valeur du paramètre $\left(\frac{\partial X}{\partial \beta}\right)$. Dans le cas d'un modèle différentiel, la sensibilité varie dans le temps. Ainsi, il est possible de calculer la sensibilité d'une variable par rapport à un paramètre à n'importe quel moment dans du temps. Cependant pour des aspects pratiques la sensibilité est calculée en quelques points. La durée de l'expérience et l'intervalle entre les points où la sensibilité est calculée peuvent influencer l'analyse de l'impact des paramètres, en effet, une mauvaise sélection des points testés peut entraîner une surestimation ou une sous-estimation de l'impact de cinétiques associées sur les prédictions du modèle.

Par exemple, certaines réactions de transports actifs peuvent être significatives uniquement lorsque la concentration de substrat est faible. Si les données expérimentales ne couvrent que des conditions expérimentales où la concentration du substrat est élevée, ou si la majorité des points expérimentaux représente des conditions où la concentration de substrat est élevée, il est possible que les paramètres associés à ces réactions soient considérés comme non significatifs.

La sensibilité est donc évaluée pour m points de la simulation. La matrice de sensibilité (Z) comporte n colonnes (nombre de paramètres du modèle) et m lignes (nombre d'observations de la variable) (équation 3.9).

$$Z = \begin{bmatrix} \frac{\partial X_1}{\partial \beta_1} & \cdots & \frac{\partial X_1}{\partial \beta_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial X_m}{\partial \beta_1} & \cdots & \frac{\partial X_m}{\partial \beta_n} \end{bmatrix} \begin{matrix} n = \text{nb. de paramètres} \\ m = \text{nb. de points} \end{matrix} \quad \text{Éq. 3.9}$$

En calculant la norme euclidienne de chaque colonne, il est possible de comparer l'importance de chaque paramètre. Plus la norme euclidienne est élevée, plus le paramètre a d'impact sur la prédiction de la variable (Wu, 2011). Ce calcul prend en compte la sensibilité de chaque variable individuellement par rapport à la valeur de l'ensemble des paramètres.

Pour considérer la sensibilité de l'ensemble des variables du modèle par rapport aux paramètres, deux options sont possibles. Il est possible de considérer chaque variable individuellement (par exemple : reformer une matrice où chaque ligne est le vecteur de norme de chaque variable). Il est aussi possible de générer une matrice contenant la sensibilité pour chaque point (m) de chaque variable (v) pour chaque paramètre (n). Cette méthode revient à placer les matrices de sensibilité des variables une après l'autre pour former une nouvelle matrice tel que démontré à l'équation 3.10. Si le nombre de points considérés pour chaque variable est différent,

cela accordera proportionnellement plus d'importance aux variables ayant plus de points.

$$Z = \begin{bmatrix} \frac{\partial X_1}{\partial k_1} & \dots & \frac{\partial X_1}{\partial k_n} \\ \dots & \dots & \dots \\ \frac{\partial X_m}{\partial k_1} & \dots & \frac{\partial X_m}{\partial k_n} \\ \dots & \dots & \dots \\ \frac{\partial X_1}{\partial k_1} & \dots & \frac{\partial X_1}{\partial k_n} \\ \dots & \dots & \dots \\ \frac{\partial X_m}{\partial k_1} & \dots & \frac{\partial X_m}{\partial k_n} \\ \dots & \dots & \dots \\ \vdots & & \vdots \\ \frac{\partial X_1}{\partial k_1} & \dots & \frac{\partial X_1}{\partial k_n} \\ \dots & \dots & \dots \\ \frac{\partial X_m}{\partial k_1} & \dots & \frac{\partial X_m}{\partial k_n} \\ \dots & \dots & \dots \end{bmatrix} \begin{matrix} \text{Var 1} \\ \\ \text{Var 2} \\ \\ \\ \\ \text{Var } v \end{matrix} \quad \text{Éq. 3.10}$$

$n = \text{nb. de paramètres}$
 $m = \text{nb. de points (peut différer d'une variable à l'autre)}$
 $v = \text{nb. de variables}$

Dû à la grande complexité des modèles utilisés, il faut calculer numériquement les éléments de la matrice Z . Cela peut être fait en résolvant les équations de sensibilité (Leis, 1988) ou par approximation à partir de plusieurs simulations avec des perturbations dans la valeur des paramètres (Wu, 2011), ce qui peut être coûteux en temps de calcul pour des modèles complexes. À partir de la matrice Z , Wu (2011) a développé une méthode où chaque paramètre est estimé un par un à partir du plus estimable (paramètre dont la colonne a la norme euclidienne la plus élevée). Chaque paramètre est calculé à l'aide de la technique de régression des moindres carrés. La méthode de Wu s'arrête lorsque l'estimation de nouveau paramètre ne permet pas d'améliorer significativement le modèle. Ainsi, cette méthode utilise l'analyse de sensibilité pour développer une technique d'estimation de la valeur des paramètres. Les autres paramètres conservent la valeur qui leur a été posée initialement. Cette méthode ne comprend pas d'étape de simplification, car elle n'enlève aucun paramètre.

3.5.3 Critère d'information d'Akaike

Le critère d'information d'Akaike (AIC) (Akaike, 1974) est utilisé pour comparer des modèles afin de quantifier l'impact du nombre de paramètres du modèle sur la qualité de l'estimation. Le critère d'Akaike pénalise l'ajout de paramètres de façon à sélectionner un modèle offrant une bonne estimation tout en ayant un minimum de paramètres. Ce critère est utilisé après que l'algorithme Levenberg-Marquardt (Marquardt, 1963) ait permis d'estimer la valeur des paramètres par minimisation de l'erreur d'estimation. Le calcul du critère d'information d'Akaike s'effectue de la façon suivante (Posada, 2004) :

$$AIC = -2 * \mathcal{L} + 2n \quad \text{Éq. 3.11}$$

où n est le nombre de paramètres et \mathcal{L} est la fonction de vraisemblance logarithmique. Lorsqu'on suppose que les erreurs expérimentales ont une distribution normale, la fonction de vraisemblance logarithmique prend la forme de l'équation 3.12 (Burnham, 2002), où m est la taille de l'échantillon (nombre de données) et RSS est la somme des carrés résiduels.

$$\mathcal{L} = -\frac{1}{2} * m * \ln\left(\frac{RSS}{m}\right) \quad \text{Éq. 3.12}$$

Le critère d'information d'Akaike devient donc :

$$AIC = m * \ln\left(\frac{RSS}{m}\right) + 2n \quad \text{Éq. 3.13}$$

Lorsque la taille de l'échantillon m est petite par rapport au nombre de paramètres n ($m/n < 40$), il est recommandé d'utiliser un critère d'information corrigé (AIC_c) (Hurvich, 1989).

$$AIC_c = AIC + 2n \frac{n-1}{m-n-1} \quad \text{Éq. 3.14}$$

La différence Δ_i entre le critère d'Akaike pour le modèle avec le AIC minimal et le modèle courant est présentée dans l'équation 3.15. Une heuristique est que $\Delta_i < 2$ suggère un très bon modèle, $3 < \Delta_i < 7$ indique un modèle présentant des incertitudes élevées, $\Delta_i > 10$ suggère que le modèle est inadapté à décrire la situation modélisée (Burnham, 2002). La pondération d'Akaike (ω_i) (équation 3.15) est la probabilité que chaque sous-modèle soit le meilleur de l'ensemble considéré. La pondération d'Akaike (ω_i) sert à comparer les modèles entre eux. Plus la pondération d'Akaike est élevée, plus le modèle est simple et précis (DiStefano, 2015).

$$\Delta_i = AIC_{Ci} - AIC_{Cmin}$$

$$\omega_i = \frac{\exp(-\frac{\Delta_i}{2})}{\sum_{r=1}^R \exp(-\frac{\Delta_r}{2})} \quad \text{Éq. 3.15}$$

Le critère d'information d'Akaike est utilisé en biologie des systèmes pour caractériser des modèles. Par exemple, Chen (2005) utilise le critère d'information d'Akaike pour quantifier les mécanismes de régulation de 20 gènes chez *Saccharomyces cerevisiae*. Peskov (2012) utilise le critère d'Akaike pour comparer différents modèles de contrôle du métabolisme central du carbone chez *Escherichia coli*.

Il est à noter que le critère d'Akaike n'est pas adapté pour évaluer les modèles idéaux (modèles représentant parfaitement les données expérimentales). En effet, le RSS de ces modèles est près de 0. Cela fait en sorte que la valeur du critère d'Akaike tend vers moins l'infini. Ainsi, très peu de simplifications seront favorisées puisqu'une petite modification au RSS entraîne une forte augmentation de la valeur du critère. Alors que ces modèles sont souvent surparamétrés (ils modélisent l'erreur de mesure), ils seraient plus représentatifs de la réalité avec moins de paramètres.

3.5.4 Critère d'information Bayésien

Le critère d'information Bayésien est similaire au critère d'information d'Akaike. Ce critère tient compte du nombre de paramètres du modèle et de la précision de l'estimation du modèle. L'équation permettant de calculer le critère d'information Bayésien est (Schwarz, 1978) :

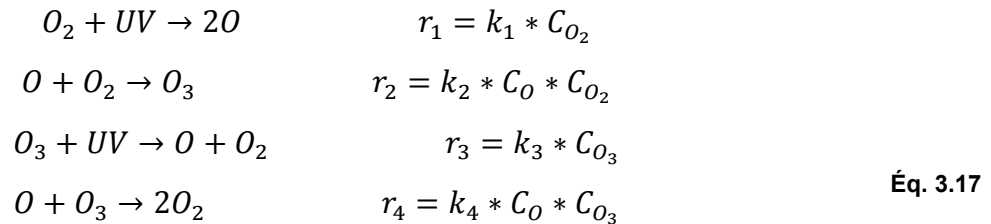
$$\begin{aligned}
 BIC &= -2 * \mathcal{L} + n * \ln(m) \\
 \mathcal{L} &= -\frac{1}{2} * m * \ln\left(\frac{RSS}{m}\right) \\
 BIC &= m * \ln\left(\frac{RSS}{m}\right) + n * \ln(m)
 \end{aligned}
 \tag{Éq. 3.16}$$

Le critère d'information Bayésien a été utilisé pour comparer plusieurs modèles présentant des voies métaboliques alternatives en biologie des systèmes (Vyshemirsky, 2008) ainsi que les voies de régulation de l'expression de gènes (Oates, 2014 et Perin, 2003). En général, le critère d'information Bayésien est plus sévère que le critère d'information d'Akaike puisque, si le nombre de données est plus grand ou égal à 8, $n * \ln(m) > 2 * n$, la pénalisation pour l'ajout de paramètres est plus sévère pour ce critère. Cependant, il ne change pas fondamentalement le classement des modèles. De la façon dont le critère d'information Bayésien est utilisé, cette méthode n'est pas à proprement parler une méthode de simplification, mais plutôt une méthode de classification des modèles.

3.5.5 Analyse de niveau temporel

La dynamique des systèmes complexes contient souvent plusieurs échelles de temps. Après une phase initiale de latence, les réactions se produisant rapidement peuvent être considérées comme instantanées par rapport aux réactions lentes. Cette situation est à la base de l'hypothèse de régime permanent utilisée dans plusieurs modèles biologiques (ex. : Leduc, 2006). L'hypothèse de régime permanent s'applique lorsque le procédé réagit à des échelles de temps différentes. Autrement dit, il est possible que certaines réactions se produisent tellement vite par rapport à d'autres qu'elles semblent instantanées ou encore qu'elles se produisent si lentement

qu'elles semblent avoir une vitesse nulle. Lorsque la vitesse des réactions est très élevée par rapport à l'échelle de temps observée, la simplification du modèle suivante peut être réalisée telle que décrite par Okino (1998) : tout apport de masse à une espèce chimique (variable) posée comme étant en régime permanent, est considéré comme consommé instantanément. Autrement dit, l'équation différentielle décrivant l'évolution de la variable ne s'applique plus, mais est remplacée par une équation algébrique. Cela ne veut pas dire que la dérivée de la variable vaut 0, mais plutôt que l'évolution de la variable est définie par une équation algébrique reliant cette variable aux autres variables du système. L'exemple suivant provient de Seinfeld (2006).



Où

$$\frac{dC_{O_3}}{dt} = k_2 * C_O * C_{O_2} - k_3 * C_{O_3} - k_4 * C_O * C_{O_3}$$

Posons l'hypothèse de régime permanent sur la variable O_3 .

$$\begin{aligned}
 \frac{dC_{O_3}}{dt} &= k_2 * C_O * C_{O_2} - k_3 * C_{O_3} - k_4 * C_O * C_{O_3} = 0 \\
 C_{O_3} &= \frac{k_2 * C_O * C_{O_2}}{k_3 + k_4 * C_O}
 \end{aligned}
 \tag{Éq. 3.18}$$

Ainsi, l'équation $\frac{dC_{O_3}}{dt}$ est remplacée par l'équation algébrique sur O_3 (équation 3.18). Toutefois, cela ne veut pas dire que $\frac{dC_{O_3}}{dt} = 0$, puisque O_3 varie en fonction de O_2 et O . L'hypothèse de régime permanent peut être aussi utilisée lorsqu'une ou plusieurs réactions affectant une variable se produisent très lentement de telle sorte que la variation engendrée par ces réactions sur la variable peut être considérée

comme négligeable. Les éléments de la cinétique associés à ces réactions peuvent alors être estimés comme valant zéro.

L'utilisation de l'hypothèse de régime permanent permet de réduire le niveau de complexité d'un modèle en diminuant le nombre de cinétiques du modèle. Cela entraîne aussi une diminution du nombre de paramètres à estimer. L'implémentation de cette technique pose cependant certaines difficultés. Lorsque plusieurs réactions affectent une espèce chimique, il est difficile de déterminer quelles sont les réactions rapides et les réactions lentes. Cela peut faire en sorte que le modèle réduit ne soit pas une bonne approximation des cinétiques réactionnelles.

3.6 Réduction de l'ordre d'un modèle

L'objectif de la réduction d'ordre d'un modèle est de diminuer la dimension de l'espace des variables d'état dans lequel se situe le modèle. Les techniques de réduction de l'ordre d'un modèle ont principalement été développées dans le domaine du contrôle avec pour objectif de réduire l'effort de calcul requis afin d'exercer le contrôle d'un système ou d'un procédé (Zhou, 1996). Les systèmes étudiés (Σ) sont définis comme suit :

$$\Sigma \begin{cases} \frac{dx(t)}{dt} = f(x(t), u(t), \beta) \\ y(t) = h(x(t), u(t), \beta) \end{cases} \quad \text{Éq. 3.19}$$

où

$$x(t) \in \mathbb{R}^v, y(t) \in \mathbb{R}^p, u(t) \in \mathbb{R}^w, \beta \in \mathbb{R}^n$$

Dans cette représentation, x représente le vecteur de variables d'état de dimension v , u représente les fonctions d'ajout ou d'excitation du système de dimension w et y représente les sorties ou les variables d'intérêt du système de dimension p . La réduction d'ordre du modèle a pour but d'obtenir un système simplifié ($\hat{\Sigma}$) permettant d'approximer le système original (Σ), où :

$$\hat{x}(t) \in \mathbb{R}^k, \hat{y}(t) \in \mathbb{R}^p, u(t) \in \mathbb{R}^w \quad \text{Éq. 3.20}$$

Le système simplifié doit permettre de réduire le nombre de variables d'état ($k < v$) idéalement de façon importante ($k \ll v$). Cela, tout en respectant les conditions suivantes (Antoulas, 2001) :

1. L'erreur d'approximation ($\|y - \hat{y}\|$) doit être faible pour toutes les valeurs de u .
2. La stabilité du système doit être préservée (une variation de $u(t)$ entraîne une variation finie de $y(t)$).
3. La procédure doit être efficace au niveau de la puissance de calcul requise.
4. La procédure doit être automatique (basée sur la tolérance de l'erreur d'approximation).

Deux éléments sont problématiques en ce qui a trait à l'utilisation des techniques de réduction d'ordre d'un modèle dans le cadre de ce projet. Tout d'abord, plusieurs de ces techniques ont été élaborées afin de simplifier des modèles dynamiques linéaires (Antoulas, 2001), c'est à dire sous la forme présentée à l'équation 3.21.

$$\Sigma \begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad \text{Éq. 3.21}$$

Or, les modèles étudiés dans ce projet comportent souvent de fortes non-linéarités. De plus, les modèles d'ordre réduit sont souvent une représentation abstraite du système modélisé. Ils ont donc perdu la signification physique ou biologique du modèle original (Antoulas, 2001). Or, dans le cadre de ce projet, il est souhaité que les modèles simplifiés conservent une signification quant aux phénomènes présents dans le procédé étudié.

3.6.1 Décomposition orthogonale propre

La technique de décomposition orthogonale propre peut être appliquée à la fois à des systèmes linéaires très complexes et à des systèmes non linéaires (Antoulas, 2001). Cette méthode fait appel à la décomposition en valeurs singulières (« Singular Value Decomposition » - SVD) selon laquelle toute matrice peut être décomposée en un produit de trois matrices :

$$X = UDV^T \quad \text{pour } X \in \mathbb{R}^{m \times v} \quad \text{Éq. 3.22}$$

où U est une matrice unitaire de dimension m par m dont les colonnes sont les vecteurs propres de $X^T X$, V est une matrice unitaire de dimension v par v dont les colonnes sont les vecteurs propres de XX^T et D est une matrice rectangulaire diagonale où la racine carrée des valeurs de la diagonale sont appelées valeurs propres des matrices $X^T X$ et XX^T (Golub, 1996). Il est possible d'estimer la matrice X avec un ordre réduit grâce à l'équation 3.23 en incluant uniquement les valeurs propres représentant une portion significative de la somme des valeurs propres (ex. : *Valeur propre* > 2% * \sum *Valeurs propres*).

$$X = \sum_{i=1}^k \sigma_i * u_i * v_i^T \quad \text{Éq. 3.23}$$

où k est l'ordre utilisé (par rapport à v qui est l'ordre de la matrice X) (idéalement, $k \ll v$), σ sont les éléments de la diagonale de la matrice D, u sont les vecteurs colonnes de la matrice U et v sont les vecteurs colonnes de la matrice V. La méthode de décomposition orthogonale propre consiste à appliquer la décomposition en valeurs singulières sur la matrice d'échantillons des variables d'état à des temps donnés (χ) suite à une perturbation fixe du système (u(t)).

$$\chi = [x(t_1) \ x(t_2) \ x(t_3) \ \dots \ x(t_N)]' \in \mathbb{R}^{m \times v} \quad \text{Éq. 3.24}$$

En général, le nombre d'échantillons (m) est beaucoup plus grand que le nombre de variables d'état (v). Une approximation d'ordre réduit de la matrice peut être réalisée :

$$X = UDV^T \sim U_k D_k V_k^T \quad \text{Éq. 3.25}$$

$x(t)$ peut être approximé comme suit : $x(t) \sim U_k * \xi(t)$, $\xi(t) \in \mathbb{R}^k$. Le système d'ordre réduit peut donc être estimé par les équations suivantes :

$$\frac{d\xi(t)}{dt} = U_k^T f(U_k \xi(t), u(t)), \quad y(t) = h(U_k \xi(t), u(t)) \quad \text{Éq. 3.26}$$

Cette simplification est toutefois restreinte à l'étude de la réponse du système à la perturbation $u(t)$ appliquée au système.

3.6.2 Techniques de troncature équilibrée

La technique de troncature équilibrée permet de simplifier des modèles dynamiques linéaires. Tel que mentionné précédemment, la structure de ces systèmes est la suivante :

$$\Sigma = \begin{cases} \frac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad \text{Éq. 3.27}$$

Où A , B , C et D sont des matrices de coefficients. La notation suivante est souvent utilisée pour décrire ce type de modèle (Penzl, 2006) :

$$\Sigma = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \in \mathbb{R}^{(n+p) \times (n+m)} \quad \text{Éq. 3.28}$$

Les équations de Lyapunov permettent d'obtenir de l'information sur la dynamique du système.

$$\begin{aligned} AP + PA^T + BB^T &= 0 \\ A^T Q + QA + C^T C &= 0 \end{aligned} \quad \text{Éq. 3.29}$$

Les matrices P et Q obtenues par la résolution de ce système d'équations sont appelées Gramiens d'observabilité (P) et Gramiens de contrôlabilité (Q) (Gugercin, 2004). La racine carrée des valeurs propres du produit des matrices P et Q sont appelées les valeurs singulières de Hankel du système ($\sigma_i(\Sigma)$).

$$\sigma_i(\Sigma) = \sqrt{\lambda_i(PQ)} \quad \text{Éq. 3.30}$$

Les k plus grandes valeurs singulières de Hankel sont conservées pour former le modèle réduit. Le système est donc divisé en fonction de la valeur des paramètres de Hankel. Le nombre de valeurs singulières à inclure est déterminé selon la relation suivante (Gugercin, 2004) :

$$\|\Sigma - \Sigma_r\| \leq 2 * \sum_{i=k+1}^q \sigma_i(\Sigma) \quad \text{Éq. 3.31}$$

Cette technique permet de réduire l'ordre d'un modèle. Toutefois, elle ne s'applique pas dans le présent projet puisqu'elle ne s'applique qu'aux modèles linéaires.

Ainsi, aucune des méthodes de simplification identifiées dans la littérature n'a permis de répondre aux objectifs de ce projet, c'est-à-dire d'être en mesure de simplifier des modèles non linéaires tout en conservant la signification du modèle.

L'objectif d'un modèle est de décrire comment un système évolue et comment il réagit à des perturbations externes. Pour ce faire, une structure générale du modèle est posée comprenant les réactions composant le schéma réactionnel ainsi que les vitesses de réaction. Ces éléments constituent l'ensemble des hypothèses sur les phénomènes représentés dans le modèle et qui agissent sur le comportement du système. Des données expérimentales permettent d'estimer les paramètres du modèle par minimisation de l'erreur d'estimation du modèle. En analysant l'impact des différents paramètres du modèle sur les prédictions de celui-ci, il est possible d'obtenir de l'information sur l'importance relative des différents phénomènes pris en compte par le modèle. Il est aussi possible de vérifier si les différents phénomènes intégrés au modèle ont bel et bien été observés lors des scénarios d'expériences utilisées pour la simulation des variables. Lors de la simplification du modèle, la forme des cinétiques de réaction de même que les paramètres du modèle sont connus. Il est alors possible d'estimer l'impact d'éléments du modèle sur les prédictions de ce dernier. L'objectif de l'algorithme de simplification est de retirer du modèle les fonctions unitaires et les impacts réactionnels dont l'effet est négligeable sur les prédictions du modèle.

4.1 Données

Les modèles sont analysés dans les conditions expérimentales disponibles dans les articles présentant ces modèles. Une grande partie de l'analyse est réalisée par l'observation de l'ordre de grandeur d'un phénomène par rapport à l'ordre de grandeur des données. Cela est représenté par la plage de données modélisées par rapport à la zone d'impact d'un phénomène (caractéristique d'un modèle).

4.1.1 Données expérimentales

Les données expérimentales proviennent d'expériences où des valeurs des différentes variables sont mesurées en ligne ou hors-ligne, par échantillonnage. Elles sont utilisées pour l'estimation de la valeur des paramètres par minimisation de l'erreur du modèle. Les modèles étudiés ici proviennent de la littérature. L'estimation des valeurs des paramètres à l'aide de données expérimentales y est déjà réalisée. Dans la cadre de ce projet, les données expérimentales provenant de la littérature sont utilisées pour définir la plage de données expérimentales de même que les scénarios expérimentaux auxquels seront soumis les modèles.

4.1.2 Variables modélisées

Les variables modélisées sont les variables générées par le modèle. Les transferts avec l'extérieur du système (entrées et sorties) de même que les conditions initiales du système sont considérés comme parfaitement connus. Lors de la simplification du modèle, des éléments de ce dernier sont évalués en fonction de leurs impacts sur les variables modélisées. Il est considéré que si une modification au modèle n'a pas d'impact significatif sur le comportement des variables modélisées, cette modification n'affectera pas significativement l'erreur d'estimation du modèle. Ainsi, il est possible d'effectuer la simplification sans avoir recours aux données expérimentales.

4.1.3 Plage de données

La plage de données représente l'étendue des valeurs observées lors des expériences pour chacune des variables. La plage de données peut être évaluée de deux façons : par rapport aux données expérimentales (plage de données expérimentales) et par rapport aux données modélisées (plage de variables modélisées). La plage de données expérimentales est définie comme l'ensemble des

valeurs entre le minimum et le maximum observé. Les données aberrantes sont considérées comme déjà retirées de l'ensemble des données. Par exemple, les données expérimentales suivantes (Cardin-Bernier, 2010) représentent les valeurs de trois variables (concentration cellulaire, concentration de glucose et concentration d'ammonium) pour une expérience :

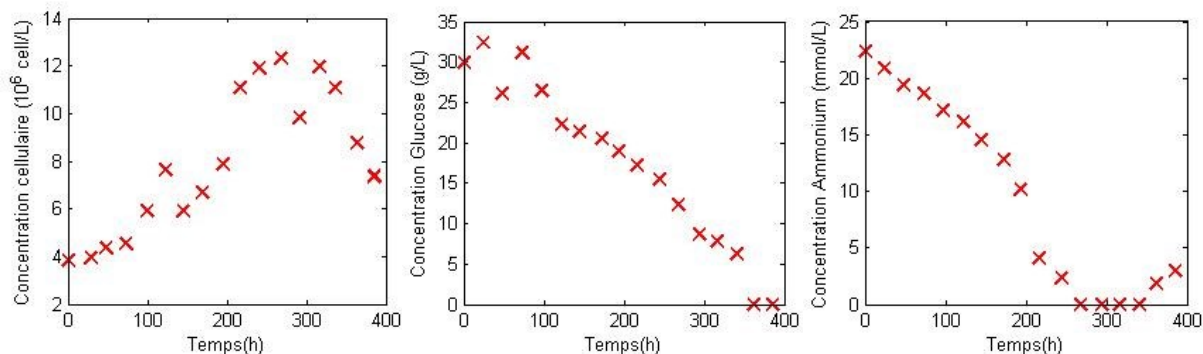


Figure 4.1 : Données expérimentales associées à trois variables (Cardin-Bernier, 2010)

On constate que la plage de données expérimentales pour la concentration cellulaire est de 3.8 à 12.3 ($\times 10^6$ cellules/L), pour la concentration de glucose, elle est de 0 à 32.4 g/L et pour l'ammonium, elle est de 0 à 22.3 mmol/L. Il peut être utile de représenter les valeurs de données expérimentales sous forme de distribution continue. Cette représentation permet de voir quelles sont les valeurs des variables observées de nombreuses fois et quelles sont les valeurs ayant été peu observées. Pour représenter les données expérimentales sous forme de distribution, on fait appel à un algorithme d'estimation par noyau (Kernel Density Estimation) (Wand, 1995).

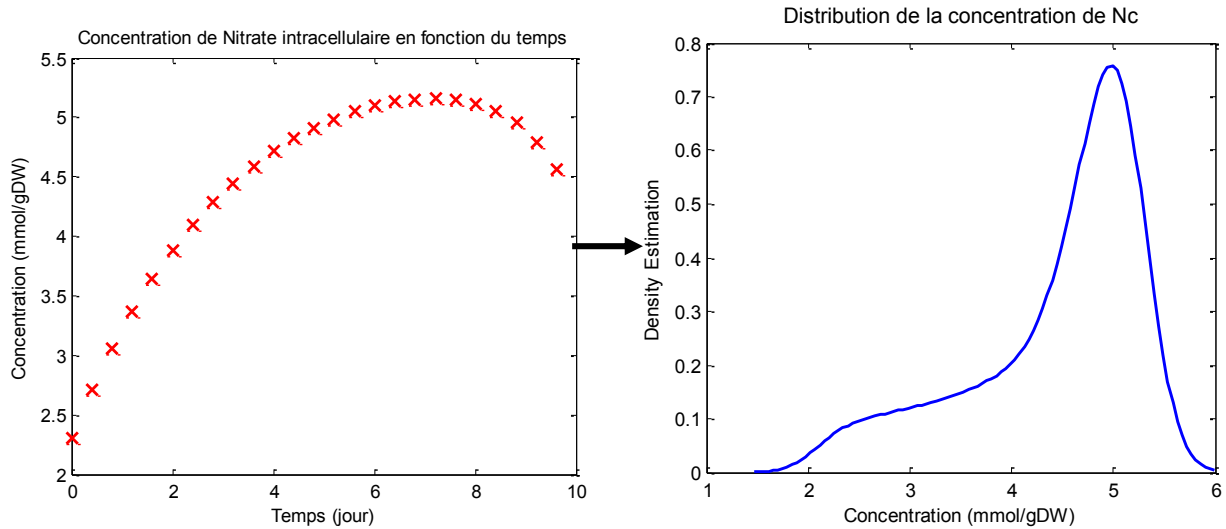


Figure 4.2 : Distribution de données expérimentales

La plage de données expérimentales représente l'étendue des valeurs observées, soit l'intervalle entre les valeurs minimale et maximale des données expérimentales considérées comme valide. La plage est représentée par un carré jaune dans la figure 4.3.

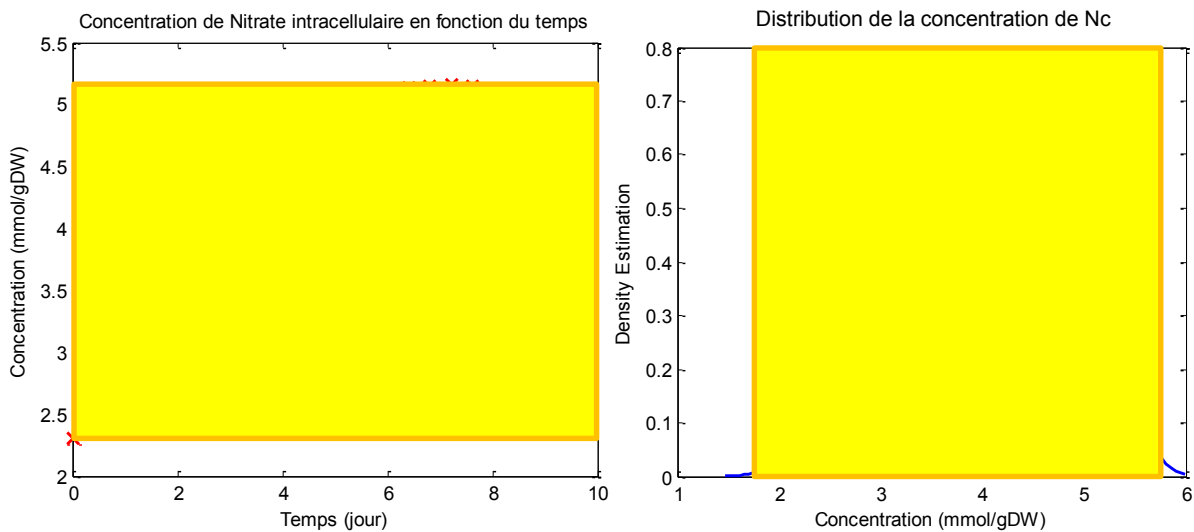


Figure 4.3 : Plage de données expérimentales

La plage de données modélisées est définie de la même façon, mais appliquée aux données générées par le modèle dans les mêmes conditions plutôt qu'aux

données expérimentales. Si le modèle est une bonne représentation du système modélisé, la plage de données modélisées devrait être similaire à la plage de données expérimentales. Cela peut facilement être vérifié en comparant les ensembles de données.

4.2 Variables d'état

La majorité des variables d'état des modèles sont les concentrations de différentes espèces chimiques ou biologiques dans le procédé, complétées si nécessaire, par le pH, la température ainsi que le volume de culture.

4.2.1 Variables d'intérêt et variables non essentielles

Dans le cadre de ce projet, les variables d'état sont divisées en deux groupes : les variables d'intérêt et les variables non essentielles. Les variables d'intérêt sont choisies préalablement à l'analyse du modèle et sont sélectionnées en fonction de leur importance sur le procédé. Parmi les variables d'intérêt, on retrouve souvent la concentration cellulaire puisque cette variable influence grandement le procédé de même que la concentration de la molécule qu'on cherche à produire grâce au procédé puisque c'est cette variable qu'on cherche à optimiser et donc à connaître. Les variables qui ne sont pas des variables d'intérêt sont appelées variables non essentielles. Comme le rôle des variables non essentielles est d'aider le système à représenter le comportement des variables d'intérêt, la qualité du modèle est alors caractérisée par sa capacité à prédire les variables d'intérêt plutôt que l'ensemble des variables du système. Cette distinction peut être appliquée au début du processus de modélisation lors de l'estimation des paramètres du modèle en calculant la réduction de l'erreur du modèle uniquement sur les variables d'intérêt. Dans le cadre de ce projet, puisque le processus de simplification est réalisé à partir de modèles dont les paramètres ont déjà été estimés, cette distinction est réalisée au moment de la simplification des variables.

4.3 Modèles

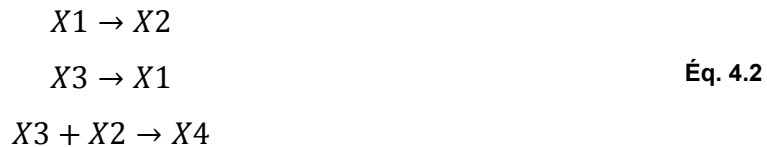
Le modèle mathématique représentant la culture cellulaire est composé d'un ensemble d'équations différentielles représentant l'évolution des variables d'état dans le temps. Cette section présente la structure de ces équations de même que certains éléments (réactions et fonctions unitaires) qui seront utilisés pour des fins de simplification des modèles.

4.3.1 Structure du modèle

Les modèles issus d'un schème réactionnel sont souvent représentés de la façon suivante :

$$\Sigma: \frac{dx}{dt} = \text{Matrice Stochiométrique} * \text{Vecteur de Taux de réactions.} \quad \text{Éq. 4.1}$$

Par exemple, pour le schème réactionnel suivant :



Le modèle est :

$$\Sigma: \begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dx_3}{dt} \\ \frac{dx_4}{dt} \end{bmatrix} = \begin{bmatrix} -K_{x1r1} & K_{x1r2} & 0 \\ K_{x2r1} & 0 & -K_{x2r3} \\ 0 & -K_{x3r2} & -K_{x3r3} \\ 0 & 0 & K_{x4r3} \end{bmatrix} * \begin{bmatrix} \mu_{r1} \\ \mu_{r2} \\ \mu_{r3} \end{bmatrix} \quad \text{Éq. 4.3}$$

où K est la matrice de coefficients stœchiométriques (K est un sous-ensemble de β , l'ensemble des paramètres du modèle) et où μ est la matrice de taux de réactions. Les taux de réactions associés aux modèles biologiques ne sont généralement pas tous constants. Ainsi, dans la plupart des modèles, les termes μ sont des fonctions dépendant des variables du modèle et des transferts avec l'extérieur du système $\mu(x, \beta, u, y)$

4.3.2 Réactions

Les réactions sont chacune des étapes du schème réactionnel. Les étapes représentent des transformations entre les différentes variables du système. Comme mentionné, elles peuvent être de nature chimique, physique ou biologique.



Dans l'équation 4.4, $X1$, $X2$ et $X3$ sont des concentrations chimiques qui interviennent dans les trois réactions $rx1$, $rx2$, $rx3$. Chaque réaction se produit à une certaine vitesse appelée taux de réaction. En combinant l'impact de l'ensemble des réactions sur chacune des variables, ainsi que la vitesse de chaque réaction, il est possible de prédire l'évolution de chacune des variables d'état du système par un système d'équations différentielles. Par exemple, le système d'équations suivant représente l'évolution des variables du schème réactionnel de l'équation 4.4.

$$\begin{bmatrix} \frac{dX_1}{dt} \\ \frac{dX_2}{dt} \\ \frac{dX_3}{dt} \\ \frac{dX_4}{dt} \end{bmatrix} = \begin{bmatrix} -K_{x1r1} * \mu_{r1} + K_{x1r2} * \mu_{r2} \\ K_{x2r1} * \mu_{r1} - K_{x2r3} * \mu_{r3} \\ -K_{x3r2} * \mu_{r2} - K_{x3r3} * \mu_{r3} \\ K_{x4r3} * \mu_{r3} \end{bmatrix}
 \tag{Éq. 4.5}$$

Chacune des réactions affecte plusieurs variables. On retrouve donc leur impact dans les cinétiques associées à chaque variable qu'elles affectent.

4.3.3 Impact réactionnel

Dans ce projet, on considère l'impact des réactions sur chacune des variables individuelles plutôt que sur l'ensemble des variables qu'elles affectent. Le terme impact réactionnel a été utilisé pour définir l'impact d'une réaction sur une variable spécifique. Selon cette définition, l'impact de la réaction chimique est analysé indépendamment sur chacune des variables. Pratiquement, cela fait en sorte que chaque élément qui est additionné ou soustrait à la dérivée de chaque variable est un impact réactionnel. La raison pour laquelle cette définition est choisie est que cela simplifie l'analyse en enlevant les connexions entre différents éléments de la cinétique. En contrepartie, cela fait en sorte qu'il est possible que l'impact d'une réaction chimique soit analysé différemment pour certaines variables. Par exemple, pour la réaction $A + B \rightarrow C$. Il est possible que l'analyse démontre que la réaction a un impact significatif sur les variables A et C, mais négligeable sur la variable B. L'impact réactionnel sera donc retiré sur la variable B, mais les impacts réactionnels sur A et C seront conservés. Concrètement, cela fait en sorte que la réaction modélisée sera $A \rightarrow C$. Les deux modèles (le modèle original et le modèle simplifié) fournissent des prédictions équivalentes. De plus, le fait de séparer l'impact réactionnel de la réaction permet d'enlever un paramètre du modèle (le coefficient stœchiométrique K) qui n'est pas identifiable. Suivant cette approche, une réaction est considérée comme négligeable (et retirée), si son impact réactionnel est négligeable sur l'ensemble des réactifs qui la composent.

