

École polytechnique de Louvain

Modélisation de la production des molécules organoleptiques du vin

Auteur: **Sacha GEORGES**

Promoteurs: **Denis DOCHAIN, Robert DAVID**

Lecteur: **Patricia LUIS ALCONERO**

Année académique 2020–2021

Master [120] : ingénieur civil en chimie et science des matériaux

Résumé

La modélisation de bioprocédés est une discipline qui a vu le jour il y a quelques décennies et qui capte toujours plus d'intérêt. L'envie de comprendre et faire des prévisions sur des systèmes biologiques s'accompagne d'intérêts industriels d'optimiser des productions de bioréacteurs.

La fermentation, et plus précisément ici la vinification, est un bioprocédé dont la mise en pratique ne date pas d'hier. Cependant, les mécanismes de production des molécules organoleptiques (qui affectent le goût, le parfum,...) ne sont toujours pas parfaitement connus. Plusieurs modélisations de la fermentation ont déjà vu le jour, mais ne prennent pas en compte ces molécules aromatiques d'intérêt.

L'objectif de ce travail est d'étendre un modèle préconçu à ces molécules d'intérêt, à l'aide de nouvelles données expérimentales. La première étape sera donc de réimplémenter le modèle et d'entreprendre une nouvelle identification des paramètres. Ce modèle sera également étendu afin de prendre en compte un comportement nouveau de certaines variables, propre aux nouvelles données expérimentales. Un nouveau paramètre ainsi qu'une nouvelle variable seront introduits à cet effet.

Par après, le but est d'intégrer au modèle les molécules organoleptiques choisies, afin de mieux comprendre les mécanismes inhérents à leur synthèse ainsi que pour tenter de former, à long terme, un modèle prédictif des propriétés organoleptiques du vin. Cependant, seule une analyse préliminaire de leur cinétique sera menée car la description des cinétiques principales a pris plus de temps que prévu.

Remerciements

Le contenu de ce mémoire est le fruit de longues heures de labeur, cependant il n'aurait pu voir le jour sans l'aide de nombreuses personnes que je n'oserais manquer de remercier.

Tout d'abord, j'aimerais remercier Robert David et Denis Dochain pour m'avoir guidé tout au long de l'année avec beaucoup de patience. Sans leur disponibilité, leurs conseils avisés et leur support indéfectible, je n'aurais jamais su venir à bout de ce mémoire.

J'aimerais remercier l'équipe de l'INRAE Montpellier, tout d'abord pour m'avoir donné accès à leurs données et ensuite pour leur excellente humeur malgré une distanciation pénible.

J'aimerais ensuite remercier Patricia Luis Alconero pour avoir accepté de lire ce mémoire et faire partie du jury, j'espère que la lecture de mon travail se révélera intéressante.

Enfin, comment ne pas remercier les gens de mon entourage, ma famille, pour m'avoir supporté durant ces longs mois de travail et m'avoir aidé, à leur manière, à persévérer.

Table des matières

Résumé	i
Remerciements	ii
Introduction	1
1 Modélisation de bioprocédés : Etat de l'art	3
1.1 Introduction	3
1.2 Description du procédé de fermentation	4
1.3 Types de bioréacteurs et modes opératoires	5
1.3.1 Introduction	5
1.3.2 Réacteur batch	7
1.3.3 Mode fed-batch	7
1.3.4 Mode continu (ou chémostat)	8
1.4 Modéliser des bilans de matière	8
1.4.1 Introduction	8
1.4.2 Principes généraux	8

1.4.3	Exemple et représentation matricielle	9
1.5	Modéliser la cinétique	10
1.5.1	Expressions cinétiques communes	11
2	Identification de paramètres	13
2.1	Introduction	13
2.2	Identifiabilité structurelle	14
2.3	Identifiabilité pratique	15
2.3.1	Cadre théorique	15
2.3.2	Intervalle de confiance des estimations	16
2.3.3	Fonctions de sensibilité	17
2.4	Algorithmes d'estimation	19
2.4.1	Minimisation globale	19
2.4.2	Minimisation locale	20
2.4.3	Conclusion sur les stratégies de minimisation	22
2.5	Validation du modèle	22
3	Modélisation du processus de vinification	24
3.1	Introduction	24
3.2	Modèles antérieurs	24
3.3	Modèle actuel	25
3.3.1	Equations du modèle	26
3.3.2	Hypothèses du modèle	27
3.3.3	Identification des paramètres du modèle	29
4	Résultats	31
4.1	Introduction	31

4.2	Données disponibles	32
4.3	Implémentation et extension du modèle	33
4.3.1	Dérivée seconde du CO_2	33
4.3.2	δ et $N_{x,\delta}$	33
4.4	Identification des paramètres	36
4.4.1	Méthodologie	36
4.4.2	Paramètres μ_{max} , K_x , k_1 et δ	37
4.4.3	Paramètres η_{max} et K_{tr}	41
4.4.4	Paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ	42
4.5	Résumé de la calibration	45
4.6	Validation	55
4.7	Molécules aromatiques	55
4.7.1	Description des courbes de production des arômes	56
4.7.2	Modélisation de la production des arômes	60
5	Conclusion	61
	Table des Figures	vi
	Bibliographie	x
A	ANNEXE I : Exemples de données	xvi
B	ANNEXE II : Comparatif ancien modèle/nouvelle implémentation	xix
C	ANNEXE III : Quelques fonctions de sensibilité non-normées	xx
D	ANNEXE IV : Coefficients des polynômes des lois d'évolutions des paramètres	xxv

Introduction

Les produits et services dépendant de bioprocédés peuvent grossièrement être regroupés en 3 catégories [1] :

- Les biopharmaceutiques (vaccins, protéines thérapeutiques,...)
- La gestion environnementale (agriculture, traitement des eaux usées,...)
- Denrées fermentées et certains produits chimiques

De manière générale, l'intérêt porté à l'automatisation de bioprocédés n'a fait qu'augmenter ces dernières années mais rencontre plusieurs obstacles inhérents à la nature des systèmes biologiques. Tout d'abord, l'implication d'organismes vivants, au comportement complexe donc, rend difficile la dérivation d'un modèle de prédiction ainsi que sa validation expérimentale. La deuxième complication provient de l'absence fréquente de capteurs permettant de mesurer l'état interne du système biologique, les variables mesurées sont donc limitées.

Le procédé étudié dans ce mémoire, la vinification, est un procédé de fermentation du moût de raisin en vin. Sa phase principale est la cuvaison, durant laquelle le moût subit une fermentation alcoolique produisant le vin. De nombreuses opérations en aval ou en amont de cette vinification permettent de créer une impressionnante gamme de vins différents. Cependant, prédire les propriétés organoleptiques (qui affectent son goût, son odeur,...) de sa production est un véritable défi. En effet, de nombreux facteurs affectent les processus de conversions biologiques (T° , pH, concentrations des différents substrats, ventilation, agitation,...). De plus, bien que la fermentation consiste majoritairement en la conversion d'hexoses en éthanol et dioxyde de carbone, de nombreux produits dérivés (alcools supérieurs, esters, composés sulfurés,...) sont formés et impactent grandement les propriétés organo-

leptiques du vin, bien qu'ils ne représentent au total moins de 4% de la production de la levure.

L'objectif premier de ce travail est d'enrichir la compréhension des cinétiques de fermentation via le développement de modèles mathématiques. Plus particulièrement, la finalité visée est d'étendre la portée des modèles actuels à certaines molécules organoleptiques d'intérêt.

Le premier chapitre de ce mémoire présente l'état de l'art de la modélisation de bioprocédés. Le second se concentre sur la théorie derrière l'identification de paramètres. Le troisième présente la modélisation du processus de vinification. Le quatrième chapitre porte sur les résultats du travail mené cette année. Finalement, le mémoire finit par une conclusion mettant en exergue les points abordés lors de ce travail.

Modélisation de bioprocédés : Etat de l'art

1.1 Introduction

La modélisation de procédés biochimiques n'est pas triviale. Le comportement des composants biologiques est complexe et parfois méconnu. La construction de modèles prédictifs nécessite donc de simplifier ces comportements sous certaines hypothèses faisant ressortir les caractéristiques majeures du procédé. Les hypothèses formulées ainsi doivent être identifiées dès le début du processus de modélisation et si possible, vérifiées.

Le modèle érigé doit refléter son objectif. En effet, un modèle peut être développé avec différentes intentions, qui affectent la manière dont il sera structuré. Par exemple, si l'on cherche à expliquer des hétérogénéités spatiales dans un réacteur, il sera nécessaire de dériver un modèle spatial régi par des équations aux dérivées partielles.

De plus, la complexité du modèle doit être adéquate par rapport au nombre de variables mesurées. Ensuite, la précision de l'identification des paramètres du modèle obtenue avec la simulation dépendra de la qualité informative des données. Les paramètres peuvent également être fortement corrélés et difficilement identifiables. Dans certains cas, le modèle se révélera structurellement non identifiable et plusieurs sets de valeurs différentes des paramètres engendreront des résultats équivalents.

Finalement, parce qu'il est fondé sur des hypothèses et des lois empiriques, la validation du modèle est cruciale. Pour démontrer son efficacité, il est primordial de prouver la concordance entre les prédictions du modèle et des données expérimentales (différentes de celles avec lesquelles le modèle a été construit).

1.2 Description du procédé de fermentation

Le contenu des sections 1.2, 1.3 et 1.4 est fortement inspiré de [2].

La réaction biologique commune qui fait l'objet de ce mémoire est la fermentation microbienne. Cette réaction consiste en la croissance d'une population de micro-organismes (bactéries, levures,...) dans des conditions favorables (température, pH, ventilation,...) via la consommation de certains éléments nutritifs afin de profiter du résultat de leur activité métabolique, généralement la transformation de substances en produits d'intérêt.

Les composantes principales de cette réaction sont :

- Des substrats, notés S_i , nécessaires à la croissance des micro-organismes ou précurseurs de composants dont la production est désirée. Ce sont généralement des sources de carbone (sucre, éthanol,...), d'azote (NO_3 , NH_4 ,...), de phosphore (PO_4 ,...) ou d'oxygène pour les réaction aérobies.
- Une ou plusieurs biomasses, notées X_j ou X dans le cas d'une population microbienne unique.
- Des produits finis, notés P_k . Une grande variété de produits peuvent être la finalité d'un bioprocédé, que ce soit des denrées alimentaires (vin, fromage, bière, levure de boulangerie,...), des produits chimiques (solvants,...), pharmaceutiques (hormones, vitamines,...) ou des dérivés énergétiques (biogaz,...).

Dans ces notations, les indices i , j et k permettent de distinguer des substrats, biomasses ou produits différents. Bien qu'ils n'apparaissent que rarement dans les modèles, des sels minéraux et vitamines sont également ajoutés car essentiels à la croissance de la biomasse.

Ces fermentations peuvent avoir différents objectifs :

1. Croissance microbienne : L'objectif est simplement d'accroître la population, c'est le cas des cultures de levures par exemple.
2. Synthèse de produit : L'objectif est de synthétiser un produit de l'activité métabolique du micro-organisme (pénicilline, éthanol,...)
3. Consommation de substrat : L'objectif est la dégradation d'un substrat, comme dans le traitement des eaux usées.
4. Etude phénoménologique : L'objectif est l'étude du micro-organisme lui-même.

Au niveau industriel, les cultures sont généralement menées avec une seule espèce de micro-organisme. Il est possible de cultiver plusieurs espèces à la fois

tant qu'elles n'entrent pas trop en compétition.

1.3 Types de bioréacteurs et modes opératoires

1.3.1 Introduction

D'un point de vue mathématique, l'une des premières choses à considérer est l'état homogène ou non au sein du réacteur. En effet, les réacteurs tels que les lits fixes ou fluidisés nécessitent de prendre en compte des gradients spatiaux de concentration via des équations aux dérivées partielles. Les bioréacteurs parfaitement mélangés cependant peuvent être modélisés via des équations différentielles ordinaires. Les modèles présentés dans ce mémoire se concentreront sur cette seconde catégorie.

Il y a de nombreuses manières de procéder à une fermentation. Cependant, dans le cadre de la modélisation mathématique, des configurations standards sont généralement utilisées afin de simplifier les équations. Voici le modèle standard d'un réacteur à milieu mélangé homogène (voir [Figure 1.1](#)) [3] :

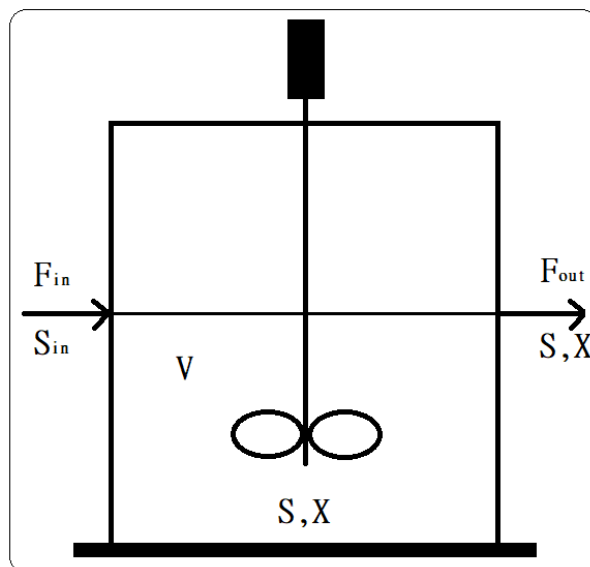


FIGURE 1.1 – Schéma de réacteur à milieu mélangé homogène.

L'accumulation nette de biomasse au sein du réacteur est alors :

$$\frac{d(VX)}{dt} = \mu VX - F_{out}X \quad (1.1)$$

L'accumulation nette de substrat dans le réacteur :

$$\frac{d(VS)}{dt} = -k_1\mu VX + F_{in}S_{in} - F_{out}S \quad (1.2)$$

La variation du volume :

$$\frac{dV}{dt} = F_{in} - F_{out} \quad (1.3)$$

avec X et S respectivement la concentration de la biomasse et du substrat dans le réacteur et le flux de sortie, S_{in} la concentration du substrat dans le flux d'entrée, F_{in} et F_{out} les débits d'entrée et de sortie, μ la vitesse de croissance spécifique de la biomasse, k_1 le coefficient de rendement de la consommation de substrat par la biomasse et V le volume du réacteur.

Dans ce modèle, la seule hypothèse posée est que les termes de croissance de la biomasse (μX) et de la consommation de substrat ($k_1\mu X$) sont proportionnels à la concentration de la biomasse X. Cette hypothèse a été vérifiée de nombreuses fois et est communément acceptée.

Suivant ce modèle, voici les types de bioréacteurs les plus communément rencontrés :

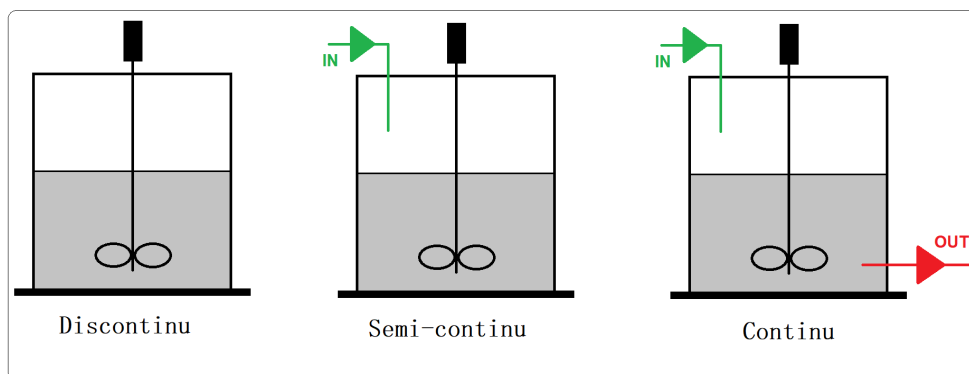


FIGURE 1.2 – Schéma des types de bioréacteurs les plus fréquents.

1.3.2 Réacteur batch

C'est le système le plus simple. La biomasse et tous les nutriments nécessaires sont introduits au temps initial. Le volume au sein du réacteur est constant :

$$F_{in} = F_{out} = 0 \quad (1.4)$$

L'opérateur peut uniquement modifier les variables environnementales (pH, température,...). Sa facilité d'application rend ce système très populaire dans l'industrie, où l'on récupère simplement biomasse et produits d'intérêts à la fin du procédé. L'apport unique d'une large quantité de substrat en début de procédé peut cependant entraîner des phénomènes d'inhibition, ce qui peut limiter les taux de croissance obtenus. En recherche, ce système clos a l'avantage de presque garantir la non-contamination du réacteur.

La production de *Lactobacillus bulgaricus* (ingrédient du yogourt) par exemple peut se faire en batch [4].

1.3.3 Mode fed-batch

Ce cas diffère du précédent par l'addition d'un flux d'entrée, le volume au sein du réacteur évolue de V_0 à l'instant initial jusqu'à V_f au temps final :

$$F_{in} \neq F_{out} = 0 \quad (1.5)$$

La gestion de l'alimentation du bioréacteur tout au long de la réaction permet un meilleur contrôle sur la croissance de la biomasse. Il est par exemple possible d'éviter les effets d'inhibition associés au mode opératoire précédent et ainsi atteindre des taux de croissance proches de l'optimum. Cette méthode est particulièrement adaptée pour une culture intermittente, où des prélèvements sont effectués à la fin de chaque fermentation.

Par exemple, le fed-batch est utilisé pour optimiser la production de lipides par des micro-algues, ce qui améliore la viabilité économique de ce type de procédé [5] ou encore pour la culture de *Saccharomyces cerevisiae*, une levure fréquemment rencontrée dans l'industrie alimentaire [6].

1.3.4 Mode continu (ou chémostat)

Ce procédé fonctionne en continu avec des débits d'entrée et de sortie fixés afin de conserver un état stationnaire (dans le cas de systèmes liquides, ce qui est généralement le cas des bioprocédés) :

$$F_{in} = F_{out} = F \quad (1.6)$$

$$\frac{dV}{dt} = 0 \quad (1.7)$$

Un premier avantage de ce mode opératoire est la possibilité de productions importantes dans un espace limité étant donné que l'alimentation en nutriments et le prélèvement de la production se fait en continu. Il est également possible de travailler en boucle fermée et de commander l'effluent en modifiant les débits et compositions des flux d'entrée.

Ce système est communément utilisé par exemple pour le traitement des eaux usées [7], bien que dans ce cas on ne peut pas modifier les débits et les concentrations d'entrée.

1.4 Modéliser des bilans de matière

1.4.1 Introduction

Le comportement des micro-organismes est difficile à représenter à l'aide d'équations mathématiques, ce qui s'oppose à la modélisation. Cependant, il est possible de rattacher ces systèmes à des lois physiques absolues telles que la conservation de la masse. Les modèles peuvent donc être construits de manière relativement robuste en se basant sur l'analyse de balances de flux de matière [8].

1.4.2 Principes généraux

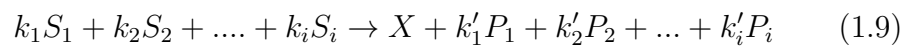
Le schéma réactionnel est alors décrit via un ensemble de réactions chimiques et biologiques. Pour la notation de ces équations, on peut adopter un formalisme proche de la chimie [3] et définir celles-ci comme la simple transformation de réactifs A et B en produit C :



Ce schéma réactionnel est en fait une écriture synthétique de toutes les réactions supposées déterminer la dynamique du système et il se base donc sur les hypothèses créant la fondation du modèle. De nombreux composants intervenant dans les réactions n'y sont donc pas représentés et cela simplifie la démarche.

Les composantes décrites dans le schéma réactionnel synthétisé sont donc cruciales à la réaction et au mieux, faciles à observer. On y retrouvera les transformations principales des bioprocédés étudiés, en voici quelques exemples :

- Croissance de la biomasse et biosynthèse suite à la consommation de nutriments :



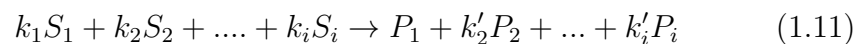
où S_i indique les substrats, X la biomasse, P_i les produits synthétisés, k_i et k'_i les taux de consommation et production correspondants et μ la vitesse de croissance de la biomasse.