4.3.4 Fonctions unitaires

Les fonctions unitaires sont des fonctions servant à décrire l'impact des variables du modèle sur les cinétiques des réactions. Leur valeur peut varier entre 0

et 1 (d'où le nom fonction unitaire) pour l'ensemble des valeurs possibles de la variable associée. Ainsi, selon cette définition, les fonctions unitaires ne peuvent que réduire le taux de réaction (le multiplier par une valeur plus petite que 1). Les interactions positives (ex. : accélération d'une réaction grâce à une enzyme) sont représentées par une absence d'interaction négative (diminution de la vitesse de réaction si l'enzyme n'est pas présente). Il existe deux types de fonctions unitaires : les fonctions de limitation et les fonctions d'inhibition. Les fonctions de limitation sont des fonctions qui ont pour effet de diminuer la vitesse de réaction lorsque la valeur de la variable associée diminue. Parmi les fonctions unitaires de limitation, on retrouve :

Monod (1949)	$u = u_{max} * \frac{[X]}{K + [X]}$	
Tessier (1936)	$u = u_{max} * \left(1 - \exp\left(\frac{-[X]}{k}\right) \right)$	Éq. 4.6
Moser (1958)	$u = u_{max} * \frac{[X]^n}{K + [X]^n}$	

Les fonctions unitaires d'inhibition sont des fonctions qui ont pour effet de diminuer la vitesse de réaction lorsque la valeur de la variable associée augmente. Par exemple, la fonction $u = u_{max} * \frac{K_1}{K_1 + [X]^{K_2}}$ est une fonction unitaire d'inhibition lorsque les valeurs des paramètres sont positives.

4.3.5 Zone d'impact d'une fonction unitaire

La zone d'impact d'une fonction unitaire représente la plage de concentration dans laquelle la fonction unitaire est active c'est-à-dire dans laquelle sa valeur varie. La zone d'impact est définie comme l'intervalle de la variable associée où la fonction unitaire a une valeur comprise entre 0.05 et 0.95; ceci est considéré pour les deux types de fonctions unitaires. Ainsi pour une fonction unitaire quelconque (F), on obtient les limites suivantes.

Limitation

$$\lim_{X \rightarrow 0} F \rightarrow 0$$

$$\lim_{X \rightarrow \infty} F \rightarrow 1$$

Inhibition

$$\lim_{X \rightarrow 0} F \rightarrow 1$$

$$\lim_{X \rightarrow \infty} F \rightarrow 0$$

Éq. 4.7

La figure 4.4 trace un exemple de fonction unitaire de limitation et d'inhibition. Dans ces exemples, la zone d'impact de la fonction de limitation est de 0 à 35 et la zone d'impact de la fonction d'inhibition est de 45 à 90.

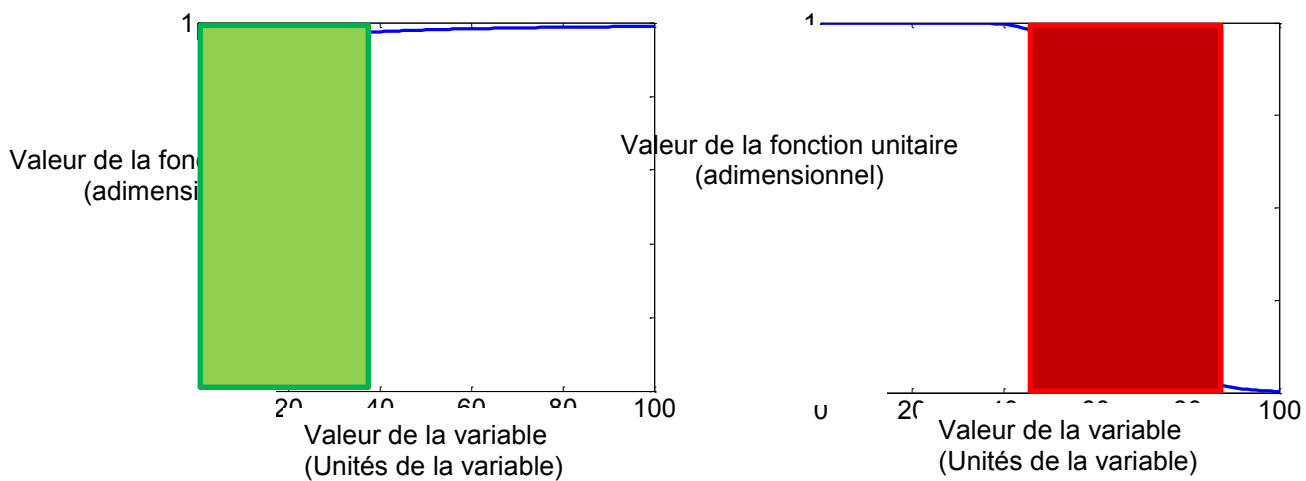


Figure 4.4 : Exemples de la zone d'impact de fonctions unitaires (limitation et inhibition)

Il est intéressant de comparer la zone d'impact d'une fonction unitaire avec la plage de données expérimentales et/ou la plage de valeurs simulées. La figure 4.5 présente la comparaison entre la plage de données et la zone d'impact de fonctions unitaires. Cette comparaison permet de voir pour quelles données la fonction unitaire est active. La plage de données est représentée par un rectangle bleu, la zone d'impact de la fonction unitaire de limitation est représentée par un rectangle vert et la zone d'impact de la fonction d'inhibition est un rectangle rouge. On constate que dans ce cas, les fonctions unitaires sont actives dans les conditions étudiées.

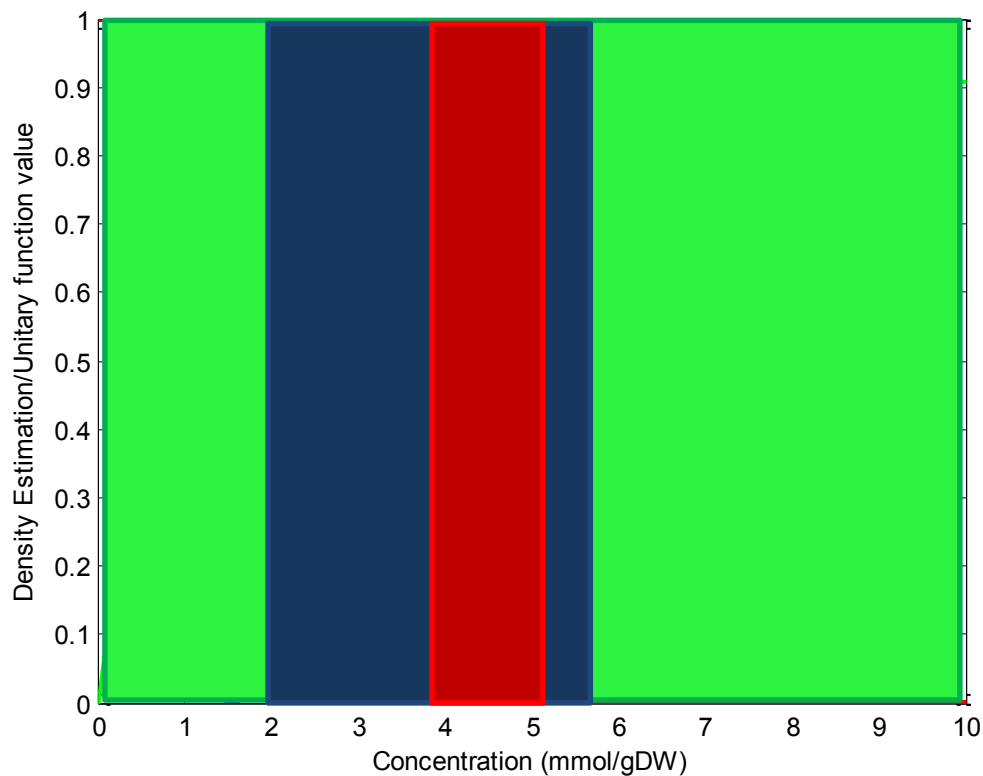


Figure 4.5 : Comparaison entre la plage de données d'une variable et la zone d'impact de fonctions unitaires

La figure 4.6 présente un scénario dans lequel les fonctions unitaires ne sont pas actives, car les zones d'impacts (rouge et vert) des fonctions unitaires ne couvrent pas la plage de données considérée (bleu). Dans ce cas, la valeur estimée des paramètres n'est pas fiable et le phénomène représenté par la fonction unitaire n'est pas observé dans les conditions expérimentales étudiées.

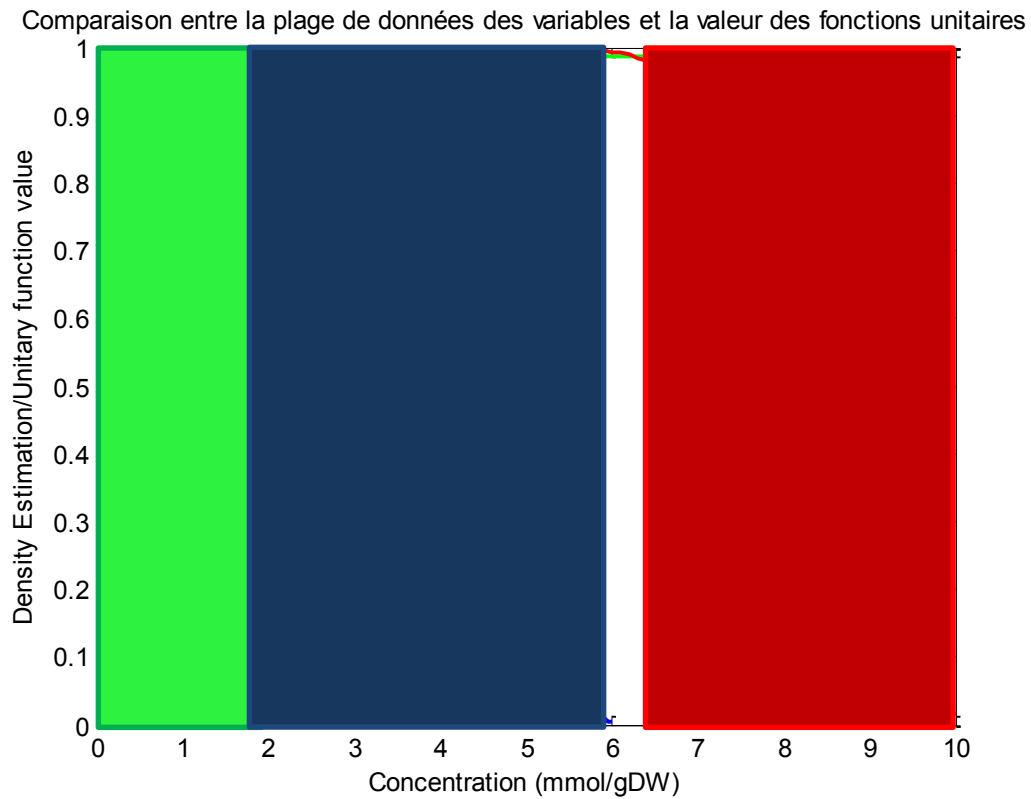


Figure 4.6 : Exemple de plage de données ayant un recoupement faible avec la plage de données.

En résumé, les différents éléments des modèles mathématiques sont utilisés afin de comprendre comment les paramètres impactent les prédictions du modèle et comment il est possible de les retirer sans affecter les prédictions du modèle.

Ce chapitre présente les algorithmes développés afin de réaliser la simplification des modèles mathématiques. Ces algorithmes ont pour objectif de diminuer le nombre de paramètres et/ou de variables d'état du modèle sans affecter sa capacité à prédire l'évolution des variables d'état du système. Pour ce faire, les effets des différents éléments des modèles mathématiques (fonctions unitaires et impacts réactionnels) sont étudiés en relation avec les plages de données dans lesquelles sont appliqués les modèles. Selon cette approche, les paramètres associés à des fonctions unitaires où la zone d'impact de la fonction unitaire et la plage des données modélisées se superposent sont les phénomènes où la cinétique du modèle devrait être plus précise puisque dans ces zones, les variations des conditions du système ont un impact important sur les prédictions du modèle. Une élimination des éléments moins significatifs du modèle permet de concentrer l'effort de modélisation sur ces éléments du modèle.

Trois algorithmes ont été développés. Le premier est basé sur l'application de critères aux structures du modèle. Ces critères ont pour objectif de simplifier les éléments du modèle représentant des phénomènes qui n'ont pas été observés dans la plage de données modélisées ou représentant des réactions ayant peu d'impact sur les prédictions du modèle tout en s'assurant que cette simplification n'affecte pas significativement l'erreur d'estimation. Le second algorithme utilise le critère d'information d'Akaike pour simplifier le modèle. Les deux algorithmes sont appliqués sur les fonctions unitaires et les impacts réactionnels. Le troisième algorithme sépare les variables d'état du modèle en deux groupes : les variables d'intérêt et les variables non essentielles. Les variables d'intérêt sont celles dont on souhaite connaître l'évolution. Les variables non essentielles sont les autres variables. L'algorithme évalue l'impact des variables non essentielles sur la prédiction des variables d'intérêt et les retire si cet impact est faible.

Lorsque des données provenant de plusieurs expériences sont disponibles, les algorithmes de simplification sont appliqués indépendamment sur chacune des expériences. Ce n'est que lorsque l'ensemble des expériences suggère de retirer un élément du modèle que celui-ci est retiré. Dès que les données d'une expérience suggèrent que l'impact de la réaction ou de la fonction unitaire est significatif, la réaction ou la fonction unitaire est conservée.

5.1 Simplification des paramètres par application de critères sur les structures du modèle

Dans le cadre de ce projet, cinq critères servant à simplifier les modèles sont développés. Les critères sont appliqués sur les données modélisées et non sur les données expérimentales. Il est posé comme hypothèse que si une simplification n'affecte pas les données modélisées, elle n'affectera pas la qualité du modèle. En effet, le modèle est caractérisé par l'erreur d'estimation (équation 5.1) qui est calculée à partir de la différence entre les données expérimentales et les données modélisées.

$$RSS(\beta) = \sum_{i=1}^m [X_{exp,i} - X_{modèle,i}(X_{ini}, \beta, u, y)]^2 \quad \text{Éq. 5.1}$$

Si une simplification au modèle a un impact négligeable sur les données modélisées ($X_{modèle}$), cette simplification aura un impact négligeable sur l'erreur d'estimation du modèle (RSS).

5.1.1 Critère A : Fonction unitaire ayant une valeur près de 1

Le critère A vise à éliminer les fonctions unitaires dont la zone d'impact n'est pas observée dans les données. Dans le cas où une fonction unitaire n'est pas active (zone d'impact disjointe dans la plage de données ou de valeurs modélisées) l'algorithme de réduction de l'erreur d'estimation du modèle, qui sert à évaluer la

valeur des paramètres, modifie les valeurs des paramètres associés à cette fonction unitaire jusqu'à ce que la fonction unitaire n'ait plus d'impact sur les prédictions du modèle. La zone d'impact de la fonction unitaire devrait ainsi être déplacée (par modification de la valeur des paramètres) jusqu'à ce que la modification de l'erreur du modèle soit sous le seuil défini comme significatif dans l'algorithme d'estimation des paramètres du modèle. La figure 5.1 présente la zone d'impact d'une fonction unitaire théorique avant (a) et après optimisation (b) de la valeur des paramètres par Levenberg-Marquardt. Dans cet exemple, les valeurs initiales des paramètres font en sorte que la fonction unitaire est active; la zone d'impact de la fonction unitaire (rouge) se superpose à la courbe bleue de la distribution des données. La figure 5.1 b montre qu'après optimisation, la fonction unitaire n'est pas active, car il n'y a plus de chevauchement entre la zone d'impact de la fonction unitaire et la distribution des données.

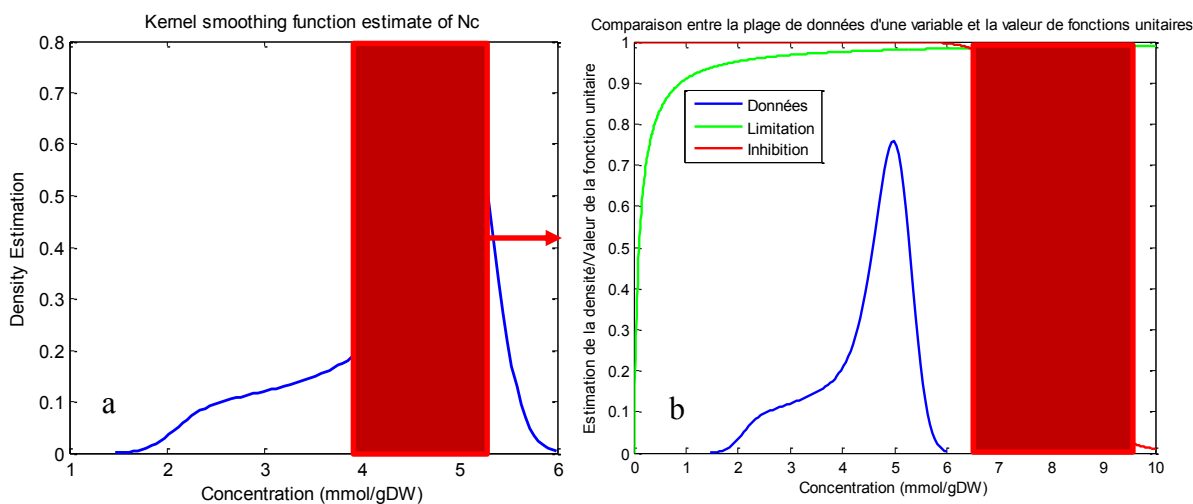


Figure 5.1 : Déplacement de la zone d'impact de fonctions unitaires aux limites de la plage de données modélisées lorsque la fonction unitaire a une valeur près de 1.

Ce comportement fait en sorte que lorsqu'on calcule la valeur de la fonction unitaire pour les données, celle-ci a une valeur près de 1. Puisque la fonction unitaire est multipliée au taux de réaction, cela fait en sorte que la fonction unitaire n'affecte pas les prédictions du modèle. Le critère est défini comme suit :

Si pour l'ensemble des expériences, la moyenne de la valeur d'une fonction unitaire calculée sur les données simulées est égale ou supérieure à 0.95, la fonction unitaire est éliminée.

Les paramètres associés aux fonctions unitaires identifiées par ce critère ne sont pas pratiquement identifiables, car une variation de leur valeur entraîne une modification très faible de l'erreur d'estimation du modèle.

5.1.2 Critère B : Fonction unitaire ayant une valeur près de 0

Il peut aussi arriver que la zone d'impact de la fonction unitaire soit déplacée hors de la plage de données modélisées par modification de la valeur des paramètres jusqu'à ce que la fonction unitaire ait une valeur près de 0 sur l'ensemble de la plage expérimentale tel que montré à la figure 5.2. Cela correspond à un déplacement de la zone d'impact de la fonction unitaire par modification des valeurs des paramètres dans la direction opposée au cas précédent. Le critère B vise à retirer les fonctions ayant une valeur près de 0 sur l'ensemble de la plage expérimentale.

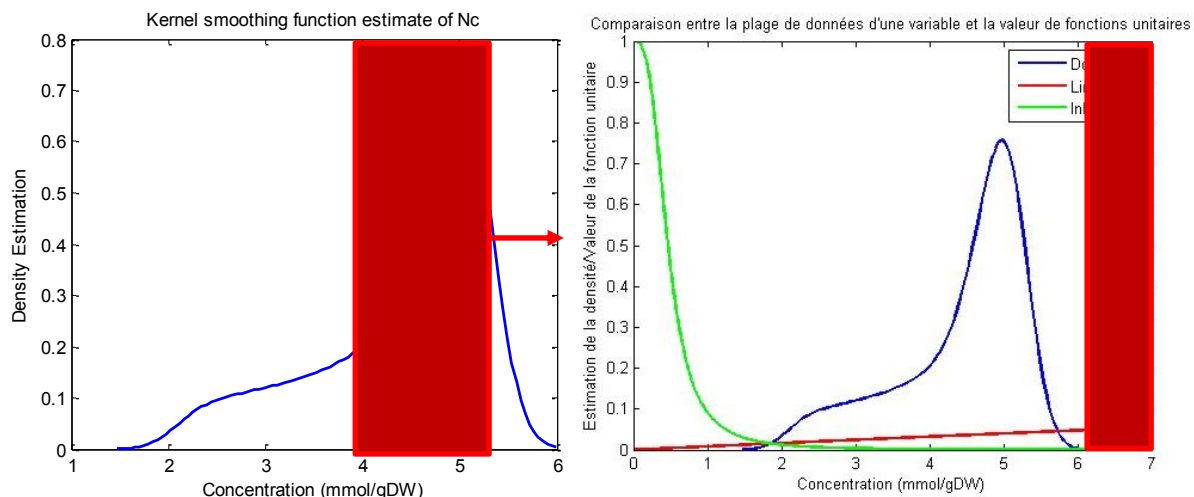


Figure 5.2 : Déplacement de la zone d'impact de fonctions unitaires aux limites de la plage de données modélisées lorsque la fonction unitaire a une valeur près de 0.

Le critère B est défini comme suit :

Si pour l'ensemble des expériences, la moyenne de la valeur d'une fonction unitaire est inférieure ou égale à 0.05, l'impact réactionnel associé à la fonction unitaire est éliminé.

Puisque les fonctions unitaires sont multipliées au taux de réaction, si la valeur d'une fonction unitaire est près de 0 pour l'ensemble des données expérimentales, la réaction qui y est associée devrait avoir une valeur près de 0. Si c'est le cas, cette réaction ne devrait pas affecter les variables d'état. Il serait donc possible de retirer l'ensemble des impacts réactionnels associés à la réaction.

5.1.3 Critère C : Impact réactionnel ayant une valeur faible par rapport aux autres

Le critère C vise à retirer les impacts réactionnels ayant une valeur faible. Tel que défini dans le chapitre précédent, l'impact réactionnel est un segment d'équation représentant l'influence d'une réaction sur une variable. Les impacts réactionnels ayant une valeur faible peuvent être dus à la situation décrite précédemment (réaction multipliée par une fonction unitaire ayant une valeur près de 0), mais n'y sont pas limités (ex. : réaction ayant un taux de réaction maximal $[\mu_{\max}]$ avec une valeur faible). Afin de définir si l'effet de l'impact réactionnel sur l'estimation de la variable est significatif, celui-ci est comparé à l'ensemble des impacts réactionnels affectant la variable. Le critère est défini comme suit :

Si la valeur moyenne d'un impact réactionnel sur l'ensemble des points pour chaque expérience a une valeur égale ou inférieure à 5% de la somme des valeurs absolues de l'ensemble des impacts réactionnels affectant une variable, cet impact est éliminé.

5.1.4 Critère D : Impact réactionnel ayant un impact faible sur l'estimation des variables

Le critère D vise à retirer les impacts réactionnels ayant peu d'influence sur l'estimation des variables. Le critère est défini comme suit :

Si le retrait d'un impact réactionnel entraîne une erreur relative à la valeur maximale d'une variable égale ou inférieure à 5% pour chaque expérience, l'impact réactionnel est retiré.

Ce critère consiste à tester le modèle avec et sans chacun des impacts réactionnels et à observer si ce retrait affecte les prédictions du modèle. Les impacts réactionnels qui n'affectent pas les prédictions du modèle sont retirés. Cela peut être dû à un taux de réaction trop faible ou à une fonction unitaire multipliant l'impact réactionnel par une valeur très faible.

5.1.5 Critère E : Fonctions unitaires ayant une valeur constante

Le critère E vise à simplifier les fonctions unitaires ayant une valeur constante pour l'ensemble des données simulées. Les fonctions unitaires jouent alors le rôle de constante multipliant l'impact réactionnel. Dans cette situation, plusieurs constantes ont le même effet sur l'impact réactionnel. La constante réelle affectant la cinétique de la variable est la combinaison de ces constantes. L'objectif de ce critère est donc de rassembler ces constantes en une seule. Pour ce faire, ce critère est défini comme suit :

Si la différence entre la valeur maximale et la valeur minimale observée pour la fonction unitaire est plus petite que 5% de la valeur moyenne de la fonction unitaire, elle est retirée et le taux de réaction maximal est multiplié par la moyenne de la fonction unitaire.

5.2 Simplification des paramètres du modèle par application du critère d'information d'Akaike

Un second algorithme de simplification a été développé basé sur le critère d'information d'Akaike. Le critère d'information d'Akaike n'est pas un algorithme de simplification en lui-même, mais un outil servant à comparer différents modèles. Le critère d'information d'Akaike est composé de deux parties. La première est liée à l'erreur d'estimation du modèle et est appelée « poids lié à la précision du modèle ». La seconde partie du critère est liée au nombre de paramètres du modèle et est appelée « poids lié à la simplicité du modèle ». Les modèles sont comparés entre eux selon la valeur de ce critère. Plus la valeur du critère associé au modèle est petite, meilleur est le modèle.

$$AIC = \underbrace{m * \ln\left(\frac{RSS}{m}\right)}_{\text{Poids lié à la précision du modèle}} + \underbrace{2 * n}_{\text{Poids lié à la simplicité du modèle}}$$

Éq. 5.2

L'équation 5.2 présente le critère d'information d'Akaike où m est le nombre d'échantillons, RSS est la somme des carrés résiduels et n est le nombre de paramètres.

La méthode de simplification développée utilise le critère d'information d'Akaike afin d'évaluer l'impact du retrait d'un ou plusieurs paramètres du modèle. Ce retrait est effectué de la façon suivante : si un paramètre appartient à une fonction unitaire, cette fonction est retirée de l'ensemble des cinétiques du modèle. Si le paramètre n'appartient pas à une fonction unitaire, l'impact réactionnel auquel il appartient est retiré de l'ensemble des cinétiques du modèle. Le modèle simplifié peut être comparé au modèle original à l'aide de l'AIC. La première étape de cet algorithme consiste à trier les paramètres en ordre croissant d'AIC du modèle lorsque ce paramètre est

retiré du modèle. Une fois les paramètres triés en ordre croissant d'AIC, les paramètres sont retirés cumulativement selon l'ordre défini ci-dessus et un nouvel AIC est calculé. Ainsi, l'algorithme calcule l'AIC pour le modèle original sans le premier paramètre. Ensuite, l'AIC est calculé pour le modèle original sans le premier et le deuxième paramètre, puis sans le premier, le deuxième et le troisième paramètre et ainsi de suite. Le modèle ayant l'AIC le plus petit est sélectionné comme modèle simplifié optimal.

5.2.1 Utilisation d'un modèle bruité

Comme les méthodes précédentes (critères A-E, AIC) ont été appliquées sur des données idéales, il est nécessaire de mesurer l'importance des erreurs expérimentales sur l'efficacité de ces méthodes. Pour cela, on génère des « données expérimentales » en ajoutant un bruit aux données idéales du modèle original. L'utilisation de données bruitées est plus représentative de la réalité puisque les données mesurées sur les procédés biotechnologiques comportent toujours une certaine erreur. Le bruit ajouté a une distribution normale standard (moyenne de 0 et écart-type de 1). Ce bruit est multiplié par une fraction de la valeur maximale de la variable sur l'ensemble des scénarios étudiés. Les valeurs bruitées sont décrites par l'équation 5.3.

$$\begin{aligned} \text{Valeur bruitée} &= \text{Valeur originale} + \\ &(\text{Erreur normale} * \text{Valeur maximale} * \text{fraction}) \end{aligned} \quad \text{Éq. 5.3}$$

La fraction de l'erreur devrait être choisie de façon à représenter l'erreur expérimentale associée à la mesure de chacune des variables. Dans le cadre de ce projet, la fraction choisie est généralement 5% de la valeur maximale de la variable observée. Cette fraction est élevée en ce qui concerne les variables extracellulaires (Concentration de nutriments, pH, etc.), mais peut être considérée comme faible pour les variables intracellulaires dû au nombre d'expériences requises pour mesurer ces variables (extraction, purification, quantification). Il peut être difficile d'estimer cette

erreur, puisque pour les variables intracellulaires, le nombre de répliquas est généralement faible. Si ces données sont disponibles, l'utilisation de la valeur réelle est recommandée, sinon, la valeur de 5% est utilisée.

De plus, lorsque la valeur modifiée d'une variable est négative alors que par définition cette variable ne peut avoir de valeurs négatives (ex. : concentration), la valeur modifiée de la variable est posée à 0. Ce sont les valeurs bruitées qui sont utilisées comme référence pour le calcul du RSS dans le critère d'information d'Akaike.

5.2.2 Utilisation d'un RSS relatif

Afin d'équilibrer l'impact de chaque variable dans le calcul du critère d'information, il est possible d'utiliser un RSS relatif. De cette façon, chaque variable possède un poids équivalent. La façon choisie pour relativiser le RSS est de le diviser par la valeur maximale de la variable observée telle que présentée par l'équation 5.4. Une alternative à ce choix est de diviser le RSS par la valeur présente de la variable, mais cela fait en sorte d'augmenter les poids des valeurs faibles de la variable par rapport aux valeurs élevées. Le RSS utilisé pour calculer l'AIC est donc :

$$RSS_{Rel2max} = \sum \left(\frac{\text{Données} - \text{Valeurs prédites}}{\max(\text{Données})} \right)^2 \quad \text{Éq. 5.4}$$

Et l'équation pour calculer l'AIC est :

$$AIC_{Rel} = m \ln \left(\frac{RSS_{Rel2max}}{m} \right) + 2n \quad \text{Éq. 5.5}$$

5.2.3 Utilisation d'un critère d'information d'Akaike modifié

Le critère d'Akaike permet de mettre en relation la précision d'un modèle (déterminée par le RSS) et la simplicité d'un modèle (déterminée par le nombre de paramètres). Le poids relatif de ces deux caractéristiques tel que défini par le critère d'Akaike original peut être modifié de façon à accorder plus d'importance à la précision ou à la simplicité du modèle. Ainsi, il est possible d'ajuster la forme ou la valeur des paramètres du critère pour favoriser un modèle plus simple ou un modèle plus précis. Parmi les alternatives étudiées, notons le critère d'information Bayésien (Schwarz, 1978) (équation 5.6), la modification du critère par le retrait du logarithme naturel (équation 5.7), ainsi que l'augmentation de la valeur de la constante associée au nombre de paramètres (remplacer $2*n$ par $3*n$, $3.5*n$, ...).

$$BIC = m * \ln\left(\frac{RSS}{m}\right) + n * \ln(m) \quad \text{Éq. 5.6}$$

$$AIC_m = m * RSS + 2n \quad \text{Éq. 5.7}$$

5.2.4 Algorithme de simplification basé sur le critère d'information d'Akaike

Différentes formes du critère d'information d'Akaike ont été envisagées pour le classement des paramètres et la détermination du modèle simplifié optimal. Toutefois, l'algorithme développé pour obtenir le modèle simplifié est similaire pour les différentes formes du critère d'Akaike utilisées.

L'algorithme est bâti de la façon suivante :

1. Pour chacun des paramètres, si le paramètre appartient à une fonction unitaire, la fonction unitaire est retirée; sinon, l'impact réactionnel auquel appartient le

paramètre est retiré pour l'ensemble du modèle. L'AIC du modèle ainsi simplifié est calculé.

2. Les paramètres sont mis en ordre croissant selon leur AIC.
3. Les paramètres sont retirés successivement et l'AIC du modèle ainsi simplifié est calculé.
4. Le modèle ayant l'AIC le plus petit est le modèle simplifié optimal.

5.3 Réduction d'ordre du modèle par retrait de variables non essentielles

À cette étape, les différentes variables du système sont divisées entre les variables d'intérêt et les variables non essentielles. L'impact du retrait des variables non essentielles est alors évalué. Pour ce faire, un algorithme a été développé afin d'évaluer l'impact de poser les cinétiques des variables non essentielles à 0 sur les prédictions de modèles. C'est l'équivalent de transformer ces variables en constantes.

5.3.1 Ordre de simplification des variables

La première étape de la simplification des variables est de déterminer dans quel ordre les variables seront retirées du modèle. Afin de déterminer l'ordre de retrait des variables non essentielles, le retrait de chaque variable non essentielle est évalué en fonction de son impact sur l'estimation des variables d'intérêt. Pour déterminer l'ordre de retrait des variables, la cinétique de chacune des variables non essentielle est posée à 0 individuellement. L'impact de cette modification sur l'estimation des variables d'intérêt est mesuré à l'aide du RSS calculé uniquement sur les variables d'intérêt. Différentes formules de RSS ont été évaluées :

1. le RSS absolu

$$RSS = \sum (Valeurs bruitées - Valeurs prédites)^2$$

Éq. 5.8

2. le RSS relatif

$$RSS_{Rel} = \sum \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\text{Valeurs bruitées}} \right)^2$$

3. le RSS relatif à la valeur maximale de la variable.

$$RSS_{Rel2max} = \sum \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\max(\text{Valeurs bruitées})} \right)^2$$

Les variables non essentielles sont ensuite placées en ordre croissant selon la valeur du RSS choisi.

5.3.2 Limite de simplification

Une fois les variables non essentielles placées en ordre croissant de RSS, celles-ci sont retirées cumulativement, jusqu'à ce que le modèle ne soit plus assez précis. Pour définir le seuil de précision, trois options ont été envisagées :

1. La variable sera enlevée si l'erreur d'estimation relative au maximum de la variable du modèle par rapport à l'ensemble des données des variables d'intérêt ne dépasse pas 5%.

$$Err_{rel2data} = \frac{1}{vi} \sum_{i=1}^{vi} \frac{1}{m} \sum_{j=1}^m \frac{|\text{Données}_{i,j} - \text{Modèle simplifié}_{i,j}|}{\max(\text{Données}_i)} * 100\% \quad \text{Éq. 5.9}$$

2. La variable sera enlevée si le maximum parmi les variables d'intérêt de l'erreur d'estimation relative au maximum de la variable du modèle simplifié par rapport au modèle original ne dépasse pas 5%,

$$Err_{rel2model}$$

Éq. 5.10

$$= \max \left(\frac{1}{m} \sum_{j=1}^m \frac{|Modèle\ original_{i,j} - Modèle\ simplifié_{i,j}|}{\max(Modèle\ original_i)} * 100\% \right)$$

3. La variable sera enlevée si la valeur du critère d'information modifié est plus petite que la valeur du critère d'information modifié précédente. Pour ce faire, un critère d'information de type d'Akaike appliqué à des variables est utilisé où n , le nombre de paramètres, est remplacé par v , le nombre de variables.

Ainsi, trois méthodes de simplification ont été présentées. Chacune de ces méthodes comporte plusieurs options. Dans la méthode de simplification par critères, cinq critères sont envisagés. La méthode de simplification par AIC envisage l'utilisation de données modélisées bruitées ainsi que la modification de la formulation du critère d'Akaike. La réduction d'ordre du modèle propose trois méthodes de classification des variables, de même que trois limites de simplification possibles. Chacune de ces options sera étudiée dans le chapitre suivant.

CHAPITRE 6 ANALYSE DES MÉTHODES DE SIMPLIFICATION

Chacune des méthodes de simplification présentées dans le chapitre précédent a été analysée afin de vérifier qu'elles répondent aux objectifs de simplification suivants:

- La méthode doit permettre de simplifier le modèle par réduction du nombre de paramètres affectant le modèle ou par réduction du nombre de variables d'état du modèle.
- L'application de la méthode de simplification ne doit pas modifier significativement l'erreur d'estimation du modèle. Puisque les modifications aux modèles influencent les valeurs modélisées des variables d'état (et non les données expérimentales associées aux variables d'état), cet objectif est défini comme suit : l'application de la méthode de simplification ne doit pas modifier significativement les valeurs modélisées des variables d'état conservées dans le modèle.

Ainsi, les méthodes de simplification rejetées le sont parce qu'elles ne répondent pas aux objectifs de simplification (elles affectent les valeurs des données modélisées ou elles ne permettent pas de réaliser la simplification des modèles). Afin de représenter l'impact des méthodes de simplification, le modèle de Cloutier (2007) est utilisé. Les équations formant ce modèle peuvent être retrouvées à l'annexe A.1. Deux scénarios sont considérés se différenciant par la concentration initiale en biomasse (0.17 gDW/L dans le scénario 1 et 0.5 gDW/L dans le scénario 2).

6.1 Simplification des paramètres par application de critère sur les structures du modèle

Chacun des critères proposés a été évalué individuellement afin de vérifier s'ils répondaient aux objectifs de simplifications.

6.1.1 Critère A : Fonction unitaire ayant une valeur près de 1

Si pour l'ensemble des expériences, la moyenne de la valeur d'une fonction unitaire calculée sur les données simulées est égale ou supérieure à 0.95, la fonction unitaire est éliminée.

Observons l'impact de ce critère sur les fonctions unitaires liées à l'équation de croissance cellulaire du modèle de Cloutier (2007) :

$$\frac{dX}{dt} = u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] * [1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X \quad \text{Éq. 6.1}$$

La valeur moyenne des fonctions unitaires de même que la décision de les conserver ou non selon le critère A sont présentées dans le tableau 6.1.