- Mortalité de la biomasse :

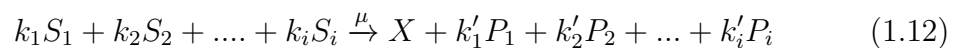


avec X_d la biomasse morte et k_d le taux de mortalité.

- Synthèse de produit via un métabolisme secondaire :



Il faut encore compléter ces équations en ajoutant la vitesse de réaction. Par exemple, pour la croissance de la biomasse on écrit



avec μ la vitesse spécifique de croissance de la biomasse.

Finalement, en combinant ce schéma réactionnel avec les équations différentielles ordinaires décrivant la dynamique du bioréacteur choisi, on obtient un modèle de bilan de matière.

1.4.3 Exemple et représentation matricielle

La croissance de levure *Saccharomyces cerevisiae* consomme du glucose S et produit de l'éthanol P, selon ce schéma réactionnel :



Dans un réacteur mélangé à milieu homogène, cette réaction peut donc être modélisée via les équations différentielles ordinaires :

$$\begin{cases} \frac{d(VX)}{dt} = \mu V - F_{out}X \\ \frac{d(VS)}{dt} = -k_1\mu V + F_{in}S_{in} - F_{out}S \\ \frac{d(VP)}{dt} = k_2\mu V - F_{out}P \\ \frac{dV}{dt} = F_{in} - F_{out} \end{cases} \quad (1.14)$$

Ce modèle peut ensuite être étoffé par l'ajout de termes de maintenance, de mortalité, des réactions intermédiaires, etc... Il est également possible de synthétiser ces systèmes d'équations via une représentation matricielle équivalente :

$$\frac{d\xi}{dt} = Kr + D(\xi_{in} - \xi) - Q(\xi) + F \quad (1.15)$$

où ξ représente le vecteur des composés du procédé, ξ_{in} le vecteur des éléments d'entrée du réacteur, r le vecteur des vitesses de réaction, K la matrice des coefficients de rendement et $D = \frac{F_{in}}{V}$ le taux de dilution. $Q(\xi)$ représente les débits des gaz sortant produits dans le réacteur et F l'apport de matière. Ces termes deviennent nécessaire lorsque de tels échanges interviennent dans le schéma réactionnel.

Ce modèle met bien en évidence que la dynamique d'un bioprocédé (et c'est valable pour tout système réactionnel) est la somme d'un terme de conversion (Kr) et de termes de transport ($D(\xi_{in} - \xi) - Q(\xi) + F$).

Appliquée au schéma réactionnel présenté plus haut (1.1), dans le cas d'une réaction continue ($F_{out} = F_{in}$), le système peut donc être ré-écrit sous forme matricielle :

$$\xi = \begin{bmatrix} X \\ S \\ P \end{bmatrix}, \quad k = \begin{bmatrix} 1 \\ -k_1 \\ k_2 \end{bmatrix}, \quad \xi_{in} = \begin{bmatrix} 0 \\ S_{in} \\ 0 \end{bmatrix} \quad (1.16)$$

1.5 Modéliser la cinétique

La modélisation peut aussi se pencher sur la cinétique reliant certaines variables du système. Cependant, les expressions permettant de décrire la cinétique pro-

viennent généralement d'observations empiriques. Il est important de s'en rappeler lors de l'analyse de résultats obtenus avec des modèles cinétiques.

Les modèles cinétiques sont généralement construits en observant l'évolution d'une vitesse en fonction d'une quantité pertinente. Dans le cas de la croissance microbienne, on observe par exemple l'évolution de la vitesse de croissance de la biomasse en fonction de la concentration en substrat.

Il faut aussi s'attendre à ce que ces lois soient affectées de manière assez complexe par la température et le pH puisque l'activité microbienne en dépend.

1.5.1 Expressions cinétiques communes

De nos jours, l'expression cinétique la plus souvent utilisée pour caractériser la croissance microbienne est celle de Monod, analogue à celle de Michaelis-Menten mais dérivée de manière empirique [9] :

$$\mu(S) = \mu_{max} \frac{S}{K_S + S} \quad (1.17)$$

où μ est la vitesse de croissance de la biomasse, μ_{max} la vitesse de croissance maximale, S la concentration en substrat et K_S la constante de saturation, c'est-à-dire la valeur de S lorsque $\mu = \frac{\mu_{max}}{2}$ (voir [Figure 1.3](#)).

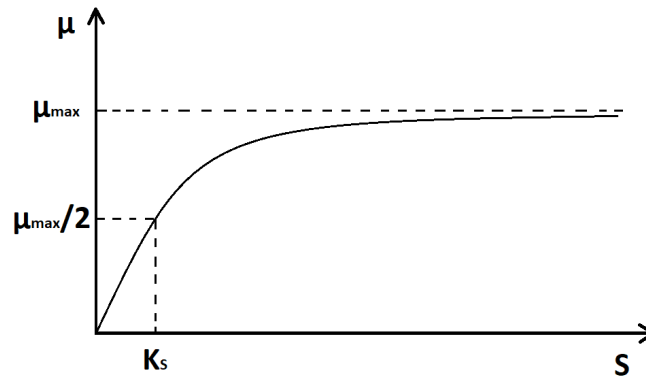


FIGURE 1.3 – Courbe de Monod, représentant la vitesse de réaction en fonction de la concentration en substrat.

Lorsque des phénomènes d'inhibition sont à l'oeuvre, il devient intéressant d'introduire une seconde constante, celle d'inhibition K_i . Dans ce cas, le modèle

de Haldane, dérivé de l'expression de Michaelis-Menten, est le plus communément utilisé [10] :

$$\mu(S) = \mu_{max} \frac{S}{K_S + S + \frac{S^2}{K_I}} \quad (1.18)$$

De nombreuses expressions peuvent être trouvées dans la littérature (plus de soixante!) et gardent chacune un caractère empirique (Table 1.1). Dans tous les cas, prouver le bien-fondé du choix de l'expression utilisée est difficile, d'autant plus que le comportement de la biomasse reste généralement variable. Il devient alors fondamental de valider proprement le modèle pour établir sa pertinence.

Nom	Remarque	Expression
Monod	Seule l'absorption en substrat est limitante	$\frac{S}{K_S+S}$
Haldane	Inhibition	$\frac{S}{K_S+S+\frac{S^2}{K_I}}$
Teissier	Equation empirique	$1 - e^{(-K_S S)}$
Moser	Monod avec des réactions de plus haut ordre	$\frac{S^N}{K_S^N+S^N}$
Vavilin[11]	Extension de Moser, inactivation initiale de substrat	$\frac{S^N}{K_S^{N-M} S_0^M + S^N}$

TABLE 1.1 – Expressions normalisées de cinétiques de croissances microbiennes [12].

Identification de paramètres

2.1 Introduction

Une fois le schéma réactionnel établi, il devient nécessaire de donner des valeurs numériques aux paramètres inconnus. La qualité de cette phase d'identification dépend évidemment de la quantité et de la qualité des données disponibles.

L'identification se déroule en deux phases. La première phase de calibration va déterminer des valeurs aux paramètres inconnus. La seconde phase de validation va confirmer que les valeurs obtenues restent adéquates avec de nouvelles données expérimentales.

Tout d'abord, il faut séparer les données disponibles en deux jeux distincts, l'un pour la calibration et l'autre pour la validation. Les données de validation doivent être suffisamment indépendantes des données de calibration afin que la validation soit plus robuste et donc crédible. Il faut ainsi répartir les données disponibles en deux jeux relativement équitables, bien qu'on favorisera généralement légèrement le jeu de calibration afin de s'assurer qu'il possède suffisamment d'information pour une bonne identification des paramètres. Ainsi, deux tiers des données pour la calibration et un tiers pour la validation est une règle de bonne pratique.

Ensuite, en plus de l'information disponible, d'autres problèmes émergent de la non-linéarité intrinsèque des modèles dynamiques biologiques ainsi que des corrélations parfois fortes entre certains paramètres. Une première étude de l'identifiabilité des paramètres avant leur estimation semble donc nécessaire.

Cette analyse se déroule *a priori*, via une étude structurale de l'identifiabilité. *A posteriori*, une analyse statistique peut être menée via des méthodes plus pratiques

prenant en compte les données disponibles [13].

L'identifiabilité structurelle se concentre sur les propriétés mathématiques du modèle pour déterminer à l'avance si certains paramètres sont identifiables ou non. L'identifiabilité pratique quant à elle repose toujours sur la qualité et la quantité des données disponibles. Un paramètre déclaré identifiable structurellement peut ne pas s'avérer identifiable en pratique. Il n'y a pas de méthodologie unique applicable à toutes les situations, ainsi de nombreuses techniques différentes ont été mises au point. En plus de déceler les corrélations entre paramètres, ces études permettent d'établir des intervalles de confiance afin de connaître la précision attendue de l'estimation des paramètres [14].

Enfin, dans une optique d'optimisation de cette phase d'identification des paramètres, il est aussi proposé de procéder de manière itérative et d'utiliser le savoir acquis après une première phase d'analyse de l'identifiabilité pour concevoir des "expériences optimales" dont les données seront encore plus riches dans le cadre de l'estimation des paramètres [15].

2.2 Identifiabilité structurelle

La notion d'identifiabilité est liée à la possibilité de donner une valeur unique à chaque paramètre d'un modèle mathématique. L'identifiabilité théorique (ou structurelle) cherche donc à vérifier si tous les paramètres du modèle sont identifiables, en supposant le cas académique de données idéales (qui correspondent parfaitement au modèle). On ne s'intéresse donc ici qu'à la structure du modèle.

Selon cette structure mathématique, il est en effet possible que seules des combinaisons ou des ratios de paramètres soient identifiables. Dans ces situations, l'identification ne pourra avoir lieu qu'en connaissant à l'avance la valeur de certains paramètres.

Pour illustrer, considérons le modèle simple : $y = ax_1 + bx_2 + c(x_1 + x_2)$. Seules les combinaisons de paramètres $a + c$ et $b + c$ sont identifiables et il faudra donc connaître à l'avance la valeur d'un des 3 paramètres a , b ou c pour pouvoir identifier les deux autres [16].

2.3 Identifiabilité pratique

En contraste avec l'identifiabilité théorique, l'identifiabilité pratique dépend de la qualité informative des données disponibles. On se demande si les données disponibles sont suffisantes pour identifier les paramètres du modèle et d'en donner des valeurs précises.

Pour illustrer, le modèle simple $y = ax_1 + bx_2$ est structurellement identifiable mais peut ne pas être identifiable en pratique, si les valeurs des variables x_1 et x_2 observées sont proportionnelles. Si $x_1 = Cx_2$, alors seule la combinaison $a + c$ est identifiable [16].

Ainsi, l'identification des paramètres peut être compromise par une quantité de données insuffisante ou de mauvaise qualité. Par exemple, dans le cas de l'utilisation du modèle de Monod (équation 1.17) pour de simple réactions de croissance microbienne, une variation du paramètre μ_{max} peut être presque parfaitement compensée par une variation proportionnelle du paramètre K_S .

De plus, l'algorithme utilisé pour l'estimation des paramètres peut avoir de mauvaises propriétés de convergence et les valeurs estimées peuvent être forts sensibles aux conditions initiales choisies. Ainsi, les valeurs obtenues varient parfois fortement et rendent difficiles l'interprétation de leur sens physique.

2.3.1 Cadre théorique

L'estimation des paramètres se base généralement sur la minimisation d'un critère J , nommé la fonction objectif. Ce critère compare les données aux résultats du modèle et représente donc l'erreur entre les deux. Voici une représentation mathématique typique de ce critère :

$$J(\theta) = \sum_{i=1}^N (y_i(\theta) - y_i)^T Q_i (y_i(\theta) - y_i) \quad (2.1)$$

Le fonction objectif J prend comme variable θ qui est un vecteur rassemblant les paramètres à estimer. J est donc la somme pondérée des carrés des erreurs entre la simulation $y_i(\theta)$ et le point de donnée y_i . La matrice de pondération Q_i permet de donner des poids différents aux points de données, en fonction de leur qualité. Par exemple, il est commun d'utiliser l'inverse de la matrice de covariance des erreurs de mesure comme matrice Q_i [17].

On peut donc ensuite calculer comment évolue J lorsqu'on modifie les valeurs des différents paramètres θ [18] :

$$E[J(\theta + \delta\theta)] \simeq \delta\theta^T \left[\sum_{i=1}^N \left[\frac{\partial y}{\partial \theta}(t_i) \right]^T Q_i \left[\frac{\partial y}{\partial \theta}(t_i) \right] \right] \delta\theta + \sum_{i=1}^N \text{tr}(C_i C_i^{-1}) \quad (2.2)$$

où C_i représente la matrice de la covariance des erreurs de mesure (Q_i est souvent défini égale à C_i^{-1}). Afin d'améliorer l'identifiabilité pratique, on cherche à maximiser la différence entre $J(\theta + \delta\theta)$ et $J_{opt}(\theta)$. On essaie donc de maximiser le terme entre crochets, qui n'est autre que la matrice d'information de Fischer. Cette matrice quantifie l'information relative à un paramètre contenue dans une distribution. :

$$F = \sum_{i=1}^N \left[\frac{\partial y}{\partial \theta}(t_i) \right]^T Q_i \left[\frac{\partial y}{\partial \theta}(t_i) \right] \quad (2.3)$$

Les termes $\frac{\partial y}{\partial \theta}$ sont les fonctions de sensibilité des variables d'output. Elles quantifient la dépendance des prédictions du modèle par rapport à chacun des paramètres. L'inverse de la matrice de Fisher est la matrice de covariance des erreurs d'estimations de la meilleure estimation linéaire non-biaisée [19] :

$$V = F^{-1} = \left[\sum_{i=1}^N \left[\frac{\partial y}{\partial \theta}(t_i) \right]^T Q_i \left[\frac{\partial y}{\partial \theta}(t_i) \right] \right]^{-1} \quad (2.4)$$

Cette approximation permet de tracer des courbes de valeurs de la fonction pour des valeurs données de l'espace des paramètres. Dans le cas d'un modèle à deux paramètres, comme c'est le cas par exemple du modèle de Monod (équation 1.17), ces courbes sont des ellipses. Les axes de ces ellipses correspondent aux vecteurs caractéristiques de la matrice de Fisher et leur longueur correspond à la racine carrée de l'inverse de la valeur propre correspondante. Ainsi, on peut visualiser de manière graphique si l'estimation de ces deux paramètres est bien conditionnée ou non (voir [Figure 2.1](#)).

2.3.2 Intervalles de confiance des estimations

Un des résultats importants de l'identifiabilité pratique est la possibilité de calculer la variance des paramètres, qui permet de leur calculer des intervalles de

confiance. En connaissant la matrice de covariance V et en considérant $Q_i = V^{-1}$, on peut approximer les écarts types des paramètres [20] :

$$\sigma(\theta_i) = \sqrt{V_{ii}} \quad (2.5)$$

On peut ensuite attribuer au paramètre θ_i un intervalle de confiance

$$\theta_i \pm 1/2/3 \sigma(\theta_i) \quad (2.6)$$

où la valeur effective de ce paramètre a 68/95/99,7% de chances de se situer.

Les intervalles de confiances permettent également de tracer des régions de confiance :

$$\theta \pm t_{\alpha;N-p}\sigma(\theta_i) \quad (2.7)$$

pour un niveau de confiance de $100(1 - \alpha)\%$ avec les valeurs t obtenues par la distribution de Student, N le nombre d'observations et p le nombre de paramètres. Les courbes de la [Figure 2.1](#) ont été formées de cette manière, pour des valeurs croissantes de α .

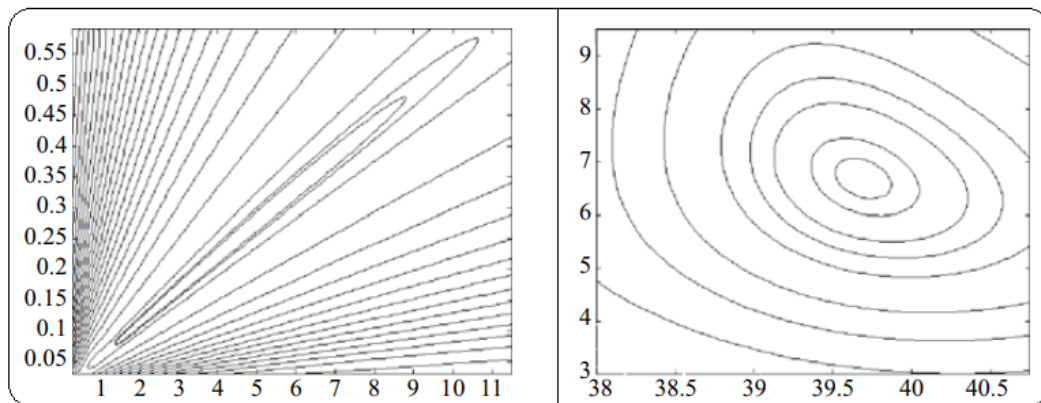


FIGURE 2.1 – Contour tracé par une fonction objectif mal conditionnée (gauche) et bien conditionnée (droite) dans le cas de l'estimation de deux paramètres [21].

2.3.3 Fonctions de sensibilité

Les fonctions de sensibilité jouent un rôle clé dans l'analyse de l'identifiabilité pratique. Si les fonctions de sensibilité de deux paramètres différents sont propor-

tionnelles, alors la matrice de covariance V sera singulière et le modèle ne sera pas identifiable [22].

Il s'avère que dans de nombreux cas de modélisation de bioprocédés, les fonctions de sensibilité sont quasi-linéairement dépendantes, ce qui entraîne des paramètres fortement corrélés. A nouveau, cela peut être visualisé sur la Figure 2.1 (gauche), où de nombreuses combinaisons des deux paramètres décrivent les données de manières équivalentes.

Différentes approches sont possibles pour déterminer les fonctions de sensibilité $\frac{\partial y_j}{\partial \theta}$. La manière la plus précise est la dérivation analytique. Pour des modèles plus complexes, une approche numérique peut être envisagée. Dans ce cas, en partant d'une valeur initiale du paramètre, la sensibilité d'une fonction y_j à une variation $\Delta\theta_i$ du paramètre θ_i est :

$$\frac{\partial y_j}{\partial \theta_i} = \frac{y_j(\theta_i) - y_j(\theta_i + \Delta\theta_i)}{\Delta\theta_i} \quad (2.8)$$

En traçant ces courbes ensemble sur un graphe, on peut ensuite déceler si certaines sont linéairement dépendantes, ce qui nous informe sur la possible corrélation de certains paramètres. Par exemple, on observe sur la Figure 2.2 une corrélation entre les paramètres K_M et V_M .

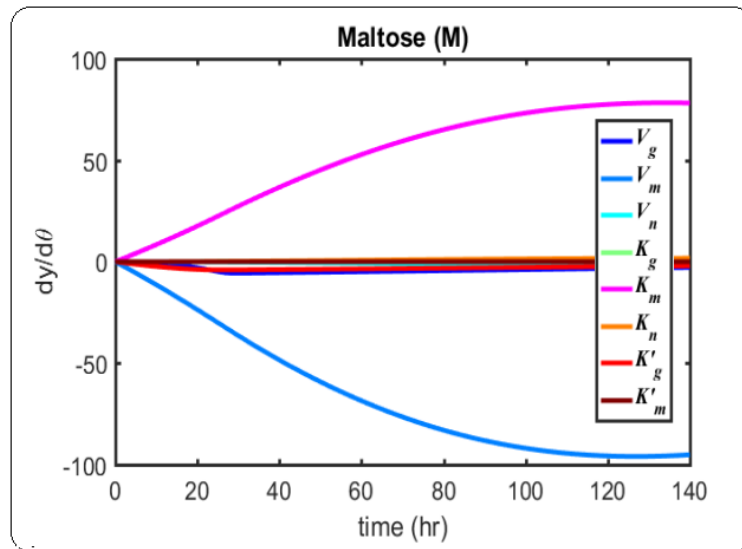


FIGURE 2.2 – Cas d'analyse de fonctions de sensibilité. On observe une corrélation entre les paramètres K_M (en rose) et V_M (en bleu), qui évoluent de manière proportionnelle [23].