Fonction unitaire	Valeur moyenne		Conservée ?
	Exp 1	Exp 2	
$(1 - \exp(-K_P * (P_C - P_{C,min})))$	0.9869	0.9066	Oui
$(1 - \exp(-K_S * (S_C - S_{C,min})))$	0.987	0.9869	Non
$(1 - \exp(-K_N * (N_C - N_{C,min})))$	0.6486	0.5541	Oui
$K_{pp} / (K_{pp} + P_i^{inhi,P})$	0.9941	0.9978	Non

Tableau 6.1 : Valeur moyenne des fonctions unitaires liées à la croissance du modèle de Cloutier (2007) ainsi que leur évaluation selon le critère A

On constate que le critère A suggère de retirer les fonctions unitaires liées au sucre (S_C) et au phosphate extracellulaire (P_i). De plus, le tableau 6.1 indique que pour la fonction unitaire liée au phosphate intracellulaire (P_C), la valeur moyenne est supérieure à 0.95 dans le premier scénario, mais inférieure à 0.95 dans le second scénario. Dans ce cas, la fonction unitaire est conservée dans le modèle. La figure 6.1 présente les valeurs dans le temps de ces quatre fonctions unitaires. Cela permet de constater la valeur quasi constante des fonctions unitaires dont le retrait est suggéré.

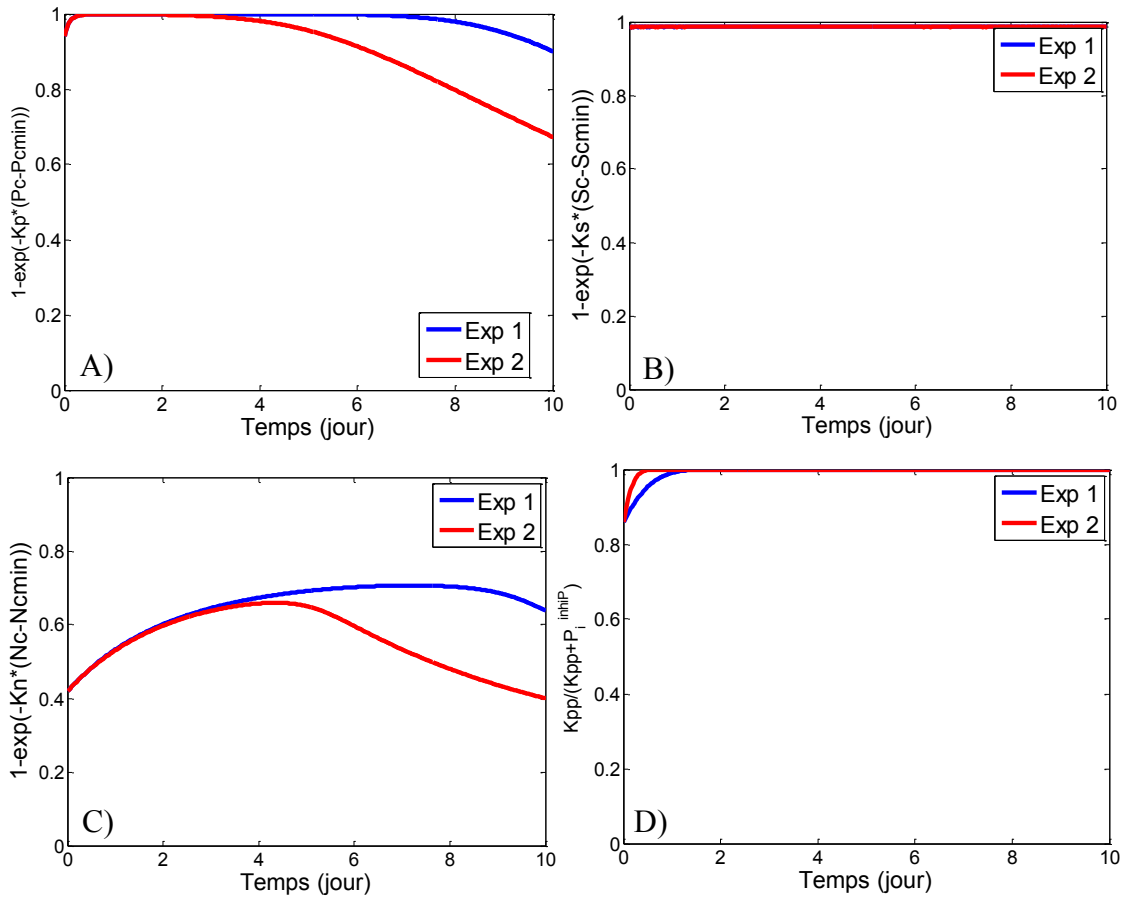


Figure 6.1 : Valeurs dans le temps des fonctions unitaires liées à la croissance cellulaire du modèle de Cloutier (2007) A) limitation associée au phosphate intracellulaire (conservée), B) limitation associée au glucose intracellulaire (retirée), C) limitation associée à l'azote intracellulaire (conservée), D) inhibition associée au phosphate extracellulaire (retirée).

Ce critère a permis d'obtenir des modèles simplifiés sans affecter significativement l'erreur d'estimation du modèle. Ce critère ayant atteint les objectifs de la simplification, il est conservé.

6.1.2 Critère B : Fonction unitaire ayant une valeur près de 0

Le critère B est défini comme suit :

Si pour l'ensemble des expériences, la moyenne de la valeur d'une fonction unitaire est inférieure ou égale à 0.05, l'impact réactionnel associé à la fonction unitaire est éliminé.

L'objectif de ce critère est de retirer un impact réactionnel si celui-ci est affecté par une fonction unitaire où le phénomène limitant ou inhibant la réaction est dans une zone très restrictive. Cependant, lors de tests effectués sur des modèles, il a été constaté que parfois, une valeur très faible des fonctions unitaires est compensée par une valeur élevée d'un autre terme (ex. : le taux de réaction maximal (μ_{max})). Le retrait de ces impacts réactionnels avait un impact significatif sur les prédictions du modèle. Par exemple, observons la deuxième fonction de limitation associée au NH_4 ($\frac{NH_4}{K_{2,NH_4} + NH_4}$) dans la cinétique de l'ammonium du modèle de Cloutier (2007) représentée par l'équation 6.2 :

$$\begin{aligned} \frac{dNH_4}{dt} = & -V_{max,1,NH_4} \frac{NH_4}{K_{1,NH_4} + NH_4} \frac{X}{V} - V_{max,2,NH_4} \frac{NH_4}{K_{2,NH_4} + NH_4} \frac{X}{V} \\ & + \frac{NH_4}{V} DWr * u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] \\ & * [1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X \end{aligned} \quad \text{Éq. 6.2}$$

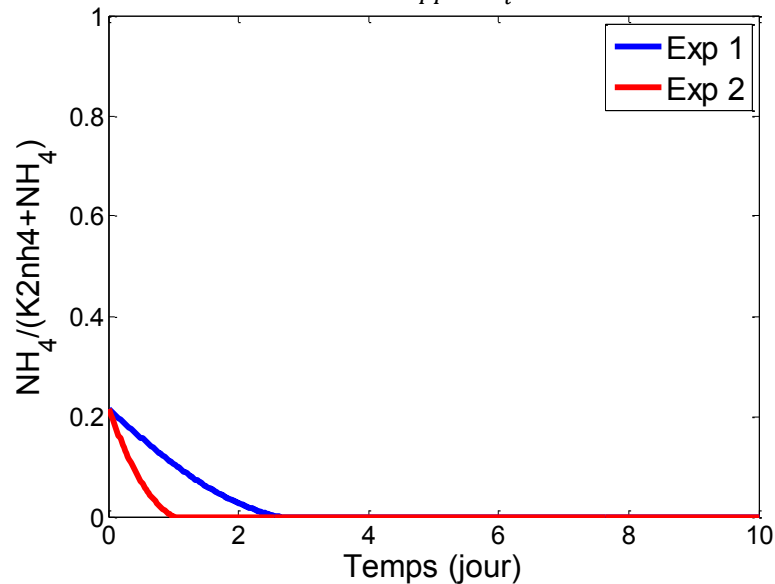


Figure 6.2 : Valeurs dans le temps de la fonction unitaire ($NH_4/(K_{2,NH_4} + NH_4)$) du modèle de Cloutier (2007)

La valeur de cette fonction dans le temps est présentée à la figure 6.2. Le tableau 6.2 montre que la valeur moyenne de la fonction unitaire est inférieure à 0,05. Selon le critère B, cette fonction unitaire ainsi que l'impact réactionnel associé devraient être retirés du modèle.

Fonction unitaire	Valeur moyenne		Conservée ?
	Exp 1	Exp 2	
$*NH_4/(K_2nh_4+NH_4)$	0.0158	0.0231	Non

Tableau 6.2 : Valeur moyenne de la fonction unitaire ($NH_4/(K_{2,NH_4}+NH_4)$) du modèle de Cloutier (2007) ainsi que son évaluation selon le critère B

La figure 6.3 permet de constater que l'application du critère B affecte significativement l'estimation du NH_4 associé au modèle de Cloutier (2007), car les courbes de NH_4 intracellulaire sont très différentes pour les deux expériences. Ainsi, ce critère ne répond pas aux objectifs fixés puisqu'il a un impact significatif sur les données modélisées et donc sur l'erreur d'estimation du modèle. Il a donc été rejeté, car cette cinétique doit être conservée.

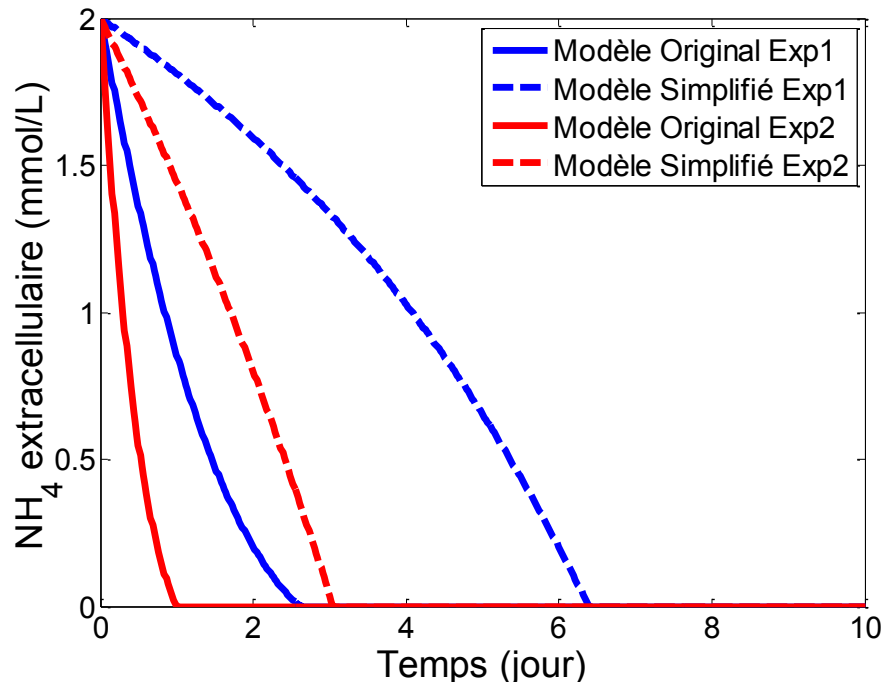


Figure 6.3 : Estimation du NH_4 selon le modèle de Cloutier (2007) avec et sans l'impact réactionnel ($-V_{max,2,NH_4}*(NH_4/(K_{2,NH_4}+NH_4))*(X/V)$) pour les deux expériences testées

Il est toutefois important de noter que le phénomène représenté par la combinaison de l'impact réactionnel avec la fonction unitaire ayant une valeur près de 0 n'est peut-être pas bien représenté par le modèle.

Théoriquement, si la cinétique et les paramètres estimés sont exacts, il serait possible d'augmenter significativement la vitesse de réaction en opérant le procédé dans des conditions différentes (en augmentant la valeur de la variable associée jusqu'à ce que celle-ci soit dans la zone d'impact de la fonction unitaire). L'alternative est que la combinaison entre l'impact réactionnel et la fonction unitaire représente mal le phénomène en jeu. Dans ce cas, les paramètres ne sont pas identifiables (au moins pratiquement). Il est possible de faire varier conjointement la valeur des paramètres sans faire varier l'erreur d'estimation du modèle. De nouvelles données expérimentales dans la zone d'impact de la fonction unitaire permettraient de mieux estimer des paramètres du modèle et éventuellement les rendre identifiables.

6.1.3 Critère C : Impact réactionnel ayant une valeur faible par rapport aux autres

Le critère C vise à retirer les impacts réactionnels ayant une valeur faible. Cela peut être dû à la situation décrite précédemment, mais n'y est pas limité. Ce critère est défini comme suit :

Si la valeur moyenne d'un impact réactionnel sur l'ensemble des points pour chaque expérience a une valeur égale ou inférieure à 5% de la somme des valeurs absolues de l'ensemble des impacts réactionnels affectant une variable, cet impact est éliminé.

Afin de démontrer l'application de ce critère, la cinétique du phosphate intracellulaire provenant du modèle de Cloutier (2007) est utilisée.

$$\frac{dP_C}{dt} = IR1 + IR2 + IR3 + IR4 + IR5$$

où

$$IR1 = V_{max,1,P} \frac{P_i}{K_{1,P} + P_i}$$

$$IR2 = V_{max,2,P} \frac{P_i}{K_{2,P} + P_i}$$

$$IR3 = \frac{1}{Y_X} * u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] *$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$IR4 = -P_C u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] *$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$IR5 = -m_P$$

Éq. 6.3

La définition de chacun des paramètres est présentée dans l'article de Cloutier (2007). Les valeurs dans le temps de ces cinq impacts réactionnels sont présentées à la figure 6.4 où IR représente les cinq impacts réactionnels de l'équation 6.3.

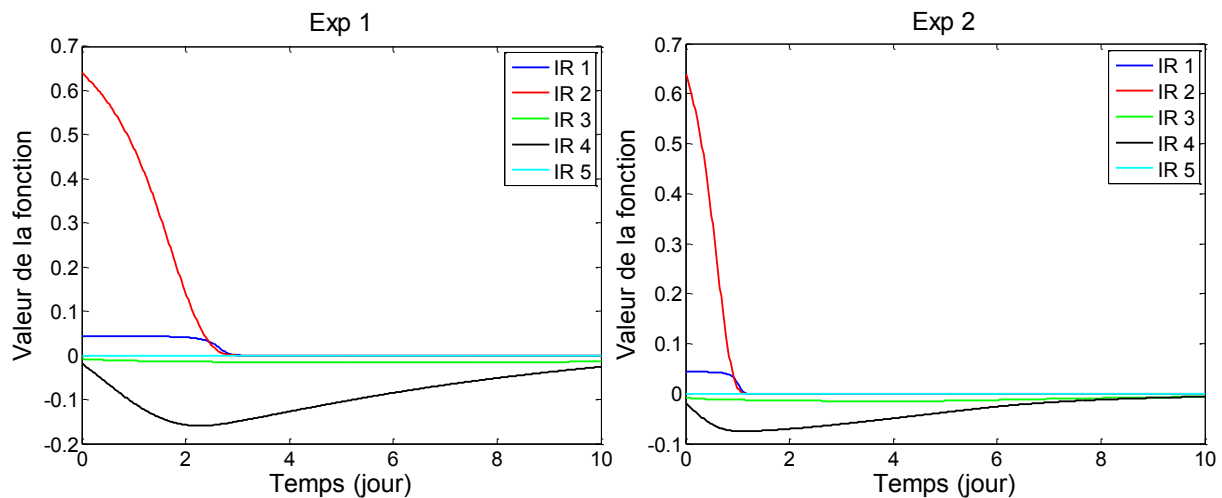


Figure 6.4 : Impacts réactionnels liés à la cinétique du phosphate intracellulaire dans le modèle de Cloutier (2007)

Le tableau 6.3 présente l'analyse et les conclusions selon le critère C. On constate que l'application de ce critère permet la simplification de deux impacts

réactionnels : une réaction d'adsorption du phosphate (IR1) et la réaction de maintenance cellulaire (IR5).

Impact réactionnel	Conserver ?	% moyen de l'impact réactionnel sur la somme des impacts réactionnels	
		Exp1	Exp 2
$V_{max1p} * P_i / (K_{1p} + P_i)$	Non	2.43	1.26
$+V_{max2p} * P_i / (K_{2p} + P_i)$	Oui	15.74	6.77
$-(1/Y_{xp}) * u_{max} * (1 - \exp(-K_p * (P_c - P_{cmin}))) * (1 - \exp(-K_s * (S_c - S_{cmin}))) * (1 - \exp(-K_n * (N_c - N_{cmin}))) * K_{pp} / (K_{pp} + P_i^{inhiP})$	Oui	14.41	28.17
$-P_c * u_{max} * (1 - \exp(-K_p * (P_c - P_{cmin}))) * (1 - \exp(-K_s * (S_c - S_{cmin}))) * (1 - \exp(-K_n * (N_c - N_{cmin}))) * K_{pp} / (K_{pp} + P_i^{inhiP})$	Oui	67.37	63.66
-mp	Non	0.05	0.14

Tableau 6.3 : Analyse de la cinétique du phosphate intracellulaire du modèle de Cloutier (2007) selon le critère C

Ce critère répond aux objectifs fixés, c'est-à-dire simplifier le modèle sans affecter sa qualité. Toutefois, un problème a été constaté. La sévérité du critère varie beaucoup selon le nombre de réactions qui affecte la variable puisque la valeur acceptée dépend de la somme des impacts réactionnels. Ce critère ne permet pas de simplifier l'impact réactionnel d'une variable affectée par une seule réaction, même si cet impact peut être négligeable, comme cela a été constaté pour la variable volume dans le modèle de Cloutier (2007). Ce problème peut être corrigé en fixant une valeur limite considérée comme significative pour un impact réactionnel à partir de la valeur maximale de la variable et de la durée de l'expérience plutôt que de définir cette valeur à partir de la somme des impacts réactionnels. Cette constatation a mené à la définition du critère D. Une fois le critère D mis en place, il a été constaté que tous les impacts réactionnels simplifiés par le critère C l'étaient aussi par le critère D. Le critère C n'a donc pas été conservé puisqu'il était redondant avec le critère D.

6.1.4 Critère D : Impact réactionnel ayant un impact faible sur l'estimation des variables

Le critère D a pour objectif de retirer les impacts réactionnels ayant peu d'influence sur l'estimation des variables. Le critère est défini comme suit :

Si le retrait d'un impact réactionnel entraîne une erreur relative à la valeur maximale d'une variable égale ou inférieure à 5%, l'impact réactionnel est retiré.

Observons l'impact de l'application de ce critère sur la cinétique du nitrate (équation 6.4) dans le modèle de Cloutier (2007). On peut constater que cette cinétique comporte 3 impacts réactionnels.

$$\frac{dNO3}{dt} = IR1 + IR2 + IR3$$

Où

$$IR1 = -V_{max,1,NO3} \frac{NO3}{K_{1,NO3} + NO3} \frac{X}{V}$$

$$IR2 = -V_{max,2,NO3} \frac{NO3}{K_{2,NO3} + NO3} \frac{X}{V}$$

$$IR3 = \frac{NO3}{V} DWr * u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] * [1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X$$

Éq. 6.4

Chacun de ces trois impacts réactionnels a été retiré individuellement du modèle. L'impact de ce retrait a été évalué sur l'estimation de l'ensemble des variables du modèle par le calcul de l'erreur relative au maximum (équation 6.5).

Erreur Relative au maximum

$$= \frac{1}{n} \sum_{i=1}^n \frac{|X_{Modèle\ original,i} - X_{modèle\ simplifié,i}|}{\max(X_{Modèle\ original})} * 100\%$$

Éq. 6.5

Les résultats de cette analyse sont présentés dans le tableau 6.4.

	Erreur relative à la valeur maximale (%)					
	IR 1		IR 2		IR 3	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
X	0.2	0.8	4.1	21.9	0.1	0.3
Pi	0.0	0.0	0.0	0.0	0.0	0.0
Pc	0.1	0.2	1.0	5.8	0.0	0.1
NO3	1.4	1.4	98.6	98.9	0.5	0.5
NH4	0.0	0.0	0.0	0.0	0.0	0.0
Nc	1.8	1.9	28.2	91.2	0.5	0.4
S	0.0	0.3	0.4	5.5	0.0	0.1
Sc	3.2	3.4	3.4	3.4	3.0	3.7
SCc	0.4	0.7	4.0	16.7	0.2	0.3
V	0.0	0.0	0.1	1.0	0.0	0.0

Tableau 6.4 : Impact du retrait de chacun des impacts réactionnels sur l'erreur relative au maximum de chaque variable

Cette analyse permet de constater que seul le retrait du deuxième impact réactionnel (IR 2) cause une variation de l'erreur relative au maximum supérieure à 5% sur certaines variables (X, Pc, NO₃, Nc, S et SCc). Ainsi, selon le critère D, les impacts réactionnels IR1 et IR3 peuvent être enlevés et seul l'impact réactionnel IR2 devrait être conservé dans le modèle. Le tableau 6.5 résume les décisions prises relativement aux trois impacts réactionnels.

Impact réactionnel	Conserver ?	Variable ayant une variation > 5%
$-V_{max1no3} * X / V * NO3 / (K1no3 + NO3)$	Non	Aucune
$-V_{max2no3} * X / V * NO3 / (K2no3 + NO3)$	Oui	X, Pc, NO ₃ , Nc, S, SCc
$+NO3 * X / V * DW_r * u_{max} * (1 - \exp(-K_p * (P_c - P_{cmin}))) * (1 - \exp(-K_s * (S_c - S_{cmin}))) * (1 - \exp(-K_n * (N_c - N_{cmin}))) * K_{pp} / (K_{pp} + P_i^{inhIP})$	Non	Aucune

Tableau 6.5 : Analyse de la cinétique du nitrate du modèle de Cloutier (2007) selon le critère D

L'impact du retrait de chacun des impacts réactionnels est présenté à la figure 6.5.

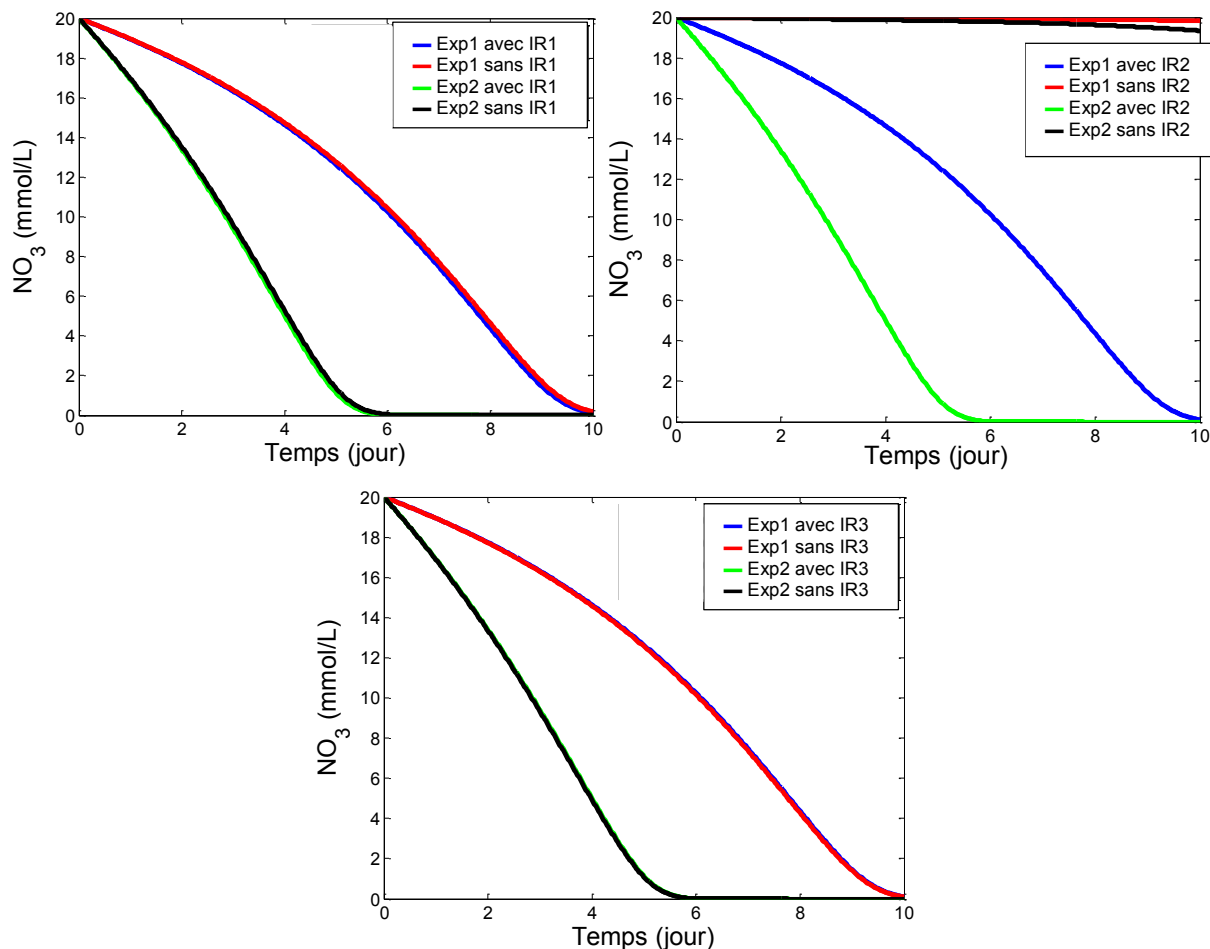


Figure 6.5 : Impact sur l'estimation du nitrate du retrait des 3 impacts réactionnels

La figure 6.5 montre que le retrait des impacts réactionnels IR1 et IR3 ne change pas la valeur de NO_3 au cours du temps. Par contre, le retrait de IR2 modifie considérablement la valeur de NO_3 . Par conséquent, l'utilisation du critère est pertinente, car il indique que l'on peut retirer IR1 et IR3, mais que IR2 doit être conservé.

6.1.5 Critère E : Fonctions unitaires ayant une valeur constante

Le rôle du critère E est de simplifier les fonctions unitaires ayant une valeur constante. L'objectif de ce critère est donc de rassembler en une seule constante le

taux de réaction maximal et la fonction unitaire ayant une valeur constante. Pour ce faire, ce critère a tout d'abord été défini comme suit :

Si la différence entre la valeur maximale et la valeur minimale observée pour la fonction unitaire est plus petite que 0.05, la fonction unitaire est retirée et le taux de réaction maximal est multiplié par la moyenne de la fonction unitaire.

Lors de tests effectués pour vérifier que le critère n'affectait pas les prédictions du modèle, il a été constaté que lorsque la moyenne de la fonction unitaire est faible (près de 0), la valeur de la fonction unitaire peut varier significativement tout en respectant la limite de 0.05. Cette situation n'a pas été observée dans le modèle de Cloutier. Cependant, il est possible de le constater en observant la fonction unitaire $\frac{C_{6pg}}{C_{6pg}+K_{PGDH,6pg}}$ issue du modèle de Chassagnole (2002) dont la valeur dans le temps est présentée à la figure 6.6.

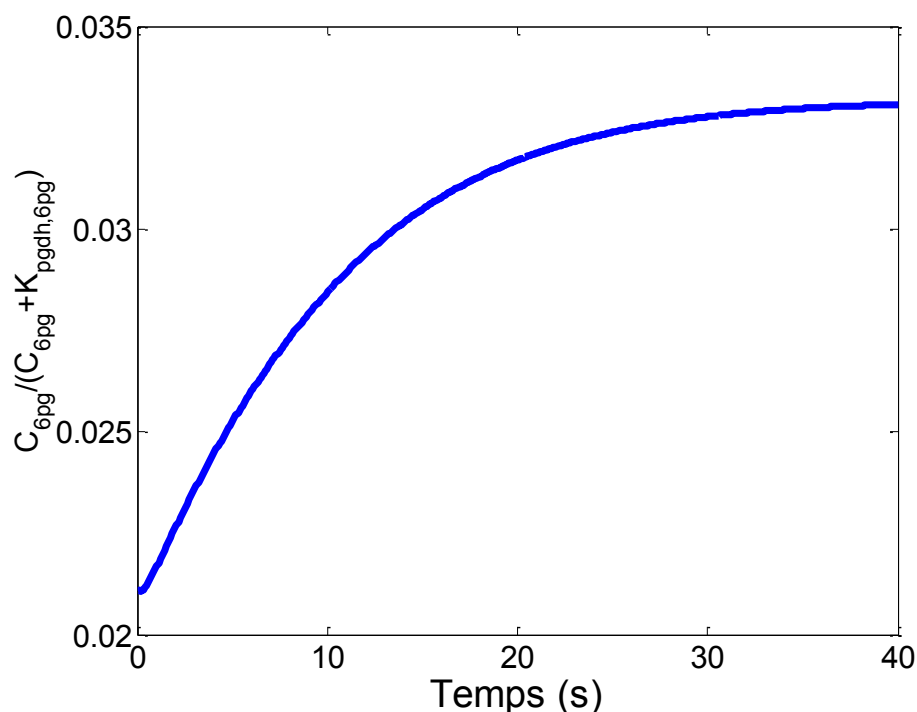


Figure 6.6 : Valeur de la fonction $C_{6pg}/(C_{6pg}+K_{PGDH,6pg})$

La valeur minimale de cette fonction unitaire est 0.021 et la valeur maximale de cette fonction est 0.033. L'écart entre la valeur maximale et minimale de cette fonction unitaire est donc de 0.012. Ainsi, selon la définition du critère E, cette fonction devrait être remplacée par sa moyenne (0.030). Or, cette fonction unitaire varie de plus de 40% dans l'intervalle de temps étudié, ce qui affecte significativement la valeur de l'impact réactionnel qui y est associé. Afin de résoudre ce problème, le critère a été modifié pour prendre en considération la variation relative à la moyenne plutôt qu'une valeur absolue. Ainsi, si la moyenne de la fonction unitaire est faible, la différence entre la valeur maximale et la valeur minimale acceptée est plus petite. Le critère a été redéfini comme suit :

Si la différence entre la valeur maximale et la valeur minimale observée pour la fonction unitaire est plus petite que 5% de la valeur moyenne de la fonction unitaire, elle est retirée et le taux de réaction maximal est multiplié par la moyenne de la fonction unitaire.

Ce critère a été testé sur les fonctions unitaires présentes dans la cinétique des sucres de réserve du modèle de Cloutier (2007). Cette cinétique est présentée à l'équation 6.6. Il est possible de constater que la forme de la cinétique change en fonction de la valeur des sucres extracellulaires telle que définie par l'utilisation de « if ». Dans le cadre de cette analyse, cela est représenté par l'utilisation de fonctions logiques (ex. : $S_C > S_{C,crit}$) valant 0 si elle est fausse et 1 si elle est vraie. Ces fonctions sont traitées comme des fonctions unitaires, puisqu'elles en possèdent les caractéristiques (valeurs variant entre 0 et 1 pour l'ensemble des valeurs possibles de la variable associée).

$$\begin{aligned}
\frac{dSCc}{dt} &= V_{max,1,SC} * \frac{S}{K_{1,S} + S} * (S_C > S_{C,crit}) \\
&- \frac{1}{Y_X^S} * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] * [1 - e^{-K_N(N_C - N_{C,min})}] \\
&\quad * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * (S_C < S_{C,min}) \\
&- S_C * u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] * \\
&[1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}
\end{aligned}
\tag{Éq. 6.6}$$

Le tableau 6.6 présente la valeur moyenne des fonctions unitaires, de même que la valeur maximale de l'écart entre la valeur de la fonction et la moyenne des fonctions. Quatre fonctions unitaires, dont les valeurs moyennes étaient plus grandes que 0.95, ont été retirées par le critère A et ne sont donc pas prises en compte dans l'application de ce critère. Elles sont indiquées comme retirées dans le tableau 6.6.

Fonction unitaire	Moyenne			Écart Max (%)		Recommandation
	Global	Exp 1	Exp 2	Exp 1	Exp 2	
S/(K1s+S)	0.94	0.95	0.92	0.45%	3.01%	*0.94
If(Sc>Sccrit)	0.95	0.98	0.92	100%	100%	If(Sc>Sccrit)
(1-exp(-Kp*(Pc-Pcmin)))	0.95	0.99	0.91	1.71%	8.85%	*(1-exp(-Kp*(Pc-Pcmin)))
(1-exp(-Ks*(Sc-Scmin)))	0.99	0.99	0.99	0.07%	0.06%	Retirée
(1-exp(-Kn*(Nc-Ncmin)))	0.60	0.65	0.55	5.47%	7.31%	*(1-exp(-Kn*(Nc-Ncmin)))
Kpp/(Kpp+Pi^inhiP)	1.00	0.99	1.00	1.03%	0.41%	Retirée
If(Sc<Scmin)	0	0	0	0%	0%	*0
(1-exp(-Kp*(Pc-Pcmin)))	0.95	0.99	0.91	1.71%	8.85%	*(1-exp(-Kp*(Pc-Pcmin)))
(1-exp(-Ks*(Sc-Scmin)))	0.99	0.99	0.99	0.07%	0.06%	Retirée
(1-exp(-Kn*(Nc-Ncmin)))	0.60	0.65	0.55	5.47%	7.31%	*(1-exp(-Kn*(Nc-Ncmin)))
Kpp/(Kpp+Pi^inhiP)	1.00	0.99	1.00	1.03%	0.41%	Retirée

Tableau 6.6 : Analyse des fonctions unitaires de la cinétique des sucres de réserve selon le critère E

On constate que ce critère suggère de remplacer deux fonctions unitaires (*S/(K1s+S) et If(Sc<Scmin)) par la valeur de leurs moyennes, respectivement 0.94 et 0. Ces modifications n'ont pas un impact significatif sur l'erreur d'estimation du modèle tel que démontré à la figure 6.7 sur la concentration des sucres de réserves.

De plus, lorsqu'appliquée à la fonction unitaire $\frac{C_{6pg}}{C_{6pg} + K_{PGDH,6pg}}$ issue du modèle de

Chassagnole (2002) la valeur du maximum de l'erreur sur la valeur maximum de la variable est de 39.7%. Cette fonction unitaire n'est donc pas simplifiée par ce critère. La nouvelle définition du critère E a permis de corriger le problème associé à la formulation initiale du critère. Ce critère a donc été retenu.

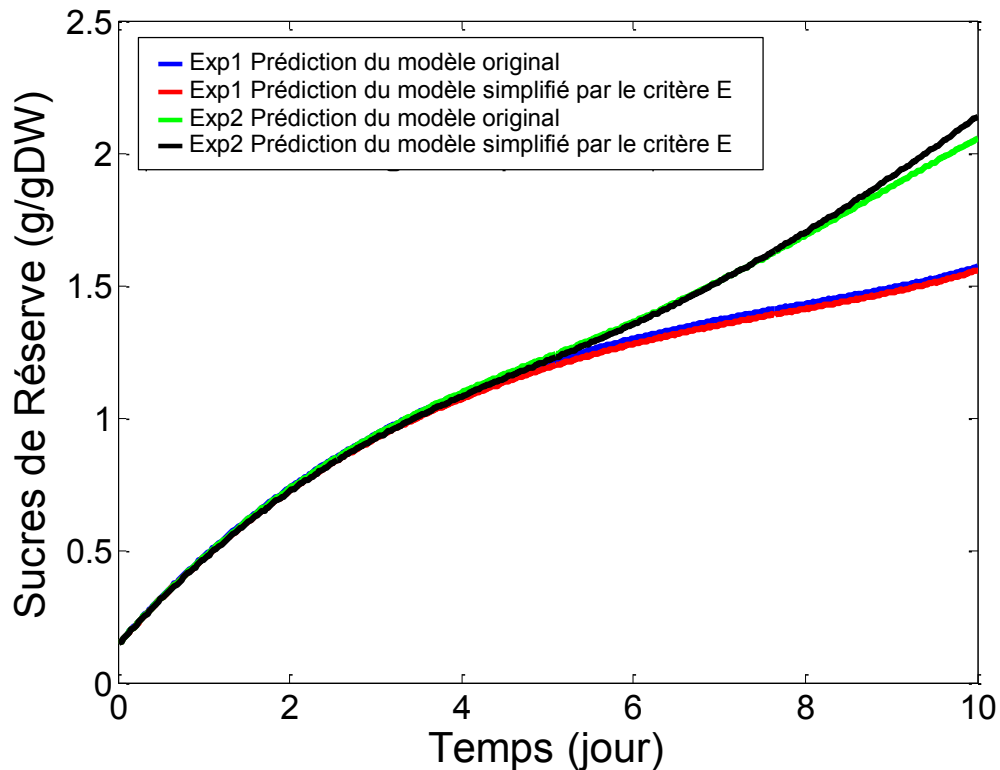


Figure 6.7 : Impact de l'application du critère E sur l'estimation de la concentration des sucres de réserve dans le modèle de Cloutier (2007)

6.1.6 Critères retenus

Suite à l'analyse des différents critères proposés, les critères suivants ont été retenus :

Critère A : Si pour l'ensemble des expériences, la moyenne de la valeur d'une fonction unitaire est égale ou supérieure à 0.95, la fonction unitaire est retirée

Critère D : Si le retrait d'une réaction entraîne une variation moyenne de la valeur de chaque variable égale ou inférieure à 5%, la réaction est retirée.

Critère E : Si la différence entre la valeur maximale et la valeur minimale observée pour la fonction unitaire est plus petite que 5% de la valeur moyenne de la fonction unitaire, la fonction unitaire est retirée et le taux de réaction maximal est multiplié par la moyenne de la fonction unitaire.

6.2 Simplification des paramètres du modèle par application du critère d'information d'Akaike

L'algorithme de simplification basé sur le critère d'information d'Akaike (AIC) a été évalué pour différentes formes du AIC.

6.2.1 Critère d'information d'Akaike original

Le critère d'information tel que décrit par Akaike (1974) ne peut pas toujours être utilisé tel quel, pour générer le modèle simplifié.

$$AIC = m * \ln\left(\frac{RSS}{m}\right) + 2n \quad \text{Éq. 6.7}$$

Le AIC varie grandement selon l'ordre de grandeur des variables. En effet, le RSS associé à une variable dont la plage de données est comprise entre 0 et 10^{-2} est beaucoup plus faible que le RSS d'une variable dont la plage de données inclut des valeurs entre 0 et 25. Cela fait en sorte qu'un modèle très précis au niveau des variables ayant des valeurs élevées, mais peu performant au niveau des variables ayant des valeurs faibles, est favorisé par rapport à un modèle ayant une bonne performance au niveau des variables ayant des valeurs faibles et élevées. Par exemple, pour le modèle de Iwamoto (2007), le modèle ayant le meilleur critère d'information d'Akaike enlevait l'ensemble des cinétiques associées aux variables

dont la plage de données avait une étendue faible puisque l'impact de la réduction du nombre de paramètres était beaucoup plus important que l'impact sur le RSS. Il est à noter que si cette modification avait un impact significatif sur les autres variables du modèle, cette simplification ne serait pas effectuée, puisque le RSS du modèle serait affecté à travers l'impact sur les autres variables.

Comme indiqué dans la section 3.5.3, le AIC n'est pas adapté pour des modèles parfaits, car le RSS est nul et le AIC tend vers moins l'infini. Ainsi, lorsque le AIC est utilisé pour simplifier un modèle parfait, aucune simplification proposée ne serait acceptée, car le AIC du modèle simplifié aurait une valeur beaucoup plus grande que celui du modèle parfait. La même situation se produit avec le critère d'information Bayésien puisqu'il contient lui aussi un logarithme du RSS. Dans le cas où les mesures expérimentales présentent une assez grande incertitude, il n'est généralement pas possible de trouver un modèle parfait (RSS=0) et le critère d'Akaike peut être utilisé tel quel. Un autre élément problématique avec le critère d'information d'Akaike est le fait que le nombre d'échantillons (m) affecte la sévérité du critère. En effet, pour des données expérimentales ayant un RSS moyen (RSS/ m) relativement constant, le fait d'augmenter le nombre de points (m) augmente la valeur du poids lié à la précision du modèle et donc, nuit à la simplification du modèle (poids lié à la simplicité du modèle). C'est pourquoi le critère d'information d'Akaike doit être modifié.

6.2.2 Utilisation d'un RSS relatif et Modèle bruité

Pour résoudre le problème des variables ayant des échelles différentes, un RSS relatif est utilisé tel que décrit à la section 5.2.2. Le RSS relatif utilisé est le RSS relatif au maximum (équation 6.8).

$$RSS_{Rel2max} = \sum \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\max(\text{Valeurs bruitées})} \right)^2 \quad \text{Éq. 6.8}$$

Et l'équation du AIC_{Rel} est :

$$AIC_{Rel} = m \ln \left(\frac{RSS_{Rel2max}}{m} \right) + 2n \quad \text{Éq. 6.9}$$

On constate que le fait d'utiliser un RSS relatif au maximum plutôt que le RSS régulier a tendance à diminuer le poids associé à la précision du modèle $\left(m * \ln \left(\frac{RSS}{n} \right) \right)$ du critère d'Akaike. En effet, si la valeur maximum d'une variable est supérieure à 1, la valeur de $RSS_{Rel2max}$ sera plus faible que la valeur du RSS. De plus, généralement, la valeur maximale de la variable est beaucoup plus élevée que la valeur de l'erreur (*Valeurs bruitées – Valeurs prédites*). Lorsque la valeur de $\left(\frac{RSS_{Rel2max}}{n} \right)$ est près de 0, le poids associé à la précision est négatif. Ainsi, cette modification a un impact sur l'équilibre entre la précision et la simplicité du modèle intrinsèque au critère d'information d'Akaike.

La première étape consiste à ordonner les paramètres en fonction de leur AIC_{rel} . Ce résultat est rapporté dans le tableau 6.7.

Paramètre	AIC rel	RSS rel
Vmax1p	-3.39E+03	1.9E-03
DWr	-3.39E+03	1.9E-03
mn	-3.39E+03	1.9E-03
K1nh4	-3.39E+03	1.9E-03
K1no3	-3.39E+03	1.9E-03
mp	-3.39E+03	1.9E-03
ms	-3.34E+03	2.0E-03
Vmax1nh4	-3.34E+03	2.1E-03
Vmax1no3	-3.33E+03	2.1E-03
Yxp	-3.31E+03	2.2E-03
Kp	-3.25E+03	2.4E-03
Pcmin	-3.25E+03	2.4E-03
Yxn	-3.12E+03	3.1E-03
K1s	-3.09E+03	3.3E-03
K1p	-3.06E+03	3.5E-03
Kpp	-2.98E+03	4.0E-03
inhiP	-2.98E+03	4.0E-03
Ks	-2.97E+03	4.0E-03
K2no3	-2.89E+03	4.7E-03
K2nh4	-2.84E+03	5.1E-03
Yxs	-2.74E+03	6.1E-03
Scmin	-2.72E+03	6.4E-03
Vmax2nh4	-2.25E+03	1.5E-02
Kn	-1.73E+03	3.9E-02
Ncmin	-1.73E+03	3.9E-02
Vmax2p	-1.54E+03	5.4E-02
Vmax2no3	-1.25E+03	9.2E-02
umax	-701	0.26
Vmax1s	-654	0.27
K2p	-168	0.66
Sccrit	1.09E+03	6.4

Tableau 6.7 : Liste des paramètres du modèle de Cloutier (2007) en ordre croissant selon leur AIC_m individuel

Une fois les paramètres ordonnés, ceux-ci sont retirés cumulativement. Le modèle simplifié optimal possède 25 paramètres alors que le modèle original en avait 31. Les caractéristiques des 15 meilleurs modèles sont présentées au tableau 6.8. Les paramètres retirés sont les 6 premiers. Le tableau 6.8 montre que le modèle simplifié optimal (AIC_{rel} minimum) a la même précision que le modèle original (RSS_{Rel}

presque identique) avec 7 paramètres en moins. Si des paramètres supplémentaires sont retirés, on constate que la précision se dégrade, mais de façon peu significative.

Paramètre	AIC rel	Nb Par	RSS rel
Vmax1p	-3.39E+03	30	0.0019
DWr	-3.39E+03	29	0.0019
mn	-3.39E+03	28	0.0019
K1nh4	-3.40E+03	27	0.0019
K1no3	-3.40E+03	26	0.0019
mp	-3.40E+03	25	0.0019
ms	-3.36E+03	24	0.002
Vmax1nh4	-3.31E+03	23	0.0022
Vmax1no3	-3.07E+03	22	0.0034
Yxp	-3.04E+03	21	0.0037
Kp	-3.04E+03	19	0.0037
Pcmin	-3.04E+03	19	0.0037
Yxn	-3.06E+03	18	0.0036
K1s	-3.10E+03	17	0.0034
K1p	-3.10E+03	17	0.0034

Tableau 6.8 : AIC_{rel} des modèles simplifiés après le retrait cumulatif des paramètres

L'erreur relative au maximum de chacune des variables pour le modèle simplifié optimal est présentée dans le tableau 6.9 pour l'expérience 1. Les résultats sont similaires pour les deux expériences. On peut constater que l'erreur relative au maximum est faible (<4%). Toutefois, elle est plus élevée que celle observée pour la méthode de simplification par critère présentée dans la section 6.1. Cela est dû au fait que les données utilisées pour cette méthode contiennent une erreur expérimentale (représenté par un bruit ayant une distribution normale).

Variable	Erreur relative au max (%)
X	4.02
Pi	2.92
Pc	3.61
NO3	3.97
NH4	3.32
Nc	3.82
S	3.43
Sc	3.85
SCc	3.77
V	3.43

Tableau 6.9 : Erreur relative au maximum de chacune des variables après la simplification du modèle par l'algorithme basé sur le critère d'information d'Akaike

La figure 6.8 présente l'impact de la simplification sur la biomasse pendant la durée de l'expérience.

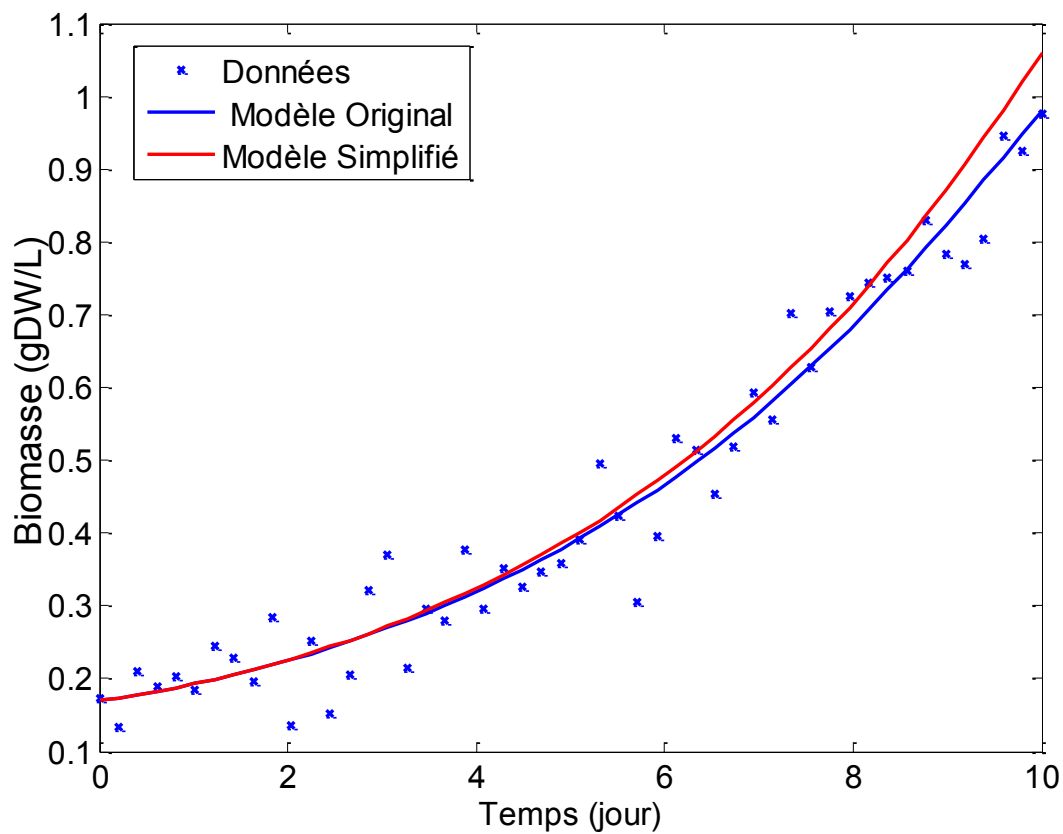


Figure 6.8 : Impact de la simplification par la méthode de l'AICrel avec une erreur sur l'estimation de la biomasse

La figure 6.8 confirme que le modèle simplifié décrit le comportement de cette variable presque aussi bien que le modèle original, car les courbes sont presque identiques durant la majorité de l'expérience avec une faible surestimation de la variable à la fin de l'expérience. L'erreur relative au maximum du modèle simplifié par rapport aux données est de 4 % alors que l'erreur relative au maximum du modèle original est de 3.6 %. Cet algorithme a permis d'obtenir un modèle simplifié ayant une erreur d'estimation faible. Cependant, les simplifications obtenues pourraient être améliorées. En effet, comme l'algorithme basé sur les critères a permis de retirer plus de paramètres sans affecter les prédictions du modèle de façon significative, il est possible d'améliorer cette méthode pour permettre plus de simplification.

6.2.3 Critère d'information d'Akaike modifié

Le critère d'information a été modifié pour obtenir un meilleur équilibre entre le poids accordé à la simplicité du modèle et le poids accordé à la précision du modèle. Tout d'abord, l'effet selon lequel la valeur du critère augmente lorsque le nombre de points augmente est indésirable. Il a donc été décidé de retirer le « m » multipliant le logarithme dans l'équation 6.9. Bien entendu, cela a un impact sur la relation entre le poids lié à la précision du modèle et le poids lié à la simplicité du modèle.

Afin d'équilibrer la relation entre le poids lié à la précision du modèle et le poids lié à la simplicité, des tests ont été effectués pour ajuster le coefficient multiplicatif du nombre de paramètres. Le tableau 6.10 présente les résultats de cette analyse sur le scénario 1.

Valeur du multiplicateur	2	0.5	0.1	0.075	0.07	0.065	0.06	0.05
Modèle Original								
Nb. Paramètres	31	31	31	31	31	31	31	31
RSSrel2max	0.0021	0.0021	0.0024	0.0025	0.0025	0.0025	0.0020	0.0023
AIcRel2max	55.85	9.32	-2.92	-3.68	-3.84	-3.97	-4.38	-4.51
Modèle Simplifié								
Nb. Paramètres	0	0	15	15	16	17	24	25
RSSrel2max	0.3121	0.2935	0.0037	0.0039	0.003	0.0038	0.002	0.0022
AIcRel2max	-1.16	-1.23	-4.10	-4.42	-4.69	-4.47	-4.77	-4.87
Erreur Relative au max (%)								
X	30.5	30.0	4.3	4.0	3.9	4.8	3.3	3.7
Pi	61.7	51.4	4.8	7.2	3.9	4.3	3.1	5.2
Pc	33.9	33.8	4.8	4.9	5.9	4.7	3.6	3.9
NO3	36.5	37.7	4.7	3.3	3.3	4.2	3.0	4.5
NH4	57.1	56.6	7.9	7.5	7.4	4.1	8.7	4.1
Nc	38.6	38.7	6.1	4.5	3.3	5.6	3.3	3.6
S	11.2	11.7	4.1	4.5	4.3	4.2	4.2	3.0
Sc	17.3	16.6	8.1	9.0	5.3	7.3	4.2	3.8
SCc	59.3	59.3	4.1	5.0	4.3	5.1	3.5	4.1
V	3.7	3.5	3.6	3.6	3.8	3.8	3.7	3.9
Erreur moyenne	35.0	33.9	5.2	5.3	4.5	4.8	4.0	4.0

Tableau 6.10 : Résultat de l'analyse de la valeur du multiplicateur du nombre de paramètres du critère d'information d'Akaike modifié

La figure 6.9 présente l'erreur relative au maximum moyenne en fonction de la valeur du paramètre. On constate que cette valeur diminue rapidement jusqu'à ce que le paramètre ait une valeur inférieure à 0.1. Après, l'erreur est relativement constante et près de l'erreur relative au maximum du modèle original.

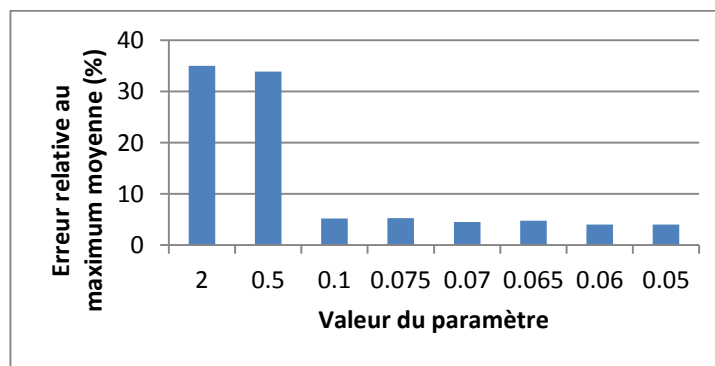


Figure 6.9 : Valeur de l'erreur relative au maximum moyenne en fonction de la valeur du paramètre

La valeur retenue pour le multiplicateur du nombre de paramètres est de 0.07. Cette valeur est choisie, car les résultats obtenus représentent un bon équilibre entre la précision des prédictions du modèle et la simplicité du modèle. On constate au tableau 6.10 que la valeur de 0.07 est la première permettant d'obtenir une erreur moyenne inférieure à 5%, et ce, tout en permettant des simplifications significatives. L'équation 6.10 présente le critère modifié. C'est le critère qui sera utilisé dans les simplifications suivantes

$$AIC_m = \ln\left(\frac{RSS_{Rel2max}}{m}\right) + 0.07 * n \quad \text{Éq. 6.10}$$

L'algorithme de simplification avec le critère modifié est le même qu'avec le critère original présenté à la section 5.2.4. Les résultats de la simplification sont présentés dans le tableau 6.11. Il est à noter que l'information sur les modèles où plus de paramètres sont retirés n'est pas montrée. Ces modèles ont un AIC_m de moins en moins bon.

Paramètre	AIC_m	Nb. Par	$RSS_{Rel2max}$
Vmax1nh4	-4.091	30	0.002
Vmax1p	-4.1664	28	0.0022
K1nh4	-4.1664	28	0.0022
K1no3	-4.2482	27	0.0022
mp	-4.3184	26	0.0022
mn	-4.3883	25	0.0022
DWr	-4.4464	24	0.0022
Vmax1no3	-4.2926	23	0.0027
Kp	-4.4545	21	0.0027
Pcmin	-4.4545	21	0.0027
Yxp	-4.4547	20	0.0029
ms	-4.1995	19	0.004
Kpp	-4.6567	16	0.0031
inhiP	-4.6567	16	0.0031
Yxn	-4.4733	15	0.004
Ks	-4.4433	14	0.0044
K1s	-4.398	13	0.005
K1p	-4.398	13	0.005

Tableau 6.11 : AIC_m des modèles simplifiés après le retrait cumulatif des paramètres

Selon le critère AIC_m retenu, le modèle optimal est donc le modèle à 16 paramètres. La modification du critère a donc permis le retrait de 15 paramètres par rapport au modèle original soit 9 de plus que le retrait obtenu avec le AIC_{Rel} . Tous les paramètres retirés par le AIC_{Rel} sont aussi retirés par le AIC_m . L'impact de ces retraits sur l'erreur relative au maximum du modèle simplifié est comparé au modèle original dans le tableau 6.12. Cela montre que l'erreur relative au maximum du modèle simplifié optimal est très près de celle du modèle original. La simplification a donc conservé la précision du modèle.

Variable	Erreur relative au max (%)				Écart (%)	
	Modèle Original		Modèle Simplifié		Écart (%)	
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1	Exp 2
X	6.94	4.05	7.02	5.23	0.08	1.18
Pi	4.38	2.6	4.37	3.44	-0.01	0.84
Pc	3.79	8.15	5.87	9.24	2.08	1.09
NO3	3.65	3.9	3.68	3.9	0.03	0
NH4	3.13	4.98	4.46	6.21	1.33	1.23
Nc	3.52	3.91	3.74	4.35	0.22	0.44
S	3.75	3.77	3.83	3.89	0.08	0.12
Sc	4.54	5.81	5.9	7.14	1.36	1.33
SCc	5.07	3.14	5.04	3.74	-0.03	0.6
V	3.49	4.77	3.38	4.91	-0.11	0.14

Tableau 6.12 : Erreur relative au maximum de chacune des variables du modèle original ainsi que du modèle simplifié par l'algorithme basé sur le critère d'information d'Akaike modifié

La figure 6.10 présente l'impact de la simplification sur la biomasse et le phosphate intracellulaire (Pc). Le phosphate intracellulaire est la variable la plus affectée par cette simplification. Il est possible de constater que le modèle optimal a tendance à surestimer légèrement la concentration de phosphate intracellulaire. Toutefois, l'erreur relative au maximum entre les variables générées par le modèle original les variables générées par le modèle simplifié (tableau 6.12) montre que les performances de cet algorithme sont satisfaisantes, car l'erreur entre le modèle original et le modèle simplifié est inférieure à 5%.

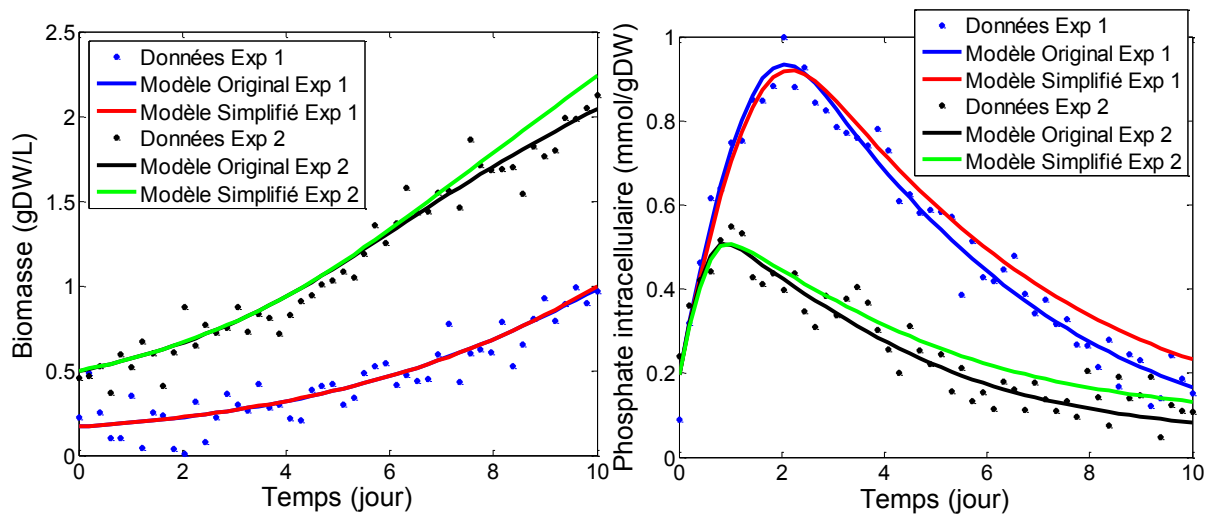


Figure 6.10 : Impact de la simplification par l'AIC_m sur l'estimation de la biomasse et du phosphate intracellulaire

En conclusion, le choix final de l'algorithme de simplification des paramètres basé sur le critère d'information d'Akaike s'est donc porté sur un algorithme utilisant le critère d'information d'Akaike avec une formulation modifiée, tout en utilisant un RSS relatif au maximum et en ajoutant une erreur aux données générées par le modèle original. Le critère d'information utilisé est donc calculé comme suit :

$$RSS_{Rel2max} = \sum \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\max(\text{Valeurs bruitées})} \right)^2 \quad \text{Éq. 6.11}$$

$$AIC_m = \ln \left(\frac{RSS_{Rel2max}}{n} \right) + 0.07 * p$$

L'algorithme de simplification basé sur le critère AIC_m a permis d'obtenir des simplifications significatives tout en conservant une très bonne précision.

6.3 Réduction d'ordre du modèle par retrait de variables non essentielles

Dans la méthode de simplification des variables telle que présentée à la section 5.3, il est nécessaire d'identifier les variables d'intérêt et les variables non essentielles. Deux scénarios sont présentés et se distinguent par le choix des variables d'intérêt. La variable d'intérêt du scénario 1 est la biomasse (X) uniquement, les variables d'intérêt du scénario 2 sont la biomasse (X), le phosphate intracellulaire (Pc) et les sucres de réserve (SCc). Ces deux scénarios sont appliqués sur deux simulations se différenciant par la concentration initiale en biomasse (0.17 gDW/L dans la simulation 1 et 0.5 gDW/L dans la simulation 2), car la biomasse est une variable critique du procédé de culture cellulaire.

6.3.1 Ordre de simplification des variables

Pour déterminer l'ordre de retrait des variables, la cinétique de chacune des variables non essentielles est posée à 0 individuellement. L'impact de cette modification sur l'estimation des variables d'intérêt est mesuré à l'aide du RSS calculé uniquement sur les variables d'intérêt. Suite à la réflexion réalisée quant à la forme du RSS utilisé dans le critère d'Akaike pour la simplification de paramètre, trois formes de RSS ont été envisagées, le RSS absolu (équation 6.12), le RSS relatif (équation 6.13) et le RSS relatif au maximum (équation 6.14) :

1. le RSS absolu

$$RSS = \sum_{\substack{\text{variables} \\ \text{d'intérêt}}} (\text{Valeurs bruitées} - \text{Valeurs prédites})^2 \quad \text{Éq. 6.12}$$

2. le RSS relatif

$$RSS_{Rel} = \sum_{\substack{\text{variables} \\ \text{d'intérêt}}} \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\text{Valeurs bruitées}} \right)^2 \quad \text{Éq. 6.13}$$

3. le RSS relatif à la valeur maximale de la variable.

Éq. 6.14

$$RSS_{Rel2max} = \sum_{\substack{\text{variables} \\ \text{d'intérêt}}} \left(\frac{\text{Valeurs bruitées} - \text{Valeurs prédites}}{\max(\text{Valeurs bruitées})} \right)^2$$

Le tableau 6.13 présente le résultat pour les deux scénarios. On constate que les trois méthodes de calcul du RSS produisent le même ordre de simplification. Ainsi, ce choix a un impact mineur pour l'algorithme de simplification. Parce que le $RSS_{rel2max}$ accorde un poids similaire à l'ensemble des variables, peu importe leur plage de données, sans accorder un impact plus grand aux valeurs faibles de chacune des variables, il sera utilisé pour ordonner les variables non essentielles.

Scénario 1				Scénario 2			
	RSS absolu moyen	RSS relatif moyen	RSS relatif au max moyen		RSS absolu moyen	RSS relatif moyen	RSS relatif au max moyen
Sc	0.0132	0.003	0.0029	NH4	0.0074	0.0022	0.0022
S	0.0132	0.003	0.003	V	0.0074	0.0022	0.0022
NH4	0.0132	0.003	0.003	S	0.0077	0.0024	0.0023
SCc	0.0132	0.003	0.003	NO3	0.0149	0.004	0.004
V	0.0132	0.003	0.003	Nc	0.0723	0.0213	0.0212
Pc	0.0136	0.0031	0.0031	Sc	0.438	0.1075	0.1043
Pi	0.0182	0.0041	0.0041	Pi	1.4844	1.4817	1.4817
NO3	0.026	0.0064	0.0058				
Nc	0.077	0.0178	0.0173				

Tableau 6.13 : Classification des variables non essentielles en fonction des différents calculs de RSS

6.3.2 Critères de simplification

Afin de réduire l'ordre du modèle différentiel, trois options ont été présentées dans la section 5.3.2. :

1. Les variables non essentielles peuvent être enlevées si l'erreur d'estimation relative au maximum de la variable du modèle par rapport à l'ensemble des données des variables d'intérêt ne dépasse pas 5%.

$$Err_{rel2data} = \frac{1}{v_i} \sum_{i=1}^{v_i} \frac{1}{m} \sum_{j=1}^m \frac{|Données_{i,j} - Modèle simplifié_{i,j}|}{\max(Données_{i,j})} * 100\% \quad \text{Éq. 6.15}$$

2. Les variables non essentielles peuvent être enlevées si l'erreur d'estimation relative au maximum du modèle simplifié par rapport au modèle original ne dépasse pas 5%, pour aucune des variables d'intérêt.

$$Err_{rel2model} = \max \left(\frac{1}{m} \sum_{j=1}^m \frac{|Modèle original_{i,j} - Modèle simplifié_{i,j}|}{\max(Modèle original_{i,j})} * 100\% \right) \quad \text{Éq. 6.16}$$

3. Le modèle sélectionné est celui dont la valeur du critère AIC_{Var} (équation 6.17) est la plus faible.

$$AIC_{Var} = \ln(RSS_{Rel2max}) + 0.07 * v \quad \text{Éq. 6.17}$$

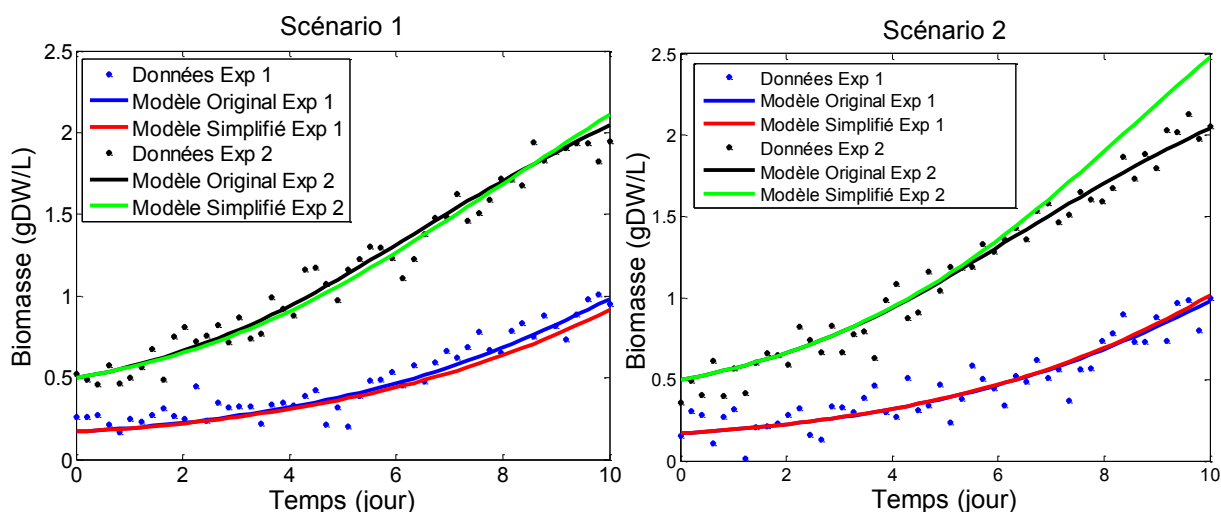
Les résultats de simplification obtenus pour les trois options sont présentés dans le tableau 6.14.

	Err _{rel2data} 5%		Err _{rel2model} 5%		AIC _{var}	
	Scénario 1	Scénario 2	Scénario 1	Scénario 2	Scénario 1	Scénario 2
X	Var d'intérêt	Var d'intérêt	Var d'intérêt	Var d'intérêt	Var d'intérêt	Var d'intérêt
Pi	Conservée	Conservée	Conservée	Conservée	Conservée	Conservée
Pc	Retirée	Var	Retirée	Var	Conservée	Var

		d'intérêt		d'intérêt		d'intérêt
NO3	Conservée	Retirée	Conservée	Retirée	Conservée	Conservée
NH4	Retirée	Retirée	Retirée	Retirée	Retirée	Retirée
Nc	Conservée	Conservée	Conservée	Conservée	Conservée	Conservée
S	Retirée	Retirée	Retirée	Retirée	Retirée	Conservée
Sc	Retirée	Conservée	Retirée	Conservée	Retirée	Conservée
SCc	Retirée	Var d'intérêt	Retirée	Var d'intérêt	Retirée	Var d'intérêt
V	Retirée	Retirée	Retirée	Retirée	Retirée	Retirée

Tableau 6.14 : Résultat de l'application de l'algorithme de simplification des variables non essentielles selon les trois critères testés.

Le tableau 6.14 cumule les résultats des deux simulations correspondant à des conditions initiales de biomasse différentes. Pour qu'une variable soit retirée, elle doit être retirée dans les deux simulations. Les deux premières options identifient le même modèle simplifié. Ayant le même modèle, ces deux options produisent les mêmes prédictions. Elles retirent 6 variables dans le premier scénario et 4 variables dans le second. La troisième option retire 5 variables dans le premier scénario et 2 variables dans le second. L'impact de la simplification sur les variables d'intérêt est présenté à la figure 6.11 (1^{ère} et 2^e options) et à la figure 6.12 (3^e option).



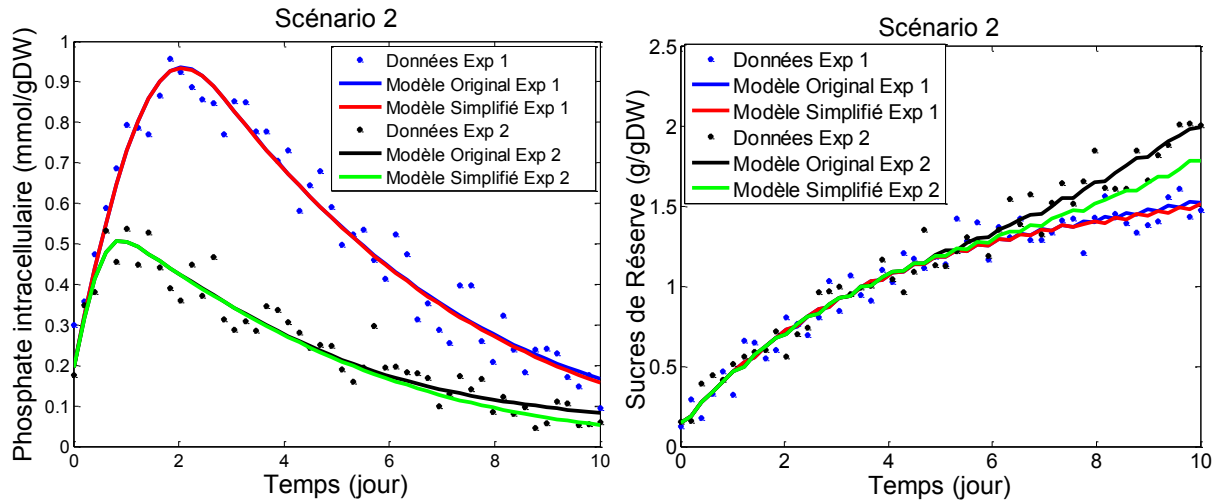


Figure 6.11 : Résultat de l'algorithme de simplification grâce aux critères 1 et 2 sur les variables d'intérêt

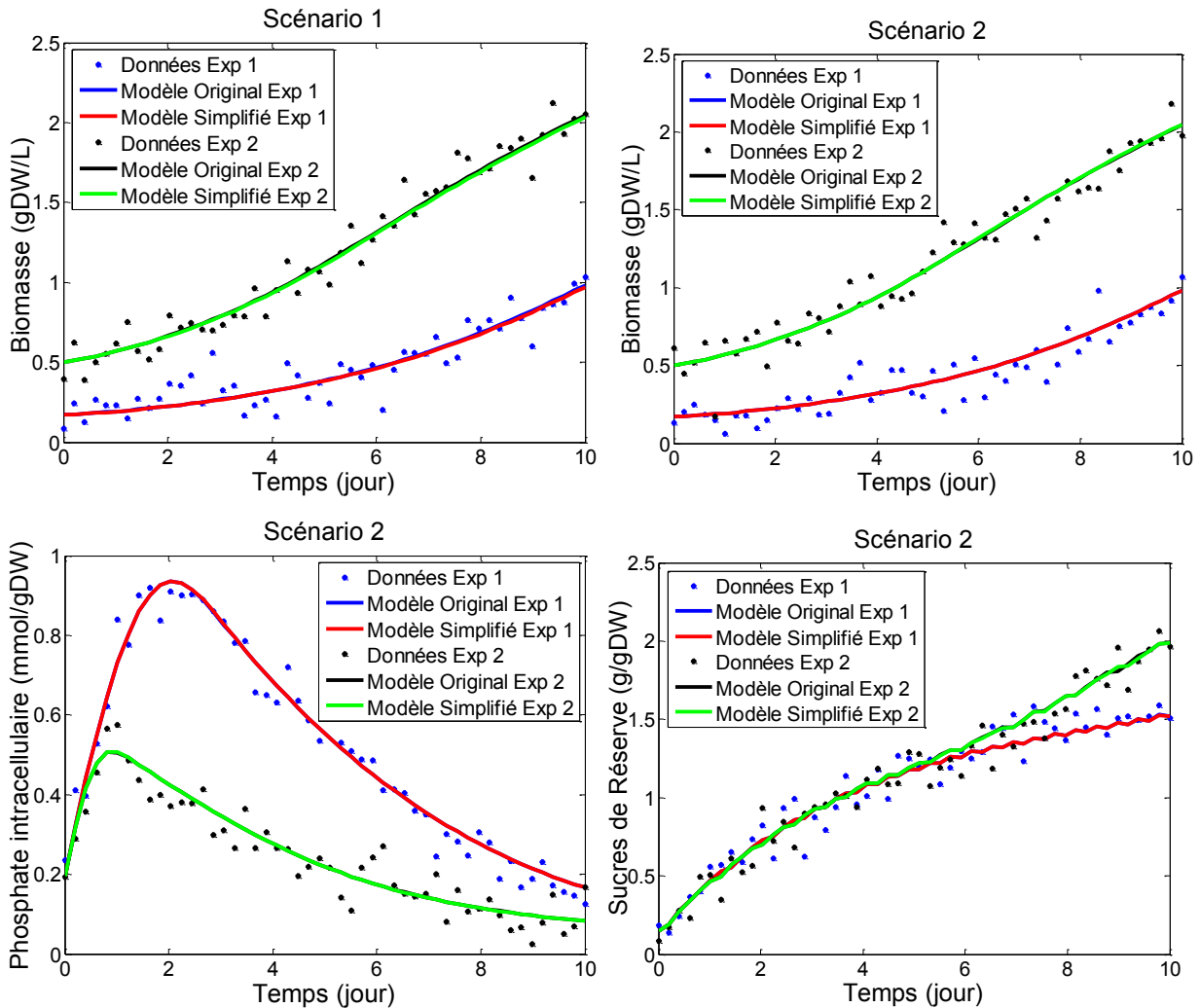


Figure 6.12 : Résultat de l'algorithme de simplification grâce au critère 3 sur les variables d'intérêt

Les résultats obtenus sont similaires, bien que la 1^{ère} et la 2^e option surestiment légèrement la biomasse et sous-estiment légèrement les sucres de réserves dans le 2^e scénario. Ces options permettent de simplifier une variable de plus que la 3^e option, dans le 1^{er} scénario et deux variables de plus dans le 2^e. Autrement dit, les 2 premières options permettent d'obtenir des modèles plus simples, alors que la 3^e option permet d'obtenir des modèles plus précis. Entre les deux premières options, la 2^e a été préférée puisque l'évaluation se fait variable par variable plutôt que sur l'ensemble des variables. Ainsi, avec la première option, il était possible d'obtenir un modèle où une variable d'intérêt avait une erreur assez élevée avec plusieurs variables ayant une erreur faible. Parmi les trois options testées, l'option sélectionnée pour l'algorithme est la deuxième.