2.4 Algorithmes d'estimation

Comme cité plus haut, les stratégies d'estimation sont généralement basées sur la minimisation d'un critère J , nommé la fonction objectif. Ce critère compare les données aux résultats du modèle et représente donc l'erreur entre les deux. Rappelons la représentation mathématique typique de ce critère :

$$J(\theta) = \sum_{i=1}^N (y_i(\theta) - y_i)^T Q_i (y_i(\theta) - y_i) \quad (2.9)$$

Le problème d'estimation revient alors à trouver le minimum global de la fonction objectif, visualisé sur la [Figure 2.3](#). Les algorithmes de minimisation de la fonction objectif peuvent être divisés en deux catégories : les minimisations globales et locales. Un minimum local est le minimum de la fonction dans un certain intervalle alors que le minimum global est absolu.

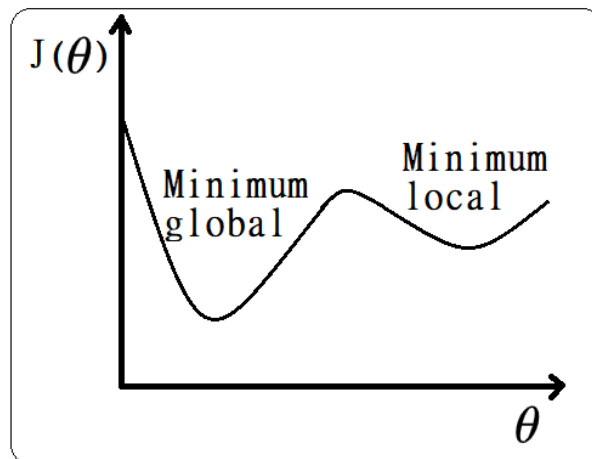


FIGURE 2.3 – Fonction objectif et ses minimums.

2.4.1 Minimisation globale

Pour explorer tout l'espace des solutions, deux types de solutions sont envisageables.

La première option est d'échantillonner cet espace et d'évaluer la fonction objectif en chaque point de la grille obtenue et ainsi de découvrir, par chance, le minimum global après un certain nombre d'évaluations. Ce procédé n'est pas le

plus efficace, bien qu'il soit possible de commencer par une grille plus étalée afin d'ensuite affiner la recherche sur certaines zones d'intérêt.

La deuxième option est de faire intervenir des changements purement aléatoires dans les essais de valeurs de paramètres afin d'éviter de s'enfermer dans un minimum local. Les méthodes qui utilisent ces procédés stochastiques tiennent cependant compte des résultats obtenus précédemment afin d'affiner la recherche et de ne pas perdre trop de temps dans des culs-de-sac.

La méthode la plus connue est celle du "simulated annealing", ou recuit simulé en français, qui s'inspire comme son nom l'indique du procédé de métallurgie où l'on cherche à minimiser l'énergie d'un système cristallin après refroidissement. L'algorithme emprunte en effet une formule de probabilité inspirée de la loi de distribution de Boltzmann-Gibbs [24] :

$$P = \exp - \frac{\Delta E}{T} \quad (2.10)$$

On commence avec une "température" T élevée et un état (valeur des paramètres) arbitraire. A chaque itération on effectue une variation de l'état élémentaire, qui engendre une variation de "l'énergie" $\Delta E = \Delta J$. Si $\Delta E < 0$, l'état est appliqué à l'état actuel. Si ΔE est positif, l'état est accepté avec une probabilité P . A haute température, le système va donc fort se déplacer dans l'espace des paramètres, tandis qu'à plus basse température il aura plus tendance à choisir les états qui baissent l'énergie du système tout en gardant la possibilité de sortir de minimums locaux. Différentes stratégies de "refroidissement" sont envisageables et il a été prouvé que cette méthodologie converge à un minimum global.

2.4.2 Minimisation locale

Les méthodes d'optimisation locales sont nombreuses. S'il est possible de calculer la dérivée de la fonction objectif, de nombreuses méthodes se servent du gradient obtenu pour "simplement" chercher une direction où le gradient est négatif et glisser le long de cette direction jusqu'à trouver le minimum. Evidemment, on parle ici de minimisation locale et donc le minimum obtenu n'est pas garanti d'être celui recherché. Les valeurs initiales des paramètres jouent alors un rôle crucial dans la recherche du minimum global.

Plusieurs algorithmes usant de cette stratégie sont largement employés, tels que l'algorithme de Gauss-Newton, de Levenberg-Marquardt ou encore la méthode des régions de confiances. Cette dernière utilise une approximation de la fonction

objectif et détermine un rayon de confiance où l'approximation reste valable. Dans la région de confiance ainsi formée, on trouve le point où la fonction est minimale et on réitère l'opération. Cette méthode s'approche des techniques de minimisations globales car la région de confiance peut permettre à l'algorithme de s'affranchir des minimums locaux et le choix des valeurs initiales des paramètres est moins critique.

L'algorithme de Gauss-Newton résout les problèmes de moindres carrés, où $J = \sum r_i^2(\theta)$ avec $r_i = y(\theta) - y_i$. On procède par itération à partir des valeurs initiales θ_0 ,

$$\theta_{s+1} = \theta_s - (J_r^T J_r)^{-1} J_r r \quad (2.11)$$

r dénote le vecteur des fonctions r_i et J_r la matrice jacobienne de r par rapport à θ . Cette méthode est très rapide lorsqu'elle converge mais la convergence n'est pas garantie.

L'algorithme de Levenberg-Marquardt en est une altération, avec l'addition d'un terme λI :

$$\theta_{s+1} = \theta_s - (J_r^T J_r + \lambda I)^{-1} J_r r \quad (2.12)$$

λ est un coefficient arbitraire et I la matrice identité. C'est un compromis entre la méthode de Newton (rapide) et une méthode du gradient (convergente), dont les résultats dépendent du choix de la valeur de λ .

Les méthodes de recherches directes permettent de trouver des minimums locaux sans utiliser la dérivée de la fonction objectif. La première méthode de ce type qui fut élaborée est l'algorithme de Hooke-Jeeves [25]. La méthodologie est assez simple, on essaye des valeurs de paramètres proches des initiales et lorsque la fonction objectif décroît, on se déplace dans ce sens.

Depuis, l'algorithme de Nelder-Mead est plus populaire, il utilise la notion de simplexes, des polytopes à $N+1$ sommets dans un espace de N dimensions. Le concept est relativement identique, on prend $N+1$ points formés de valeurs différentes des N paramètres puis on les classe suivant la valeur de la fonction objectif J en ces points. On prend alors la réflexion (par symétrie par rapport au centre de gravité des N autres points) du point $N+1$, où J est maximum. Si la valeur de J de ce nouveau point est plus faible que certains points du simplexe, on intègre le nouveau point au simplexe. On procède ainsi de suite jusqu'à trouver le minimum.

2.4.3 Conclusion sur les stratégies de minimisation

De manière générale, les inconvénients des optimisations globales sont des implémentations hardues et une exécution fastidieuse. Les minimisations locales quant à elles sont plus rapides mais n'atteignent pas forcément le minimum recherché.

Il est donc commun d'employer ces méthodes en parallèle, c'est-à-dire d'utiliser une minimisation globale pour avoir une idée de la localisation des différents minimums puis un algorithme de minimisation local afin de rapidement converger vers ceux-ci. De nombreuses combinaisons de minimisations globale et locale sont possibles et forment les stratégies hybrides.

Enfin, il est bon de se rappeler que les paramètres recherchés ont souvent des significations biologiques et que certaines valeurs jugées "optimales" par les algorithmes ne reflètent pas la réalité attendue. Ce problème peut être réglé en posant des contraintes sur l'espace des paramètres, pour n'admettre que des solutions jugées cohérentes [26].

2.5 Validation du modèle

La validation du modèle est une étape cruciale, qui permet de définir à quel point un modèle a su atteindre ses objectifs et de renforcer sa crédibilité. Quelle que soit la discipline pour laquelle un modèle est développé, son usage postérieur et l'impact de ses résultats dépend du niveau de validation qu'il atteint. Les résultats d'un modèle non-validé peuvent être balayés du revers de la main et le modèle lui-même déchu de toute appréciation scientifique [27].

Malgré tout cela, la validation des modèles biologiques est souvent négligée car il n'existe pas encore de méthodologie standard de validation. Cette standardisation se heurte en effet au manque récurrent de données disponibles. De plus, les contextes dans lesquels les modèles sont construits varient énormément et il est donc difficile de proposer une méthodologie unique et incontestable pour chaque cas [28].

Rappelons ici qu'un modèle n'est jamais validé de manière absolue, il ne peut qu'être falsifié. La validation est plutôt un témoignage de sa robustesse et chaque étape de la validation sert à renforcer celle-ci. De nombreuses validations sont envisageables par exemple en vérifiant les propriétés mathématiques du système réactionnel, qualitativement et quantitativement [2].

Plus globalement, la validation la plus classique reste de comparer quantitativement les résultats du modèle sur des données différentes de celles avec lesquelles il a

été construit. Ainsi un modèle dont les résultats restent cohérents lorsqu'il est utilisé avec d'autres conditions expérimentales peut se targuer d'une certaine robustesse. Cependant, dans le cas d'une inadéquation entre modèle et nouvelles données, il reste difficile de déterminer si le problème provient des équations mathématiques utilisées, des valeurs des paramètres estimées ou bien des hypothèses fondatrices du modèle.

En pratique, la comparaison peut même être qualitative et on vérifie dans ce cas si le comportement des courbes se rapporte bien au comportement attendu. Une bonne adéquation visuelle peut être un argument suffisamment encourageant en biologie pour continuer à développer un modèle bien qu'assurer la corrélation entre les prédictions et les mesures reste plus prudent.

Modélisation du processus de vinification

3.1 Introduction

Dans ce travail, l'objectif était d'étendre la portée d'un modèle à certains marqueurs aromatiques. Le modèle utilisé est fondé sur des bilans de masse d'azote, de biomasse, de sucre, d'éthanol et de dioxyde de carbone. Il a déjà été validé de nombreuses fois par des données expérimentales. Un bref rappel de son développement permet de mieux comprendre ses tenants et aboutissants. Le contenu de ce chapitre provient majoritairement de l'article décrivant la création du modèle, par R. David et al. [29].

3.2 Modèles antérieurs

Les premiers modèles cinétiques de fermentations alcooliques décrivaient l'influence du sucre et des niveaux d'éthanol, ainsi que de la température et de l'utilisation du sucre, afin de capturer les tendances macroscopiques générales observées ([30][31][32][33][34]). D'autres modèles ont été développés où la cinétique est basée sur l'azote comme substrat limitant, en conditions isothermes [35], et plus tard en incluant la dépendance en température de certains paramètres [36].

Malheureusement, la pertinence de ces modèles n'a jamais été validée dans de véritables conditions de vinification. Plusieurs modèles empiriques et non-paramétriques ont été publiés et parmi ceux-ci, le plus validé est le modèle de Malherbe et al.[37] qui considère les mécanismes physiologiques principaux de la levure en prédisant les quantités d'éthanol ou de CO_2 produites.

Ce modèle inclut également les effets des facteurs principaux : la température T , qui peut varier de 18 à 30°C, et l'azote N , qui a un impact majeur sur l'activité de la levure et varie beaucoup en fonction du moût utilisé. De plus, l'azote est la source de précurseurs de composés organoleptiques, ce qui lui donne un rôle essentiel dans le cas présent.

Cependant, le modèle de Malherbe n'inclut pas de bilans de matières et son extension aux composés aromatiques serait délicate car la formulation des équations est complexe et le nombre important de paramètres complique leur identification (si une modification due aux conditions opérationnelles est nécessaire).

3.3 Modèle actuel

Le schéma réactionnel communément utilisé en fermentation considère la croissance de la biomasse par consommation de sucre et d'azote, ainsi que la synthèse d'éthanol et de dioxyde de carbone [38]. Les observations expérimentales et la révision de la littérature ont amené à la conclusion que la fermentation s'effectue comme étant la croissance de biomasse avec l'azote comme substrat limitant. Le sucre est également dégradé comme substrat non-limitant pour former de l'éthanol et du dioxyde de carbone.

Cependant, il a été observé [39] que l'azote n'est pas uniquement consommé pour la croissance de la biomasse mais qu'une autre réaction importante avait lieu. Dans des conditions de vinification, une partie de l'azote est assimilée afin de synthétiser de nouvelles cellules de levure mais la partie restante est majoritairement utilisée pour la synthèse et la réparation de protéines essentielles.

En effet, la cellule de levure peut assimiler le sucre et l'azote grâce à des protéines dédiées appelées transporteurs qui permettent la traversée de la membrane cellulaire. Durant la fermentation ces protéines de transport doivent subir une inactivation catabolique par l'éthanol et l'absorption de sucre devient l'étape déterminante de vitesse de la fermentation ([40][41][31]).

Ce phénomène augmente avec la concentration d'éthanol. Par conséquent, la cinétique de fermentation est ralentie (décroissance de la vitesse de production du CO_2). Des expériences ont montré que lorsque la concentration initiale en azote N_0 est faible, la levure se concentre principalement sur la production de cellules. Lorsque N_0 augmente, la production de cellules croît également mais plus d'azote est utilisé pour la synthèse de transporteurs de manière à prévenir leur inhibition catabolique [42]. Cet effet a été illustré expérimentalement par l'addition d'azote

assimilable dans la réacteur une fois que la biomasse est à l'état stationnaire : la croissance de la biomasse est faible mais l'activité de la levure est stimulée (pic de production de CO_2) ([43][32]).

3.3.1 Equations du modèle

Les considérations mentionnées ci-dessus peuvent être résumées par ce schéma réactionnel :



où la biomasse X (levure) croit par consommation d'azote dédié à la croissance de la biomasse N_x . Tr représente la concentration de transporteurs d'hexoses en excès, tel qu'expliqué précédemment. N_{tr} est l'azote dédié à la synthèse de transporteurs en excès. Le sucre S est dégradé par les enzymes en éthanol E et en dioxyde de carbone CO_2 .

Il a aussi été démontré [44] que

$$\begin{aligned}
 E(t) = CO_2(t) &\longleftrightarrow \dot{E}(t) = C\dot{O}_2(t) \\
 \dot{S}(t) &= -2.17 C\dot{O}_2(t)
 \end{aligned} \tag{3.2}$$

Les résultats cumulés de [39] et [44] et la validation expérimentale ont mené au bilan de masse suivant :

$$\left\{ \begin{array}{l}
 \dot{X} = \mu_{max}(T) \frac{N_x}{K_x X + N_x} X \quad \text{et} \quad \dot{N}_x = -k_1 \dot{X} \\
 \dot{Tr} = \eta_{max}(T) \frac{N_{tr}}{K_{tr} + N_{tr}} X \quad \text{et} \quad \dot{N}_{tr} = -k'_1 \dot{Tr} \\
 \dot{E} = C\dot{O}_2 \\
 = \left(\beta_{max,1}(T) \frac{N_x}{K_x X + N_x} + \beta_{max,2}(T) \frac{K_E(T)}{K_E(T) + E} \right) \frac{S}{K_S + S} X (1 + \phi(T) Tr) \\
 \dot{S} = -k_2 \dot{E}
 \end{array} \right. \tag{3.3}$$

avec

$$\begin{aligned}
 X_0 &= X(t=0), & N_{x,0} &= N_x(t=0) \\
 Tr_0 &= 0, & N_{tr,0} &= N_{tr}(t=0) \\
 E_0 = CO_{2,0} &= 0, & S_0 &= S(t=0)
 \end{aligned} \tag{3.4}$$

$\mu_{max}(T)$, $\eta_{max}(T)$ et $\beta_{max}(T)$ sont les vitesses spécifiques de réaction. K_x est une constante de Contois tandis que K_{tr} et K_S sont des constantes de saturation. $K_E(T)$ représente l'inhibition de l'éthanol. k_1 , k'_1 et k_2 sont des coefficients stoechiométriques. k_2 vaut 2.17, voir (3.2). $\phi(T)$ est un paramètre de rendement lié à Tr et associé à l'activité de la fermentation (vitesse de production d'éthanol ou de CO_2). La concentration initiale d'azote (totale) est définie comme $N_0 = N_{x,0} + N_{tr,0}$

3.3.2 Hypothèses du modèle

La variable Tr étant une nouveauté, la littérature ne dispose d'aucune information à son sujet. Des hypothèses doivent donc être formulées quand à son comportement dynamique. Il faut ainsi définir la distribution d'azote en N_x et N_{tr} , les valeurs d'équilibre et le comportement transitoire de Tr .

Valeurs initiales de N_x et N_{tr}

La variable Tr représente la concentration de transporteurs en excès par rapport à une expérience de référence. Cette expérience est caractérisée par une concentration initiale très faible en azote : 0.071g/l, qui permet tout de même une fermentation normale. On considère que la quantité de transporteurs synthétisés afin de dégrader tout le sucre en éthanol et dioxyde de carbone est minimale.

Le rapport $\frac{N_0}{X_{max}}$ ($X_{max} = X$ à l'équilibre) est calculé pour cette expérience et informe de la proportion d'azote initial dédiée à la croissance de la biomasse et la synthèse d'un minimum de transporteurs pour la fermentation. Ce ratio peut ensuite être appliqué aux X_{max} obtenus lors d'expériences où $N_0 > 0.071$ g/l afin d'obtenir $N_{x,0}$, la quantité initiale d'azote nécessaire à la croissance de la biomasse et la synthèse du minimum de transporteurs.

$$N_{x,0} = \frac{0.071}{X_{max,N_0=0.071}} \cdot X_{max,N_0} \tag{3.5}$$

On obtient ensuite $N_{tr,0}$, la quantité d'azote dédiée à la synthèse de transporteurs en excès, par simple soustraction (voir [Figure 3.1](#)) :

$$N_{tr,0} = N_0 - N_{x,0} \quad (3.6)$$

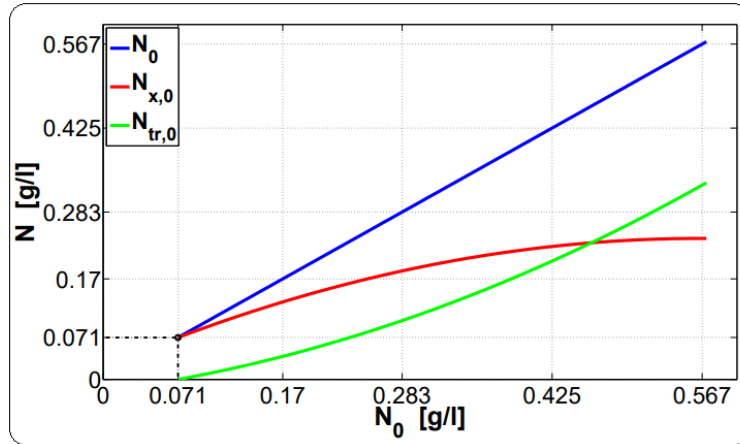


FIGURE 3.1 – Distribution d'azote initial en fonction de son usage, pour $N_0 = 0.071 \rightarrow 0.57g/l$ [29].

Valeur à l'équilibre de Tr , Tr_{max}

Dans la littérature, le contenu d'azote dans les protéines est généralement de 16% [45]. Ce ratio peut être appliqué dans notre cas pour déterminer Tr_{max} , la valeur d'équilibre de Tr , correspondant à un certain $N_{tr,0}$:

$$Tr_{max} = \frac{N_{tr,0}(N_0)}{0.16}, \text{ pour } N_0 = 0.071 \rightarrow 0.57g/l \quad (3.7)$$

Ceci est illustré à la [Figure 3.2](#). Cela implique que la valeur de k'_1 est égale à 0.16.

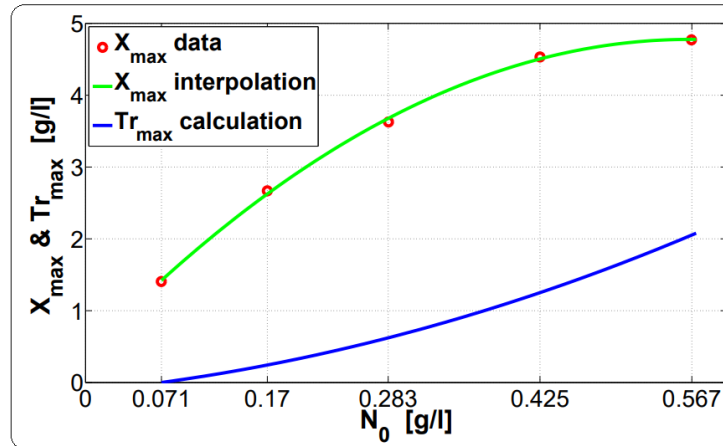


FIGURE 3.2 – Valeurs d’équilibres de X et Tr correspondant aux valeurs d’azote initial N_0 [29].