En conclusion, la méthode de simplification retenue utilise le $RSS_{rel2max}$ pour ordonner les variables et utilise l'option selon laquelle la simplification ne doit pas entraîner une erreur relative au maximum supérieure à 5% sur aucune variable. Avec cette méthode, les modèles ont été simplifiés de façon significative. La simplification dépend évidemment du choix des variables d'intérêt. Ce choix peut être guidé par des conditions de production et d'optimisation, mais il laisse place à interprétation.

Ce chapitre présente l'impact des différents algorithmes de simplification développés sur des modèles provenant de la littérature.

7.1 Cloutier, 2007

Cloutier (2007) présente trois modèles mathématiques décrivant la croissance de trois souches de cellules : *Catharanthus roseus*, *Daucus carota* et *Eschscholzia californica* en suspension. Le modèle utilisé pour cette analyse est le modèle présentant la croissance d'*Eschscholzia californica*. Ce modèle comporte 10 variables, 31 paramètres, 30 impacts réactionnels et 15 fonctions unitaires uniques dont certaines se retrouvent à plusieurs endroits dans le modèle (73 fonctions unitaires totales). Deux simulations sont considérées (les mêmes qu'au chapitre 6). Ces simulations se différencient par la concentration initiale en biomasse (0.17 gDW/L dans la simulation 1 et 0.5 gDW/L dans la simulation 2).

7.1.1 Simplification par critères

Le modèle a été analysé selon les trois critères de la section 6.1.6. Le tableau 7.1 présente le nombre de fois où chaque critère a identifié une simplification possible. Il est à noter que lorsqu'un élément (fonction unitaire ou impact réactionnel) se retrouve à plusieurs endroits dans le modèle et qu'un critère suggère de le retirer, cette simplification est comptée à chaque fois. De plus, certaines simplifications peuvent se recouper. Par exemple, si le critère A suggère une simplification sur une fonction unitaire appartenant à un impact réactionnel dont le retrait est suggéré par le critère D, les simplifications suggérées sont comptabilisées pour les deux critères, même si l'application du critère D entraîne le retrait de la fonction unitaire. Toutefois, lorsque les critères A et E suggèrent tous les deux le retrait de fonctions unitaires, ces retraits ne sont pas appliqués de façon concurrente. Si la moyenne d'une fonction

unitaire est supérieure à 0.95, le critère E n'est pas appliqué sur cette fonction unitaire.

Critère A	29
Critère D	12
Critère E	4

Tableau 7.1 : Nombre de fois où chacun des critères a suggéré une simplification dans le modèle de Cloutier (2007)

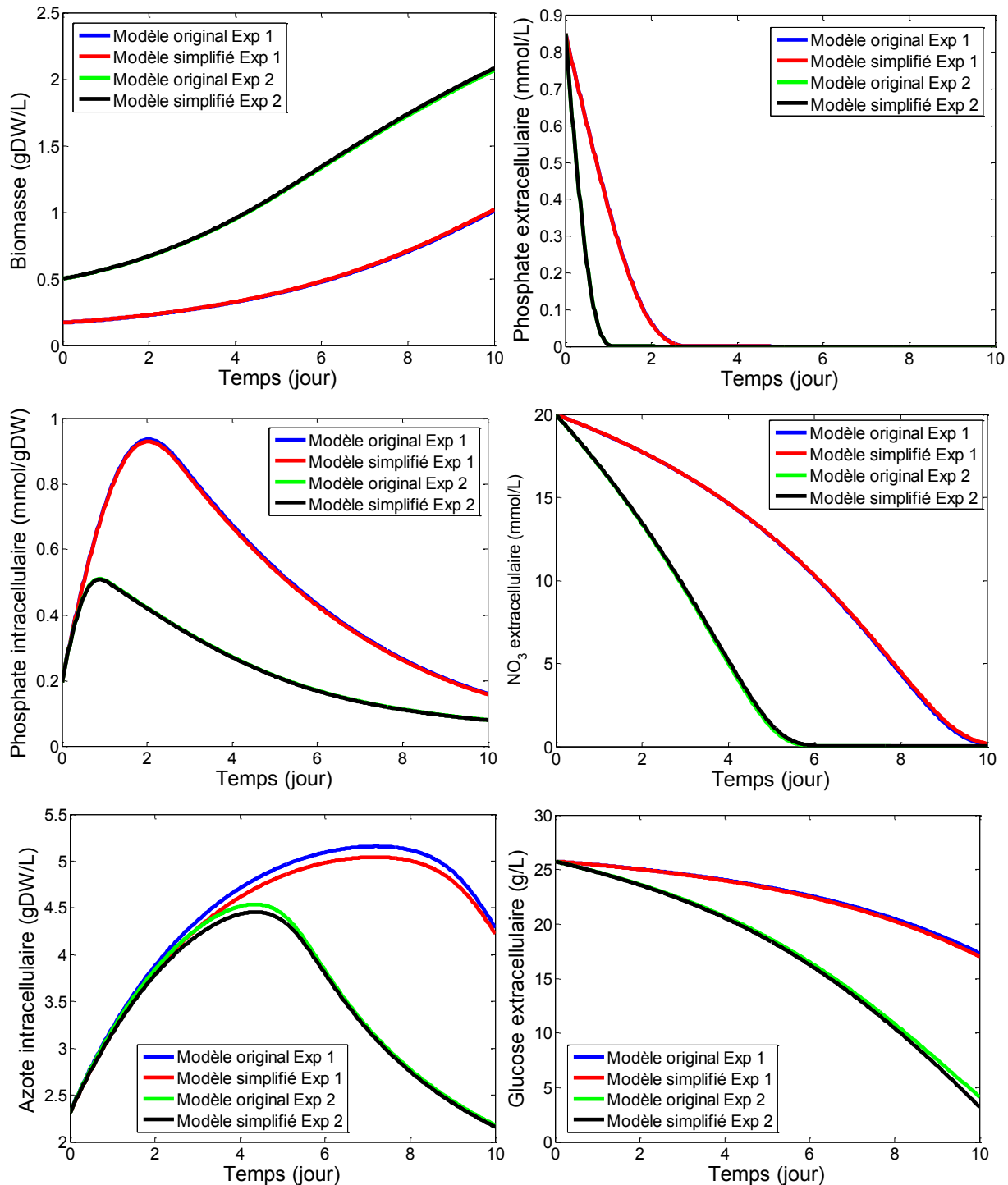
Le tableau 7.2 suivante présente les caractéristiques du modèle original ainsi que du modèle simplifié. On constate que l'algorithme de simplification par critère a permis de réduire significativement la complexité du modèle.

Caractéristiques	Cloutier 2007	Modèle simplifié	%
Nombre de variables	10	9	-10%
Nombres de paramètres	31	20	-35%
Nombre d'impacts réactionnels	30	18	-40%
Nombre de fonctions unitaires (uniques)	15	9	-40%
Nombre de fonctions unitaires (total)	73	24	-67%

Tableau 7.2 : Caractéristiques du modèle de Cloutier (2007) original et du modèle simplifié

On remarque que le critère D a permis de retirer une variable, le volume (V), parce que cette variable a une valeur quasi constante. Bien qu'il suggère quatre simplifications, le critère E ne permet pas de simplification effective du modèle puisque les fonctions unitaires retirées appartiennent à des impacts réactionnels retirés par le critère D. Puisque le critère A suggère plusieurs simplifications, cela veut dire que plusieurs fonctions unitaires ont des valeurs près de 1 sur l'ensemble des simulations. Par conséquent, cela signifie que ces fonctions unitaires n'étaient pas actives. Ce qui donne une information sur les mécanismes biologiques présents, c'est-à-dire, quelles sont les limitations et les inhibitions affectant les vitesses de réaction.

La figure 7.1 présente les prédictions obtenues par le modèle original et par le modèle simplifié. On peut constater que les prédictions obtenues des deux modèles sont très similaires. L'erreur relative au maximum entre le modèle original et le modèle simplifié est inférieure à 5% pour l'ensemble des variables.



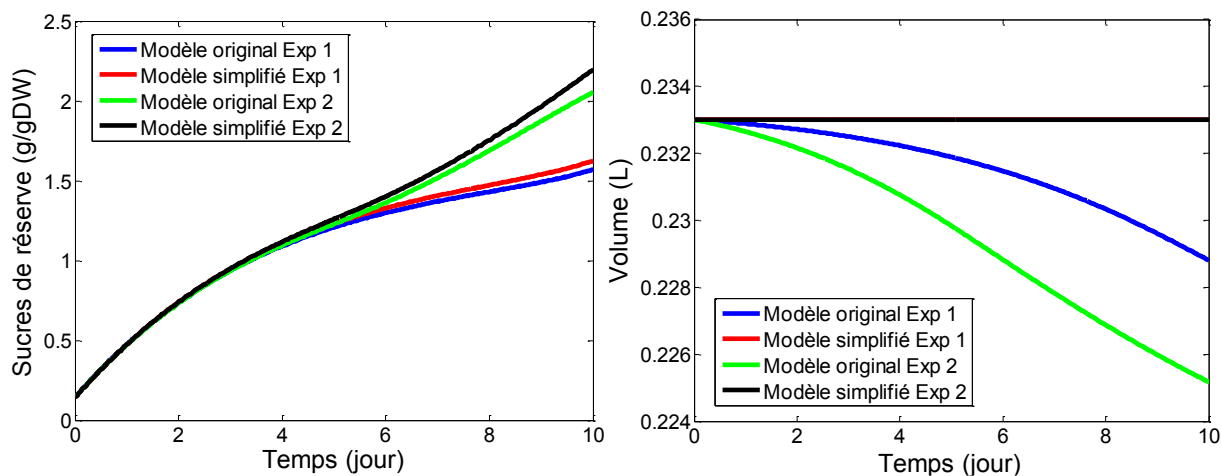


Figure 7.1 : Données modélisées par le modèle de Cloutier (2007) original et par le modèle simplifié

Afin de caractériser la précision de l'estimation fournie par le modèle simplifié, l'erreur relative au maximum entre les données générées par modèle original et les données générées par le modèle simplifié est calculée à l'aide de l'équation 7.1 pour chacune des variables. Les résultats obtenus sont présentés dans le tableau 7.3.

$$Erreur\ Relative\ au\ max = \frac{1}{m} \sum \frac{|X_{Modèle\ original,i} - X_{modèle\ simplifié,i}|}{max(X_{Modèle\ original})} * 100\% \quad \text{Éq. 7.1}$$

Variable	Erreur relative au max (%)	
	Simulation 1	Simulation 2
X	0.45	0.37
Pi	0.21	0.08
Pc	0.51	0.30
NO3	0.23	0.50
NH4	0.20	0.07
Nc	1.73	0.97
S	0.42	0.94
Sc	1.24	1.21
SCc	1.51	1.89
V	0.62	1.48
Erreur moyenne (%)	0.71	0.78

Tableau 7.3 : Erreur relative au max associée à la simplification du modèle de Cloutier (2007) pour les deux simulations (simulation 1 : Biomasse initiale :0.17 gDW/L, simulation 2 : Biomasse initiale :0.5 gDW/L)

On constate que le modèle simplifié effectue des prédictions similaires à celles obtenues par le modèle original. Il est intéressant d'observer les différences au niveau des cinétiques entre le modèle original et le modèle simplifié. Par exemple, le modèle original utilise la cinétique suivante pour décrire la croissance cellulaire :

$$\frac{dX}{dt} = u_{max} * [1 - e^{-K_P(P_C - P_{C,min})}] * [1 - e^{-K_S(S_C - S_{C,min})}] * [1 - e^{-K_N(N_C - N_{C,min})}] * \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} \quad \text{Éq. 7.2}$$

La croissance cellulaire est affectée par des limitations sur le phosphate (Pc), le sucre (Sc), ainsi que l'azote intracellulaire (Nc) et par une inhibition sur le phosphate extracellulaire (Pi). Le modèle simplifié ne contient qu'une limitation sur le phosphate et l'azote. Ainsi, il est possible de conclure que les phénomènes de limitation sur le sucre intracellulaire et d'inhibition sur le phosphate extracellulaire ne sont pas observés dans les conditions étudiées. L'analyse des impacts réactionnels permet de constater que les réactions de maintenance (consommation d'azote de phosphate et de sucre pour le maintien de la concentration cellulaire) sont trop faibles pour avoir un impact significatif sur les concentrations. De plus, l'algorithme de simplification retire l'ensemble de la cinétique associée au volume, transformant cette variable en constante. Ainsi, l'application de l'algorithme de simplification par critères permet de faire ressortir les phénomènes importants.

7.1.2 Simplification par AIC_m

L'algorithme de simplification basé sur le critère d'information d'Akaike modifié utilisant un RSS relatif au maximum a été appliqué sur le modèle de Cloutier. Le modèle simplifié ainsi obtenu comporte 16 paramètres, comme indiqué à la section 6.2. L'erreur relative au maximum de chacune des variables est présentée dans le tableau 7.4 pour le modèle original et le modèle simplifié.

Variable	Erreur relative au max par rapport aux données		Erreur relative au max par rapport au modèle original	
	Simulation 1	Simulation 2	Simulation 1	Simulation 2
X	7.0%	5.2%	0.2%	1.9%
Pi	4.4%	3.4%	0.8%	0.3%
Pc	5.9%	9.2%	4.4%	7.1%
NO3	3.7%	3.9%	0.4%	0.3%
NH4	6.5%	8.4%	3.7%	2.1%
Nc	3.7%	4.4%	1.4%	2.1%
S	3.8%	3.9%	0.5%	1.1%
Sc	5.9%	7.1%	1.1%	1.5%
SCc	5.0%	3.7%	0.5%	1.8%
V	3.4%	4.9%	0.6%	1.5%

Tableau 7.4 : Erreur relative au maximum par rapport aux données et par rapport au modèle original du modèle de Cloutier (2007) simplifié

On constate que l'erreur est élevée pour le phosphate intracellulaire (Pc) et relativement élevée pour l'ammonium extracellulaire (NH₄). La figure 7.2 présente les graphiques associés à ces variables.

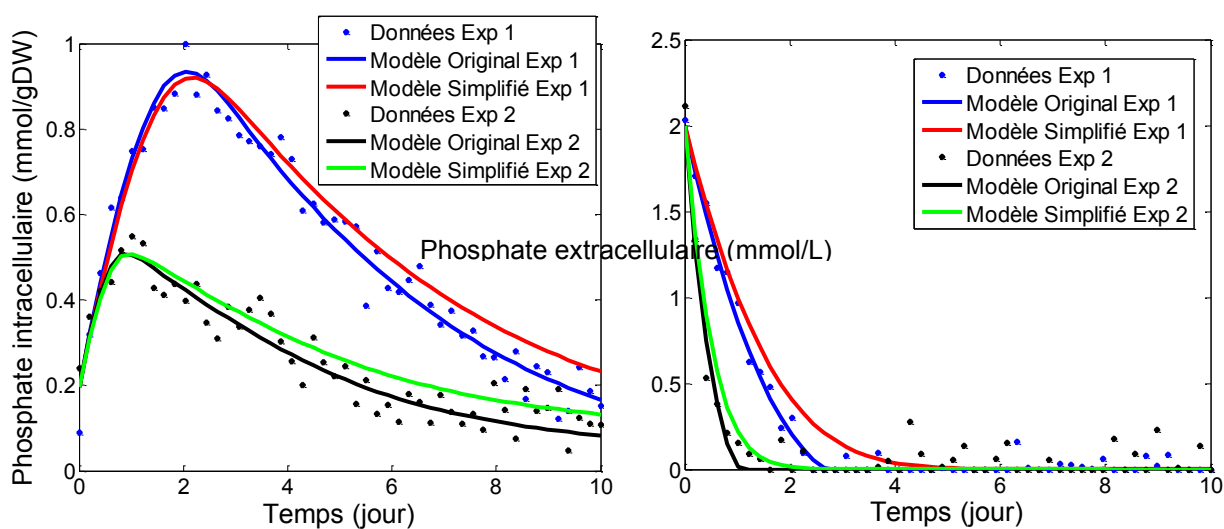


Figure 7.2 : Impact de la simplification par l'AICm sur l'estimation du phosphate intracellulaire et extracellulaire

Il est possible de constater que le modèle simplifié surestime légèrement les deux variables (NH₄ et Pc). Cette différence est jugée acceptable, car elle

n'augmente pas l'erreur relative au maximum de plus de 5% de celle liée au modèle original tel que présenté au tableau 7.4. Les caractéristiques du modèle simplifié et leurs comparaisons avec le modèle original sont présentées au tableau 7.5.

Caractéristiques	Cloutier 2007	Modèle simplifié	%
Nombre de variables	10	9	-10%
Nombres de paramètres	31	16	-48%
Nombre d'impacts réactionnels	30	16	-47%
Nombre de fonctions unitaires (uniques)	15	10	-33%
Nombre de fonctions unitaires (total)	73	28	-62%

Tableau 7.5 : Caractéristiques du modèle de Cloutier (2007) original et du modèle simplifié par l'algorithme de simplification basé sur le critère d'information d'Akaike modifié

Lorsqu'on compare le modèle obtenu avec cet algorithme avec le modèle simplifié par l'algorithme des critères, on constate que cet algorithme permet de retirer plus de paramètres que l'algorithme précédent. Certaines conclusions sont communes. Par exemple, les deux algorithmes considèrent que le volume est constant, que les réactions de maintenance sont négligeables et que les fonctions unitaires significatives sur la réaction de croissance cellulaire sont les limitations sur le phosphate et l'azote. Par contre, cette méthode suggère de simplifier une des deux réactions d'adsorption du phosphate (les deux impacts réactionnels de cette réaction), alors que l'algorithme basé sur les critères conservait les deux réactions. Cette simplification modifie la prédiction du phosphate intracellulaire, mais de façon faible.

7.1.3 Simplification des variables non essentielles

Tel que présenté à la section 6.3.1, la simplification des variables non essentielles a été testée sur deux simulations qui se différencient par la concentration initiale en biomasse combinées à deux scénarios quant à la composition des variables d'intérêt. Dans le premier scénario, la seule variable d'intérêt est la biomasse. On cherche alors à identifier le modèle le plus simple possible permettant

de décrire la croissance de la biomasse. Dans le second scénario, les variables d'intérêt sont la biomasse et les variables intracellulaires Pc, Nc et SCc. Le modèle doit alors pouvoir décrire la biomasse ainsi que les concentrations de nutriments intracellulaires.

	Scénario 1	Scénario 2
X	Var d'intérêt	Var d'intérêt
Pi	Conservée	Conservée
Pc	Retirée	Var d'intérêt
NO3	Conservée	Retirée
NH4	Retirée	Retirée
Nc	Conservée	Conservée
S	Retirée	Retirée
Sc	Retirée	Conservée
SCc	Retirée	Var d'intérêt
V	Retirée	Retirée

Tableau 7.6 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles

On constate que dans les deux scénarios, plusieurs variables sont retirées dans le modèle simplifié. Ainsi, ces variables ne sont pas nécessaires pour décrire la croissance de la biomasse. La figure 7.3 présente les résultats obtenus pour les variables d'intérêt par le modèle simplifié pour les deux scénarios de variables d'intérêt.

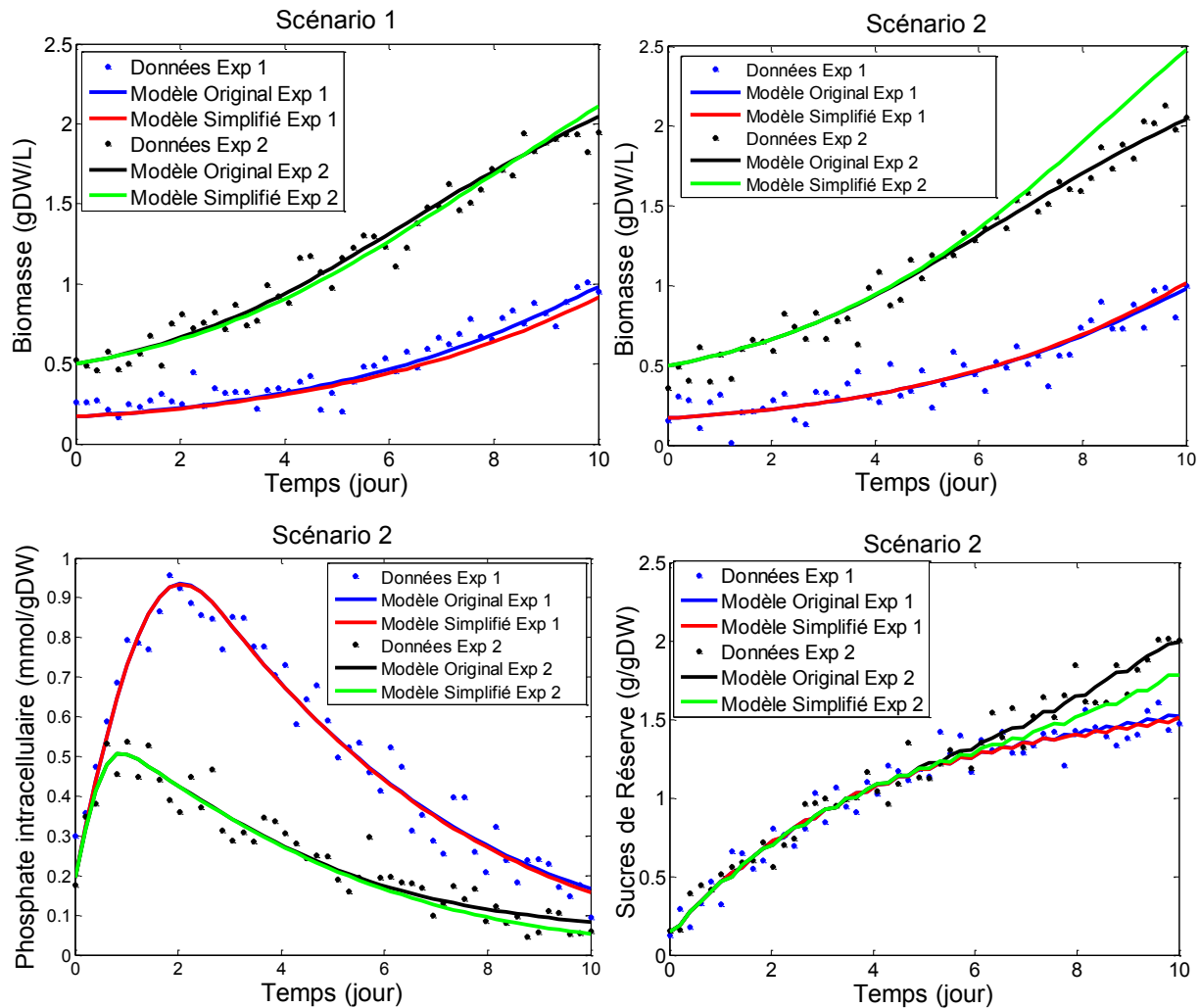


Figure 7.3 : Impact de la simplification des variables non essentielles sur les variables d'intérêt

On constate que le modèle simplifié surestime la biomasse et sous-estime légèrement les sucres de réserve dans la simulation 2. L'erreur sur la biomasse (4.45%) est d'ailleurs près de la limite posée par le critère (5%). L'erreur sur les sucres de réserves est due à la surestimation de la biomasse. L'erreur sur la biomasse est due à l'effet de la simplification de NO_3 et NH_4 sur l'azote intracellulaire (N_c) qui a une erreur assez élevée (~30%). Cet effet se répercute sur la biomasse à travers la fonction unitaire de limitation de l'azote intracellulaire. Les résultats obtenus sont tout de même considérés comme bons. Le tableau 7.7 présente l'erreur relative au maximum par rapport aux données et par rapport au modèle original. Bien sûr, l'erreur par rapport aux données est supérieure à l'erreur par rapport au modèle

original puisque les données possèdent un bruit. On constate que ces erreurs sont faibles pour l'ensemble des variables d'intérêt.

	Erreur relative au max par rapport aux données (%)		Erreur relative au max par rapport au modèle original (%)	
	Scénario 1		Scénario 1	
	Simulation 1	Simulation 2	Simulation 1	Simulation 2
X	8.20%	3.30%	2.50%	1.28%
	Scénario 2		Scénario 2	
	Simulation 1	Simulation 2	Simulation 1	Simulation 2
X	7.97%	5.97%	0.54%	4.45%
Pc	3.57%	5.88%	0.25%	1.79%
SCc	3.61%	4.00%	0.61%	1.45%

Tableau 7.7 : Erreurs relatives au maximum par rapport aux données simulées et par rapport au modèle original associées à la simplification du modèle de Cloutier (2007)

Le tableau 7.8 présente les résultats de la simplification pour les deux scénarios ayant des variables d'intérêt différentes. Le modèle ayant une variable d'intérêt est significativement plus simple que le modèle ayant trois variables d'intérêt, ce qui correspond au résultat attendu.

Caractéristiques	Cloutier 2007	Scénario 1		Scénario 2	
		Modèle simplifié	%	Modèle simplifié	%
Nombre de variables	10	4	-60%	6	-40%
Nombres de paramètres	31	20	-35%	27	-13%
Nombre d'impacts réactionnels	30	12	-60%	21	-30%
Nombre de fonctions unitaires (uniques)	15	8	-47%	13	-13%
Nombre de fonctions unitaires (total)	73	26	-64%	52	-29%

Tableau 7.8 : Caractéristiques de modèles simplifiés pour les deux scénarios de variables d'intérêt

7.2 Chassagnole, 2002

Le modèle de Chassagnole (2002) décrit la dynamique du métabolisme central du carbone chez *Escherichia coli* en réponse à une impulsion de glucose pendant une courte durée de temps (40 secondes). Ce modèle est beaucoup plus détaillé que le modèle de Cloutier au niveau du métabolisme cellulaire. Il comporte 122 paramètres, 86 impacts réactionnels, 39 fonctions unitaires (28 fonctions unitaires uniques) et 26 variables dont une seule est une molécule extracellulaire (glucose), les autres étant des métabolites intracellulaires. Ce modèle est étudié en fonction d'un seul scénario expérimental, c'est-à-dire en batch avec les conditions expérimentales décrites dans l'article de Chassagnole (2002). Ce modèle contient une variable temporelle (t) servant à décrire l'évolution de certaines variables représentées par une dynamique non reliée aux conditions du système. Ces variables sont principalement des variables énergétiques (atp, adp, amp, nadph, nadp, nadh et nad). Leur utilisation par des systèmes non représentés par le schème réactionnel explique que leur comportement ne peut être entièrement décrit par le modèle.

7.2.1 Simplification par critères

Le modèle a été analysé selon les trois critères retenus. Le nombre de fois où chaque critère suggère une simplification est présenté au tableau 7.9.

Critère A	5
Critère D	34
Critère E	16

Tableau 7.9 : Nombre de simplifications suggérées par chacun des critères pour le modèle de Chassagnole (2002)

Il est possible de constater que contrairement au modèle de Cloutier, le critère E permet de simplifier plus de fonctions unitaires que le critère A. Plusieurs des fonctions unitaires ont des valeurs très faibles (valeur moyenne inférieure à 0.01)

combinées à des vitesses de réaction élevées. Les caractéristiques du modèle original et du modèle simplifié sont présentées au tableau 7.10. Contrairement au modèle de Cloutier, aucune variable n'a été retirée.

Caractéristiques	Chassagnole 2002	Modèle simplifié	%
Nombre de variables	26	26	0%
Nombres de paramètres	122	73	-40%
Nombre d'impacts réactionnels	86	52	-40%
Nombre de fonctions unitaires (uniques)	28	12	-58%
Nombre de fonctions unitaires (total)	39	18	-54%

Tableau 7.10 : Caractéristiques du modèle de Chassagnole (2002) original et du modèle simplifié selon l'algorithme de simplification par critère.

L'erreur relative au maximum entre les prédictions du modèle original et du modèle simplifié pour chacune des variables est présentée dans le tableau 7.11. La variable Cfdp (fructose-1,6-biphosphate) est la variable ayant l'erreur la plus élevée. L'évolution de cette variable est présentée à la figure 7.4.

Variabes	Erreur (%)	Variabes	Erreur (%)
Cextglc	0.16	Cxyl5p	0.18
Cg6p	0.53	Csed7p	0.37
Cf6p	0.00	Crib5p	0.63
Cfdp	3.69	Ce4p	0.20
Cgap	0.97	Cglp	0.58
Cdhap	0.11	t	1.66
Cpgp	0.00	Catp	0.00
C3pg	0.44	Cadp	0.00
C2pg	0.00	Camp	0.00
Cpep	0.00	Cnadph	0.00
Cpyr	0.43	Cnadp	0.00
C6pg	0.52	Cnadh	0.00
Cribu5p	0.16	Cnad	0.18
Erreur moyenne (%)	0.44		

Tableau 7.11 : Erreur relative au max associée à la simplification par critère du modèle de Chassagnole (2002)

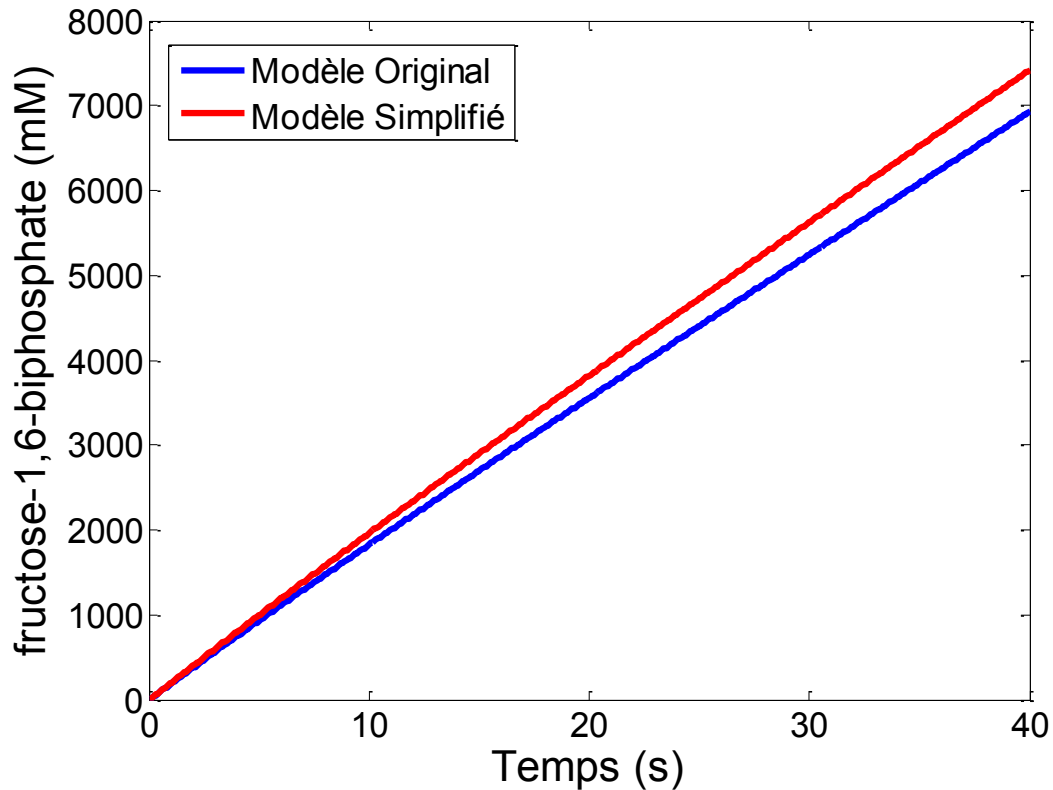


Figure 7.4 : Impact de la simplification par critère sur la variable Cfdp

On constate que la simplification a un impact faible sur les prédictions du modèle, tout en réduisant significativement le nombre de paramètres de celui-ci. Parmi les simplifications intéressantes, notons le fait que seule l'alimentation en glucose est considérée comme significative, tous les autres impacts réactionnels sont retirés du modèle simplifié. Ainsi, le taux d'ajout du glucose sur une courte période est beaucoup plus important que le taux d'adsorption du glucose par les cellules.

7.2.2 Simplification par AIC_m

Lorsque l'algorithme basé sur l'AIC est utilisé sur le modèle de Chassagnole, le modèle simplifié obtenu est composé de 78 paramètres. Le tableau 7.12 présente l'erreur relative au maximum moyenne par rapport aux données et par rapport au modèle original pour chacune des variables.

	Erreur relative au max par rapport aux données	Erreur relative au max par rapport au modèle original
Cextglc	3.33%	0.15%
Cg6p	3.57%	0.52%
Cf6p	2.09%	0.00%
Cfdp	4.10%	2.60%
Cgap	2.26%	0.08%
Cdhap	3.38%	0.10%
Cpgp	2.03%	0.00%
C3pg	3.73%	0.50%
C2pg	2.62%	0.68%
Cpep	2.90%	0.57%
Cpyr	6.13%	5.92%
C6pg	3.72%	0.51%
Cribu5p	3.27%	0.23%
Cxyl5p	3.50%	0.34%
Csed7p	3.81%	0.74%
Crib5p	3.56%	0.40%
Ce4p	3.67%	0.67%
Cglp	2.62%	0.88%
t	3.65%	0.00%
Catp	3.73%	0.00%
Cadp	3.49%	0.00%
Camp	3.55%	0.00%
Cnadph	3.44%	0.00%
Cnadp	3.62%	0.00%
Cnadh	3.66%	0.00%
Cnad	3.61%	0.00%

Tableau 7.12 : Erreur relative au max associée à la simplification AIC_m du modèle de Chassignole (2002)

On constate que l'erreur relative est faible pour toutes les variables à l'exception de la concentration de pyruvate (Cpyr) et de la concentration de fructose-1,6-biphosphate (Cfdp). L'évolution de ces variables est présentée la figure 7.5. On constate que le modèle simplifié a tendance à surestimer ces deux variables.

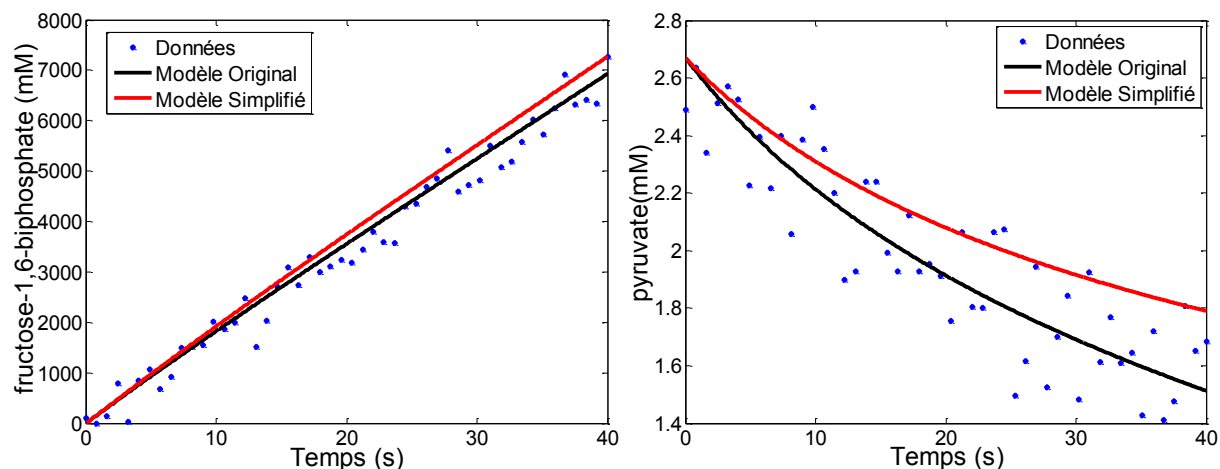


Figure 7.5 : Impact de la simplification par l'AICm sur les variables Cfdp et Cpyr

L'impact de la simplification basée sur le critère d'information d'Akaike modifié est présenté au tableau 7.13. On constate que le modèle obtenu est significativement plus simple (-36 % de paramètres), en particulier en ce qui a trait aux fonctions unitaires (~ -70 %).

Caractéristiques	Chassagnole 2002	Modèle simplifié	%
Nombre de variables	26	26	0%
Nombres de paramètres	122	78	-36%
Nombre d'impacts réactionnels	86	66	-23%
Nombre de fonctions unitaires (uniques)	28	7	-75%
Nombre de fonctions unitaires (total)	39	12	-69%

Tableau 7.13 : Caractéristiques du modèle de Chassagnole (2002) original et du modèle simplifié par l'algorithme de simplification basé sur le critère d'information d'Akaike modifié

7.2.3 Simplification des variables non essentielles

Les variables d'intérêt sélectionnées dans le modèle de Chassagnole (2002) sont les concentrations de métabolites importants du métabolisme central du carbone. Les 8 variables choisies représentent les intermédiaires importants de la voie principale de transformation de glucose en énergie allant de l'absorption du

glucose extracellulaire jusqu'à la production d'acétylcoenzyme A (accoa) (voir figure 7.6) qui est ensuite utilisé dans le cycle de Krebs. Ces variables sont le glucose extracellulaire (Cextglc), le glucose-6-phosphate (Cg6p), le fructose-6-phosphate (Cf6p), le glyceraldehyde-3-phosphate (Cgap) le 3-phosphoglycerate (C3pg), le phosphoenolpyruvate (Cpep), le pyruvate (Cpyr) et l'érythrose-4-phosphate (Ce4p).