Comportement transitoire de Tr

Puisque les transporteurs sont des protéines synthétisées par la levure, leur comportement transitoire est considéré comme étant identique à celui de la biomasse. On suppose que

- $Tr_0 = 0$ g/l
- les mesures de biomasse forment le profil d’évolution de Tr , après les avoir mis à l’échelle par rapport à Tr_{\max}

Les paramètres liés à Tr peuvent ainsi être estimés.

3.3.3 Identification des paramètres du modèle

Les expériences qui forment la base de données pour l’identification des paramètres du modèle recouvrent une gamme de températures de 18 à 30° et une gamme d’azote initial de 0.071 à 0.57 g/l (voir Figure 3.3).

L’identification des paramètres est facilitée par la structure en cascade du modèle : les paramètres $\mu_{\max}(T)$, K_x et k_1 ont été identifiés en premier, puis $\eta_{\max}(T)$ et K_{tr} avant de finalement identifier $\beta_{\max,1}(T)$, $\beta_{\max,2}(T)$, $K_E(T)$, $K_S(T)$ et $\phi(T)$. La plupart des paramètres sont dépendant de la température et la Table 3.1 montre la portée de leurs variations.

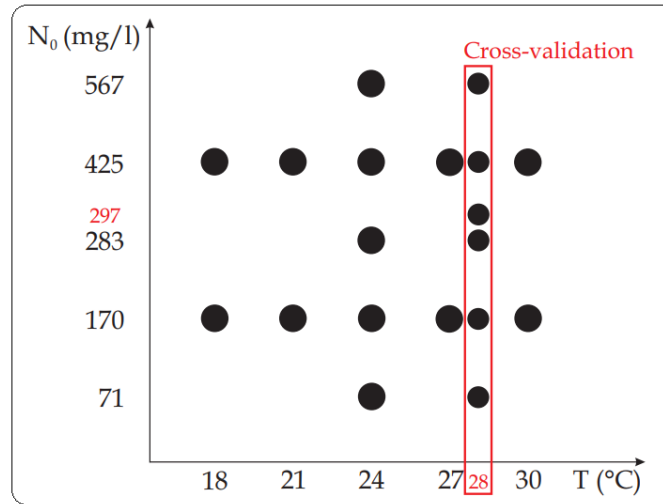


FIGURE 3.3 – Gamme de conditions initiales des expériences servant à l’identification des paramètres du modèle. Encadrées en rouge, les expérience dédiées à sa validation [29].

Paramètre	Plage de valeurs	Unités
T	18 - 30	°C
μ_{max}	0.0996 - 0.3034	h^{-1}
K_x	0.0127	-
k_1	0.0528	-
η_{max}	0.0799 - 0.2257	h^{-1}
K_{tr}	0.3134	
$\beta_{max,1}$	0.1963 - 0.3759	h^{-1}
$\beta_{max,2}$	0.3318 - 0.9205	h^{-1}
K_E	7.6560 - 2.9011	g/l
K_S	47.7440 - 26.3585	g/l
ϕ	0.3887 - 0.4727	$(g/l)^{-1}$

TABLE 3.1 – Plages de valeurs des paramètres et leurs unités [29].

4.1 Introduction

Dans ce chapitre sera décrit la démarche qui a été entreprise cette année. Comme il a déjà été mentionné, l'objectif est d'étendre un modèle de vinification préconçu [29] à certains marqueurs aromatiques, grâce à de nouvelles données expérimentales.

Rappelons la formulation mathématique du modèle utilisé, qui a été présentée dans le chapitre précédent :

$$\left\{ \begin{array}{l} \dot{X} = \mu_{max}(T) \frac{N_x}{K_x X + N_x} X \quad \text{et} \quad \dot{N}_x = -k_1 \dot{X} \\ \dot{Tr} = \eta_{max}(T) \frac{N_{tr}}{K_{tr} X + N_{tr}} X \quad \text{et} \quad \dot{N}_{tr} = -k'_1 \dot{Tr} \\ \dot{E} = C \dot{O}_2 \\ \quad = \left(\beta_{max,1}(T) \frac{N_x}{K_x X + N_x} + \beta_{max,2}(T) \frac{K_E(T)}{K_E(T) + E} \right) \frac{S}{K_S + S} X (1 + \phi(T) Tr) \\ \dot{S} = -k_2 \dot{E} \end{array} \right. \quad (4.1)$$

Dans un premier temps, l'objectif est donc de ré-identifier les paramètres en fonction des nouvelles données. Dans un second temps, on cherche à étendre le modèle aux marqueurs aromatiques choisis. Dans le cas présent, les 5 molécules organoleptiques choisies sont l'isobutanol, l'isoamyl alcool, l'isoamyl acétate, l'éthyl hexanoate et l'éthyl octanoate.

4.2 Données disponibles

Les 15 jeux de nouvelles données sont représentés sur la [Figure 4.1](#). Parmi ceux-ci, 9 serviront à la phase de calibration et 6 à la phase de validation.

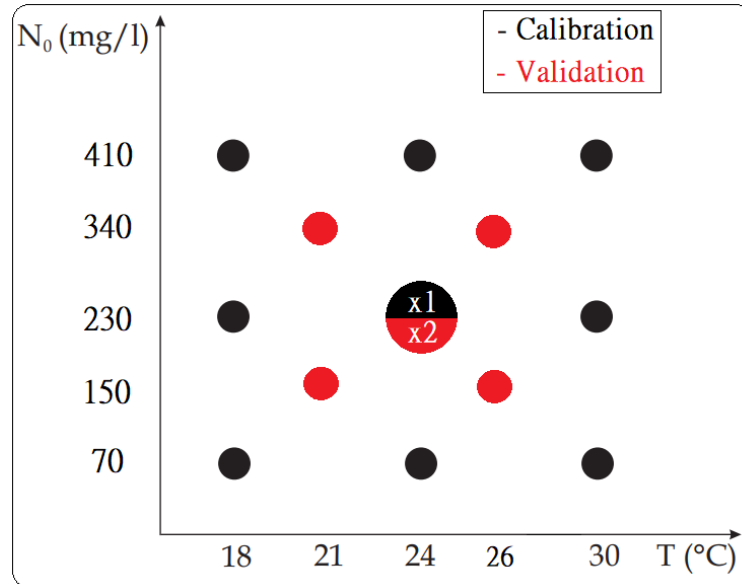


FIGURE 4.1 – Gamme de conditions initiales des expériences servant à l'identification des paramètres du modèle.

Pour la calibration, les jeux de données forment donc une gamme des combinaisons entre des températures de 18, 24 et 30 °C et des concentrations initiales d'azote de 70, 230 et 410 mg/l. La concentration initiale en sucre est de 180 g/l. Les variables mesurées sont la biomasse X , le CO_2 cumulé, le flux de CO_2 , l'azote N , le sucre S et les 5 molécules aromatiques.

Pour la validation, les jeux de données forment des combinaisons entre des températures de 21 et 26 °C et des concentrations initiales d'azote de 150 et 340 mg/l, ainsi que deux jeux de données à 24°C et 230 mg/l d'azote initial. Pour les données de validation, la concentration initiale en sucre est de 170 g/l. Les variables mesurées sont les mêmes que pour les données de calibration, à l'exception de la biomasse et de l'azote qui n'ont pas été observés.

Quelques jeux de données expérimentales peuvent être observés à l'[ANNEXE I](#)

4.3 Implémentation et extension du modèle

4.3.1 Dérivée seconde du CO_2

Tout d'abord, il fallait ajouter à la modélisation la dérivée seconde du CO_2 (celle-ci faisait en vérité déjà partie du modèle mais n'était pas mentionnée dans l'article). En effet, le flux de CO_2 est une variable critique puisque celui-ci est finement lié à l'activité microbienne. Comme son expression mathématique est assez longue, elle a été divisée en parties pour plus de clarté :

$$C\ddot{O}_2 = \dot{A}BCDF + A\dot{B}CDF + AB\dot{C}DF + ABC\dot{D}F + ABCD\dot{F} \quad (4.2)$$

avec

$$\left\{ \begin{array}{l} A = \beta_{max,1}(T) \frac{N_x}{K_x X + N_x} + \beta_{max,2}(T) \frac{K_E(T)}{K_E(T) + E} \\ \dot{A} = \beta_{max,1}(T) \frac{\dot{N}_x}{K_x X + N_x} - \beta_{max,1}(T) \frac{N_x(K_x \dot{X} + \dot{N}_x)}{(K_x X + N_x)^2} - \beta_{max,2}(T) \frac{K_E(T)C\dot{O}_2}{(K_E(T) + E)^2} \\ B = \frac{1}{K_S(T) + S} \quad \dot{B} = -\frac{\dot{S}}{(K_S(T) + S)^2} \\ C = S \quad \dot{C} = \dot{S} \\ D = X \quad \dot{D} = \dot{X} \\ F = (1 + \phi(T)Tr) \quad \dot{F} = \phi(T)\dot{Tr} \end{array} \right. \quad (4.3)$$

Ensuite, il a fallu ré-implémenter le modèle mathématique et s'assurer que la réplique est exacte à l'aide de quelques essais réutilisant les anciennes valeurs de paramètres. La simulation a été mise en oeuvre grâce au logiciel Matlab. On peut voir dans l'ANNEXE II le vis-à-vis du "nouveau" modèle par rapport à d'anciens résultats.

4.3.2 δ et $N_{x,\delta}$

Un paramètre δ a été ajouté au modèle. En effet, on observe dans les données disponibles un déphasage entre la consommation de l'azote et la croissance de la biomasse (voir Figure 4.2).

4.3. IMPLÉMENTATION ET EXTENSION DU MODÈLE

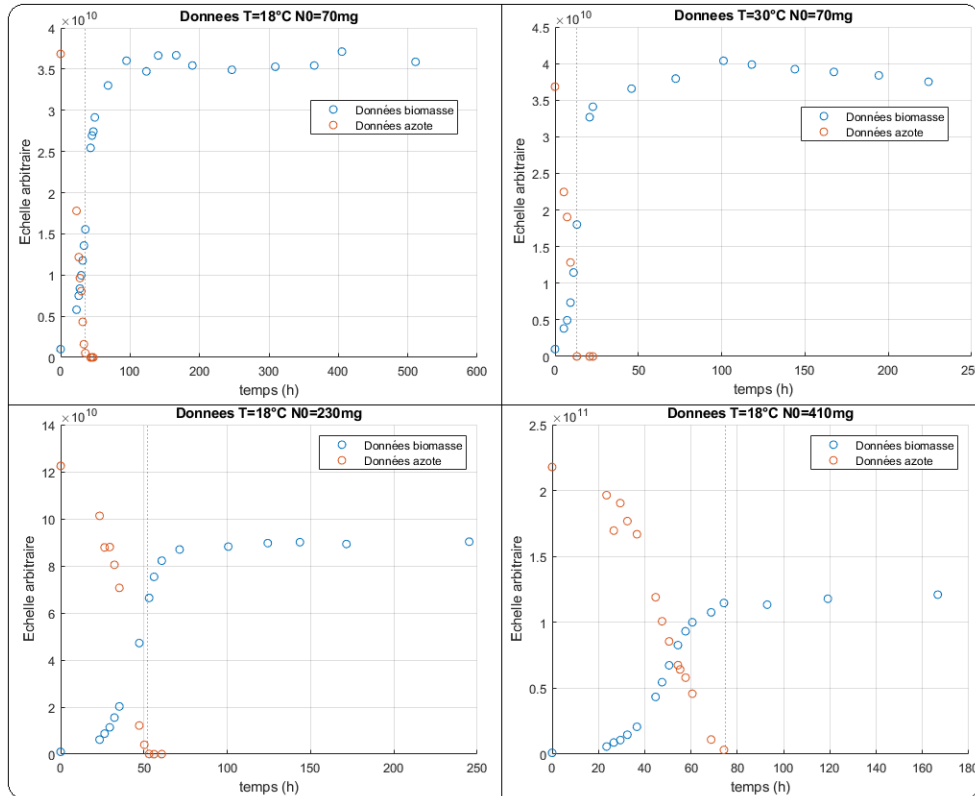


FIGURE 4.2 – On observe un déphasage entre la fin de la consommation de l’azote (traits en pointillé) et la fin de la croissance de la biomasse. Ce déphasage diminue avec la température et avec la concentration initiale d’azote. A 410mg/l, ce déphasage s’est complètement atténué.

Ce déphasage peut s’expliquer par exemple par une différence entre la consommation d’azote du milieu réactionnel et son usage intra-cellulaire effectif (croissance ou synthèse de transporteurs). L’apparition de ce phénomène peut également être une conséquence de l’utilisation d’une souche de levure différente. Afin de le prendre en compte, le modèle a donc été légèrement modifié et une variable représentant le délai dû à l’absorption de l’azote en milieu intracellulaire a été ajoutée, δ .

Les équations de X et N_x ont donc été modifiées en conséquence :

$$\begin{cases} \dot{X} = \mu_{max}(T) \frac{N_{x,\delta}}{K_x X + N_{x,\delta}} X \\ \dot{N}_{x,\delta} = \delta(N_x - N_{x,\delta}) \\ \dot{N}_x = -k_1 \mu_{max}(T) \frac{N_{x,\delta}}{K_x X + N_{x,\delta}} X \end{cases} \quad (4.4)$$

Une variable supplémentaire a donc également été ajoutée. N_x représente l'azote au sein du réacteur tandis qu' $N_{x,\delta}$ représente l'azote intracellulaire.

Le modèle final est donc :

$$\begin{cases} \dot{X} = \mu_{max}(T) \frac{N_{x,\delta}}{K_x X + N_{x,\delta}} X \\ \dot{N}_{x,\delta} = \delta(N_x - N_{x,\delta}) \\ \dot{N}_x = -k_1 \mu_{max}(T) \frac{N_{x,\delta}}{K_x X + N_{x,\delta}} X \\ \dot{Tr} = \eta_{max}(T) \frac{N_{tr}}{K_{tr} + N_{tr}} X \\ \dot{N}_{tr} = -k'_1 \dot{Tr} \\ \dot{E} = C\dot{O}_2 \\ \quad = \left(\beta_{max,1}(T) \frac{N_x}{K_x X + N_x} + \beta_{max,2}(T) \frac{K_E(T)}{K_E(T) + E} \right) \frac{S}{K_S + S} X (1 + \phi(T)Tr) \\ \dot{S} = -k_2 \dot{E} = -k_2 C\dot{O}_2 \\ C\ddot{O}_2 = \dot{A}BCDF + \dot{A}\dot{B}CDF + \dot{A}B\dot{C}DF + \dot{A}BC\dot{D}F + \dot{A}BCD\dot{F} \end{cases} \quad (4.5)$$

avec A, B, C, D, F et leur dérivée présentés plus haut, à l'équation (4.3) .

Pour compléter le modèle, les valeurs initiales de N_x et N_{tr} ainsi que la valeur à l'équilibre de Tr , Tr_{max} devaient également être re-calculées en fonction des nouvelles données. Celles-ci sont représentées à la [Figure 4.3](#), dont les courbes ont été calculées à l'identique du mode opératoire détaillé à la section 3.3.2. Puisque X_{max} dépend de N_0 et non de T , et que les données comprennent des conditions expérimentales où uniquement T varie (voir [Figure 4.1](#)), X_{max} est ici une moyenne.

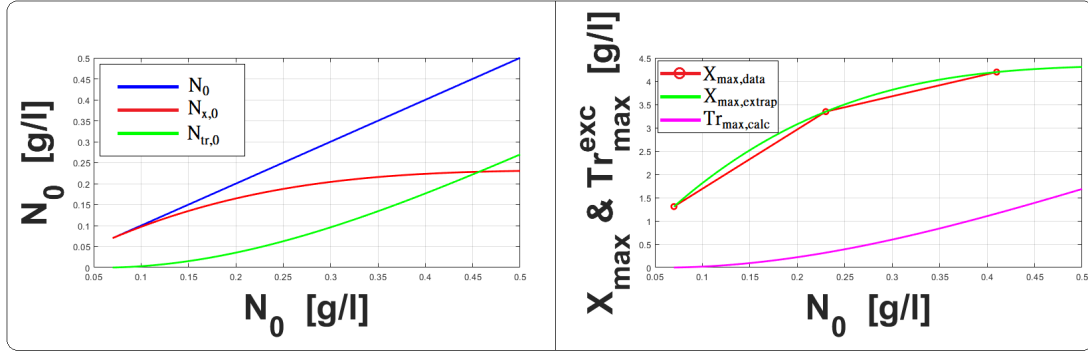


FIGURE 4.3 – A gauche, distribution de l’azote initial en fonction de son usage. A droite, valeurs d’équilibres de X et Tr correspondant aux valeurs d’azote initial N_0 .

4.4 Identification des paramètres

4.4.1 Méthodologie

La méthodologie employée lors de cette phase d’identification est identique à celle décrite à la section 3.3.3. La structure en cascade du modèle permet d’identifier d’abord μ_{max} , K_x , k_1 , puis η_{max} et K_{tr} et enfin $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ . Cette fragmentation de l’identification facilite grandement son exécution, puisque le nombre de variables à identifier simultanément est moindre.

Pour chaque phase d’identification, les équations différentielles du modèle sont résolues grâce à la fonction ODE15s de Matlab et les paramètres sont ensuite estimés en formulant un problème des moindres carrés qui peut être minimisé via les fonctions dédiées `fminsearch` (algorithme de Nelder-Mead) et `lsqnonlin` (régions de confiance et Levenberg-Marquardt), de Matlab également. Les valeurs initiales des paramètres sont fixées aux valeurs précédemment identifiées.

Une analyse des fonctions de sensibilité des paramètres et des intervalles de confiance pour les valeurs qui leurs sont obtenues est également menée :

L’analyse des fonctions de sensibilité $\frac{\partial y_j}{\partial \theta_i}$ permet de cerner les corrélations entre différents paramètres (voir section 2.3.2), lorsque leurs fonctions de sensibilité se révèlent linéairement dépendantes. Cette analyse est faite sur des fonctions de sensibilité normées, afin de plus facilement comparer leur profil. Pour chaque étape d’identification, les fonctions de sensibilité normées des paramètres d’un seul batch seront présentées ($T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$) car celles-ci sont représentatives

de l'ensemble des données (quelques fonctions non-normées sont consultables à l'[ANNEXE III](#)).

Ces fonctions de sensibilité ont été obtenues en résolvant simultanément les équations du modèle avec leurs dérivées partielles par rapport à chaque paramètre :

$$\frac{\partial y_j}{\partial \theta_i} = \int \frac{\partial}{\partial \theta_i} \left(\frac{dy_j}{dt} \right) dt \quad (4.6)$$

Les intervalles de confiance quant à eux informent sur la précision des valeurs obtenues (voir section 2.3.3). Les intervalles de confiance d'un $\sigma(\theta_i)$ seront présentés, afin de se rendre compte visuellement de la précision attendue des estimations.

Lorsque des paramètres sont corrélés, un comparatif de l'impact de chaque paramètre sur le critère J a été calculé en observant **les variations du critère J pour une variation de $\pm 5\%$ de chaque paramètre**. Cela permet de déterminer à quels paramètres le modèle est le plus sensible. Les paramètres dont les valeurs devront être fixées à l'avance, à cause de corrélations, seront choisis parmi les moins sensibles.

4.4.2 Paramètres μ_{max} , K_x , k_1 et δ

Les premiers paramètres à identifier sont donc μ_{max} , K_x , k_1 et δ . Pour l'identification de ces 4 paramètres, le critère J à minimiser est porté sur les données de biomasse X et d'azote au sein du réacteur N_x :

$$J = \sum_{i=1}^{M_X} \left(\frac{(X_i(\theta) - X_i)^2}{X_i^2} \right) + \frac{M_X}{M_N} \sum_{j=1}^{M_N} \left(\frac{(N_{x,i}(\theta) - N_{x,i})^2}{N_{x,i}^2} \right) \quad (4.7)$$

avec M_X et M_N le nombre de mesures de biomasse et d'azote, $X_i(\theta)$ et $N_{x,i}(\theta)$ les valeurs prédites de X et N_x au temps t_i et X_i et $N_{x,i}$ leur valeur mesurée. Le rapport $\frac{M_X}{M_N}$ permet de donner un poids équivalent aux deux variables si l'une ou l'autre présente plus de points de données.

Puisque seul l'azote total N est mesuré et non N_x , la proportion d'azote initial dédiée à la croissance de la biomasse et la synthèse d'un minimum de transporteurs pour la fermentation, N_x est approximé en utilisant le rapport $\frac{N_{x,0}}{N_0}$:

$$N_{x,i} = N_i \cdot \frac{N_{x,0}}{N_0} \quad (4.8)$$

Analyse des fonctions de sensibilité

Les corrélations entre différents paramètres sont visibles lorsque leurs fonctions de sensibilité sont linéairement dépendantes. En normant celles-ci, les corrélations deviennent encore plus évidentes car les fonctions se chevauchent. Pour les paramètres μ_{max} , K_x , k_1 et δ , le critère à minimiser se porte sur deux variables, X et N_x , il est donc bon d'observer leurs fonctions de sensibilité par rapport à celles-ci.

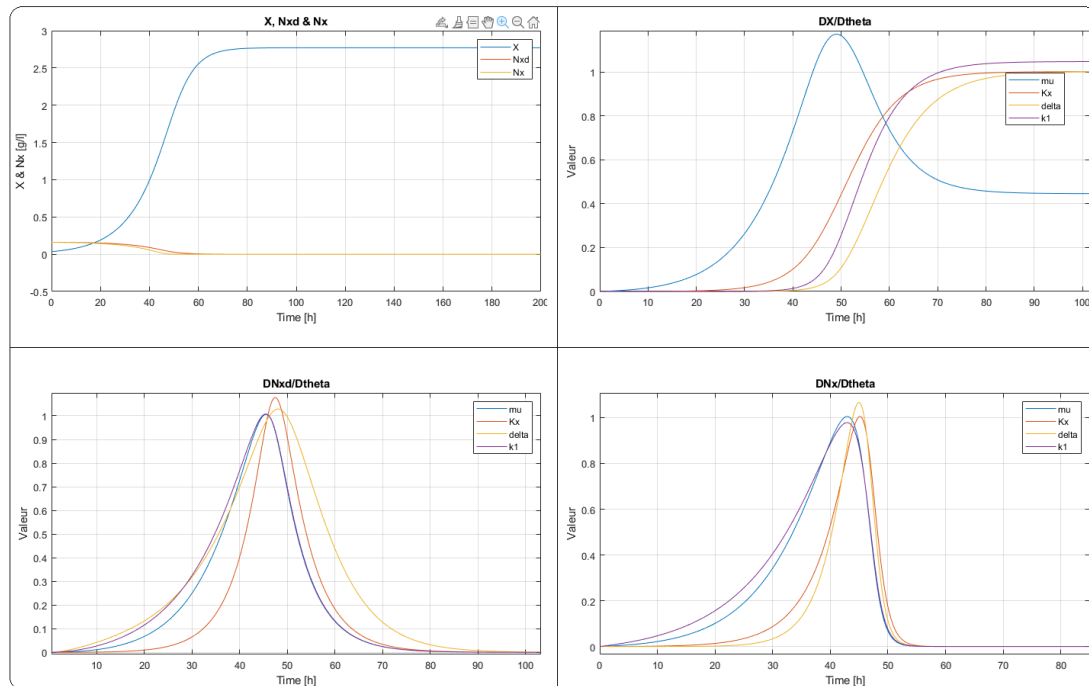


FIGURE 4.4 – Fonctions de sensibilité des paramètres μ_{max} (μ), K_x , k_1 et δ (δ), normées afin de les comparer plus facilement, pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

On observe sur la [Figure 4.4](#) que par rapport à la fonction X , seul μ_{max} semble indépendant des autres paramètres, alors que par rapport aux fonctions N_x et $N_{x,\delta}$, K_x et δ semblent corrélés ainsi que μ_{max} et k_1 .

Analyse des intervalles de confiance

La précision de l'estimation de chaque paramètre peut être visualisée grâce aux intervalles de confiance. Pour calculer ceux-ci, on calcule d'abord la matrice de Fisher et les intervalles de confiance sont alors les racines carrées des éléments diagonaux de l'inverse de la matrice de Fisher (voir section 2.3.1 et 2.3.2).

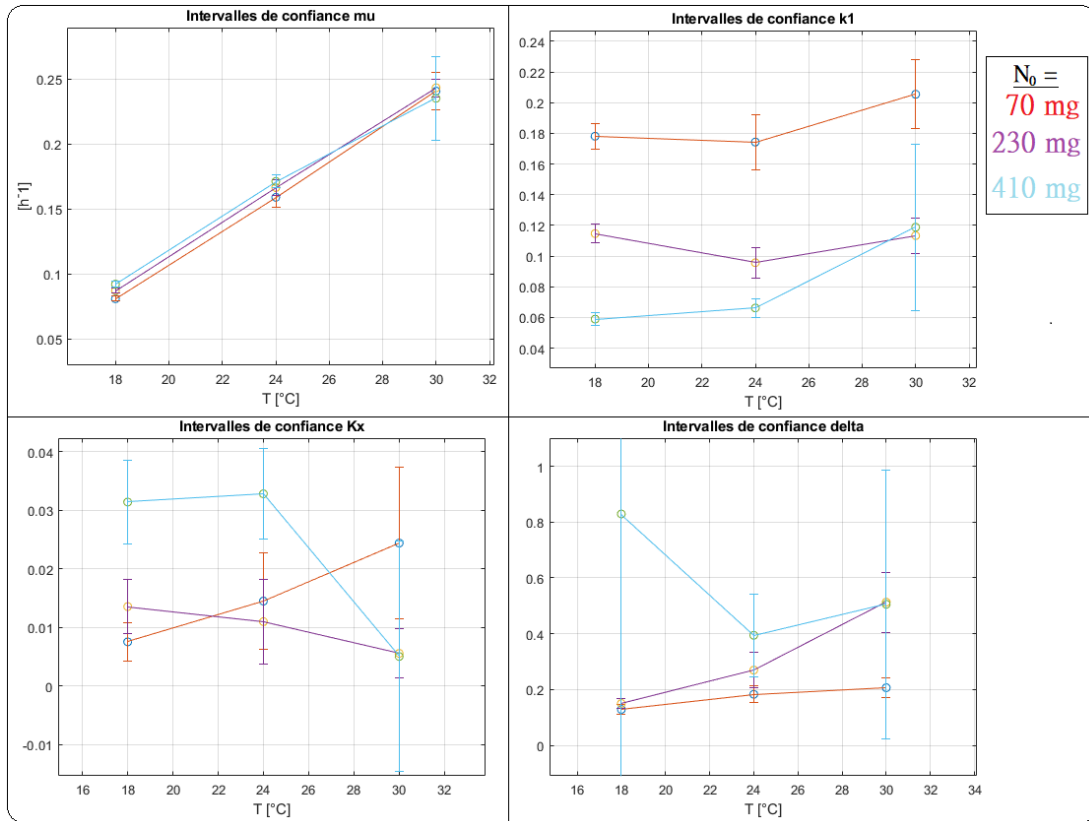


FIGURE 4.5 – Intervalles de confiance d'un écart-type σ calculés pour l'identification des paramètres μ_{max} , K_x , k_1 et δ .

On observe sur la Figure 4.5 que les estimations des paramètres μ_{max} et δ sont les plus précises alors que celles de K_x et k_1 le sont moins (généralement plus de 30% de variation sur un écart-type pour K_x). Afin de régler les problèmes de corrélation cités plus haut, les valeurs de K_x et k_1 ont été fixées car les moins sensibles (voir Table 4.1) et μ_{max} et δ ont été identifiées par rapport aux valeurs de K_x et k_1 choisies. On constate aussi que pour $N_0 = 410$ mg/l, l'estimation sur δ est très peu précise. C'est attendu puisqu'on a vu plus haut que le déphasage entre les évolutions de N_x et X disparaissait quasi complètement à ce niveau d'azote initial, δ n'a donc presque plus aucun effet et varie énormément.

En observant les valeurs de K_x et k_1 on remarque deux choses :

- Les valeurs de k_1 semblent liées à l'azote initial, ce qui peut être une conséquence du déphasage observé plus tôt. De plus, k_1 recouvre une notion de rendement fonction de N_0 et pas de la température a priori.
- Les estimations de K_x et k_1 sont peu précises pour les expériences à 30°C

	$\% \Delta J (\theta + 5\%)$	$\% \Delta J (\theta - 5\%)$
μ_{max}	307.7	86.92
K_x	10.61	23.59
k_1	42.44	17.47
δ	85.59	101.73

TABLE 4.1 – Pourcentages de variations du critère J après une variation de $\pm 5\%$ de μ_{max} , K_x , k_1 et δ .

En prenant ces deux points en compte, une étude de minimisation du critère J en fonction des valeurs de K_x et k_1 a été menée. K_x a finalement été fixé à 0.015. Les valeurs de k_1 ont été fixées à 0.18, 0.11 et 0.06 pour respectivement $N_0 = 70$, 230 et 410 mg/l. Ces valeurs de $k_1(N_0)$ sont communes aux intervalles de confiance de 2σ (95% de probabilité que la valeur correcte se situe dans cet intervalle) obtenues aux différentes températures. On constate aussi que lorsque le déphasage est absent ($N_0 = 410$ mg/l), on retrouve une valeur de k_1 proche de celle identifiée par l'ancien modèle.

La seconde identification avec K_x et k_1 fixés engendre des intervalles de confiance encore plus réduits pour μ_{max} et δ (voir Figure 4.6). Les valeurs de μ_{max} étant fort proches, il est possible d'écrire une loi de température unique pour μ_{max} en prenant une moyenne des valeurs identifiées avec des concentrations d'azote initiales différentes (voir Figure 4.13). δ quant à lui est trop influencé par l'azote initial, il faudra donc lui décrire une loi d'évolution dépendant de la température et de N_0 (voir Figure 4.12).

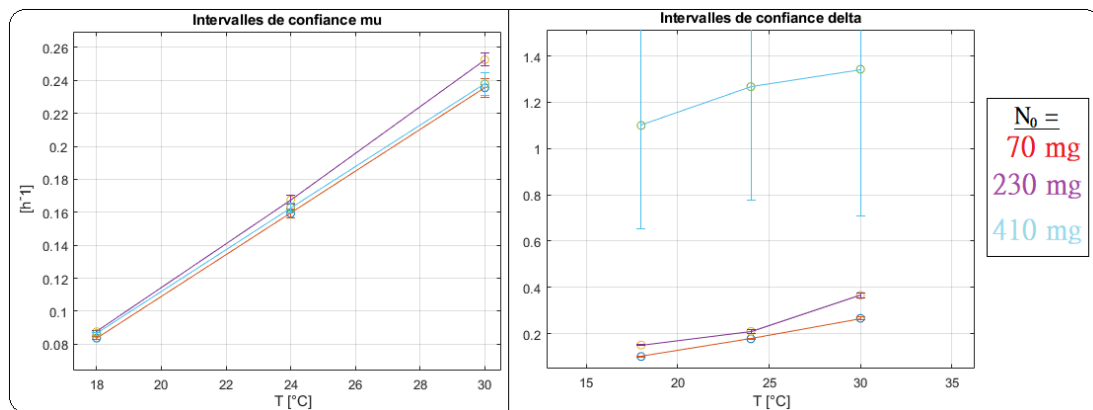


FIGURE 4.6 – Intervalles de confiance d'un écart-type σ de μ_{max} et δ , après avoir fixé K_x et k_1 .

4.4.3 Paramètres η_{max} et K_{tr}

Pour l'identification de η_{max} et K_{tr} , le critère J à minimiser est porté sur la variable N_{tr} :

$$J = \sum_{i=1}^{M_N} \frac{(N_{tr,i}(\theta) - N_{tr,i})^2}{N_{tr,i}^2} \quad (4.9)$$

avec $N_{tr,i}(\theta)$ et $N_{tr,i}$ les valeurs prédites et approximées de N_{tr} au temps t_i . A nouveau, puisque seul l'azote total N est observé, N_{tr} est approximé à partir du ratio $\frac{N_{tr,0}}{N_{x,0}}$. De plus, afin de prendre en compte le déphasage entre la consommation d'azote et la croissance de biomasse, on utilisera les valeurs de $N_{x,\delta}$ comme référence, en profitant d'avoir déjà identifié les paramètres μ_{max} , K_x , k_1 et δ :

$$N_{tr,i} = N_{x,\delta,i} \cdot \frac{N_{tr,0}}{N_{x,0}} \quad (4.10)$$

avec $N_{x,\delta,i}$ la valeur de $N_{x,\delta}$ prédite au temps t_i .

Analyse des fonctions de sensibilité

Dans le cas de η_{max} et K_{tr} , on observe sur la [Figure 4.7](#) une forte corrélation entre ces deux paramètres. Pour les identifier, il sera donc nécessaire de fixer l'un des deux paramètres. On voit sur la [Table 4.2](#) que K_{tr} est le moins sensible de ces deux paramètres.

	$\% \Delta J (\theta + 5\%)$	$\% \Delta J (\theta - 5\%)$
η_{max}	54.33	49.10
K_{tr}	24.52	27.37

TABLE 4.2 – Pourcentages de variations du critère J après une variation de η_{max} et K_{tr} .

Cependant, l'identification de η_{max} et K_{tr} ne repose sur aucune base solide puisqu'on ne dispose d'aucune véritable donnée sur les variables Tr et N_{tr} , seulement leurs points initiaux et finaux. Leurs comportements transitoires sont seulement des approximations, basées sur les évolutions de X et N_x .

Malgré cela, un bon ajustement de ces approximations est obtenu avec des valeurs de η_{max} et K_{tr} proches de celles utilisées pour le modèle de référence. Ainsi,

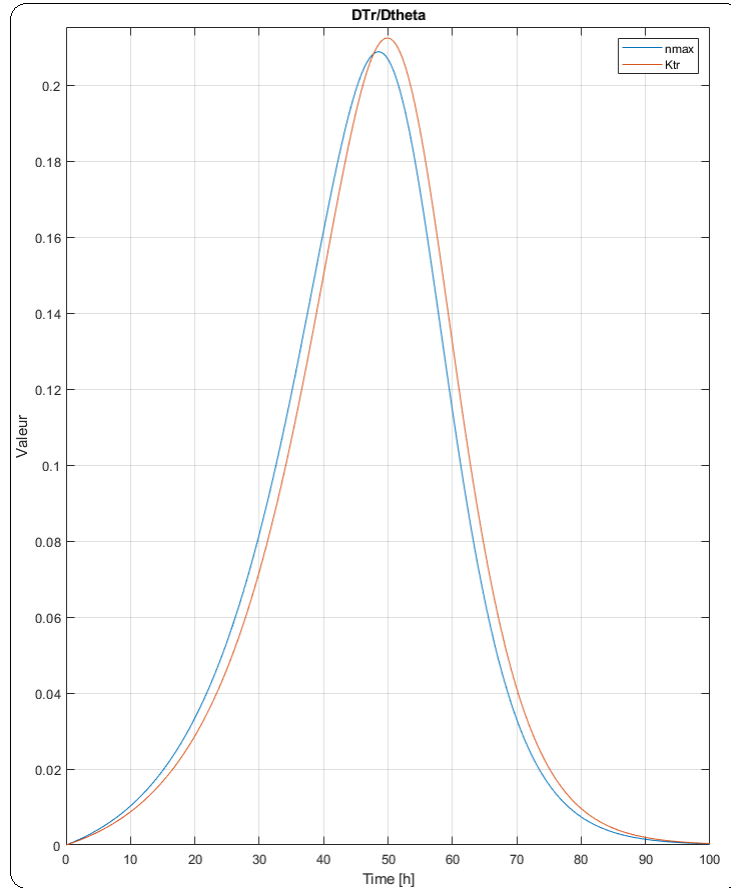


FIGURE 4.7 – Fonctions normées de sensibilité d' η_{max} et K_{tr} , pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$

K_{tr} est fixé à 0.3 et une loi de η_{max} en fonction de la température peut être déduite (voir Figure 4.14).

4.4.4 Paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ

L'identification des paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ se fait avec un critère J qui porte sur les données du flux de CO_2 :

$$J = \sum_{i=1}^{M_C} \frac{(C\dot{O}_{2,i}(\theta) - C\dot{O}_{2,i})^2}{C\dot{O}_{2,i}^2} \quad (4.11)$$

avec M_C le nombre de mesures du flux de CO_2 , $C\dot{O}_{2,i}(\theta)$ et $C\dot{O}_{2,i}$ les valeurs prédites et mesurées du flux de CO_2 au temps t_i .

De plus, la condition initiale du flux de CO_2 , $C\dot{O}_{2,0} = C\dot{O}_{2}(t = 0)$, n'est pas connue à l'avance et celle-ci dépend des valeurs des paramètres à identifier. Cette condition initiale doit donc être identifiée simultanément avec les paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ .

Analyse des fonctions de sensibilité

On observe sur la [Figure 4.8](#) que les paramètres $\beta_{max,2}$, K_E et ϕ sont fortement corrélés, tandis que $\beta_{max,1}$ et K_S semblent indépendants. Parmi ces paramètres corrélés, K_E et ϕ sont les moins sensibles (voir [Table 4.3](#)). Ce sont donc ces paramètres que l'on va décider de fixer à des valeurs proches de celles précédemment identifiées.

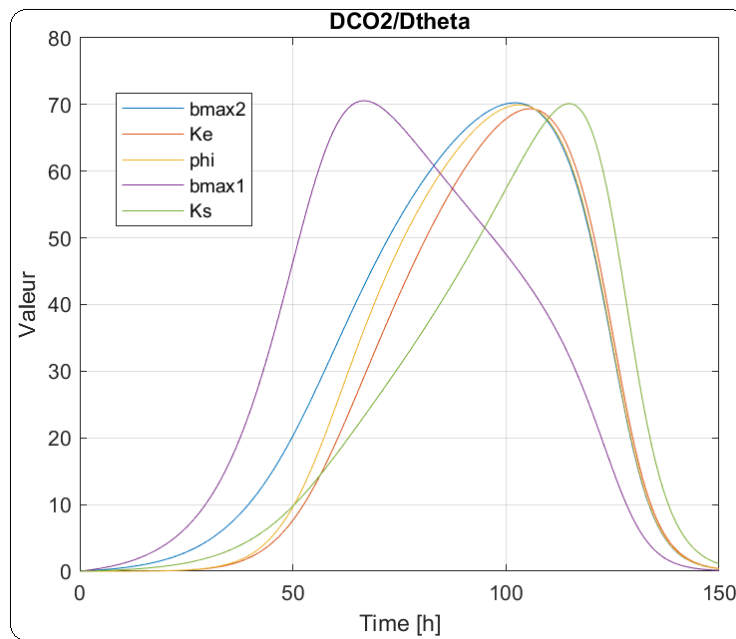


FIGURE 4.8 – Fonctions de sensibilité normées des paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ , pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

	$\% \Delta J (\theta + 5\%)$	$\% \Delta J (\theta - 5\%)$
$\beta_{max,1}$	16.61	14.15
$\beta_{max,2}$	99.03	62.81
K_E	32.05	44.2
K_S	4.99	3.42
ϕ	2.26	2.27

TABLE 4.3 – Pourcentages de variations du critère J après une variation de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ .

Intervalles de confiance

Après avoir défini les lois d'évolution de K_E et de ϕ à des valeurs proches de celles précédemment identifiées (voir Figure 4.15), des lois d'évolutions de $\beta_{max,1}$, $\beta_{max,2}$ et K_S se dégagent déjà (voir Figure 4.9).