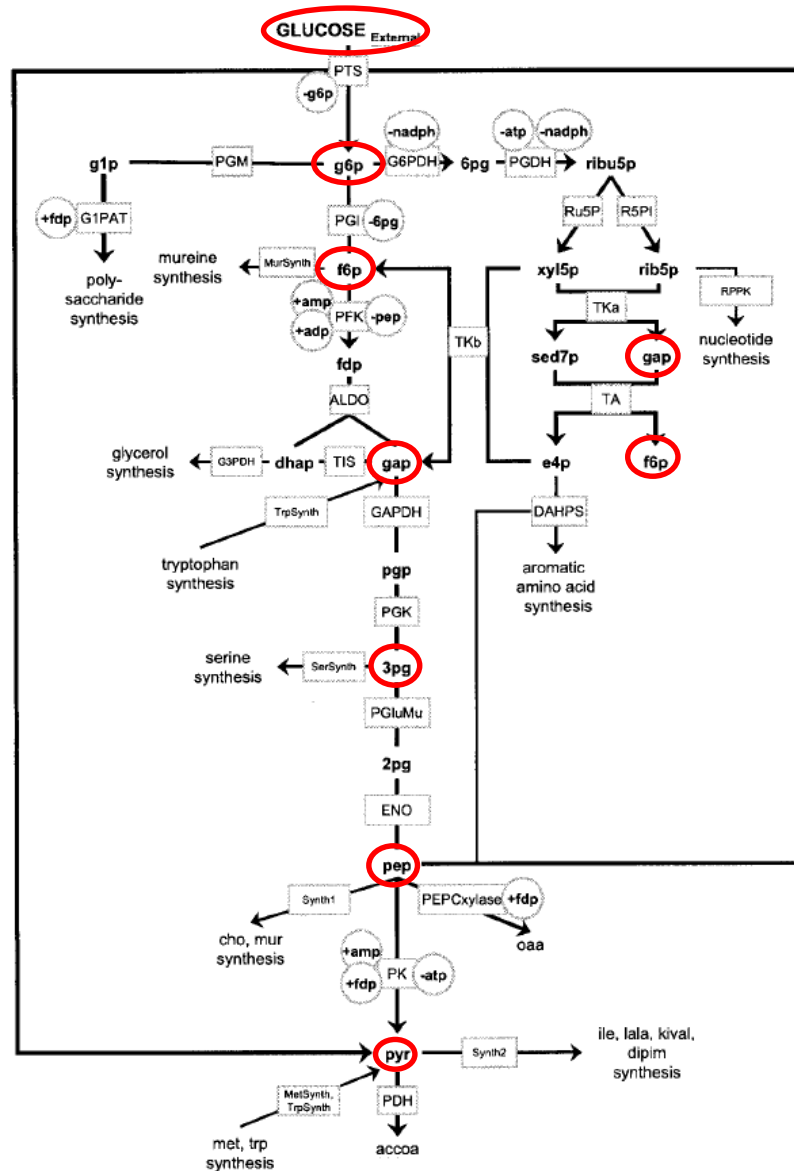


Figure 7.6 : Schéma réactionnel du métabolisme central du carbone (Chassagnole, 2002)

Les conclusions de l'algorithme de simplification des variables non essentielles sont présentées au tableau 7.14. On constate que seules 4 variables sont retirées du

modèle. Parmi ces variables, on retrouve 3 variables liées au transport de l'énergie dans la cellule (Camp, Cnadh et Cnadph), de même qu'une molécule associée à une voie métabolique secondaire (Cg1p).

Variable	Résultat	Variable	Résultat
Cextglc	Var d'intérêt	Cxyl5p	Conservée
Cg6p	Var d'intérêt	Csed7p	Conservée
Cf6p	Var d'intérêt	Crib5p	Conservée
Cfdp	Conservée	Ce4p	Conservée
Cgap	Var d'intérêt	Cg1p	Retirée
Cdhap	Conservée	t	Conservée
Cpgp	Conservée	Catp	Conservée
C3pg	Var d'intérêt	Cadp	Conservée
C2pg	Conservée	Camp	Retirée
Cpep	Var d'intérêt	Cnadph	Retirée
Cpyr	Var d'intérêt	Cnadp	Conservée
C6pg	Conservée	Cnadh	Retirée
Cribu5p	Conservée	Cnad	Conservée

Tableau 7.14 : Variables d'intérêt et résultat de la simplification du modèle de Chassagnole (2002) selon l'algorithme de simplification des variables non essentielles

Le tableau 7.15 présente l'erreur relative au maximum par rapport aux données et par rapport au modèle original des variables d'intérêt du modèle de Chassagnole suite à la simplification par retrait des variables non essentielles. On constate que l'erreur relative au maximum est très faible pour toutes les variables. Les deux variables dont l'erreur est la plus élevée (Cg6p et C3pg) sont présentées à la figure 7.7. On constate que le modèle simplifié parvient à prédire l'évolution des variables d'intérêt de façon très satisfaisante.

Variabes	Erreur relative au max par rapport aux données (%)	Erreur relative au max par rapport au modèle original (%)
Cextglc	3.52%	0%
Cg6p	3.55%	0.37%
Cf6p	2.06%	0%
Cgap	2.04%	0.03%
C3pg	3.67%	0.33%
Cpep	2.17%	0%
Cpyr	2.12%	0%

Tableau 7.15 : Erreur relative au max des variables d'intérêt du modèle de Chassagnole (2002) associées à la simplification des variables non essentielles

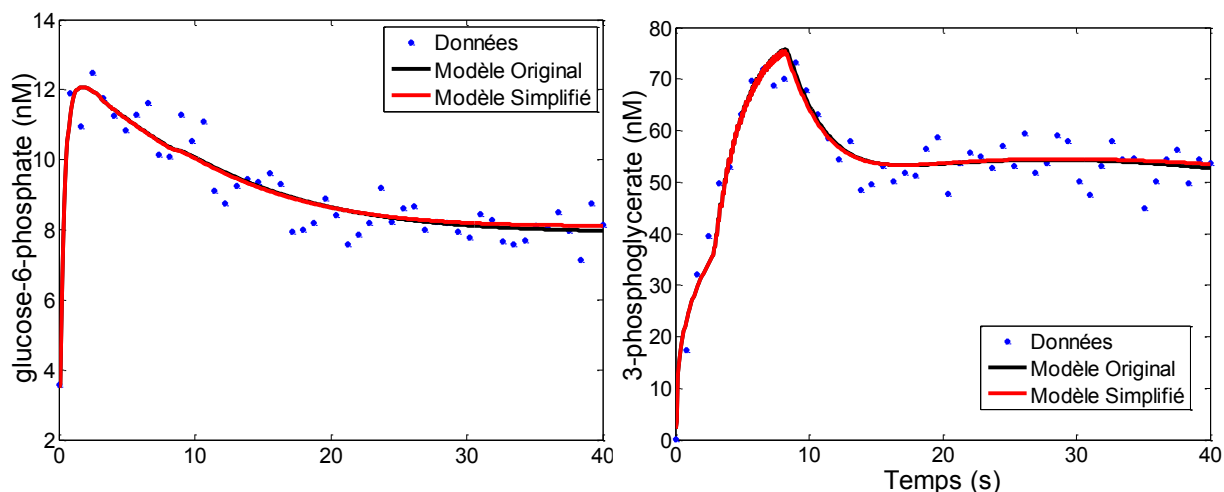


Figure 7.7 : Impact de la simplification des variables non essentielles sur les variables Cg6p et C3pg.

7.3 Iwamoto, 2007

Le modèle proposé par Iwamoto (2007) décrit l'impact de dommages à l'ADN cellulaire sur la phase de reproduction dans laquelle se retrouvent les cellules. Le cycle cellulaire est composé de 5 phases qui sont la mitose (S), la méiose (M), deux interphases (G1 et G2) et un stade de dormance (G0). Les dommages à l'ADN ont un impact important sur le procédé, puisque la phase dans laquelle se retrouve une cellule est liée à des processus cellulaires tels que la prolifération (métabolisme primaire) et la différenciation et l'apoptose (tous deux liés au métabolisme secondaire) qui influencent les voies métaboliques exprimées. Ce modèle comporte une variable temporelle (time) servant à décrire l'évolution de la dégradation de l'ADN. En effet, cette dégradation ne se produit pas instantanément, mais s'accroît dans le temps suite à l'exposition des cellules à des rayons UV. Ce modèle n'est pas affecté par des variables extracellulaires. Cette caractéristique se traduit par des cinétiques comportant beaucoup de réactions (et donc d'impacts réactionnels) et peu de fonctions unitaires. La simulation de ce modèle est réalisée pour le cas où les dommages à l'ADN sont faibles (<3%). Le modèle est composé de 31 variables d'état, de 76 paramètres, de 127 impacts réactionnels et de 6 fonctions unitaires différentes.

7.3.1 Simplification par critères

Le modèle a été analysé selon les trois critères retenus. Le nombre de fois où chaque critère a suggéré une simplification (tableau 7.16) de même que les caractéristiques du modèle original et du modèle simplifié (tableau 7.17) sont présentés ci-dessous :

Critère A	3
Critère D	8
Critère E	3

Tableau 7.16 : Nombre simplifications suggérées par chacun des critères pour le modèle de Iwamoto (2007)

Caractéristiques	Iwamoto 2007	Modèle simplifié	%
Nombre de variables	30	30	0%
Nombres de paramètres	76	61	-20%
Nombre d'impacts réactionnels	127	119	-6%
Nombre de fonctions unitaires (uniques)	6	0	-100%
Nombre de fonctions unitaires (total)	6	0	-100%

Tableau 7.17 : Caractéristiques du modèle de Iwamoto (2007) original et du modèle simplifié par l'algorithme de simplification basé sur les critères

On constate que le nombre simplifications suggérées par les critères est relativement faible. Toutefois, la simplification du modèle entraîne le retrait de l'ensemble des fonctions unitaires. Des tests avec des conditions expérimentales où les dommages à l'ADN sont plus importants permettraient de réévaluer cette conclusion. Pratiquement tous les impacts réactionnels sont nécessaires à l'estimation précise des variables. L'erreur relative au maximum entre le modèle original et le modèle simplifié pour chacune des variables est présentée dans le tableau 7.18.

Variabes	Erreur (%)	Variabes	Erreur (%)
Y0	8.10E-06	Y16	3.82E-04
Y1	1.73E-05	Y17	1.33E-05
Y2	9.48E-07	Y18	0.0019
Y3	1.53E-05	Y19	4.66E-06
Y4	3.06E-07	Y20	1.52E-05
Y5	9.89E-05	Y21	6.00E-05
Y6	1.07E-05	Y22	2.57E-05
Y7	0.85	Y23	0.041
Y8	9.54E-06	Y24	0
Y9	2.11E-05	Y25	0
Y10	1.14E-05	Y26	1.41E-05
Y11	5.47E-05	Y27	0
Y12	2.07E-04	Signal	0
Y13	2.44E-06	Degradation	0
Y14	6.21E-06	temps	0
Y15	1.06E-04		
Erreur moyenne (%)	0.0307		

Tableau 7.18 : Erreur relative au maximum du modèle de Iwamoto (2007) simplifié

On constate que la simplification a un impact très faible (moins de 1%) sur les prédictions du modèle, tout en réduisant significativement le nombre de paramètres (-20 %). La simplification du modèle par l'algorithme de simplification par critère est donc très satisfaisante.

7.3.2 Simplification par AIC_m

La simplification basée sur l' AIC_m , conduit à un modèle simplifié composé de 43 paramètres pour un modèle original composé de 76 paramètres ce qui constitue une simplification importante. Cette méthode retire davantage de paramètres que la méthode de simplification par critères. Le tableau 7.19 présente l'erreur relative au maximum par rapport au modèle original et par rapport aux données pour chacune des variables.

Erreur relative au max par rapport au modèle original				Erreur relative au max par rapport aux données			
Y0	2.01%	Y16	3.25%	Y0	3.43%	Y16	4.75%
Y1	0.85%	Y17	2.74%	Y1	3.02%	Y17	3.33%
Y2	3.41%	Y18	1.22%	Y2	5.07%	Y18	4.77%
Y3	1.61%	Y19	0.37%	Y3	4.40%	Y19	2.00%
Y4	0.15%	Y20	1.00%	Y4	3.08%	Y20	4.50%
Y5	1.17%	Y21	1.71%	Y5	3.34%	Y21	4.14%
Y6	5.71%	Y22	0.96%	Y6	6.35%	Y22	3.14%
Y7	0.00%	Y23	0.09%	Y7	1.34%	Y23	4.40%
Y8	0.47%	Y24	0.04%	Y8	3.26%	Y24	2.51%
Y9	0.30%	Y25	0.00%	Y9	1.96%	Y25	3.06%
Y10	2.23%	Y26	0.85%	Y10	4.69%	Y26	3.62%
Y11	5.05%	Y27	0.00%	Y11	7.21%	Y27	0.00%
Y12	1.17%	Signal	0.00%	Y12	4.25%	Signal	0.00%
Y13	0.90%	Degradation	0.00%	Y13	3.01%	Degradation	4.59%
Y14	0.40%	temps	0.00%	Y14	3.83%	temps	0.00%
Y15	2.66%			Y15	4.16%		

Tableau 7.19 : Erreur relative au maximum par rapport aux données et par rapport au modèle original associée à la simplification par critères du modèle de Iwamoto (2007)

On constate que le modèle simplifié a une erreur supérieure à 5% par rapport au modèle original et par rapport aux données pour deux variables (Y6 : CycE/CDK2 et Y11 : p27/CycDCDK4/6). Ces deux variables sont présentées à la figure 7.8.

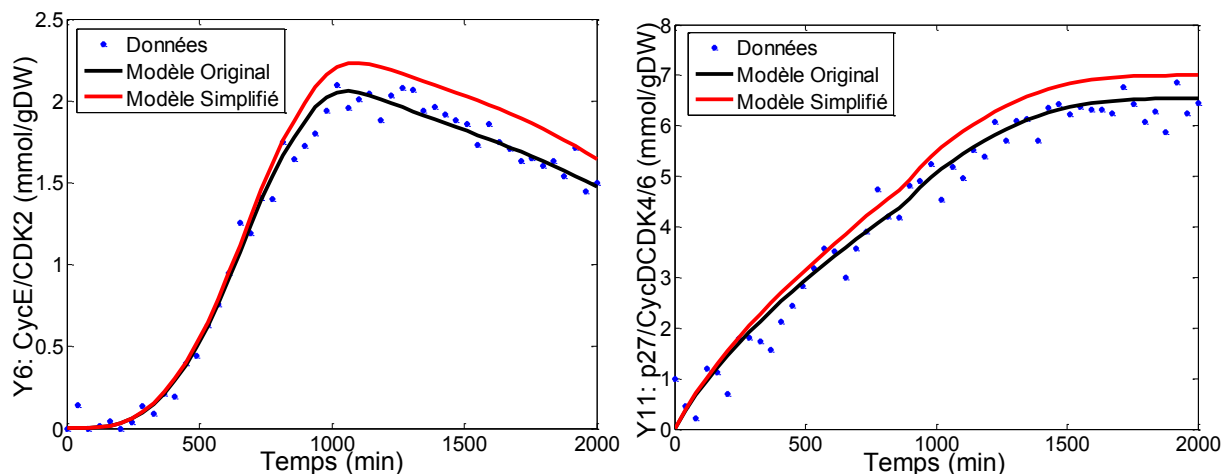


Figure 7.8 : Impact de la simplification par l'AICm sur l'estimation des variables Y2 et Y9.

On constate que les variables CycE/CDK2 et p27/CycDCDK4/6 sont toutes deux surestimées. Le tableau 7.19 montre que contrairement à l'algorithme par critères, deux variables sont retirées (considérées comme constantes). La méthode par AIC_m permet de retirer significativement plus d'impacts réactionnels (46 impacts réactionnels plutôt que 8). Comme pour la méthode par critères, l'ensemble des fonctions unitaires est retiré du modèle.

Caractéristiques	Iwamoto 2007	Modèle simplifié	%
Nombre de variables	30	28	-7%
Nombres de paramètres	76	43	-43%
Nombre d'impacts réactionnels	127	81	-36%
Nombre de fonctions unitaires (uniques)	6	0	-100%
Nombre de fonctions unitaires (total)	6	0	-100%

Tableau 7.20 : Caractéristiques du modèle de Iwamoto (2007) original et du modèle simplifié par l'algorithme de simplification basé l'AIC_m

Cette simplification a un impact plus grand sur les prédictions du modèle que la simplification réalisée par l'application de critères. Toutefois, la simplification obtenue est plus importante et l'erreur d'estimation du modèle simplifié demeure acceptable, considérant les erreurs présentées à la figure 7.8. La simplification du modèle d'Iwamoto (2007) par l'algorithme de simplification par application du critère d'information d'Akaike modifié est donc intéressante.

7.3.3 Simplification des variables non essentielles

Les variables d'intérêt du modèle de Iwamoto (2007) sont les 10 variables suivantes : Y0 (Cycline D), Y1 (Cycline E), Y2 (Cycline A), Y3 (CDK4/6), Y4 (CDK 2), Y10 (expression du gène p27), Y23 (Rb : protéine Retinoblastoma), Y24 (expression du gène p53), Y25 (Mdm2) et dégradation (niveau de dégradation de L'ADN). Ces variables ont été choisies en raison de leur importance au niveau de la régulation du cycle cellulaire. En effet, le gène p53 a pour but de réguler la croissance de tumeurs

par l'activation de processus de réparation de l'ADN ou de l'apoptose cellulaire. La protéine Mdm2 est un régulateur négatif du gène p53. Les cyclines sont des protéines contrôlant la progression de la cellule à travers les différentes phases du cycle cellulaire. Pour ce faire, les cyclines doivent former un complexe avec les enzymes CDK. Le gène p27 produit une protéine empêchant la formation du complexe entre les cyclines et les CDK afin de ralentir la division cellulaire. Les résultats de l'algorithme de simplification des variables non essentielles sont présentés au tableau 7.21. Le tableau 7.21 montre que 9 des 21 variables non essentielles sont retirées du modèle.

Variable	Résultat	Variable	Résultat
Y0	Var d'intérêt	Y16	Retirée
Y1	Var d'intérêt	Y17	Retirée
Y2	Var d'intérêt	Y18	Retirée
Y3	Var d'intérêt	Y19	Conservée
Y4	Var d'intérêt	Y20	Conservée
Y5	Conservée	Y21	Conservée
Y6	Conservée	Y22	Retirée
Y7	Conservée	Y23	Var d'intérêt
Y8	Conservée	Y24	Var d'intérêt
Y9	Conservée	Y25	Var d'intérêt
Y10	Var d'intérêt	Y26	Conservée
Y11	Conservée	Y27	Retirée
Y12	Conservée	Signal	Retirée
Y13	Conservée	Degradation	Var d'intérêt
Y14	Retirée	time	Retirée
Y15	Retirée		

Tableau 7.21 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles

Le tableau 7.22 présente l'impact de la simplification des variables non essentielles sur l'estimation des variables d'intérêt. On constate que les erreurs relatives au maximum sont faibles par rapport aux données (<5%) et par rapport au modèle original (<1%).

	Erreur relative au max par rapport au modèle original (%)	Erreur relative au max par rapport aux données (%)
Y0	0.21%	4.29%
Y1	0.16%	3.06%
Y2	0.05%	3.47%
Y3	0.18%	3.52%
Y4	0.01%	3.25%
Y10	0.55%	4.04%
Y23	0.01%	3.62%
Y24	0.00%	1.78%
Y25	0.00%	3.42%
Degradation	0.00%	3.65%

Tableau 7.22 : Impact de la simplification des variables non essentielles sur l'estimation des variables d'intérêt

7.4 Singh, 2006

Le modèle proposé par Singh (2006) décrit le mécanisme de réponse d'une cellule hépatocyte à la présence d'interleukine-6 (voir figure 7.9). La présence d'interleukine-6 est associée à la réponse d'un organisme à une blessure ou une inflammation. Ce modèle étudie deux voies métaboliques liées à cette réponse : la voie SOCS3 (à gauche de la figure 7.9) et la voie SHP-2 (à droite de la figure 7.9).

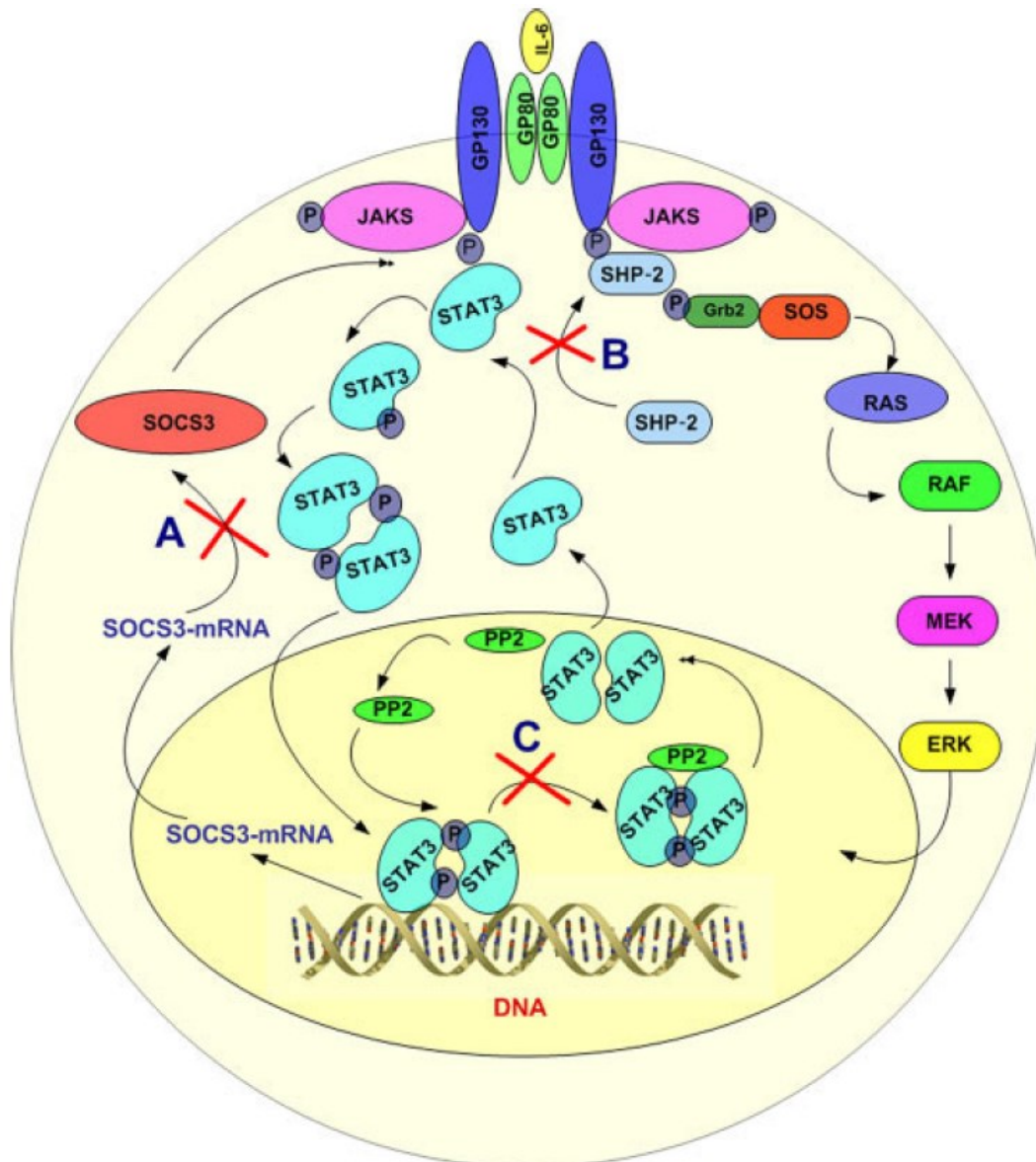


Figure 7.9 : Représentation du schéma réactionnel utilisé dans le modèle de Singh (2006)

Ce modèle ne représente pas une culture cellulaire, mais les concentrations intracellulaires de métabolites dans une cellule idéale. La seule variable extracellulaire étant considérée comme affectant le modèle est la concentration d'interleukine-6 qui est fixée à 10 ng/mL. Ce modèle est composé de 68 variables, de 107 paramètres, de 321 impacts réactionnels et de 2 fonctions unitaires uniques (3 fonctions unitaires totales).

7.4.1 Résultats de la simplification par critères

Le modèle a été analysé selon les trois critères retenus. Le nombre simplifications suggérées par chacun des critères est présenté dans le tableau 7.23. Les caractéristiques du modèle original et du modèle simplifié sont présentées dans le tableau 7.24. Il est possible de constater que seule une des trois fonctions unitaires est conservée. Cette fonction unitaire est associée à la voie de signalisation SOCS3. Les deux fonctions unitaires retirées sont associées à la réaction identifiée par la lettre B liée à la voie de signalisation SHP-2 dans la figure 7.9.

Critère A	0
Critère D	75
Critère E	2

Tableau 7.23 : Nombre de simplifications suggérées par chacun des critères pour le modèle de Singh (2006)

La méthode de simplification par critère a simplifié une variable, et a retiré un petit nombre de paramètres (12%), mais un nombre important d'impacts réactionnels. Ce qui s'explique par le fait que chaque paramètre est impliqué dans un nombre important d'impacts réactionnels.

Caractéristiques	Singh, 2006	Modèle simplifié	%
Nombre de variables	68	67	-1%
Nombres de paramètres	107	94	-12%
Nombre d'impacts réactionnels	321	246	-23%
Nombre de fonctions unitaires (uniques)	2	1	-50%
Nombre de fonctions unitaires (total)	3	1	-67%

Tableau 7.24 : Caractéristiques du modèle de Singh (2006) original et du modèle simplifié par l'algorithme de simplification basé sur les critères

Le tableau 7.25 présente l'erreur relative au maximum des 68 variables composant le modèle de Singh. Deux variables présentent une erreur relative au maximum supérieure ou égale à 2%, x4 (protéine JAK) et x20 (complexe STAT3N*-STAT3N*). L'impact de la simplification sur ces deux variables est présenté à la figure 7.10. On constate que la simplification affecte peu les prédictions du modèle, même sur ces deux variables.

Variabes	Erreur relative au max	Variabes	Erreur relative au max	Variabes	Erreur relative au max
x1	0.08%	x24	0.72%	x47	0.01%
x2	0.98%	x25	1.27%	x48	0.15%
x3	0.03%	x26	0.66%	x49	0.04%
x4	2.00%	x27	1.26%	x50	0.07%
x5	0.16%	x28	0.17%	x51	0.15%
x6	0.08%	x29	0.51%	x52	0.06%
x7	0.72%	x30	0.59%	x53	0.01%
x8	1.30%	x31	0.26%	x54	0.11%
x9	0.13%	x32	0.34%	x55	0.07%
x10	1.99%	x33	0.22%	x56	0.03%
x11	1.28%	x34	0.21%	x57	0.02%
x12	1.58%	x35	0.34%	x58	0.11%
x13	1.18%	x36	0.02%	x59	0.05%
x14	0.33%	x37	0.91%	x60	0.11%
x15	0.56%	x38	0.11%	x61	0.19%
x16	0.84%	x39	0.85%	x62	0.09%
x17	0.04%	x40	1.12%	x63	0.19%
x18	0.60%	x41	1.06%	x64	0.22%
x19	0.38%	x42	0.10%	x65	0.36%
x20	2.84%	x43	0.07%	x66	0.03%
x21	0.73%	x44	0.48%	x67	0.31%
x22	0.11%	x45	0.84%	x68	0.59%
x23	0.06%	x46	0.86%		

Tableau 7.25 : Erreur relative au max associée à la simplification du modèle de Singh (2006)

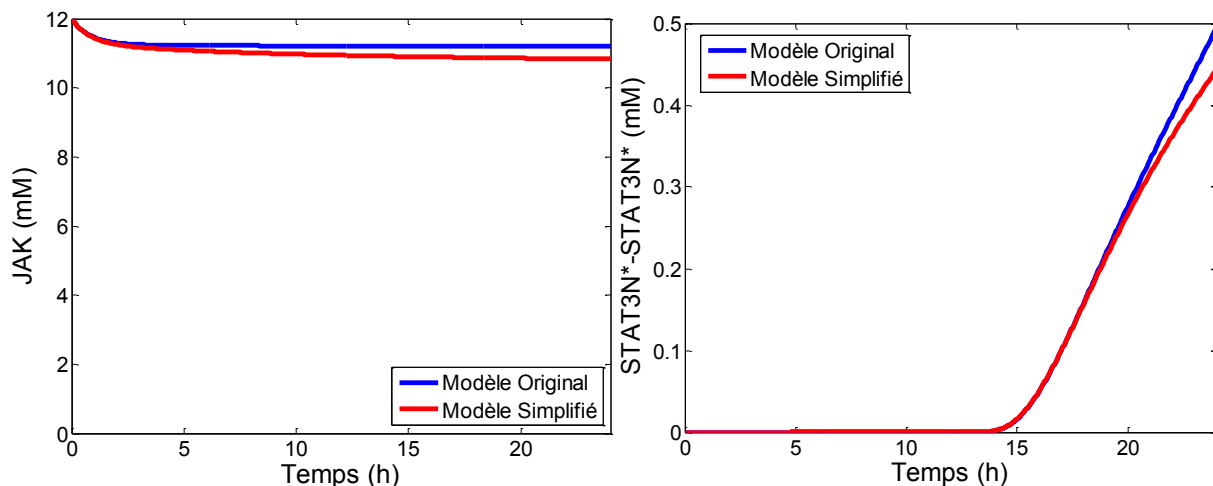


Figure 7.10 : Impact de la simplification par critères sur l'estimation des variables x_4 (YAK) et x_{20} (STAT3*-STST3N*).

Ainsi, la précision du modèle simplifié est excellente, toutefois, le nombre de simplifications proposées par cet algorithme de simplification est assez limité (une variable (x_{23}), 12 % des paramètres, 23% des impacts réactionnels et 2 des 3 fonctions unitaires).

7.4.2 Résultats de la simplification par AIC_m

Le modèle simplifié obtenu est composé de 82 paramètres pour un modèle original de 107 paramètres. Cette méthode permet donc de retirer 12 paramètres de plus que la simplification par critères. Le tableau 7.26 présente l'erreur relative au maximum par rapport aux données et par rapport au modèle original pour chacune des variables.

Erreur relative au max par rapport au modèle original				Erreur relative au max par rapport aux données			
x1	0.06%	x35	0.23%	x1	3.70%	x35	4.00%
x2	0.63%	x36	0.13%	x2	3.47%	x36	3.21%
x3	1.71%	x37	0.14%	x3	3.03%	x37	2.01%
x4	0.11%	x38	0.46%	x4	3.46%	x38	3.95%
x5	5.98%	x39	5.76%	x5	6.51%	x39	6.44%
x6	5.66%	x40	2.00%	x6	5.75%	x40	3.47%
x7	3.34%	x41	0.04%	x7	4.34%	x41	1.90%
x8	0.69%	x42	0.80%	x8	2.85%	x42	2.91%
x9	0.05%	x43	0.32%	x9	3.46%	x43	2.55%
x10	0.32%	x44	0.33%	x10	2.56%	x44	2.73%
x11	0.61%	x45	0.49%	x11	2.81%	x45	2.67%
x12	0.79%	x46	0.95%	x12	2.62%	x46	2.94%
x13	0.20%	x47	0.30%	x13	2.56%	x47	3.06%
x14	0.06%	x48	0.84%	x14	2.67%	x48	3.10%
x15	0.53%	x49	0.49%	x15	3.63%	x49	2.66%
x16	7.30%	x50	0.50%	x16	7.63%	x50	3.38%
x17	0.02%	x51	0.03%	x17	3.43%	x51	2.29%
x18	0.15%	x52	0.24%	x18	2.61%	x52	3.04%
x19	0.10%	x53	0.03%	x19	2.91%	x53	2.10%
x20	0.23%	x54	0.04%	x20	2.63%	x54	2.36%
x21	0.14%	x55	0.45%	x21	2.68%	x55	3.14%
x22	0.60%	x56	0.04%	x22	2.82%	x56	2.74%
x23	0.00%	x57	0.03%	x23	3.45%	x57	3.12%
x24	0.13%	x58	0.05%	x24	2.53%	x58	3.15%
x25	0.14%	x59	0.69%	x25	2.78%	x59	3.44%
x26	0.11%	x60	4.00%	x26	2.69%	x60	4.91%
x27	0.15%	x61	0.05%	x27	2.71%	x61	3.25%
x28	0.07%	x62	0.05%	x28	2.19%	x62	3.18%
x29	0.16%	x63	0.14%	x29	2.30%	x63	3.32%
x30	0.11%	x64	0.05%	x30	2.18%	x64	2.51%
x31	0.01%	x65	0.16%	x31	2.11%	x65	2.35%
x32	1.41%	x66	0.06%	x32	3.25%	x66	3.37%
x33	5.31%	x67	0.12%	x33	6.24%	x67	3.19%
x34	0.12%	x68	0.14%	x34	3.90%	x68	3.12%

Tableau 7.26 : Erreur relative au max associée à la simplification du modèle de Singh (2006) par l'algorithme basé sur l'AIC_m

On constate que bien que l'erreur soit relativement faible (<5%) pour la majorité des variables, certaines variables possèdent une erreur relative au max assez élevée (>5% par rapport aux données et par rapport au modèle original). Ces

variables sont x_5 (gp130-JAK), x_6 (IL6-gp80-gp130-JAK), x_{16} (le complexe (IL6-gp80-gp130-JAK) $_2$ -SHP2), x_{33} (le complexe (IL6-gp80-gp130-JAK) $_2$ -STAT3C - SHP2), x_{39} (le complexe (IL6-gp80-gp130-JAK) $_2$ -SHP2*). Ces variables sont présentées à la figure 7.11. On constate que l'erreur associée à ces variables est relativement élevée.

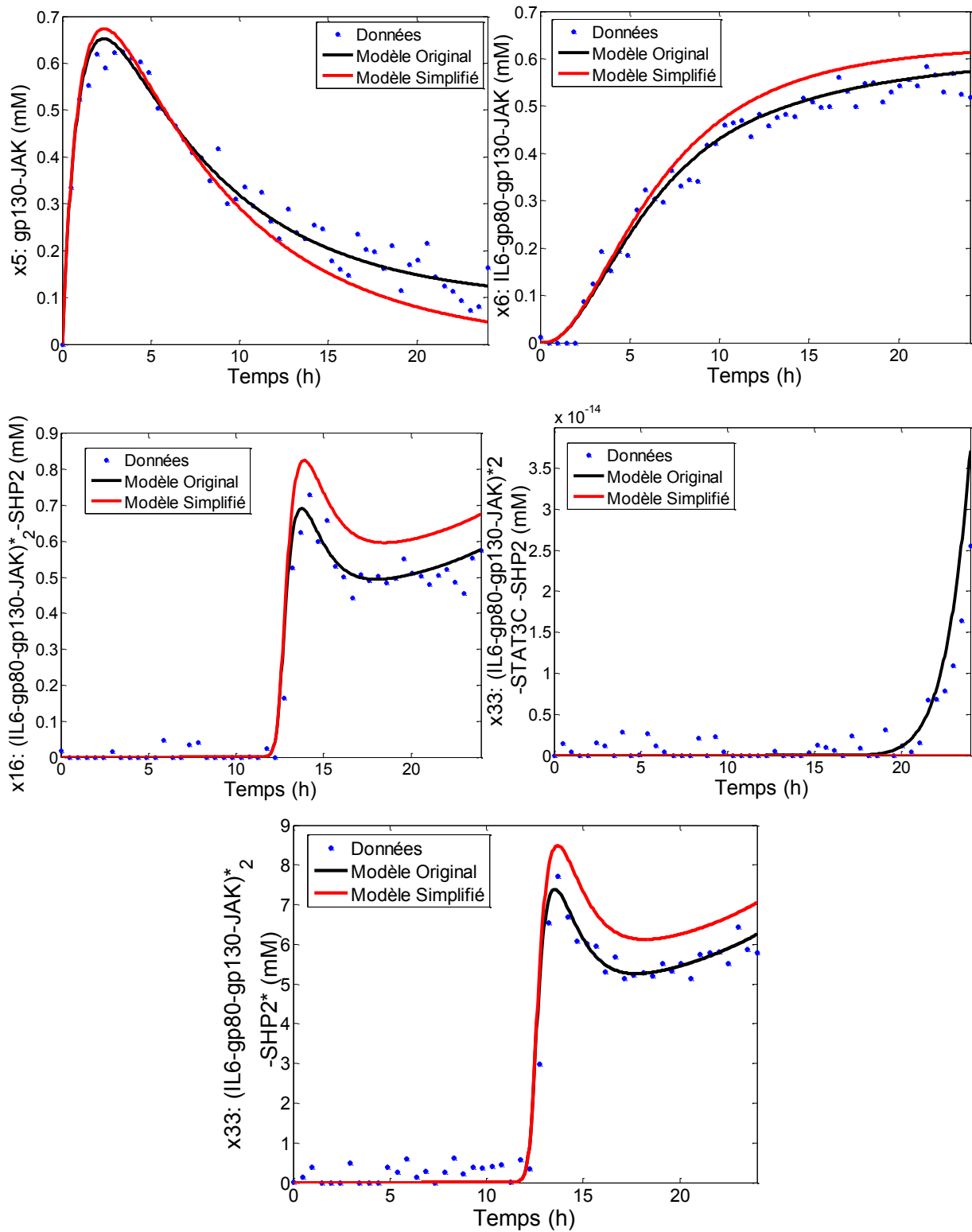


Figure 7.11 : Impact de la simplification par l'AICm sur l'estimation des variables $x_{16}((\text{IL6-gp80-gp130-JAK})^2 \cdot \text{SHP2})$ et $x_{39}((\text{IL6-gp80-gp130-JAK})^2 \cdot \text{SHP2}^*)$

Les caractéristiques du modèle simplifié sont présentées au tableau 7.27. On constate que la conclusion au niveau des fonctions unitaires est la même que celle obtenue par l'algorithme basé sur les critères soit que seule la fonction unitaire associée à la voie de signalisation SOCS3 a un impact significatif sur les prédictions du modèle.