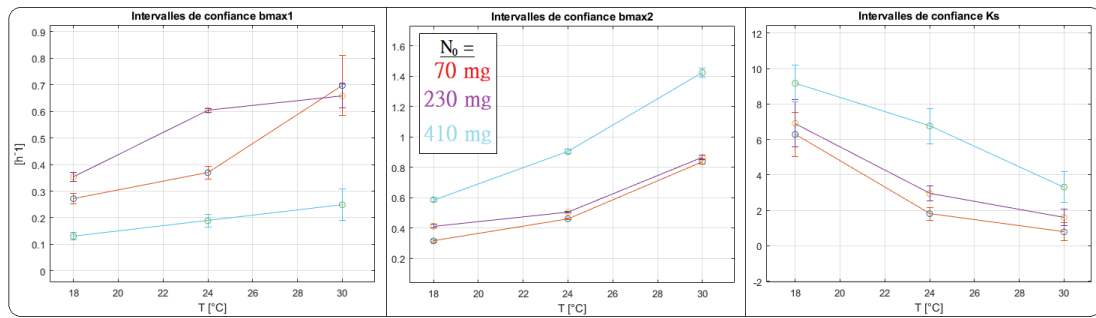


FIGURE 4.9 – Intervalles de confiance d'un écart-type σ des paramètres $\beta_{max,1}$, $\beta_{max,2}$ et K_S , 1^{ère} identification.

Cependant, ces lois d'évolutions semblent varier en fonction de N_0 , ce qui n'est pas prévu par le modèle. Puisque l'estimation de K_S est la moins précise (Table 4.3), on définit sa loi d'évolution comme étant une moyenne, pondérée en fonction de la variation du critère J , des estimations précédentes, afin d'avoir une loi d'évolution qui est uniquement fonction de la température (voir Figure 4.15). Une nouvelle identification des 2 paramètres restants montre que les évolutions de $\beta_{max,1}$ et $\beta_{max,2}$ continuent d'être fonctions de N_0 (voir Figure 4.10).

A nouveau, on définit une loi d'évolution moyenne, pondérée par rapport aux variations du critère J , pour $\beta_{max,1}$, dont les écart-types sont les plus grands. L'identification finale de $\beta_{max,2}$, sur la Figure 4.11, indique que cette fois les valeurs estimées forment bien une loi d'évolution uniquement fonction de la température. On peut donc approximer celle-ci avec la moyenne des valeurs identifiées (voir

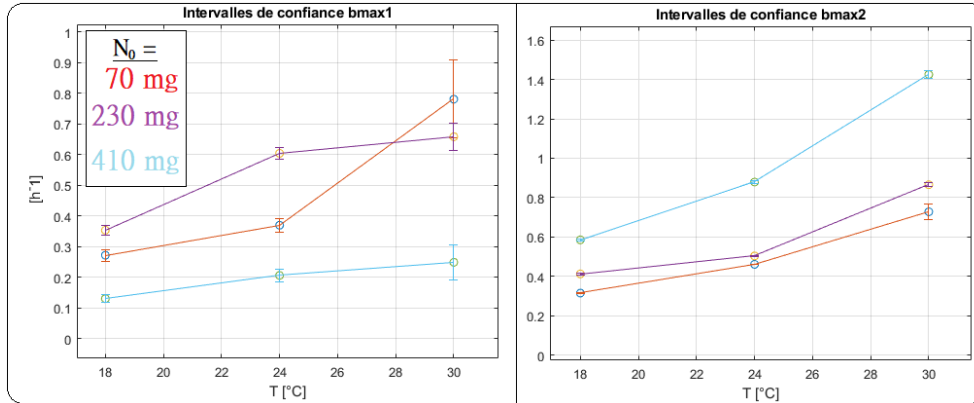


FIGURE 4.10 – Intervalles de confiance d’un écart-type σ des paramètres $\beta_{max,1}$ et $\beta_{max,2}$, 2^{ème} identification.

Figure 4.15).

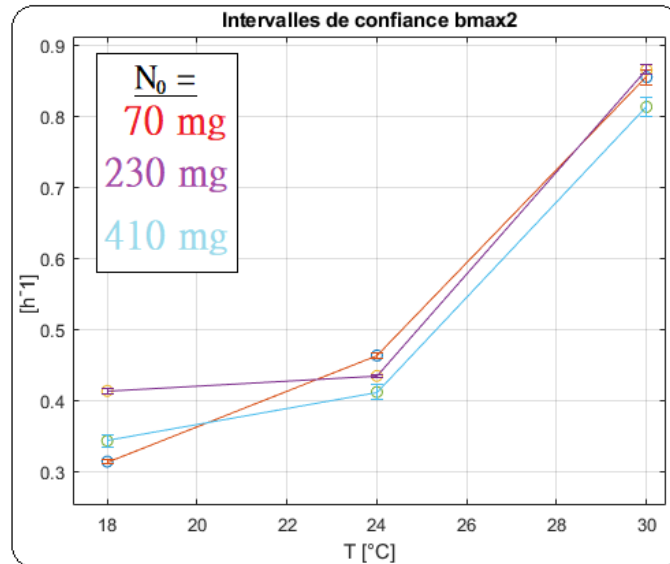


FIGURE 4.11 – Intervalles de confiance d’un écart-type σ de $\beta_{max,2}$, 3^{ème} identification.

4.5 Résumé de la calibration

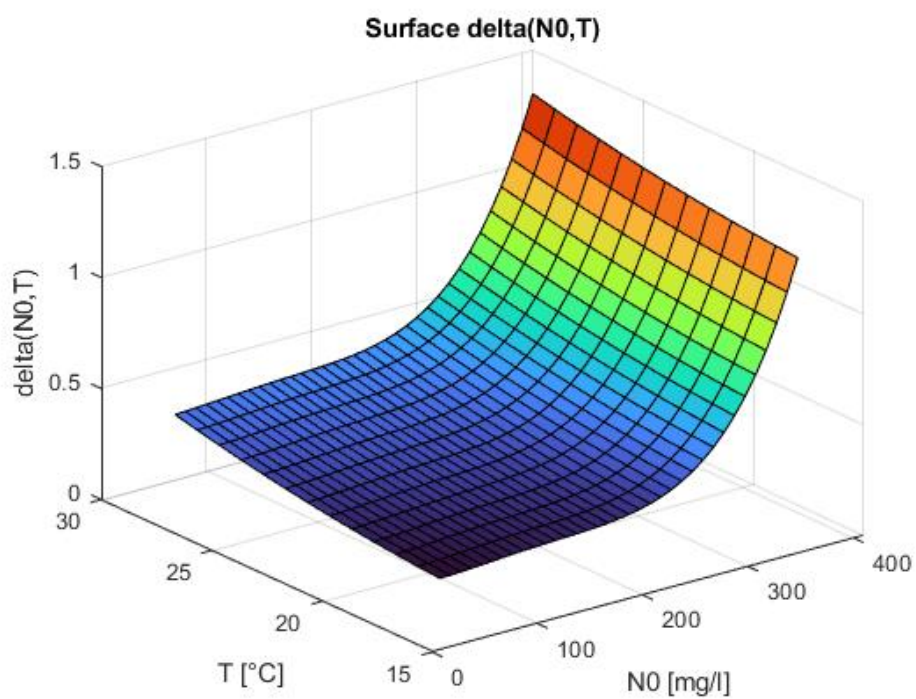
Des lois d’évolution de chaque paramètre ont été définies, indiquées dans la Table 4.4. Ces lois sont généralement fonction de la température et sont assez

similaires de celles identifiées par le modèle de référence. Il y a deux exceptions : k_1 est fonction de N_0 et delta est fonction de N_0 et T . Ces exceptions sont acceptées car nécessaires à cause du phénomène de déphasage entre la consommation d'azote et la croissance de la biomasse, observé dans la section 4.2.2 et qui est à l'origine du paramètre δ . Les lois d'évolutions elles-mêmes sont observables aux Figures 4.13, 4.12, 4.14 et 4.15 . Les coefficients des polynômes formant les lois d'évolutions peuvent être consultés à l'ANNEXE IV

Les Figures 4.16, 4.17, 4.18, 4.19, 4.20, 4.21, 4.22, 4.23 et 4.24 montrent les courbes fournies par le modèle final pour chaque batch des données de calibration. De manière générale, les courbes des différentes variables sont bien représentées, bien que lorsque $N_0 = 410\text{mg/l}$, le pic de flux de CO_2 simulé a tendance à être trop élevé.

Paramètre	Plage de valeurs	Unités
T	18 - 30	$^{\circ}\text{C}$
N_0	70 - 410	mg/l
$\mu_{max}(T)$	0.0865 - 0.2445	h^{-1}
K_x	0.015	-
$k_1(N_0)$	0.18 - 0.06	-
$\delta(N_0, T)$	0.1052 - 1.5682	-
$\eta_{max}(T)$	0.0872 - 0.2281	h^{-1}
K_{tr}	0.3	g/l
$\beta_{max,1}(T)$	0.3538 - 0.5287	h^{-1}
$\beta_{max,2}(T)$	0.335 - 0.84	h^{-1}
$K_E(T)$	45 - 26	g/l
$K_S(T)$	7.2 - 1.8	g/l
$\phi(T)$	0.4 - 0.5	$(g/l)^{-1}$

TABLE 4.4 – Lois d'évolution des paramètres.

FIGURE 4.12 – Loi d'évolution de $\delta(N_0, T)$.

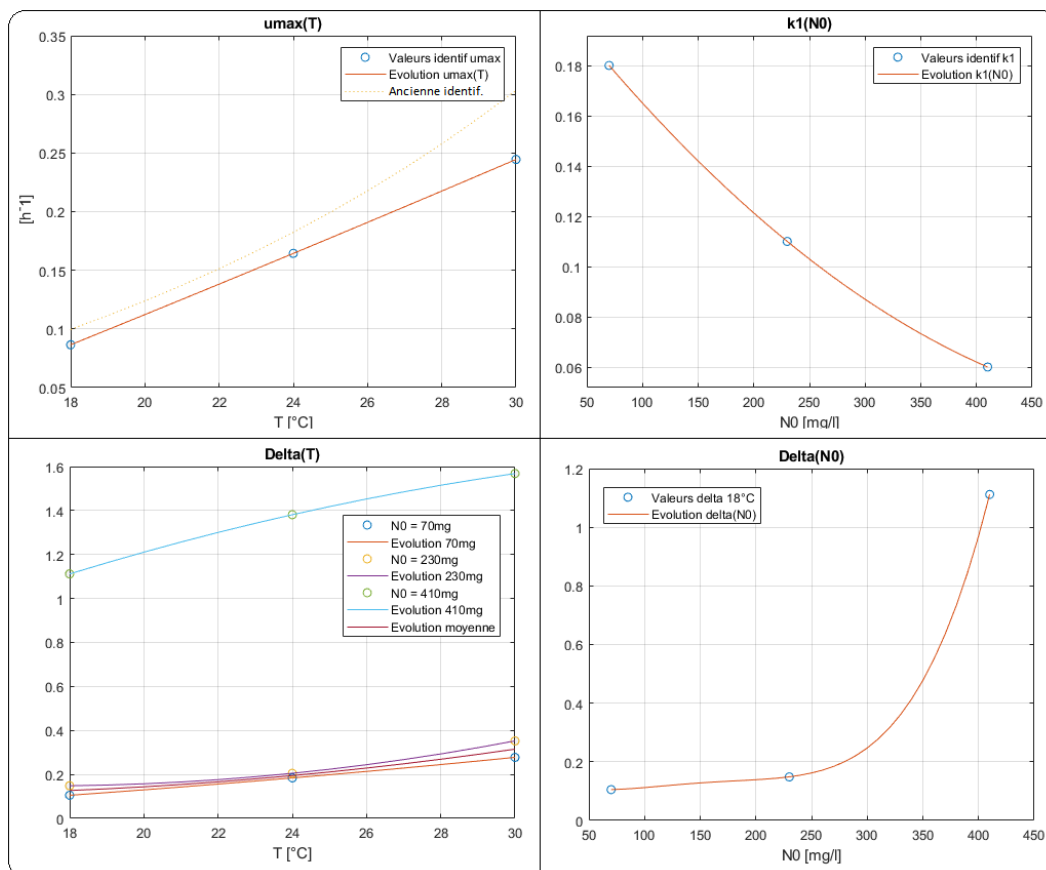


FIGURE 4.13 – Lois d'évolution de $\mu_{max}(T)$, $k_1(N_0)$ et δ . $\delta(N_0)$ détermine l'ordonnée à l'origine de la courbe $\delta(T)$, les deux forment donc ensemble $\delta(N_0, T)$ (Figure 4.12).

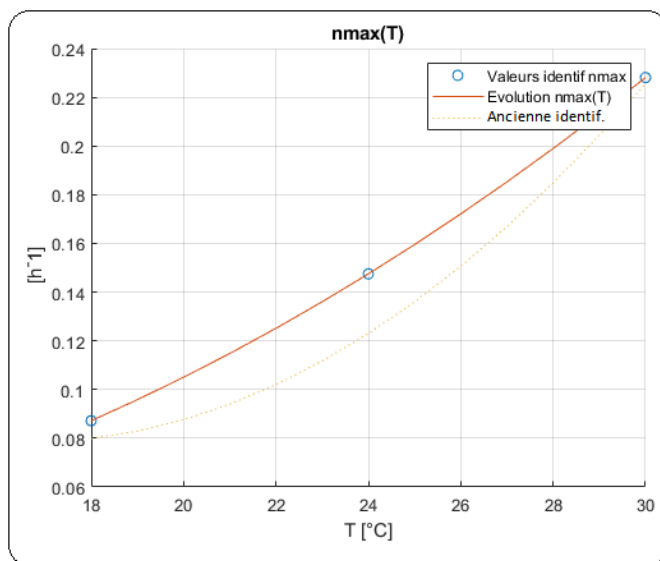


FIGURE 4.14 – Loi d'évolution d' $\eta_{max}(T)$

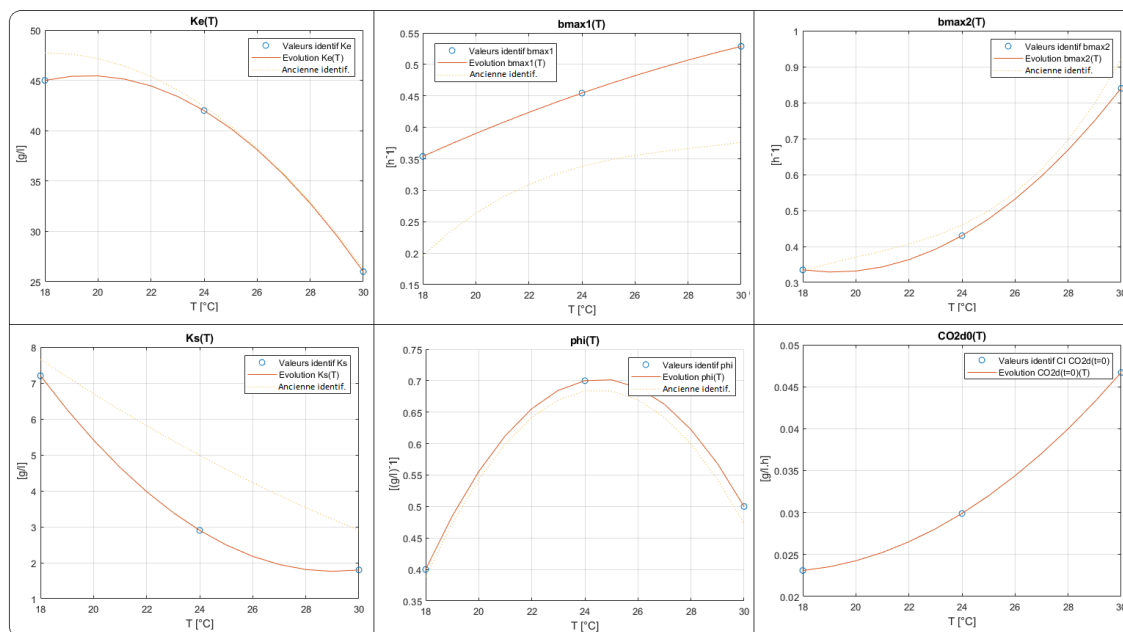


FIGURE 4.15 – Lois d'évolution de $\beta_{max,1}(T)$, $\beta_{max,2}(T)$, $K_E(T)$, $K_S(T)$ et $\phi(T)$ ainsi que de la condition initiale du flux de CO_2 , $CO_{2,0}(T)$.

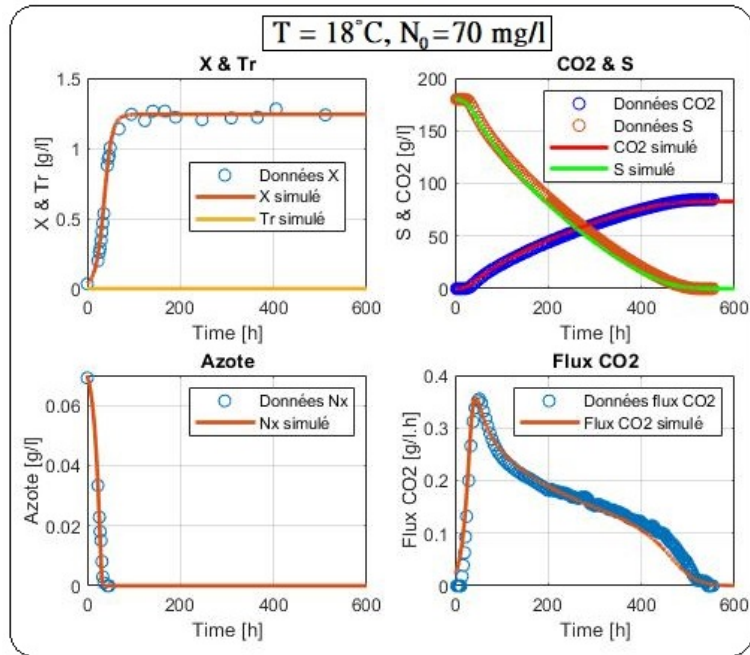


FIGURE 4.16 – Résultats finaux pour $T = 18^\circ\text{C}$ et $N_0 = 70\text{ mg/l}$.

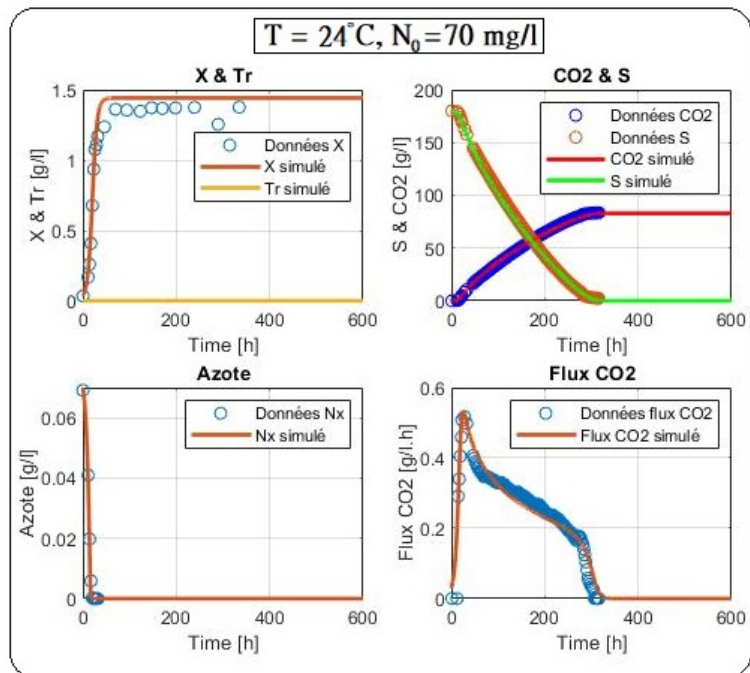


FIGURE 4.17 – Résultats finaux pour $T = 24^\circ\text{C}$ et $N_0 = 70\text{ mg/l}$.

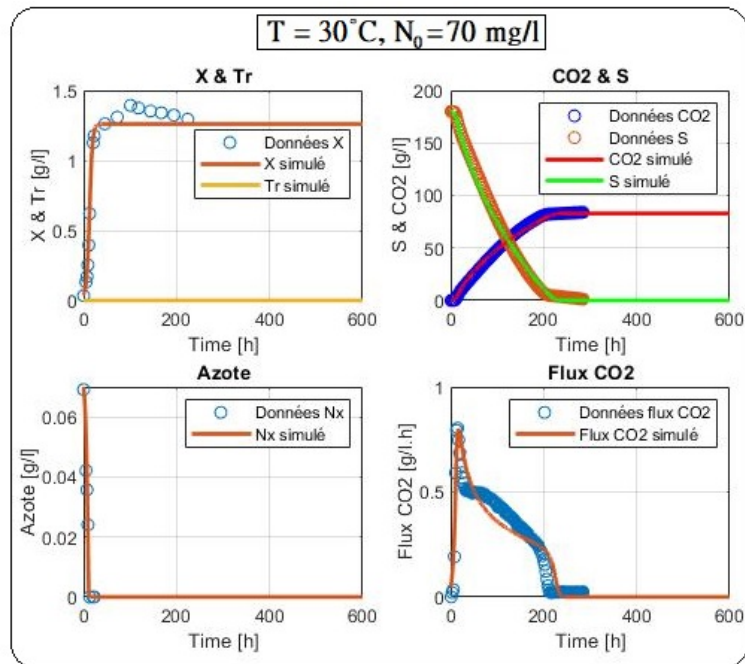


FIGURE 4.18 – Résultats finaux pour $T = 30^\circ\text{C}$ et $N_0 = 70\text{mg/l}$.

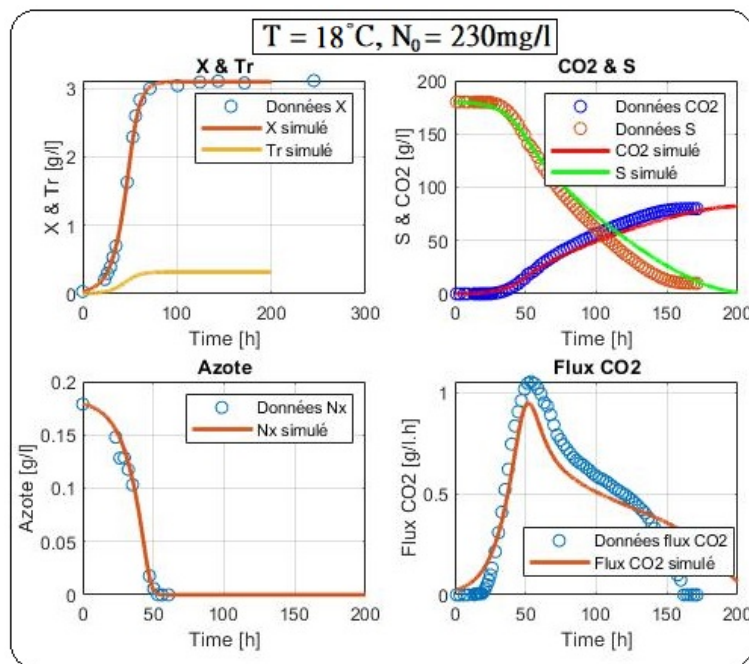
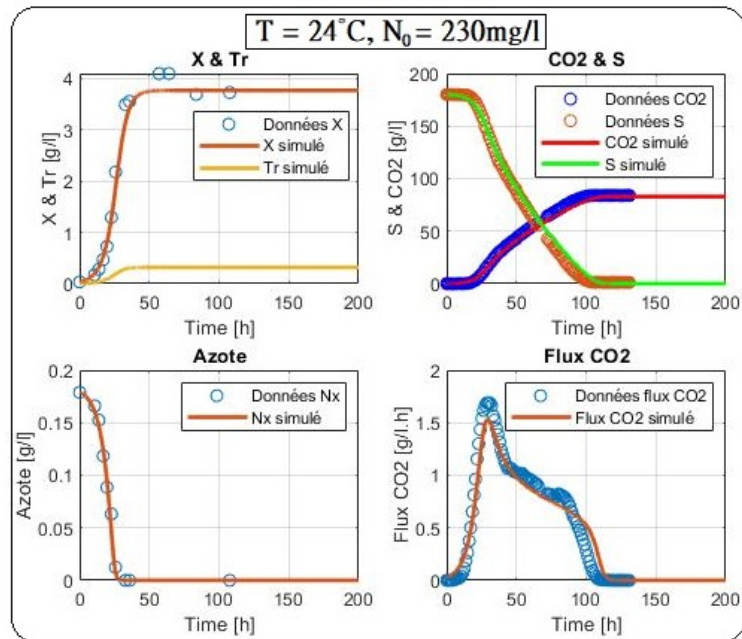
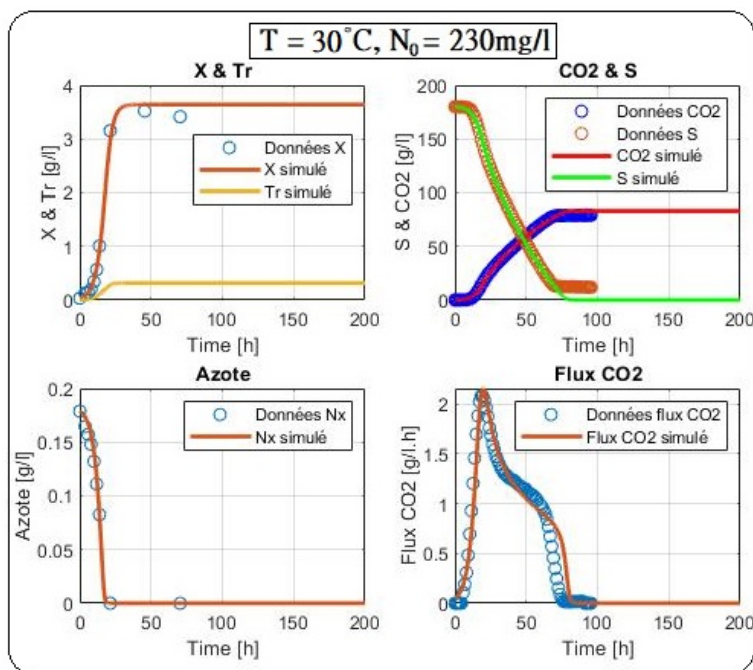
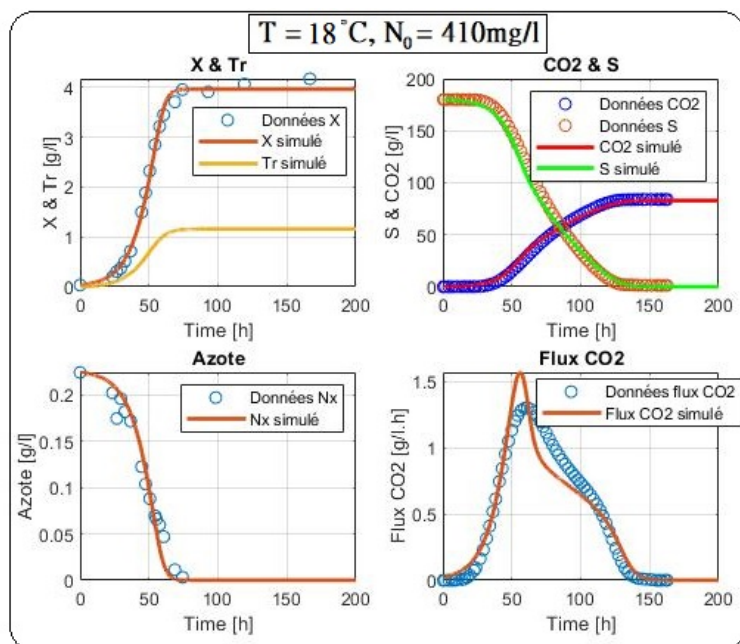
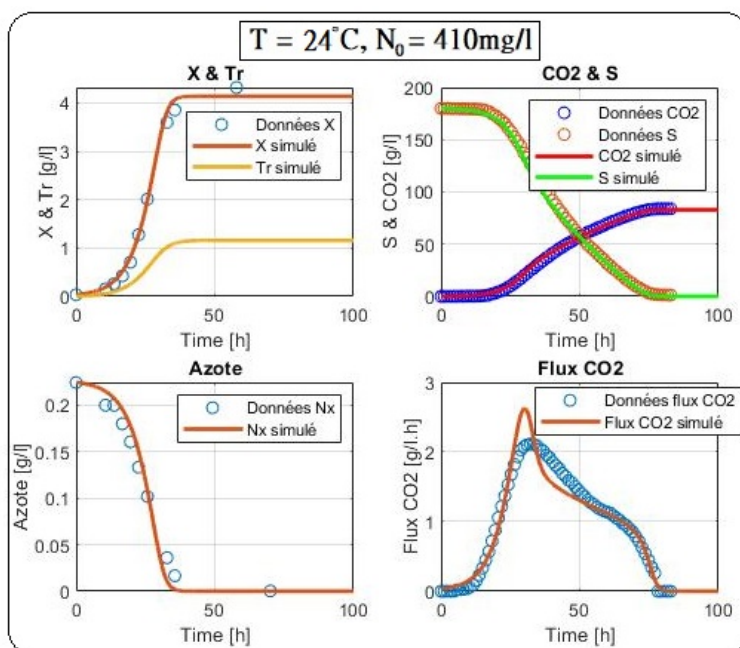
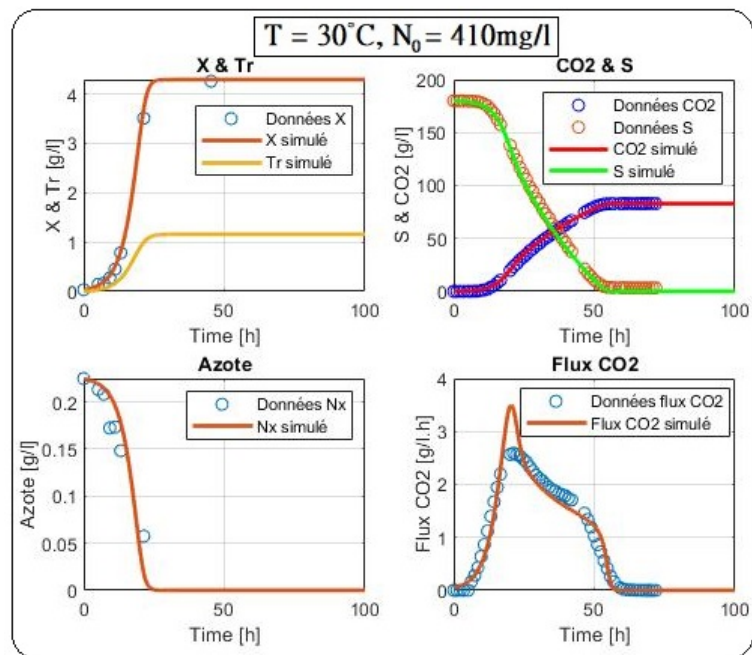


FIGURE 4.19 – Résultats finaux pour $T = 18^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

FIGURE 4.20 – Résultats finaux pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.FIGURE 4.21 – Résultats finaux pour $T = 30^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

FIGURE 4.22 – Résultats finaux pour $T = 18^\circ\text{C}$ et $N_0 = 410\text{mg/l}$.FIGURE 4.23 – Résultats finaux pour $T = 24^\circ\text{C}$ et $N_0 = 410\text{mg/l}$.

FIGURE 4.24 – Résultats finaux pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$.

4.6 Validation

La validation a été menée en comparant les résultats du modèle avec les données expérimentales conservées pour cette étape (voir [Figure 4.1](#), à la section 4.1.1). Puisque ces jeux de données n'étaient pas complets (pas d'informations sur N ni X), la comparaison s'est uniquement faite par rapport à la courbe de flux de CO_2 , qui de toute façon est la variable qui représente le mieux l'activité de la levure.

La conclusion de cette phase de validation est plutôt positive, on observe à la [Figure 4.25](#) le comparatif entre prédiction et données expérimentales. De manière générale, le pic de flux de CO_2 est assez bien représenté. Lorsqu'il est trop (ou pas assez) intense par rapport à celui observé expérimentalement, l'erreur se propage à la phase d'inhibition puisque l'intégrale de la courbe doit être égale à la quantité de CO_2 totale produite.

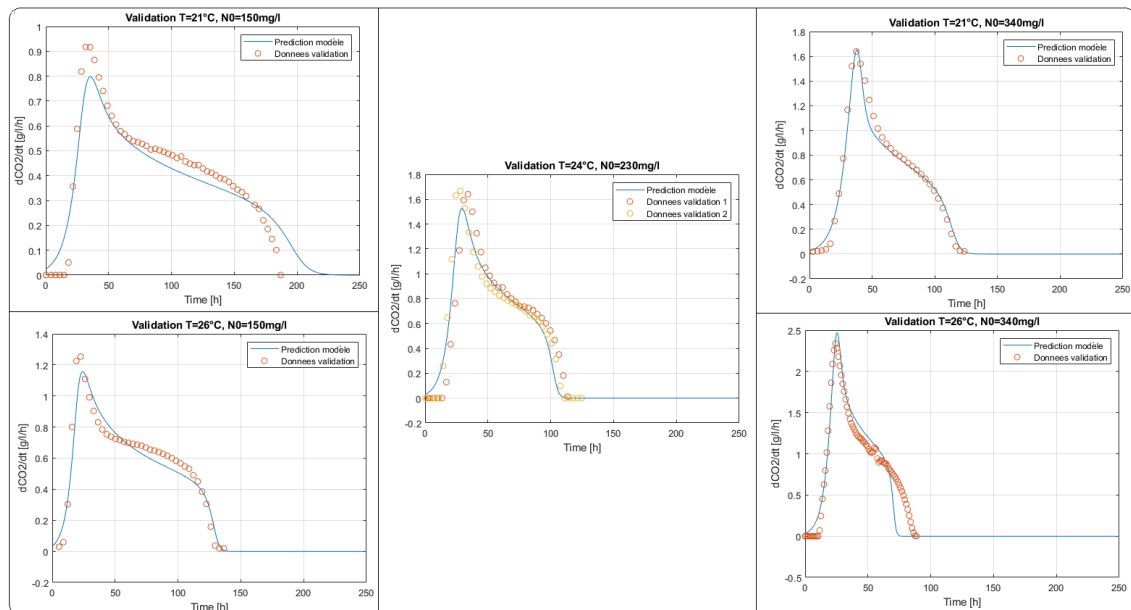


FIGURE 4.25 – Validation des lois d'évolution des paramètres identifiés.

4.7 Molécules aromatiques

Le dernier objectif de ce travail est d'étendre le modèle à certains marqueurs aromatiques. Les molécules aromatiques du vin sont très nombreuses, en concentrations diverses, mais ne représentent au total que 4% de la production des cellules

de levure. Parmi ces molécules, on retrouve des alcools supérieurs, des acides gras, des esters ou encore des composés sulfurés qui participent au parfum du vin. La [Figure 4.26](#) reprend schématiquement les voies métaboliques de ces composés.

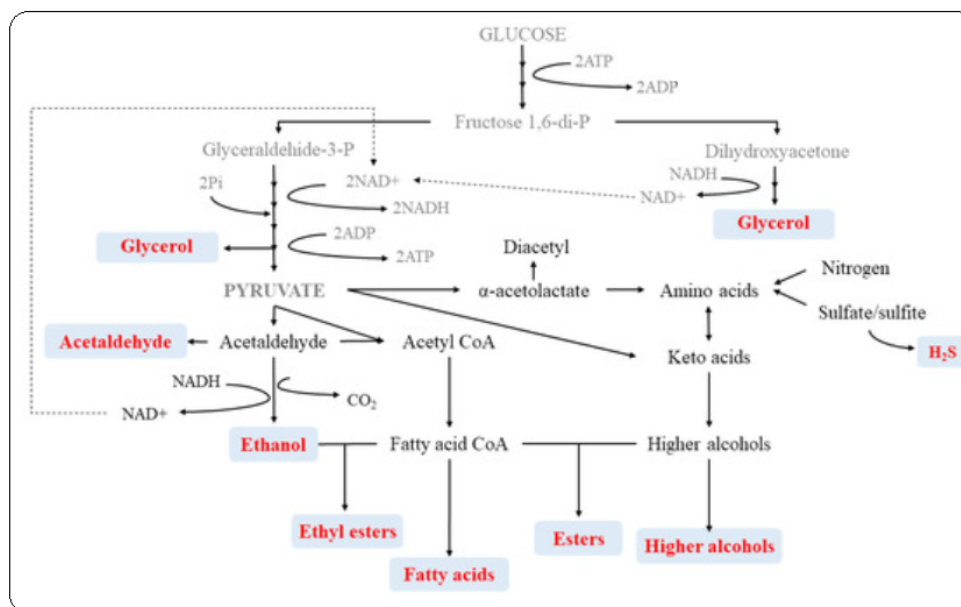


FIGURE 4.26 – Schéma des voies métaboliques des principaux composés aromatiques du vin [46].

Cependant, seules 5 molécules, qui se sont avérées mesurables expérimentalement, font l'objet de cette étude : l'isobutanol, l'isoamyl alcool, l'isoamyl acétate, l'éthyl hexanoate et l'éthyl octanoate.

4.7.1 Description des courbes de production des arômes

Dans un premier temps, on observe les données expérimentales reçues afin de tenter de discerner des caractéristiques aux courbes d'évolution de chaque molécule. Sur la [Figure 4.27](#), on observe que les productions d'arômes réagissent différemment par rapport au changement de conditions initiales. La production des alcools a surtout tendance à varier avec la température alors que la production d'esters varie en fonction de N_0 et de T , à l'exception de l'isoamyl acétate qui semble être majoritairement affecté par N_0 .

Ensuite, on peut observer ces courbes individuellement et faire le parallèle avec d'autres variables importantes, telles que la biomasse, le sucre, l'azote ou le CO_2

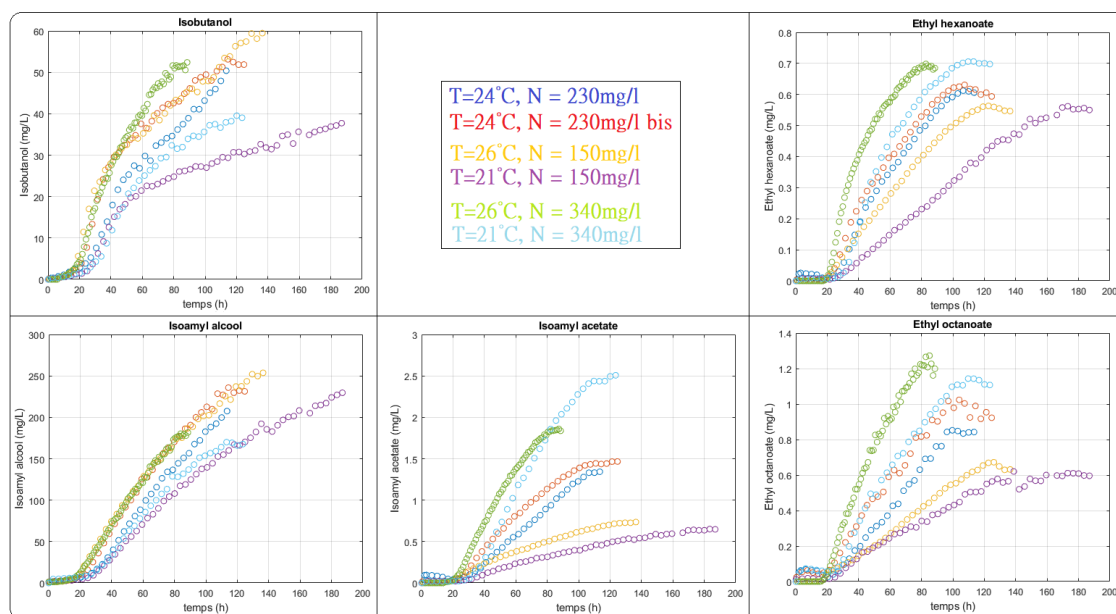


FIGURE 4.27 – Données aromatiques.

pour observer si des moments clés de leur évolution (plateaux, changements de pente,...) coïncident avec des moments clés de ces variables.

Sur les Figures 4.28, 4.29, 4.30, 4.31 et 4.32, on peut observer le comparatif entre les évolutions normées des arômes (respectivement l'isobutanol, l'isoamyl alcool, l'isoamyl acétate, l'éthyl hexanoate et l'éthyl octanoate) et des autres variables mesurées.

On remarque que les productions de ces arômes ont au final toutes un comportement assez semblable de celui de la production de CO_2 , avec un début de croissance presque simultané, un comportement transitoire similaire et un plateau lorsque le flux de CO_2 devient nul. Ces observations restent vraies lorsque T et N_0 varient.

On note cependant que les éthyls hexanoate et octanoate évoluent de manière décroissante lorsque le flux de CO_2 devient nul au lieu de former un plateau. Cela indique soit que la réaction s'inverse soit qu'un autre métabolisme entre en jeu et mobilise ces molécules.