Caractéristiques	Singh, 2006	Modèle simplifié	%
Nombre de variables	68	67	-1%
Nombres de paramètres	107	82	-23%
Nombre d'impacts réactionnels	321	249	-23%
Nombre de fonctions unitaires (uniques)	2	1	-50%
Nombre de fonctions unitaires (total)	3	1	-67%

Tableau 7.27 : Caractéristiques du modèle de Singh (2006) original et du modèle simplifié par l'algorithme de simplification basé sur l'AIC_m

7.4.3 Résultats de la simplification des variables non essentielles

Les variables sélectionnées comme variables d'intérêt dans le modèle de Singh (2006) sont l'ensemble des variables n'étant pas un complexe, c'est-à-dire une molécule formée de deux ou plusieurs molécules appartenant au modèle. Par exemple, le complexe IL6-gp80 est formé de deux molécules IL6 et gp80. Les 16 variables d'intérêt sont donc x1 (gp80) x3 (gp130), x4 (JAK), x9 (STAT3C), x15 (SPH2), x17 (PP1), x23 (PP2), x29 (SOCS3), x34 (Grb2), x35 (SOS), x47 (Raf), x50 (Phosp1), x53 (MEK), x59 (Phosp2), x61 (ERK) et x66 (Phosp3). Les résultats de l'algorithme de simplification sont présentés dans le tableau 7.28. 12 variables non essentielles ont été retirées du modèle (18% des variables du modèle).

Variable	Résultat	Variable	Résultat	Variable	Résultat
x1	Var d'intérêt	x24	Retirée	x47	Var d'intérêt
x2	Conservée	x25	Conservée	x48	Conservée
x3	Var d'intérêt	x26	Conservée	x49	Conservée
x4	Var d'intérêt	x27	Retirée	x50	Var d'intérêt
x5	Conservée	x28	Retirée	x51	Conservée
x6	Conservée	x29	Var d'intérêt	x52	Conservée
x7	Conservée	x30	Retirée	x53	Var d'intérêt
x8	Conservée	x31	Retirée	x54	Conservée
x9	Var d'intérêt	x32	Retirée	x55	Conservée
x10	Conservée	x33	Retirée	x56	Conservée
x11	Conservée	x34	Var d'intérêt	x57	Conservée
x12	Retirée	x35	Var d'intérêt	x58	Conservée
x13	Conservée	x36	Conservée	x59	Var d'intérêt
x14	Conservée	x37	Conservée	x60	Conservée
x15	Var d'intérêt	x38	Conservée	x61	Var d'intérêt
x16	Conservée	x39	Conservée	x62	Conservée
x17	Var d'intérêt	x40	Conservée	x63	Conservée
x18	Retirée	x41	Conservée	x64	Conservée
x19	Retirée	x42	Conservée	x65	Conservée
x20	Conservée	x43	Conservée	x66	Var d'intérêt
x21	Retirée	x44	Conservée	x67	Conservée
x22	Retirée	x45	Conservée	x68	Conservée
x23	Var d'intérêt	x46	Conservée		

Tableau 7.28 : Variables d'intérêt et résultat de la simplification selon l'algorithme de simplification des variables non essentielles

Le tableau 7.29 présente l'impact de la simplification des variables non essentielles sur l'erreur relative au maximum des variables d'intérêt par rapport aux données et par rapport au modèle original. Il est possible de constater que la simplification des variables non essentielles a un impact très faible sur la majorité des variables. Seule la variable x17 qui est une enzyme (PP1 : phosphatase 1) a une erreur supérieure à 1% par rapport au modèle original et de 5% par rapport aux données. Cette variable est présentée à la figure 7.12. Il est possible de constater que l'estimation de cette variable se détériore vers la fin de la simulation. Cette situation est tout de même jugée acceptable.

	Erreur relative au max par rapport au modèle original (%)	Erreur relative au max par rapport aux données (%)
x1	0.000%	3.7%
x3	0.000%	2.8%
x4	0.000%	3.6%
x9	0.048%	3.6%
x15	0.017%	3.4%
x17	2.698%	5.0%
x23	0.034%	3.4%
x29	0.512%	2.5%
x34	0.000%	3.5%
x35	0.000%	4.1%
x47	0.000%	3.1%
x50	0.000%	3.4%
x53	0.000%	2.1%
x59	0.001%	3.1%
x61	0.000%	3.6%
x66	0.000%	3.3%

Tableau 7.29 : Impact de la simplification des variables non essentielles sur l'estimation des variables d'intérêt

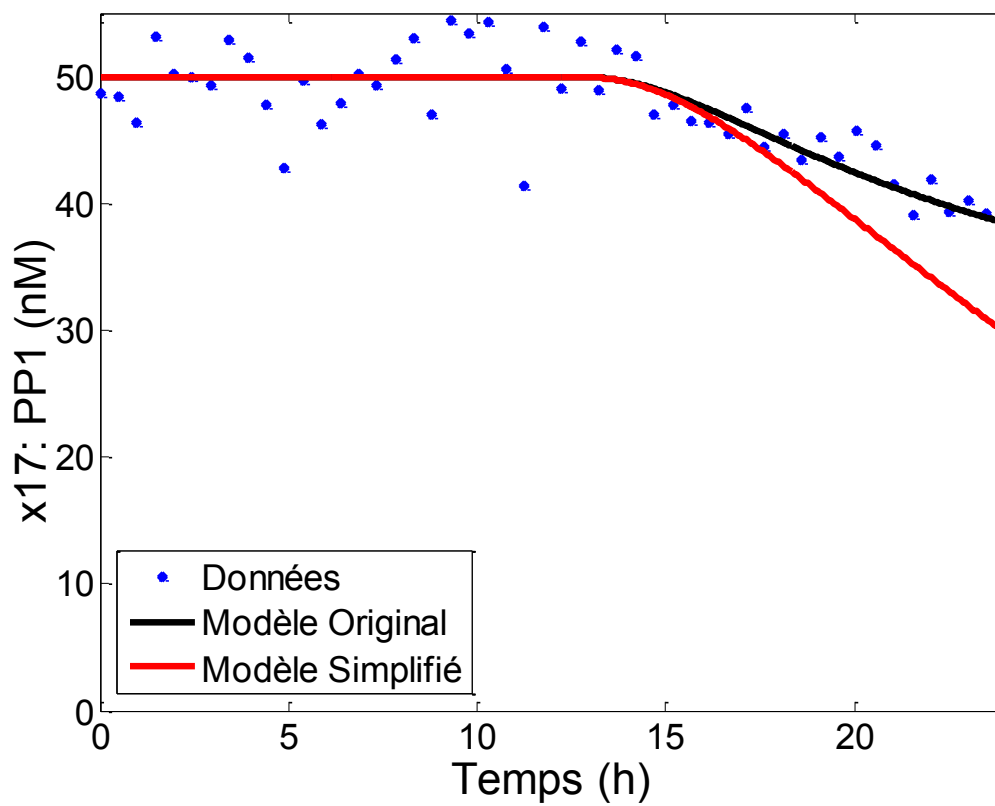


Figure 7.12 : Impact de la simplification par des variables non essentielles sur la variable x17

CHAPITRE 8 DISCUSSION ET APPLICATION COMBINÉE DES MÉTHODES

Ce chapitre analyse les méthodes de simplification à la lumière des résultats obtenus et propose une approche combinant ces méthodes pour obtenir de meilleurs résultats.

8.1 Simplification des paramètres par application de critères sur les structures du modèle

L'algorithme de simplification par application de critères a permis d'obtenir des résultats très satisfaisants au niveau de la simplification des modèles originaux et de la précision des modèles simplifiés. En effet, l'erreur relative au maximum du modèle simplifié par rapport au modèle original n'a pas dépassé 5% pour toutes les variables. L'erreur maximale observée est 3.7% dans le modèle de Chassagnole. De plus, pour la grande majorité des variables, cette erreur était inférieure à 1%. Malgré cela, cet algorithme a permis de réduire significativement la complexité des modèles mathématiques comme démontré au tableau 8.1.

Caractéristiques	Cloutier 2007			Chassagnole 2002		
	Original	Simplifié	%	Original	Simplifié	%
Nb. de variables	10	9	-10%	26	26	0%
Nb. de paramètres	31	20	-35%	122	73	-40%
Nb. d'impacts réactionnels	30	18	-40%	86	52	-40%
Nb. de fonctions unitaires (uniques)	15	9	-40%	28	12	-58%
Nb. de fonctions unitaires (total)	73	24	-67%	39	18	-54%

Caractéristiques	Iwamoto 2007			Singh 2006		
	Original	Simplifié	%	Original	Simplifié	%
Nb. de variables	30	30	0%	68	67	-1%
Nb. de paramètres	76	61	-20%	107	94	-12%
Nb. d'impacts réactionnels	127	119	-6%	321	246	-23%
Nb. de fonctions unitaires (uniques)	6	0	-100%	2	1	-50%
Nb. de fonctions unitaires (total)	6	0	-100%	3	1	-67%

Tableau 8.1 : Simplifications réalisées par l'algorithme basé sur les critères

De plus, cet algorithme est celui où il est le plus facile d'interpréter la signification des simplifications proposées. En effet, chacun des critères conservés cible des phénomènes spécifiques. Le critère A cible les inhibitions dont la zone d'impact se situe à une concentration plus élevée que la plage de données étudiées ainsi que les limitations dont la zone d'impact se situe à une concentration plus faible que la plage de données. Le critère D cible les réactions dont l'impact sur une variable est négligeable (apport ou retrait à la variable non significatif). Le critère E cible les fonctions unitaires ayant une valeur constante.

8.2 Simplification des paramètres du modèle par application du critère d'information d'Akaike modifié

L'algorithme de simplification basé sur le critère d'information d'Akaike modifié a aussi permis de réduire significativement la complexité des modèles étudiés tel que présenté au tableau 8.2. Cependant, l'application de l'algorithme tend à créer une erreur élevée ($> 5\%$) pour un faible nombre de variables dans l'ensemble des modèles. Ainsi, bien qu'il parvienne à simplifier significativement les modèles mathématiques testés, il a été jugé que cet algorithme fournissait de moins bons résultats que la simplification par critères. Toutefois, il est possible d'ajuster la sévérité de cet algorithme en modifiant la valeur du paramètre multiplicatif associé à la complexité du modèle (poids des paramètres). Dans le cadre de ce projet, cette valeur a été fixée à 0.07. Une augmentation de cette valeur aurait pu permettre d'obtenir un algorithme fournissant des modèles plus précis. Dans le cadre d'une application, il est possible de laisser l'utilisateur ajuster cette valeur en fonction de l'importance accordée à la précision du modèle ou à sa simplicité. Le tableau 8.2 présente les simplifications obtenues lors de l'application de cette méthode sur les quatre modèles étudiés.

Caractéristiques	Cloutier 2007			Chassagnole 2002		
	Original	Simplifié	%	Original	Simplifié	%
Nb. de variables	10	9	-10%	26	26	0%
Nb. de paramètres	31	16	-48%	122	78	-36%
Nb. d'impacts réactionnels	30	16	-47%	86	66	-23%
Nb. de fonctions unitaires (uniques)	15	10	-33%	28	7	-75%
Nb. de fonctions unitaires (total)	73	28	-62%	39	12	-69%

Caractéristiques	Iwamoto 2007			Singh 2006		
	Original	Simplifié	%	Original	Simplifié	%
Nb. de variables	30	28	-7%	68	67	-1%
Nb. de paramètres	76	43	-43%	107	82	-23%
Nb. d'impacts réactionnels	127	81	-36%	321	249	-23%
Nb. de fonctions unitaires (uniques)	6	0	-100%	2	1	-50%
Nb. de fonctions unitaires (total)	6	0	-100%	3	1	-67%

Tableau 8.2 : Simplifications réalisées par l'algorithme basé sur l'AIC_m

8.3 Réduction d'ordre du modèle par retrait de variables non essentielles

La réduction d'ordre du modèle pose des problèmes différents que ceux rencontrés pour la réduction des paramètres du modèle. En effet, les simplifications proposées par les deux méthodes précédentes sont entièrement déterminées par la formulation du modèle et les conditions étudiées. La réduction d'ordre du modèle demande de poser un jugement sur l'importance des variables représentées. Selon le choix réalisé, la simplification obtenue présente des résultats différents. Une solution à cette problématique devra être trouvée pour pouvoir mettre en place un algorithme de simplification des modèles automatisé.

Ainsi, l'algorithme de simplification des variables non essentielles est différent des autres algorithmes. En effet, son but n'est pas de diminuer le nombre de paramètres, d'impacts réactionnels ou de fonctions unitaires, mais de réduire le nombre de variables du modèle. Par contre, le retrait de cinétiques du modèle peut entraîner une diminution du nombre de ces éléments. Le nombre de variables simplifié est présenté au tableau 8.3.

	Nombre de Variables		
	Modèle Original	Modèle Simplifié	%
Cloutier 2007 Scénario 1	10	4	-60%
Cloutier 2007 Scénario 2	10	6	-40%
Chassagnole 2002	26	22	-15%
Iwamoto 2007	30	21	-30%
Singh 2006	68	56	-18%

Tableau 8.3 : Simplifications des variables par l'algorithme basé sur la simplification des variables non essentielles.

Cet algorithme a généralement bien fonctionné, permettant d'obtenir un modèle simplifié tout en gardant une erreur relative au maximum des variables d'intérêt faible. Toutefois, l'objectif visant à obtenir un modèle simplifié conservant une signification biologique doit être mis en contexte. Par exemple, dans le cas du scénario 2 du modèle de Cloutier (où les variables d'intérêt sont la biomasse, le phosphate intracellulaire et le glucose intracellulaire), l'algorithme conclut qu'il est possible de retirer le NO_3 et le NH_4 extracellulaire du modèle et de les remplacer par des constantes, mais conserve l'azote intracellulaire. Cela fait en sorte qu'il y a un flux entrant d'azote intracellulaire qui n'est pas expliqué par un transfert de l'extérieur de la cellule. Biologiquement, cela signifie que le flux d'azote intrant n'est pas influencé significativement par la concentration extracellulaire d'azote. Ainsi, lors de l'application de cet algorithme, il est important de porter attention à la distinction entre les variables d'intérêt et les variables non essentielles afin d'obtenir un modèle biologiquement consistant.

8.4 Application combinée des méthodes de simplification

Il est intéressant de réaliser à la fois la réduction d'ordre et la simplification de paramètres sur un modèle, afin d'obtenir le modèle le plus simple possible. Pour ce faire, l'algorithme de simplification des variables non essentielles est appliqué en combinaison avec un des deux algorithmes de simplification des paramètres. Il n'a pas été jugé nécessaire de réaliser à la fois l'algorithme basé sur les critères et l'algorithme basé sur l'AIC_m, puisque ces deux algorithmes visent les mêmes éléments, c'est-à-dire les fonctions unitaires et les impacts réactionnels. La section 8.2 a montré que l'algorithme basé sur les critères permettait d'obtenir de meilleurs résultats que l'algorithme basé sur l'AIC_m, puisqu'il génère moins de variables ayant une erreur relative au maximum par rapport au modèle original supérieure à 5%. Il a donc été choisi pour réaliser la simplification combinée des modèles.

Ainsi, la simplification combinée des modèles consiste en l'application de l'algorithme de réduction d'ordre par le retrait de variables non essentielles, suivi de l'application de l'algorithme de simplification des paramètres par application de critères. Le résultat de la simplification combinée du modèle de Cloutier pour les deux simulations étudiées est décrit pour le scénario 2, soit celui où les variables d'intérêt sont la biomasse et les variables intracellulaires (Pc et SCc). Les résultats de la première étape, l'application de l'algorithme de réduction de l'ordre, sont les mêmes que ceux présentés dans la section 7.1.3 et sont présentés dans le tableau 8.4. Les caractéristiques du modèle simplifié sont présentées au tableau 8.5.

Variable	Résultat	Erreur relative au max par rapport au modèle original	
		Simulation 1	Simulation 2
X	Var d'intérêt	0.57%	4.57%
Pi	Conservée	0.00%	0.00%
Pc	Var d'intérêt	0.26%	1.82%
NO3	Retirée	43.09%	72.13%
NH4	Retirée	90.26%	96.40%
Nc	Conservée	4.48%	31.24%
S	Retirée	11.51%	32.43%
Sc	Conservée	1.15%	0.96%
SCc	Var d'intérêt	0.36%	2.74%
V	Retirée	0.62%	1.48%

Tableau 8.4 : Résultats de l'algorithme de simplification des variables non essentielles sur le modèle de Cloutier

On constate que toutes les variables d'intérêt ont une erreur relative au maximum inférieure à 5%. De plus, parmi les variables non essentielles conservées par l'algorithme, seule Nc dans la simulation 2 a une erreur importante (31%). Cette situation est normale puisque l'algorithme de simplification des variables non essentielles ne prend pas en compte l'erreur sur les variables non essentielles. Le tableau 8.5 présente l'impact de la simplification des variables non essentielles.

Caractéristiques	Cloutier 2007	Modèle simplifié	%
Nombre de variables	10	6	-40%
Nombres de paramètres	31	27	-13%
Nombre d'impacts réactionnels	30	21	-30%
Nombre de fonctions unitaires (uniques)	15	13	-13%
Nombre de fonctions unitaires (total)	73	52	-29%

Tableau 8.5 : Caractéristiques du modèle de Cloutier après application de l'algorithme de simplification des variables non essentielles

Une fois l'algorithme de simplification des variables non essentielles utilisé, l'algorithme de simplification par critère est appliqué sur le modèle résultant. L'erreur relative au maximum est présentée dans le tableau 8.6. Elle montre que l'application

de l'algorithme de simplification par critères n'a pas un impact significatif sur cette erreur.

Variable	Résultat	Simplification des variables non essentielles		Simplification par critères	
		Simulation 1	Simulation 2	Simulation 1	Simulation 2
X	Var d'intérêt	0.57%	4.57%	1.03%	4.91%
Pi	Conservée	0.00%	0.00%	0.05%	0.01%
Pc	Var d'intérêt	0.26%	1.82%	0.80%	2.10%
NO3	Retirée	43.09%	72.13%	43.09%	72.13%
NH4	Retirée	90.26%	96.40%	90.26%	96.40%
Nc	Conservée	4.48%	31.24%	3.52%	29.44%
S	Retirée	11.51%	32.43%	11.51%	32.43%
Sc	Conservée	1.15%	0.96%	1.33%	1.25%
SCc	Var d'intérêt	0.36%	2.74%	3.41%	2.20%
V	Retirée	0.62%	1.48%	0.62%	1.48%

Tableau 8.6 : Résultats de l'algorithme de simplification combiné sur le modèle de Cloutier

Le tableau 8.7 présente l'impact de la simplification par critères sur les différents éléments du modèle de Cloutier par rapport au modèle original.

Caractéristiques	Modèle Original	Simplification des variables non essentielles		Simplification par critères	
		Modèle simplifié	%	Modèle simplifié	%
Nombre de variables	10	6	-40%	6	-40%
Nombres de paramètres	31	27	-13%	16	-48%
Nombre d'impacts réactionnels	30	21	-30%	15	-50%
Nombre de fonctions unitaires (uniques)	15	13	-13%	7	-53%
Nombre de fonctions unitaires (total)	73	26	-64%	23	-67%

Tableau 8.7 : Caractéristiques du modèle de Cloutier après application de l'algorithme de simplification des variables non essentielles

Ainsi, l'application combinée des méthodes de simplification a permis d'obtenir un modèle où près de la moitié des paramètres ont été retirés, et ce, tout en obtenant

une estimation précise des variables d'intérêt. Ainsi, l'application de l'algorithme de simplification des variables non essentielle uniquement a permis le retrait de 13% des paramètres, l'application de l'algorithme de simplification par critère uniquement permet le retrait de 35% des paramètres alors que l'application combinée est de 48%. Ainsi, dans cet exemple, le nombre de paramètres retirés équivaut à la somme des paramètres retirés par chacune des méthodes. Il est important de noter que cela n'est pas toujours le cas. Généralement, un certain nombre de paramètres est retiré à la fois par l'algorithme de simplification des variables non essentielles et par l'algorithme de simplification par critères. Cela fait en sorte que le nombre de paramètres retirés par l'application combinée de ces méthodes est inférieur à la somme des paramètres retirés par chacune de ces méthodes. C'est d'ailleurs le cas avec le modèle de Cloutier dans le scénario 1 (où la biomasse est la seule variable d'intérêt).

Ce projet a permis de définir et de tester trois différents algorithmes servant à simplifier des modèles mathématiques dans des conditions expérimentales déterminées. Ces algorithmes sont la simplification des paramètres par application de critères sur les structures du modèle, la simplification des paramètres du modèle par application du critère d'information d'Akaike modifié et la réduction d'ordre du modèle par retrait de variables non essentielles. Ces tests ont démontré qu'il était possible de réduire la complexité de modèles mathématiques provenant de la littérature de façon importante sans affecter significativement les prédictions de ces modèles dans les conditions expérimentales testées.

Parmi les algorithmes testés, deux s'appliquaient à simplifier les fonctions unitaires et les impacts réactionnels. Le premier utilise des critères tentant d'identifier quels sont les phénomènes affectant l'estimation des variables dans le cadre des conditions expérimentales étudiées. Le deuxième algorithme utilise une forme modifiée du critère d'information d'Akaike pour caractériser l'impact du retrait de ses éléments sur les prédictions du modèle. L'application de ces algorithmes sur des modèles provenant de la littérature a permis de conclure que le premier algorithme permet de garder une meilleure précision sur l'ensemble des variables du modèle. Cet algorithme est donc recommandé pour réaliser la simplification.

L'algorithme de réduction d'ordre vise à retirer les variables qui ne sont pas nécessaires à la compréhension du procédé. Les variables sont divisées en deux catégories : les variables d'intérêt et les variables non essentielles. Par la suite, l'algorithme tente de retirer les variables non essentielles et évalue l'impact de ce retrait sur l'estimation des variables d'intérêt. Cet algorithme a permis d'obtenir un résultat intéressant quant à la diminution du nombre de variables d'état du modèle. Il a été constaté que bien que cet algorithme permette d'obtenir une erreur faible pour les variables d'intérêt, il arrive que certaines des variables non essentielles conservées dans le modèle aient une erreur élevée. Dans le cadre de ce projet, cette

situation a été jugée acceptable. Cependant dans le cas où des erreurs faibles seraient souhaitées pour l'ensemble des variables du modèle, il serait possible d'utiliser une méthode itérative selon laquelle les variables non essentielles identifiées par l'algorithme sont définies comme variable d'intérêt et l'algorithme est réappliqué, et ce jusqu'à ce que le modèle simplifié ne contienne que des variables d'intérêt et des variables non essentielles retirées du modèle.

Enfin, l'algorithme de simplification des variables non essentielles a été combiné avec l'algorithme de simplification par critères afin d'obtenir un modèle le plus simple possible. Le modèle simplifié ainsi obtenu comprenait près de la moitié moins de paramètres que le modèle initial tout en fournissant une estimation des variables d'intérêt présentant une erreur inférieure à 5%. Ainsi, les méthodes de simplification obtenues répondaient aux objectifs fixés soit d'obtenir une simplification importante du modèle mathématique tout en ne modifiant pas de façon significative l'erreur d'estimation du modèle.

9.1 Perspectives

La simplification des modèles mathématiques permet de mieux comprendre les phénomènes en jeu dans les bioprocédés. Cela a pour effet de favoriser la faisabilité de ces bioprocédés de même que de rendre plus faciles leur optimisation et leur contrôle. En effet, lorsque des algorithmes d'optimisation sont appliqués sur des modèles possédant des éléments n'ayant pas été observés (ex. : une limitation sur le sucre), les conditions proposées afin d'améliorer la productivité du procédé (augmenter la quantité de sucre) n'auront peut-être pas l'effet souhaité. Seules des conditions expérimentales supplémentaires permettront de résoudre ce problème. De plus, lorsque la vitesse de réaction est importante (ex. : contrôle de procédé), l'augmentation de la vitesse de calcul par l'utilisation d'un modèle simplifié peut avoir un impact significatif.

Les modèles mathématiques ont été testés sur un nombre plutôt restreint de conditions expérimentales provenant des articles où ils sont présentés. L'analyse des modèles mathématiques bénéficierait d'un plus grand nombre d'expériences afin de valider quels sont les éléments du modèle ayant un impact significatif sur les prédictions du modèle. C'est pourquoi l'étude de données provenant de bioprocédés opérés de façon industrielle permettrait de mettre en valeur l'analyse multilots. En effet, plus les données expérimentales utilisées sont nombreuses et variées, plus il est possible d'observer l'impact de phénomènes présents dans le procédé.

Ces méthodes de simplification ont été développées par analyse de modèles provenant de la littérature. Une des avenues futures envisagées est de combiner un algorithme de simplification (celui par critères) à un algorithme de création de modèles tel un algorithme génétique afin de proposer des hypothèses sur la forme du modèle, d'estimer la valeur des paramètres associés à ces hypothèses puis d'évaluer leur pertinence à l'aide des données expérimentales disponibles. Ainsi, il serait possible à partir de données expérimentales à l'aide d'un processus automatisé de générer des modèles candidats représentant le procédé. Cela permettrait de diminuer de beaucoup l'effort de modélisation requis. Toutefois pour ce faire, une méthode servant à proposer des variables d'état valides devra être mise en place. Une possibilité serait de partir d'un modèle très complexe issu de la biologie des systèmes (plusieurs milliers de molécules et de réactions) comme base de données afin d'ajouter ou de retirer des variables et des réactions au modèle.

ANNEXE A – Modèles Mathématiques

A.1 Modèle de Cloutier (2007)

$$\frac{dX}{dt} = u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X$$

$$\begin{aligned} \frac{dP_i}{dt} = & -V_{max,1,P} \frac{P_i}{K_{1,P} + P_i} \frac{X}{V} - V_{max,2,P} \frac{P_i}{K_{2,P} + P_i} \frac{X}{V} \\ & + \frac{P_i}{V} * DWr * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X \end{aligned}$$

$$\begin{aligned} \frac{dP_C}{dt} = & V_{max,1,P} \frac{P_i}{K_{1,P} + P_i} + V_{max,2,P} \frac{P_i}{K_{2,P} + P_i} \\ & - \frac{1}{Y_X^P} u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} \\ & - P_C u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} \\ & - m_p \end{aligned}$$

$$\begin{aligned} \frac{dNO3}{dt} = & -V_{max,1,NO3} \frac{NO3}{K_{1,NO3} + NO3} \frac{X}{V} - V_{max,2,NO3} \frac{NO3}{K_{2,NO3} + NO3} \frac{X}{V} \\ & + \frac{NO3}{V} DWr * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X \end{aligned}$$

$$\begin{aligned} \frac{dNH4}{dt} = & -V_{max,1,NH4} \frac{NH4}{K_{1,NH4} + NH4} \frac{X}{V} - V_{max,2,NH4} \frac{NH4}{K_{2,NH4} + NH4} \frac{X}{V} \\ & + \frac{NH4}{V} DWr * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X \end{aligned}$$

$$\frac{dN_C}{dt} = V_{max,1,NO3} \frac{NO3}{K_{1,NO3} + NO3} + V_{max,2,NO3} \frac{NO3}{K_{2,NO3} + NO3}$$

$$-\frac{1}{Y_X \bar{N}} u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$- N_C u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$- m_N$$

$$\frac{dS}{dt} = -V_{max,1,S} \frac{S}{K_{1,S} + S} \frac{X}{V}$$

$$+ \frac{S}{V} DWr * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X$$

$$\frac{dS_C}{dt} = V_{max,1,S} \frac{S}{K_{1,S} + S} \llbracket < S_{C,crit} \rrbracket$$

$$- \frac{1}{Y_X \bar{S}} u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$\llbracket S_C > S_{C,min} \rrbracket$$

$$- S_C u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$- m_S$$

$$\frac{dSCC}{dt} = V_{max,1,SC} * \frac{S}{K_{1,S} + S} \quad Si \quad S_C \geq S_{C,crit}$$

$$- SCC * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}]$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$\frac{dSCC}{dt} = -\frac{1}{Y_X \bar{S}} u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] \quad Si \quad S_C \leq S_{C,min}$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$- SCC * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}]$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}}$$

$$\frac{dSCC}{dt} = -SCC * u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] \quad Si \quad S_{C,crit} > S_C$$

$$[1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} \quad > S_{C,min}$$

$$\frac{dV}{dt} = -u_{max} [1 - e^{-K_P(P_C - P_{C,min})}] [1 - e^{-K_S(S_C - S_{C,min})}] [1 - e^{-K_N(N_C - N_{C,min})}] \frac{K_{PP}}{K_{PP} + P_i^{inhi,P}} * X * DWr - evap$$

A.2 Modèle de Chassagnole (2002)

$$\begin{aligned} \frac{dC_{glc}^{extracellular}}{dt} &= D(C_{glc}^{feed} - C_{glc}^{extracellular}) + f_{pulse} - \frac{C_x}{\rho_x} * r_{PTS} \\ \frac{dC_{g6p}}{dt} &= r_{PTS} - r_{PGI} - r_{G6PDH} - r_{PGM} - \mu C_{g6p} \\ \frac{dC_{f6p}}{dt} &= r_{PGI} - r_{PFK} + r_{TKb} + r_{TA} - 2r_{MurSynth} - \mu C_{f6p} \\ \frac{dC_{fdp}}{dt} &= r_{PFK} - r_{ALDO} - \mu C_{fdp} \\ \frac{dC_{gap}}{dt} &= r_{ALDO} + r_{TIS} - r_{GADPDH} + r_{TKa} + r_{TKb} - r_{TA} + r_{TrpSynth} - \mu C_{gap} \\ \frac{dC_{dhap}}{dt} &= r_{ALDO} - r_{TIS} - r_{G3PDH} - \mu C_{dhap} \\ \frac{dC_{pgp}}{dt} &= r_{GADPDH} - r_{PGK} - \mu C_{pgp} \\ \frac{dC_{3pg}}{dt} &= r_{PGK} - r_{PgluMu} - r_{SerSynth} - \mu C_{3pg} \\ \frac{dC_{2pg}}{dt} &= r_{PgluMu} - r_{ENO} - \mu C_{2pg} \\ \frac{dC_{pep}}{dt} &= r_{ENO} - r_{PK} - r_{PTS} - r_{PEPCxylase} - r_{DAHPS} - r_{Synth1}^a - \mu C_{pep} \\ \frac{dC_{pyr}}{dt} &= r_{PK} + r_{PTS} - r_{PDH} - r_{Synth2}^b + r_{MetSynth} + r_{TrpSynth} - \mu C_{pyr} \\ \frac{dC_{6pg}}{dt} &= r_{G6PDH} - r_{PGDH} - \mu C_{6pg} \\ \frac{dC_{ribu5p}}{dt} &= r_{PGDH} - r_{Ru5P} - r_{R5PI} - \mu C_{ribu5p} \\ \frac{dC_{xyl5p}}{dt} &= r_{Ru5P} - r_{TKa} - r_{TKb} - \mu C_{xyl5p} \\ \frac{dC_{sed7p}}{dt} &= r_{TKa} - r_{TA} - \mu C_{sed7p} \end{aligned}$$

$$\frac{dC_{rib5p}}{dt} = r_{R5PI} - r_{TKa} - r_{RPPK} - \mu C_{rib5p}$$

$$\frac{dC_{e4p}}{dt} = r_{TA} - r_{TKb} - r_{DAHPS} - \mu C_{e4p}$$

$$\frac{dC_{g1p}}{dt} = r_{PGM} - r_{GIPAT} - \mu C_{g1p}$$

r_{PTS}

$$= \frac{r_{PTS}^{max} C_{glc}^{extracellular} \frac{C_{pep}}{C_{pyr}}}{\left(K_{PTS,a1} + K_{PTS,a2} \frac{C_{pep}}{C_{pyr}} + K_{PTS,a3} C_{glc}^{extracellular} + C_{glc}^{extracellular} \frac{C_{pep}}{C_{pyr}} \right) \left(1 + \frac{C_{g6p}^{nPTS,g6p}}{K_{PTS,g6p}} \right)}$$

$$r_{PGI} = \frac{r_{PGI}^{max} \left(C_{g6p} - \frac{C_{f6p}}{K_{PGI,eq}} \right)}{K_{PGI,g6p} \left(1 + \frac{C_{f6p}}{K_{PGI,g6p} \left(1 + \frac{C_{6pg}}{K_{PGI,f6p,6pginh}} \right)} + \frac{C_{6pg}}{K_{PGI,g6p,6pginh}} \right) + C_{g6p}}$$

$r_{PFK} =$

$$r_{PFK}^{max} C_{atp} C_{f6p}$$

$$\left(C_{atp} + K_{PFK,atp,s} \left(1 + \frac{C_{adp}}{K_{PFK,adp,c}} \right) \right) \left(C_{f6p} + K_{PFK,f6p,s} \frac{A}{B} \right) \left(1 + \frac{L_{PFK}}{\left(1 + C_{f6p} \frac{B}{K_{PFK,f6p,s} A} \right)^{n_{PFK}}} \right)$$

$$A = 1 + \frac{C_{pep}}{K_{PFK,pep}} + \frac{C_{adp}}{K_{PFK,adp,b}} + \frac{C_{amp}}{K_{PFK,amp,b}}$$

$$B = 1 + \frac{C_{adp}}{K_{PFK,adp,a}} + \frac{C_{amp}}{K_{PFK,amp,a}}$$

$r_{ALDO} =$

$$r_{ALDO}^{max} \left(C_{fdp} - \frac{C_{gap} C_{dhap}}{K_{ALDO,eq}} \right)$$

$$K_{ALDO,fdp} + C_{fdp} + \frac{K_{ALDO,gap} C_{dhap}}{K_{ALDO,eq} V_{ALDO,bif}} + \frac{K_{ALDO,dhap} C_{gap}}{K_{ALDO,eq} V_{ALDO,bif}} + \frac{C_{fdp} C_{gap}}{K_{ALDO,gap,inh}} + \frac{C_{dhap} C_{gap}}{K_{ALDO,eq} V_{ALDO,bif}}$$

$$r_{TIS} = \frac{r_{TIS}^{max} \left(C_{dhap} - \frac{C_{gap}}{K_{TIS,eq}} \right)}{K_{TIS,dhap} \left(1 + \frac{C_{gap}}{K_{TIS,gap}} \right) + C_{dhap}}$$

$$r_{GAPDH} = \frac{r_{GAPDH}^{max} \left(C_{gap} C_{nad} - \frac{C_{pgp} C_{nadh}}{K_{GAPDH,eq}} \right)}{\left(K_{GAPDH,gap} \left(1 + \frac{C_{pgp}}{K_{GAPDH,pgp}} \right) + C_{gap} \right) \left(K_{GAPDH,nadh} \left(1 + \frac{C_{nadh}}{K_{GAPDH,nadh}} \right) + C_{nad} \right)}$$

$$r_{PGK} = \frac{r_{PGK}^{max} \left(C_{adp} C_{pgp} - \frac{C_{atp} C_{3pg}}{K_{GAPDH,eq}} \right)}{\left(K_{PGK,adp} \left(1 + \frac{C_{atp}}{K_{PGK,atp}} \right) + C_{adp} \right) \left(K_{PGK,pgp} \left(1 + \frac{C_{3pg}}{K_{PGK,3pg}} \right) + C_{pgp} \right)}$$

$$r_{PGluMu} = \frac{r_{PGluMu}^{max} \left(C_{3pg} - \frac{C_{2pg}}{K_{PGluMu,eq}} \right)}{K_{PGluMu,3pg} \left(1 + \frac{C_{2pg}}{K_{PGluMu,2pg}} \right) + C_{3pg}}$$

$$r_{ENO} = \frac{r_{ENO}^{max} \left(C_{2pg} - \frac{C_{pep}}{K_{ENO,eq}} \right)}{K_{ENO,2pg} \left(1 + \frac{C_{pep}}{K_{ENO,eq}} \right) + C_{2pg}}$$

$$r_{PK} = \frac{r_{PK}^{max} C_{pep} \left(\frac{C_{pep}}{K_{PK,pep}} + 1 \right)^{(n_{PK}-1)} C_{adp}}{K_{PK,pep} \left(L_{PK} \left(\frac{1 + \frac{C_{atp}}{K_{PK,atp}}}{\frac{C_{fdp}}{K_{PK,fdp}} + \frac{C_{amp}}{K_{PK,amp}} + 1} \right)^{n_{PK}} + \left(\frac{C_{pep}}{K_{PK,pep}} + 1 \right)^{n_{PK}} \right) (C_{adp} + K_{PK,adp})}$$

$$r_{PDH} = \frac{r_{PDH}^{max} C_{pyr}^{n_{PDH}}}{K_{PDH,pyr} + C_{pyr}^{n_{PDH}}}$$

$$r_{PEPCxylase} = \frac{r_{PEPCxylase}^{max} C_{pep} \left(1 + \left(\frac{C_{fdp}}{K_{PEPCxylase,fdp}} \right)^{n_{PEPCxylase,fdp}} \right)}{K_{PEPCxylase,pep} + C_{pep}}$$

$$r_{PGM} = \frac{r_{PGM}^{max} \left(C_{g6p} - \frac{C_{g1p}}{K_{PGM,eq}} \right)}{K_{PGM,g6p} \left(1 + \frac{C_{g1p}}{K_{PGM,g1p}} \right) + C_{g6p}}$$

$$r_{GIPAT} = \frac{r_{GIPAT}^{max} C_{g1p} C_{atp} \left(1 + \left(\frac{C_{fdp}}{K_{GIPAT,fdp}} \right)^{n_{GIPAT}} \right)}{(K_{GIPAT,g1p} + C_{g1p})(K_{GIPAT,atp} + C_{atp})}$$

$$r_{RPPK} = \frac{r_{RPPK}^{max} C_{rib5p}}{K_{RPPK,rib5p} + C_{rib5p}}$$

$$r_{G3PDH} = \frac{r_{G3PDH}^{max} C_{dhap}}{K_{G3PDH,dhap} + C_{dhap}}$$

$$r_{SerSynth} = \frac{r_{SerSynth}^{max} C_{3pg}}{K_{SerSynth,3pg} + C_{3pg}}$$

$$r_{MurSynth} = r_{MurSynth}^{max}$$

$$r_{DAHPS} = \frac{r_{DAHPS}^{max} C_{e4p}^{n_{DAHPS,e4p}} C_{pep}^{n_{DAHPS,pep}}}{\left(K_{DAHPS,e4p} + C_{e4p}^{n_{DAHPS,e4p}} \right) \left(K_{DAHPS,pep} + C_{pep}^{n_{DAHPS,pep}} \right)}$$

$$r_{TrpSynth} = r_{TrpSynth}^{max}$$

$$r_{MetSynth} = r_{MetSynth}^{max}$$

$$r_{G6PDH} =$$

$$\frac{r_{G6PDH}^{max} C_{g6p} C_{nadp}}{(C_{g6p} + K_{G6PDH,g6p}) \left(1 + \frac{C_{nadph}}{K_{G6PDH,nadph,g6pinh}}\right) \left(K_{G6PDH,nadp} \left(1 + \frac{C_{nadph}}{K_{G6PDH,nadph,g6pinh}}\right) + C_{nadp}\right)}$$

r_{PGDH}

$$= \frac{r_{PGDH}^{max} C_{6pg} C_{nadp}}{(C_{6pg} + K_{PGDH,6pg}) \left(C_{nadp} + K_{PGDH,nadp} \left(1 + \frac{C_{nadph}}{K_{PGDH,nadph,inh}}\right) \left(1 + \frac{C_{ATP}}{K_{PGDH,atp,inh}}\right)\right)}$$

$$r_{Ru5p} = r_{Ru5p}^{max} \left(C_{ribu5p} + \frac{C_{xyl5p}}{K_{Ru5p,eq}}\right)$$

$$r_{R5P1} = r_{R5P1}^{max} \left(C_{ribu5p} + \frac{C_{rib5p}}{K_{R5P1,eq}}\right)$$

$$r_{TKa} = r_{TKa}^{max} \left(C_{rib5p} C_{xyl5p} + \frac{C_{sed7p} C_{gap}}{K_{TKa,eq}}\right)$$

$$r_{TKb} = r_{TKb}^{max} \left(C_{xyl5p} C_{e4p} + \frac{C_{f6p} C_{gap}}{K_{TKb,eq}}\right)$$

$$r_{TA} = r_{TA}^{max} \left(C_{gap} C_{sed7p} + \frac{C_{e4p} C_{f6p}}{K_{TA,eq}}\right)$$

$$r_{Synth1} = \frac{r_{Synth1}^{max} C_{pep}}{K_{Synth1,pep} + C_{pep}}$$

$$r_{Synth2} = \frac{r_{Synth2}^{max} C_{pyr}}{K_{Synth2,pyr} + C_{pyr}}$$

$$C_{atp} = 4.27 - 4.163 \frac{t}{0.657 + 1.43t + 0.364t^2}$$

$$C_{adp} = 0.582 + 1.73(2.731^{-0.15t})(0.12t + 0.000214t^3)$$

$$C_{amp} = 0.123 + 7.25 \frac{t}{7.25 + 1.47t + 0.17t^2} + 1.073 \frac{t}{1.29 + 8.05t}$$

$$C_{nadph} = 0.062 + 0.332(2.718^{-0.464t}) * (0.0166t^{1.58} + 0.000166t^{4.73} + 1.13 * 10^{-10}t^{7.89} + 1.36 * 10^{-13}t^{11} + 1.23 * 10^{-16}t^{14.2})$$

$$C_{nadp} = 0.159 + 0.00554 \frac{t}{2.8 + 0.271t + 0.01t^2} + 0.182 \frac{t}{4.81 + 0.526t}$$

$$C_{nadh} = 0.0934 + 0.0011(2.371^{-0.123t})(0.844t + 0.104t^3)$$

$$C_{nad} = 1.314 + 1.314(2.73^{(-0.0435t-0.342)}) - \frac{(t + 7.871)(2.73^{(-0.0218t-0.171)})}{8.481 + t}$$

A.3 Modèle de Iwamoto (2007)

$$\frac{dY_0}{dt} = k_1 + k_4Y_5 - (k_2Y_0 + k_3Y_0Y_3)$$

$$\frac{dY_1}{dt} = k_5Y_{21} + k_8Y_6 - (k_6Y_1 + k_7Y_1Y_4)$$

$$\frac{dY_2}{dt} = k_9Y_{26} + k_{12}Y_8 - (k_{10}Y_2 + k_{11}Y_2Y_4)$$

$$\frac{dY_3}{dt} = k_4Y_5 + k_{13}Y_5 - (k_3Y_0Y_3)$$

$$\frac{dY_4}{dt} = k_8Y_6 + k_{12}Y_8 + k_{14}Y_9 + k_{15}Y_8 + k_{16}Y_6 + k_{17}Y_7^2 - (k_7Y_1Y_4 + k_{11}Y_2Y_4)$$

$$\frac{dY_5}{dt} = k_3Y_0Y_3 + k_{19}Y_{15} + k_{21}Y_{11} - (k_4Y_5 + k_{13}Y_5 + k_{18}Y_5Y_{14} + k_{20}Y_5Y_{10} + k_{44}Y_5Y_{18})$$

$$\frac{dY_6}{dt} = k_7Y_1Y_4 + k_{23}k_7 - (k_8Y_6 + k_{22}Y_6Y_7 + k_{16}Y_6)$$

$$\frac{dY_7}{dt} = k_{22}Y_6Y_7 + k_{25}Y_{12} + k_{27}Y_{16} - (k_{23}Y_7 + k_{24}Y_7Y_{10} + k_{26}Y_7Y_{14} + k_{17}Y_7^2)$$

$$\begin{aligned} \frac{dY_8}{dt} &= k_{11}Y_2Y_4 + k_{29}Y_9 - (k_{12}Y_8 + k_{28}Y_8Y_9 + k_{15}Y_8) \\ \frac{dY_9}{dt} &= k_{28}Y_8Y_9 + k_{31}Y_{13} + k_{33}Y_{17} - (k_{29}Y_9 + k_{30}Y_9Y_{10} + k_{32}Y_9Y_{14} + k_{14}Y_9) \\ \frac{dY_{10}}{dt} &= k_{34} + k_{21}Y_{11} + k_{25}Y_{12} + k_{31}Y_{13} - k_{35}Y_7Y_{10} + k_{36}Y_9Y_{10} + k_{20}Y_5Y_{10} + k_{24}Y_7Y_{10} \\ &\quad + k_{20}Y_9Y_{10} \\ \frac{dY_{11}}{dt} &= k_{20}Y_5Y_{10} - (k_{21}Y_{11}) \\ \frac{dY_{12}}{dt} &= k_{24}Y_7Y_{10} - (k_{25}Y_{12}) \\ \frac{dY_{13}}{dt} &= k_{30}Y_9Y_{10} - (k_{31}Y_{13}) \\ \frac{dY_{14}}{dt} &= k_{37} + k_{38}Y_{24} + k_{19}Y_{15} + k_{27}Y_{16} + k_{33}Y_{17} \\ &\quad - (k_{39}Y_{14} + k_{18}Y_5Y_{14} + k_{26}Y_7Y_{14} + k_{32}Y_9Y_{14}) \\ \frac{dY_{15}}{dt} &= k_{18}Y_5Y_{14} - (k_{19}Y_{15}) \\ \frac{dY_{16}}{dt} &= k_{26}Y_7Y_{14} - (k_{27}Y_{16}) \\ \frac{dY_{17}}{dt} &= k_{32}Y_9Y_{14} - (k_{33}Y_{17}) \\ \frac{dY_{18}}{dt} &= k_{40} + \frac{k_{41}}{1 + k_{42}Y_{23} - (k_{43}Y_{18} + k_{44}Y_5Y_{18})} \\ \frac{dY_{19}}{dt} &= k_{45}Y_{21}Y_{23} - (k_{46}Y_5Y_{19} + k_{47}Y_{11}Y_{19} + k_{48}Y_{15}Y_{19}) \\ \frac{dY_{20}}{dt} &= k_{46}Y_5Y_{19} + k_{47}Y_{11}Y_{19} + k_{48}Y_{15}Y_{19} - (k_{49}Y_7Y_{20} + k_{50}Y_9Y_{20}) \\ \frac{dY_{21}}{dt} &= k_{49}Y_7Y_{20} + k_{50}Y_9Y_{20} + k_{51}Y_{21} + k_{52} - (k_{45}Y_{21}Y_{23} + k_{53}Y_{21} + k_{54}Y_9Y_{21}) \\ \frac{dY_{22}}{dt} &= k_{49}Y_7Y_{20} + k_{50}Y_9Y_{20} - (k_{55}Y_{24}) \\ \frac{dY_{23}}{dt} &= k_{56} + \frac{k_{58}}{1 + k_{59}Y_{18}} + k_{55}Y_{22} - (k_{57}Y_{23} + k_{45}Y_{21}Y_{23}) \\ \frac{dY_{24}}{dt} &= k_{60} + k_{61}signal - (degradationY_{24}Y_{25} + k_{62}Y_{24}) \end{aligned}$$

$$\frac{dY_{25}}{dt} = k_{63} + k_{66} \frac{Y_{27}^9}{k_{65}^9 + Y_{27}^9} - (k_{64}Y_{25})$$

$$\frac{dY_{26}}{dt} = k_{68}Y_{21} - (k_{69}Y_{26})$$

$$\frac{dY_{27}}{dt} = \frac{k_{70}Y_{24} \text{signal}}{1 + k_{71}Y_{24}Y_{25}} - (k_{67}Y_{27})$$

$$\text{signal} = \text{DDS} * \exp(-k_{72} * \text{time})$$

$$\text{Degradation} = k_{74} - k_{73}(\text{signal} - \text{DDS} * \exp(-k_{75} * \text{time}))$$

A.4 Modèle de Singh (2006)

$$\frac{dx_1}{dt} = -k_{f0}x_1u + k_{r0}x_2$$

$$\frac{dx_2}{dt} = k_{f0}x_1u - k_{r0}x_2 + k_{r2}x_6 - k_{f2}x_2x_5$$

$$\frac{dx_3}{dt} = -k_{f1}x_3x_4 + k_{r1}x_5$$

$$\frac{dx_4}{dt} = -k_{f1}x_3x_4 + k_{r1}x_5$$

$$\frac{dx_5}{dt} = k_{f1}x_3x_4 - k_{r1}x_5 - k_{f2}x_2x_5 + k_{r2}x_6$$

$$\frac{dx_6}{dt} = k_{f2}x_2x_5 - k_{r2}x_6 - 2k_{f3}x_6^2 + 2k_{r3}x_7$$

$$\frac{dx_7}{dt} = k_{f3}x_6^2 - k_{r3}x_7 - k_4x_7 + k_{10}x_{16} + k_{10}x_{32}$$

$$\begin{aligned} \frac{dx_8}{dt} = & k_4x_7 - k_{f5}x_8x_9 + k_{r5}x_{11} + k_6x_{11} - k_{f7}x_8x_{10} + k_{r7}x_{12} - k_{f9}x_8x_{15} + k_{r9}x_{16} \\ & - k_{f21}x_{29}x_8 + k_{r21}x_{30} + k_{f37}x_{39} - k_{r37}x_8x_{46} + k_{f39}x_{40} - k_{r30}x_{45}x_8 \\ & + k_{f32}x_{41} - k_{r32}x_{44}x_8 \end{aligned}$$

$$\begin{aligned} \frac{dx_9}{dt} = & -k_{f5}x_8x_9 + k_{r5}x_{11} + k_{12}x_{18} - k_{f13}x_9x_{10} + k_{r13}x_{14} + k_{17}x_{22} - k_{f5}x_9x_{30} + k_{r5}x_{31} \\ & + k_{10}x_{32} \end{aligned}$$

$$\begin{aligned} \frac{dx_{10}}{dt} = & k_6x_{11} - k_{f7}x_8x_{10} + k_{r7}x_{12} - 2k_{f8}x_{10}^2 + 2k_{r8}x_{13} - k_{f11}x_{10}x_{17} + k_{r11}x_{18} \\ & - k_{f13}x_9x_{10} + k_{r13}x_{14} \end{aligned}$$

$$\begin{aligned} \frac{dx_{11}}{dt} &= k_{f5}x_8x_9 - k_{r5}x_{11} - k_6x_{11} \\ \frac{dx_{12}}{dt} &= k_{f7}x_8x_{10} - k_{r7}x_{12} \\ \frac{dx_{13}}{dt} &= k_{f8}x_{10}^2 - k_{r8}x_{13} - k_{f11}x_{13}x_{17} + k_{r11}x_{19} - k_{14}x_{13} \\ \frac{dx_{14}}{dt} &= k_{12}x_{19} + k_{f13}x_9x_{10} - k_{r13}x_{14} \\ \frac{dx_{15}}{dt} &= -k_{f9}x_8x_{15} + k_{r9}x_{16} + k_{10}x_{16} - k_{f9}x_{15}x_{31} + k_{r9}x_{32} + k_{10}x_{32} + \frac{V_m x_{46}}{K_m + x_{46}} \\ \frac{dx_{16}}{dt} &= k_{f9}x_{14}x_8 - k_{r9}x_{16} - k_{10}x_{16} - k_{f34}x_{16} + -k_{r34}x_{39} \\ \frac{dx_{17}}{dt} &= -k_{f11}x_{10}x_{17} + k_{r11}x_{18} + k_{12}x_{18} - k_{f11}x_{13}x_{17} + k_{r11}x_{19} + k_{12}x_{19} \\ \frac{dx_{18}}{dt} &= k_{f11}x_{10}x_{17} - k_{r11}x_{18} - k_{12}x_{18} \\ \frac{dx_{19}}{dt} &= k_{f11}x_{13}x_{17} - k_{r11}x_{19} - k_{12}x_{19} \\ \frac{dx_{20}}{dt} &= k_{14}x_{13} - k_{f60}x_{20} + k_{r60}x_{21}^2 - k_{f15}x_{23}x_{20} + k_{r15}x_{27} \\ \frac{dx_{21}}{dt} &= 2k_{f60}x_{20} - 2k_{r60}x_{21}^2 - k_{f15}x_{23}x_{21} + k_{r15}x_{28} + k_{f61}x_{24} - k_{r61}x_{21}x_{22} \\ \frac{dx_{22}}{dt} &= k_{16}x_{28} + k_{f61}x_{24} - k_{r61}x_{21}x_{22} - k_{17}x_{22} \\ \frac{dx_{23}}{dt} &= -k_{f15}x_{21}x_{23} + k_{r15}x_{28} + k_{16}x_{28} - k_{f15}x_{20}x_{23} + k_{r15}x_{27} + k_{16}x_{27} \\ \frac{dx_{24}}{dt} &= k_{16}x_{27} - k_{f61}x_{24} + k_{r61}x_{21}x_{22} \\ \frac{dx_{25}}{dt} &= \frac{k_{18a}x_{20}}{k_{18b} + x_{20}} - k_{19}x_{25} \\ \frac{dx_{26}}{dt} &= k_{19}x_{25} - k_{22}x_{26} \\ \frac{dx_{27}}{dt} &= k_{f15}x_{23}x_{21} - k_{r15}x_{27} - k_{16}x_{27} \\ \frac{dx_{28}}{dt} &= k_{f15}x_{23}x_{21} - k_{r15}x_{28} - k_{16}x_{28} \end{aligned}$$

$$\frac{dx_{29}}{dt} = k_{20}x_{26} - k_{23}x_{29} - k_{f21}x_{29}x_8 + k_{r21}x_{30} + k_{23}x_{32} + k_{10}x_{32}$$

$$\frac{dx_{30}}{dt} = k_{f21}x_{29}x_8 - k_{r21}x_{30} - k_{f5}x_9x_{30} + k_{r5}x_{31}$$

$$\frac{dx_{31}}{dt} = k_{f5}x_9x_{30} - k_{r5}x_{31} - k_{f9}x_{15}x_{31} + k_{r9}x_{32}$$

$$\frac{dx_{32}}{dt} = k_{f9}x_{15}x_{31} - k_{r9}x_{32} - k_{10}x_{32} - k_{23}x_{32}$$

$$\frac{dx_{33}}{dt} = k_{23}x_{32}$$

$$\frac{dx_{34}}{dt} = -k_{f38}x_{34}x_{46} + k_{r38}x_{45} + k_{f35}x_{38} - k_{r35}x_{34}x_{35} - k_{f24}x_{39}x_{34} + k_{24r}x_{40}$$

$$\frac{dx_{35}}{dt} = -k_{f25}x_{35}x_{40} + k_{r25}x_{41} - k_{f40}x_{35}x_{45} + k_{r40}x_{44} + k_{f35}x_{38} - k_{f35}x_{34}x_{35}$$

$$\frac{dx_{36}}{dt} = -k_{f26}x_{36}x_{41} + k_{r26}x_{42} + k_{f31}x_{43} - k_{r31}x_{36}x_{41}$$

$$\frac{dx_{37}}{dt} = k_{f27}x_{42} - k_{r27}x_{37}x_{41} - k_{f28}x_{37}x_{47} + k_{r28}x_{48}$$

$$\frac{dx_{38}}{dt} = -k_{f35}x_{38} + k_{r35}x_{34}x_{35} - k_{f41}x_{38}x_{39} + k_{r41}x_{41} + k_{f33}x_{44} - k_{r33}x_{38}x_{46}$$

$$\begin{aligned} \frac{dx_{39}}{dt} &= k_{f34}x_{16} - k_{r34}x_{39} - k_{f37}x_{39} + k_{r37}x_8x_{40} - k_{f24}x_{39}x_{34} + k_{r24}x_{40} - k_{f41}x_{39}x_{38} \\ &\quad + k_{r41}x_{41} \end{aligned}$$

$$\frac{dx_{40}}{dt} = k_{f24}x_{39}x_{34} - k_{r24}x_{40} - k_{f25}x_{40}x_{35} + k_{r25}x_{41} - k_{f39}x_{40} + k_{r39}x_{45}x_8$$

$$\begin{aligned} \frac{dx_{41}}{dt} &= k_{f25}x_{35}x_{40} - k_{r25}x_{41} - k_{f32}x_{41} + k_{r32}x_{44}x_8 - k_{f26}x_{41}x_{36} + k_{r26}x_{42} + k_{f31}x_{43} \\ &\quad - k_{f26}x_{41}x_{36} + k_{r26}x_{42} + k_{f31}x_{43} - k_{r31}x_{41}x_{36} + k_{f41}x_{39}x_{38} - k_{r41}x_{41} \\ &\quad - k_{f30}x_{49}x_{41} + k_{r30}x_{43} + k_{f27}x_{42} - k_{r27}x_{41}x_{37} \end{aligned}$$

$$\frac{dx_{42}}{dt} = k_{f26}x_{36}x_{41} - k_{r26}x_{42} - k_{f27}x_{42} + k_{r27}x_{37}x_{41}$$

$$\frac{dx_{43}}{dt} = k_{f30}x_{41}x_{49} - k_{r30}x_{43} - k_{f31}x_{43} + k_{r31}x_{36}x_{41}$$

$$\frac{dx_{44}}{dt} = k_{f32}x_{41} - k_{r32}x_{44}x_8 - k_{f33}x_{44} + k_{r33}x_{38}x_{46} + k_{f40}x_{35}x_{45} - k_{r40}x_{44}$$

$$\frac{dx_{45}}{dt} = k_{f38}x_{34}x_{46} - k_{r38}x_{45} + k_{f39}x_{40} - k_{r39}x_8x_{45} - k_{f40}x_{35}x_{45} + k_{r40}x_{44}$$

$$\frac{dx_{46}}{dt} = k_{f37}x_{39} - k_{r37}x_8x_{45} - \frac{V_m x_{46}}{K_m + x_{46}} - k_{f38}x_{34}x_{46} + k_{r38}x_{45} + k_{f33}x_{44} - k_{r33}x_{38}x_{46}$$

$$\frac{dx_{47}}{dt} = -k_{f28}x_{37}x_{47} + k_{r28}x_{48} + k_{43}x_{52}$$

$$\frac{dx_{48}}{dt} = k_{f28}x_{37}x_{47} - k_{r28}x_{48} - k_{f29}x_{48} + k_{r29}x_{49}x_{51}$$

$$\frac{dx_{49}}{dt} = k_{f29}x_{48} - k_{r29}x_{49}x_{51} - k_{f30}x_{41}x_{49} + k_{r30}x_{43}$$

$$\frac{dx_{50}}{dt} = -k_{f42}x_{50}x_{51} + k_{r42}x_{52} + k_{43}x_{52}$$

$$\frac{dx_{51}}{dt} = k_{f29}x_{48} - k_{r29}x_{49}x_{51} - k_{f42}x_{50}x_{51} + k_{r42}x_{52} - k_{f44}x_{51}x_{53} + k_{r44}x_{54} + k_{45}x_{54}$$

$$+ k_{47}x_{56} - k_{f46}x_{55}x_{51} + k_{r46}x_{56}$$

$$\frac{dx_{52}}{dt} = k_{f42}x_{50}x_{51} - k_{r42}x_{52} - k_{43}x_{52}$$

$$\frac{dx_{53}}{dt} = -k_{f44}x_{51}x_{53} + k_{r44}x_{54} + k_{51}x_{60}$$

$$\frac{dx_{54}}{dt} = k_{f44}x_{51}x_{53} - k_{r44}x_{54} - k_{45}x_{54}$$

$$\frac{dx_{55}}{dt} = k_{45}x_{54} - k_{f46}x_{51}x_{55} + k_{r46}x_{56} + k_{49}x_{58} - k_{f50}x_{55}x_{59} + k_{r50}x_{60}$$

$$\frac{dx_{56}}{dt} = -k_{47}x_{56} + k_{f46}x_{51}x_{55} - k_{r46}x_{56}$$

$$\frac{dx_{57}}{dt} = k_{47}x_{56} - k_{f48}x_{57}x_{59} + k_{r48}x_{58} - k_{f52}x_{57}x_{61} + k_{r52}x_{62} + k_{53}x_{62} - k_{f54}x_{57}x_{63}$$

$$+ k_{r54}x_{64} + k_{55}x_{64}$$

$$\frac{dx_{58}}{dt} = k_{f48}x_{57}x_{59} - k_{r48}x_{58} - k_{49}x_{58}$$

$$\frac{dx_{59}}{dt} = k_{49}x_{58} - k_{f48}x_{57}x_{59} + k_{r48}x_{58} - k_{f50}x_{55}x_{59} + k_{r50}x_{60} + k_{51}x_{60}$$

$$\frac{dx_{60}}{dt} = k_{f50}x_{55}x_{59} - k_{r50}x_{60} - k_{51}x_{60}$$

$$\frac{dx_{61}}{dt} = -k_{f52}x_{57}x_{61} + k_{r52}x_{62} + k_{59}x_{68}$$

$$\frac{dx_{62}}{dt} = k_{f52}x_{57}x_{61} - k_{r52}x_{62} - k_{53}x_{62}$$

$$\frac{dx_{63}}{dt} = k_{53}x_{62} - k_{f54}x_{57}x_{63} + k_{r54}x_{64} + k_{57}x_{67} - k_{f58}x_{63}x_{66} + k_{r58}x_{68}$$

$$\frac{dx_{64}}{dt} = k_{f54}x_{57}x_{63} - k_{r54}x_{64} - k_{55}x_{64}$$

$$\frac{dx_{65}}{dt} = k_{55}x_{64} - k_{f56}x_{65}x_{66} + k_{r56}x_{67}$$

$$\frac{dx_{66}}{dt} = -k_{f56}x_{65}x_{66} + k_{r56}x_{67} + k_{57}x_{67} - k_{f58}x_{63}x_{66} + k_{r58}x_{68} + k_{59}x_{68}$$

$$\frac{dx_{67}}{dt} = k_{f56}x_{65}x_{66} - k_{r56}x_{67} - k_{57}x_{67}$$

$$\frac{dx_{68}}{dt} = k_{f58}x_{63}x_{66} - k_{r58}x_{68} - k_{59}x_{68}$$

LISTE DES RÉFÉRENCES

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transaction on Automatic Control, vol. 19, no 6, p. 716-723

Albert R. (2005) Scale-free networks in cell biology. Journal of Cell Science, vol. 119, p.4947-4957

Andrews JF. (1968) A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrate. Biotechnology Bioengineering, vol.10, no 6, p.707-724

Antoulas A et Sorensen D. (2001) Approximation of large-scale dynamical systems : an overview. International Journal of Applied Mathematics and Computer Science, vol. 11, no. 5, p.1093-1121

Archambault E. (2004) Towards a Canadian R&D strategy for bioproducts and bioprocesses. National Research Council of Canada, 141 p.

Ashlock D. (2006) Evolutionary Computation for Modeling and Optimization. Springer-Verlag New York. 572 p. ISBN : 978-0-387-22196-0

Berg JM, Tymoczko JL, Stryer L. (2002) Biochemistry. 5th edition. New York. www.ncbi.nlm.nih.gov/books/NBK22439/figure/A1942/?report=objectonly

Boghigian B, Seth G, Kiss R, Pfeifer B. (2010) Metabolic flux analysis and pharmaceutical production. Metabolic Engineering, vol. 12, p. 81-95.

Borodina I, Nielsen J. (2005) From genomes to in silico cells via metabolic networks. Current Opinion in Biotechnology, vol. 16, p. 350-355.

Briggs, G.E.; Haldane, J.B.S. (1925). A note on the kinematics of enzyme action. Biochemical Journal, vol. 19, no 2, p. 338-339.

Burnham, K. P., et Anderson D. R. (2002). Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Busetto AG, Hauser A, Krummenacher G, Sunnåker M, Dimopoulos S, Ong CS, Stelling J et Buhmann JM. (2013) Near-optimal experimental design for model selection in systems biology. Bioinformatics, vol. 29, no 20, p. 2625-2632.

Cardin-Bernier G. (2011) Observateur pour le suivi en temps réel de cultures cellulaires végétales. Mémoire de maîtrise. Université de Sherbrooke

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, vol. 79, no. 1, p. 53–73.

Chen KC, Wang TY, Tseng HH, Huang CY et Kao CY. (2005) A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics*, vol. 21, no. 12, p. 2993-2890.

Chis OT, Banga JR, Balsa-Canto E. (2011) Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. *PLoS ONE*, vol. 6, no 11.

Cintolesi A, Clomburg J, Rigou V, Zygorakis K, Gonzalez R. (2012) Quantitative Analysis of the Fermentative Metabolism of Glycerol in *Escherichia coli*. *Biotechnology and Bioengineering*, vol. 109, no. 1, p. 187-198.

Cloutier, M, Bouchard-Marchand E, Perrier M et Jolicoeur M. (2007) A predictive nutritional model for plant cells and hairy roots. *Biotechnology and Bioengineering*, vol. 99, no. 1, p. 189-200

Conolly RB, Kimbell JS, Janszen D, Schlosser PM, Kalisak D, Preston J et Miller FJ. (2003) Biologically motivated computational modeling of formaldehyde carcinogenicity in the F344 rat. *Toxicological Sciences*, vol. 75, p.432-447.

Contois DE. (1959) Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. *Journal of General Microbiology*, vol. 21, p. 40–50.

Csikasz-Nagy A. (2009) Computational systems biology of the cell cycle. *Briefings In Bioinformatics*, vol. 12, no. 4, p. 424-434.

DiStefano J. (2015) *Dynamic systems biology modeling and simulation*. Academic Press, 1st Edition, 884p. ISBN : 9780124104112

Eberlein R.L. (2006) Simplification and understanding of models. *System Dynamics Review*, vol. 5, no 1, p. 51-68.

Golub G et Van Loan F. (1996) *Matrix Computation*. Johns Hopkins University Press, 3th edition, 728 p. ISBN : 9780801854149

Gurgecin S et Antoulas A. (2004) A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, vol. 77, no 8, p. 748-766.

Haag J, Vande Wouer A et Bogaerts P. (2005) Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description. *Mathematical Biosciences*, vol. 293, p. 25-49.

Hammersley JM et Handscomb DC. (1964) Monte Carlo Methods. Monographs on Applied Probability and Statistics. 178 p. ISBN : 978-94-009-5821-0

Hasenauer J, Löhning M, Khammash M et Allgöwer F. (2012) Dynamical optimization using reduced order models : A method to guarantee performance. Journal of process control, vol. 22, p.1490-1501

Hurvich, CM, Tsai CL. (1989) Regression and time series model selection in small samples. Biometrika, vol. 76, p. 297-307.

Iwamoto K, Tashima Y, Hamada H, Eguchi Y et Okamoto M. (2008) Mathematical modeling and sensitivity analysis of G1/S phase in the cell cycle including the DNA-damage signal transduction pathway. BioSystems, vol. 94, p. 109-117

Jolicoeur M, Bouchard-Marchand E, Bécard G et Perrier M. (2003) Regulation of micorrhizal symbiosis: Development of a structured nutritional dual model. Ecological Modeling, vol. 163, no. 1-2, p. 247-267

Joshi A, Palsson O. (1989-1) Metabolic dynamics in the human red cell: Part I—A comprehensive kinetic model. Journal of Theoretical Biology, vol. 141, no 4, p. 518-528.

Joshi A, Palsson O. (1989-2) Metabolic dynamics in the human red cell: Part II—Interactions with the environment. Journal of Theoretical Biology, vol. 141, no 4, p. 529-545.

Joshi A, Palsson O. (1990-1) Metabolic dynamics in the human red cell: Part III—Metabolic reaction rates. Journal of Theoretical Biology, vol. 142, no 1, p. 41-68.

Joshi A, Palsson O. (1990-2) Metabolic dynamics in the human red cell: Part IV—Data prediction and some model computations. Journal of Theoretical Biology, vol. 142, no 1, p. 69-85.

Junker BH, Wang HY (2006) Bioprocess Monitoring and Computer Control: Key Roots of the Current PAT Initiative. Biotechnology and Bioengineering, vol. 95, no 2, p. 226-261.

Khavarpour M, Najafpour GD, Ghoreyshy A. (2011) Biodesulfurization of Natural Gas : Growth Kietic Evaluation. Middle-East Journal of Scientific Research, vol. 7, no 1, p. 22-29.

Kiran A, Jana AK. (2009) Control of continuous fed-batch fermentation process using neural network based model predictive controller. Bioprocess and biosystems engineering, vol. 32, p.801–808.

Leduc M, Tikhomiroff C, Cloutier M, Perrier M et Jolicoeur M. (2006) Development of a kinetic metabolic model: application to *Catharanthus roseus* hairy root. *Bioprocess and biosystems engineering*, vol. 28, p. 295-313

Leis J, Kramer A. (1988) The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *Transactions on Mathematical Software*, vol. 14, no 1, p. 45-60

Little MP, Heidenreich WF, Li G (2010) Parameter Identifiability and Redundancy: Theoretical Considerations. *PLoS ONE* 5(1): e8915. doi:10.1371/journal.pone.0008915

Mailier J, Vande Wouwer J. (2010) A fast and systematic procedure to develop dynamic models of bioprocesses - application to microalgae cultures. *Brazilian Journal of Chemical Engineering*, vol. 27 no. 3, p. 381-390.

Madsen, K. Nielsen HB et Tingleff O. (2004) Methods for non-linear least squares problems. *Informatics and Mathematical Modelling*, Technical University of Denmark, 60 p.

Marquardt D. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, vol. 11, no 2. p. 431-441.

Michaelis, L, Menten, M.L. (1913), Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, vol. 49, p. 333–369

Monod J. (1949) The growth of bacterial culture. *Annual review of microbiology*. vol. 3, p. 371–394

Moser H. (1958) The dynamics of bacterial populations maintained in the chemostat. Publication 614 Carnegie Institution of Washington, Publication 614.

Nelder J, Mead R. (1965) A simplex method for function minimization. *SIAM journal of scientific computing*, vol. 29, p. 2507-2530.

Nelson MI, Sidhu HS. (2009) Analysis of a chemostat model with variable yield coefficient: Tessier kinetics. *Journal of Mathematical Chemistry*, vol. 46, no 2, p. 302-321.

Niu H, Amribit Z, Fickers P, Tan W, Bogaerts P. (2013) Metabolic pathway analysis and reduction for mammalian cell cultures—Towards macroscopic modeling. *Chemical Engineering Science*, vol. 102, p. 461-473.

Oates C, Dondelinger F, Bayani N, Gray JW et Mukherjee S. (2014) Causal network inference using biochemical kinetics. *Bioinformatics*. Vol. 30, no. 17, p. i468–i474.

-
- Okino S, Mavrovouniotis ML. (1998) Simplification of mathematical models of chemical reaction systems. *Chemical Review*, vol. 98, no 2, p. 392-408
- Pargett M, Umulis D. (2013) Quantitative model analysis with diverse biological data: Applications in developmental pattern formation. *Methods*, vol. 62, p. 56-67
- Penzl T. (2006) Algorithms for model reduction of large dynamical systems. *Linear Algebra and its Applications*, vol. 415, p. 322-343
- Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J et d'Alché-Buc F. (2003) Gene networks inference using dynamic Bayesian networks. *Bioinformatics*. vol. 19, suppl. 2, ii. 138-148.
- Peskov K, Mogilecskaya E et Demin O. (2012) Kinetic modelling of central carbon metabolism in *Escherichia coli*. *Federation of European Biochemical Societies Journal*, vol.279, p.3374-3385.
- Posada D, Buckley TR. (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *System Biology*, vol. 53, no 5, p. 793-808.
- Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U et Timmer J. (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, vol. 25, no 15, p. 1923-1929.
- Rizzi M, Theobald U, Mailinger W, Baltes M, et Reuss M. (1997-1) In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: I. Experimental Observations. *Biotechnology and Bioengineering*, vol. 55, no 2, p. 305-316.
- Rizzi M, Theobald U, Mailinger W, Baltes M, et Reuss M. (1997-2) In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: II. Mathematical Model. *Biotechnology and Bioengineering*, vol. 55, no 4, p. 592-608.
- Seinfeld J, Pandis S (2006) *Atmospheric Chemistry and Physics : From Air Pollution to Climate Change*. Wiley, 2nd edition, 1232 pages.
- Schwarz G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, vol. 6, no 2, p. 461-464.
- Shi Z, Li Y, Liu Z, Mi J et Wang R. (2012) Theoretical analysis of Fas ligand-induced apoptosis with an ordinary differential equation model. *Molecular Informatics*, vol. 31, p. 793-807.
- Singh A, Jayaraman A et Hahn J. (2006) Modeling Regulatory Mechanisms in IL-6 Signal Transduction in Hepatocytes. *Biotechnology and Bioengineering*, vol. 95, no 5, p. 850-862

Sirois J. (2000) Modélisation et optimisation en continu d'un procédé de culture de cellules végétales. Thèse de Doctorat. Université de Montréal. 542 p.

Tessier G. (1936) Les lois quantitatives de la croissance. Annales de Physiologie et de Physiochimie Biologique, vol. 12, p. 527–573

United Nations (1992) Convention on biological diversity. Article 2 : « Biotechnology » <http://www.cbd.int/doc/legal/cbd-en.pdf>

Vysheirsky V et Girolami M. (2008) Bayesian ranking of biochemical system models. Bioinformatics. vol. 24, no 6, p. 833-839.

Waine JRT. (2001) Development of Mammalian Cell Bioreactor Processes. ESACT Proceedings, vol. 1, p. 491-493

Walsh R, Martin E et Darvesh S. (2007) A versatile equation to describe reversible enzyme inhibition and activation kinetics: Modeling β -galactosidase and butyrylcholinesterase. Biochimica et Biophysica Acta, vol. 1770, p. 733-746.

Wand MP et Jones MC. (1995) Kernel Smoothing. Chapman and Hall/CRC, 224 p.

Williams T, Poolman M, Howden A, Schwarzlander M, Fell D, Ratcliffe G et Sweetlove L. (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. Plant Physiology, vol. 154, p. 311-323.

Wu S, McLean K, Harris T et McAuley K. (2011) Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. International journal of advanced mechatronic systems, vol. 3, no 3, p.188-197.

Zhou K, Doyle J et Glover K. (1996) Robust and optimal control. Prentice Hall, 596p.