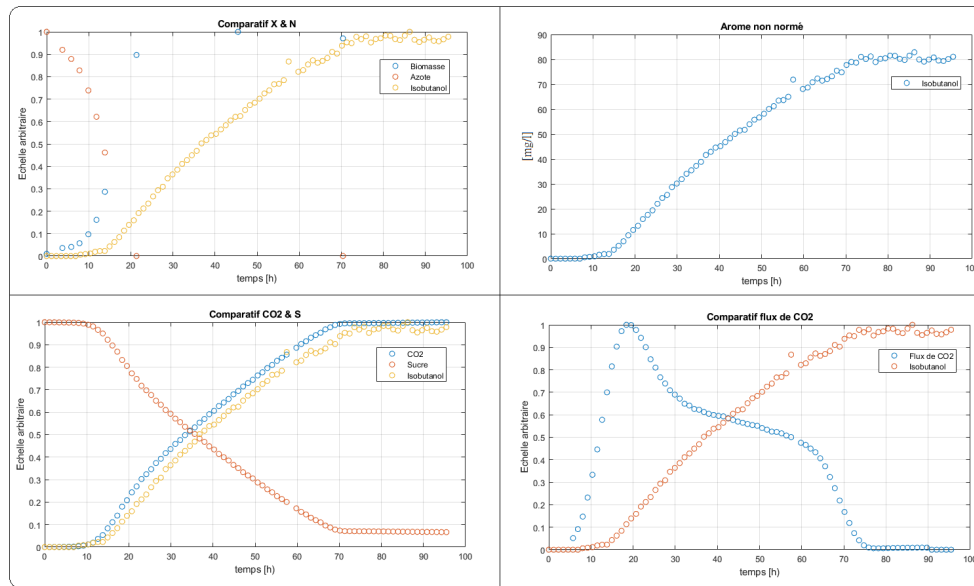


FIGURE 4.28 – Comparaison des courbes d'évolutions normées de la production d'isobutanol par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 = 230\text{mg/l}$).

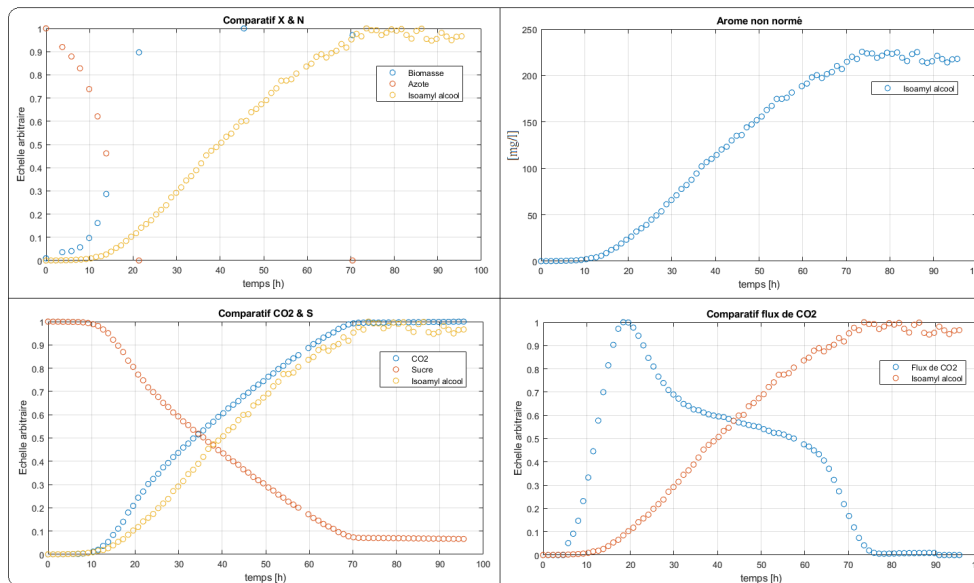


FIGURE 4.29 – Comparaison des courbes d'évolutions normées de la production d'isoamyl alcool par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 = 230\text{mg/l}$).

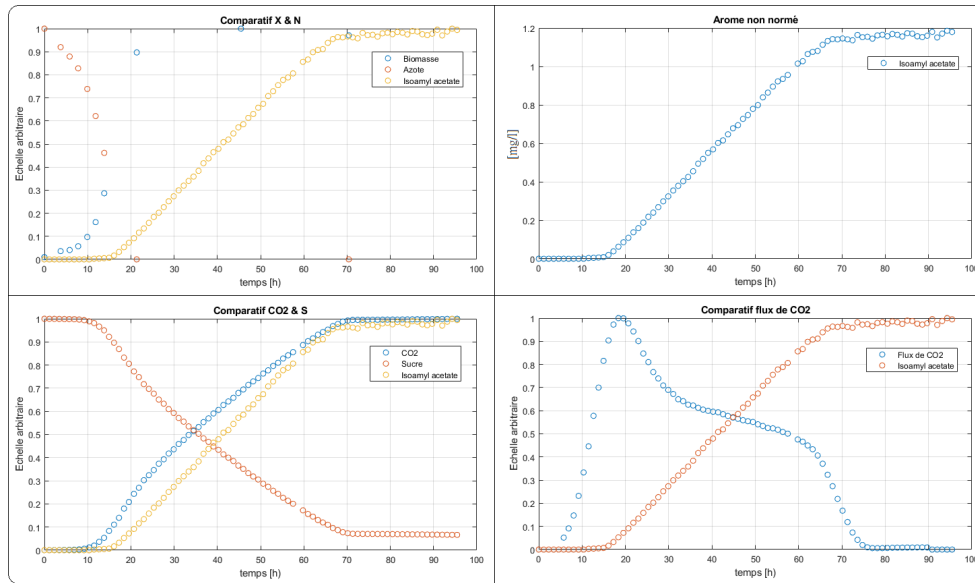


FIGURE 4.30 – Comparaison des courbes d'évolutions normées de la production d'isoamyl acétate par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 = 230\text{mg/l}$).

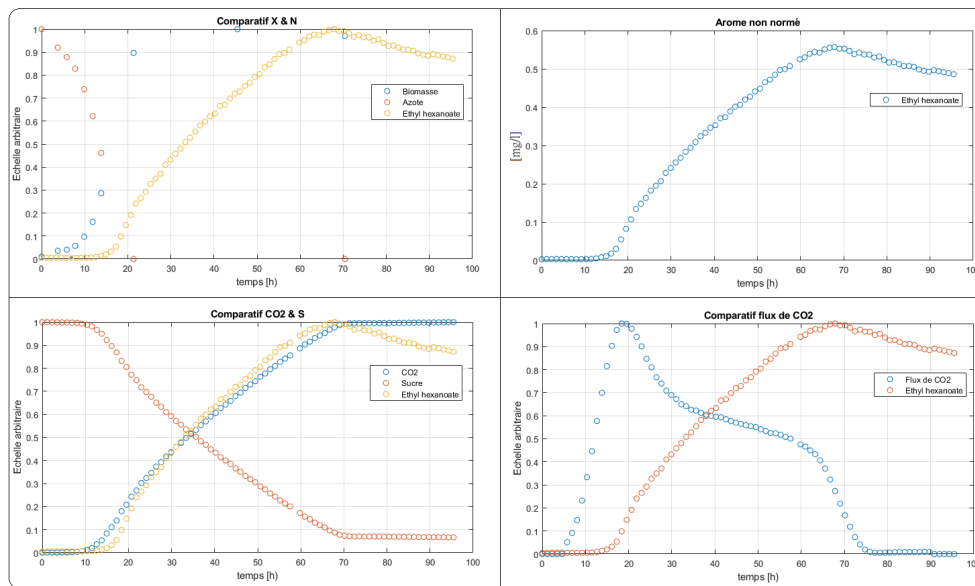


FIGURE 4.31 – Comparaison des courbes d'évolutions normées de la production d'éthyl hexanoate par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 = 230\text{mg/l}$).

4.7.2 Modélisation de la production des arômes

Finalement, en prenant en considération les éléments des deux sections précédentes, il a été établi qu'une approximation de la cinétique de production des arômes pouvait être atteinte en utilisant une expression cinétique proche de celle du CO_2 , commune à tous les arômes. Cependant, les paramètres de cette expression devraient évoluer différemment en fonction de chaque arôme, afin de prendre en compte les différentes influences des conditions initiales décrites précédemment.

Le temps a manqué pour mener cette étude à terme, cependant une première tentative d'identification de paramètres pour une loi cinétique commune à tous les arômes peut être consultée à l'**ANNEXE V**.

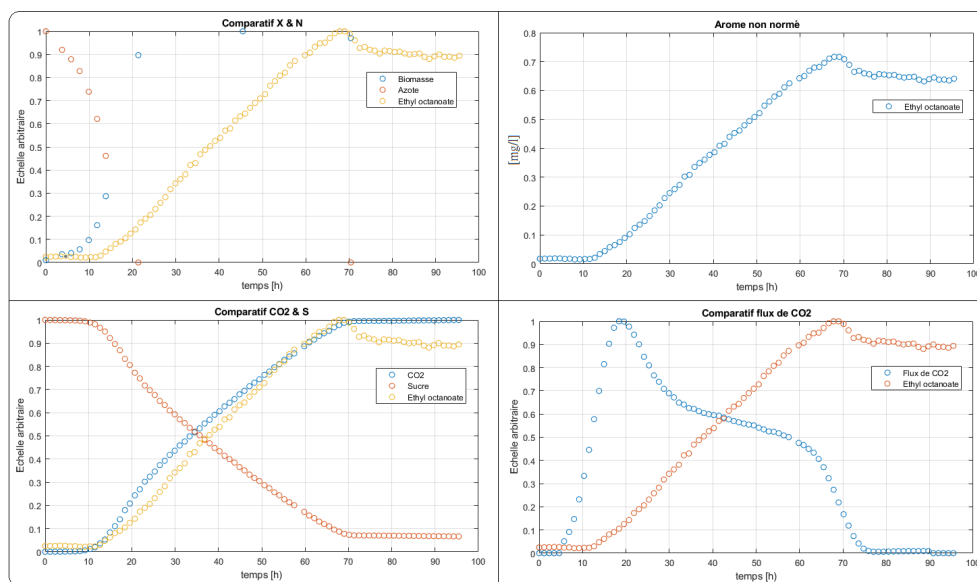


FIGURE 4.32 – Comparaison des courbes d'évolutions normées de la production d'éthyl octanoate par rapport aux autres variables ($T=30^{\circ}C$, $N_0 = 230mg/l$).

Conclusion

L'objectif de ce mémoire était d'étendre un ancien modèle de prédiction de données de fermentation à certaines composantes aromatiques du vin, à l'aide de nouvelles données expérimentales. Après une brève présentation de l'état de l'art de la modélisation des bioprocédés, la première étape de ce travail a donc été d'implémenter le modèle dans le logiciel Matlab.

Le modèle a d'abord été légèrement adapté afin d'introduire explicitement le flux de CO₂. Ensuite, une nouvelle variable a été introduite, $N_{x,\delta}$, qui représente l'azote intracellulaire de la levure. En effet, dans les nouvelles données expérimentales, un déphasage est observé entre l'azote au sein du réacteur et l'azote intracellulaire de la levure. Le déphasage à proprement parler est quant à lui représenté par le paramètre δ .

Une fois le nouveau modèle implémenté avec succès, l'identification des paramètres a suivi. Grâce à la structure en cascade du modèle, cette identification s'est effectuée par vagues, ce qui a facilité son exécution. Des lois d'évolutions des paramètres ont été définies et puis validées avec le reste des données expérimentales.

Enfin, la modélisation de la production des molécules aromatiques (isobutanol, isoamyl alcool, isoamyl acétate, éthyl hexanoate et éthyl octanoate) a été discutée. Les courbes expérimentales ont été décrites afin de tenter d'en exhiber les caractéristiques qui formeront leur modélisation. Cette modélisation n'a malheureusement pas été menée à terme, cependant une première tentative est consultable à l'**ANNEXE V**.

Par rapport aux perspectives futures, le modèle pourrait être retravaillé grâce à une meilleure compréhension du métabolisme de la levure. Les approximations

concernant les variables N_x et N_{tr} pourraient être revues. La condition initiale du flux de CO_2 pourrait également être définie à l'avance. La description du profil aromatique peut évidemment être réapprofondie et une meilleure connaissance des mécanismes de production des arômes permettrait de leur dériver des expressions cinétiques sur des fondations plus solides.

Table des figures

1.1	Schéma de réacteur à milieu mélangé homogène.	5
1.2	Schéma des types de bioréacteurs les plus fréquents.	6
1.3	Courbe de Monod, représentant la vitesse de réaction en fonction de la concentration en substrat.	11
2.1	Contour tracé par une fonction objectif mal conditionnée (gauche) et bien conditionnée (droite) dans le cas de l'estimation de deux paramètres [21].	17
2.2	Cas d'analyse de fonctions de sensibilité. On observe une corrélation entre les paramètres K_M (en rose) et V_M (en bleu), qui évoluent de manière proportionnelle [23].	18
2.3	Fonction objectif et ses minimums.	19
3.1	Distribution d'azote initial en fonction de son usage, pour $N_0 = 0.071 \rightarrow 0.57g/l$ [29].	28
3.2	Valeurs d'équilibres de X et Tr correspondant aux valeurs d'azote initial N_0 [29].	29
3.3	Gamme de conditions initiales des expériences servant à l'identification des paramètres du modèle. Encadrées en rouge, les expérience dédiées à sa validation [29].	30

4.1	Gamme de conditions initiales des expériences servant à l'identification des paramètres du modèle.	32
4.2	On observe un déphasage entre la fin de la consommation de l'azote (traits en pointillé) et la fin de la croissance de la biomasse. Ce déphasage diminue avec la température et avec la concentration initiale d'azote. A 410mg/l, ce déphasage s'est complètement atténué.	34
4.3	A gauche, distribution de l'azote initial en fonction de son usage. A droite, valeurs d'équilibres de X et Tr correspondant aux valeurs d'azote initial N_0	36
4.4	Fonctions de sensibilité des paramètres μ_{max} (mu), K_x , k_1 et δ (delta), normées afin de les comparer plus facilement, pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$	38
4.5	Intervalles de confiance d'un écart-type σ calculés pour l'identification des paramètres μ_{max} , K_x , k_1 et δ	39
4.6	Intervalles de confiance d'un écart-type σ de μ_{max} et δ , après avoir fixé K_x et k_1	40
4.7	Fonctions normées de sensibilité d' η_{max} et K_{tr} , pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$	42
4.8	Fonctions de sensibilité normées des paramètres $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ , pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$	43
4.9	Intervalles de confiance d'un écart-type σ des paramètres $\beta_{max,1}$, $\beta_{max,2}$ et K_S , 1 ^{ère} identification.	44
4.10	Intervalles de confiance d'un écart-type σ des paramètres $\beta_{max,1}$ et $\beta_{max,2}$, 2 ^{ème} identification.	45
4.11	Intervalles de confiance d'un écart-type σ de $\beta_{max,2}$, 3 ^{ème} identification.	45
4.12	Loi d'évolution de $\delta(N_0, T)$	47
4.13	Lois d'évolution de $\mu_{max}(T)$, $k_1(N_0)$ et δ . $\delta(N_0)$ détermine l'ordonnée à l'origine de la courbe $\delta(T)$, les deux forment donc ensemble $\delta(N_0, T)$ (Figure 4.12).	48
4.14	Loi d'évolution d' $\eta_{max}(T)$	49
4.15	Lois d'évolution de $\beta_{max,1}(T)$, $\beta_{max,2}(T)$, $K_E(T)$, $K_S(T)$ et $\phi(T)$ ainsi que de la condition initiale du flux de CO_2 , $CO_{2,0}(T)$	49

4.16	Résultats finaux pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	50
4.17	Résultats finaux pour $T = 24^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	50
4.18	Résultats finaux pour $T = 30^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	51
4.19	Résultats finaux pour $T = 18^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	51
4.20	Résultats finaux pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	52
4.21	Résultats finaux pour $T = 30^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	52
4.22	Résultats finaux pour $T = 18^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	53
4.23	Résultats finaux pour $T = 24^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	53
4.24	Résultats finaux pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	54
4.25	Validation des lois d'évolution des paramètres identifiés.	55
4.26	Schéma des voies métaboliques des principaux composés aromatiques du vin [46].	56
4.27	Données aromatiques.	57
4.28	Comparaison des courbes d'évolutions normées de la production d' isobutanol par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 =$ 230mg/l).	58
4.29	Comparaison des courbes d'évolutions normées de la production d' isoamyl alcool par rapport aux autres variables ($T=30^{\circ}\text{C}$, $N_0 =$ 230mg/l).	58
4.30	Comparaison des courbes d'évolutions normées de la production d' isoamyl acétate par rapport aux autres variables ($T=30^{\circ}\text{C}$, N_0 $= 230\text{mg/l}$).	59
4.31	Comparaison des courbes d'évolutions normées de la production d' éthyl hexanoate par rapport aux autres variables ($T=30^{\circ}\text{C}$, N_0 $= 230\text{mg/l}$).	59
4.32	Comparaison des courbes d'évolutions normées de la production d' éthyl octanoate par rapport aux autres variables ($T=30^{\circ}\text{C}$, N_0 $= 230\text{mg/l}$).	60
A.1	Données expérimentales pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	xvi

A.2	Données expérimentales pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	xvii
A.3	Données expérimentales pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	xvii
A.4	Quelques données aromatiques.	xviii
B.1	Comparaison des résultats de la nouvelle simulation, à droite, avec celle de référence, à gauche [29].	xix
C.1	Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	xx
C.2	Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	xxi
C.3	Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	xxi
C.4	Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	xxii
C.5	Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	xxii
C.6	Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	xxiii
C.7	Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$	xxiii
C.8	Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$	xxiv
C.9	Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$	xxiv
E.1	Valeurs identifiées de $\mu(N_0, T)$ pour l'isobutanol.	xxvii
E.2	Valeurs identifiées de $\mu(N_0, T)$ pour l'isoamyl alcool.	xxviii
E.3	Valeurs identifiées de $\mu(N_0, T)$ pour l'isoamyl acétate.	xxviii
E.4	Valeurs identifiées de $\mu(N_0, T)$ pour l'éthyl acétate.	xxix
E.5	Valeurs identifiées de $\mu(N_0, T)$ pour l'éthyl octanoate.	xxix

E.6 Comparatif entre le modèle et les données expérimentales pour l'isobutanol. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l. xxx

E.7 Comparatif entre le modèle et les données expérimentales pour l'isoamyl alcool. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l. xxxi

E.8 Comparatif entre le modèle et les données expérimentales pour l'isoamyl acétate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l. xxxii

E.9 Comparatif entre le modèle et les données expérimentales pour l'éthyl hexanoate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l. xxxiii

E.10 Comparatif entre le modèle et les données expérimentales pour l'éthyl octanoate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l. xxxiv

Bibliographie

- [1] N. R. Council *et al.*, “Putting biotechnology to work : bioprocess engineering,” 1992.
- [2] O. Bernard, “La modélisation des systèmes biologiques : allers-retours le long des fleuves qui circulent entre l’océan du réel et le lac des modèles’,” *Habilitation à Diriger les Recherches, INRIA.[34, 36]*, 2004.
- [3] G. Bastin and D. Dochain, *On-line estimation and adaptive control of bioreactors*, vol. 1. Elsevier, 2013.
- [4] C. J. Gadgil and K. Venkatesh, “Structured model for batch culture growth of lactobacillus bulgaricus,” *Journal of Chemical Technology & Biotechnology : International Research in Process, Environmental AND Clean Technology*, vol. 68, no. 1, pp. 89–93, 1997.
- [5] J. Abdollahi and S. Dubljevic, “Lipid production optimization and optimal control of heterotrophic microalgae fed-batch bioreactor,” *Chemical engineering science*, vol. 84, pp. 619–627, 2012.
- [6] J. L. Hjersted and M. A. Henson, “Optimization of fed-batch saccharomyces cerevisiae fermentation using dynamic flux balance models,” *Biotechnology progress*, vol. 22, no. 5, pp. 1239–1248, 2006.
- [7] L. Goswami, R. V. Kumar, S. N. Borah, N. A. Manikandan, K. Pakshirajan, and G. Pugazhenthii, “Membrane bioreactor and integrated membrane bioreactor systems for micropollutant removal from wastewater : a review,” *Journal of water process engineering*, vol. 26, pp. 314–328, 2018.

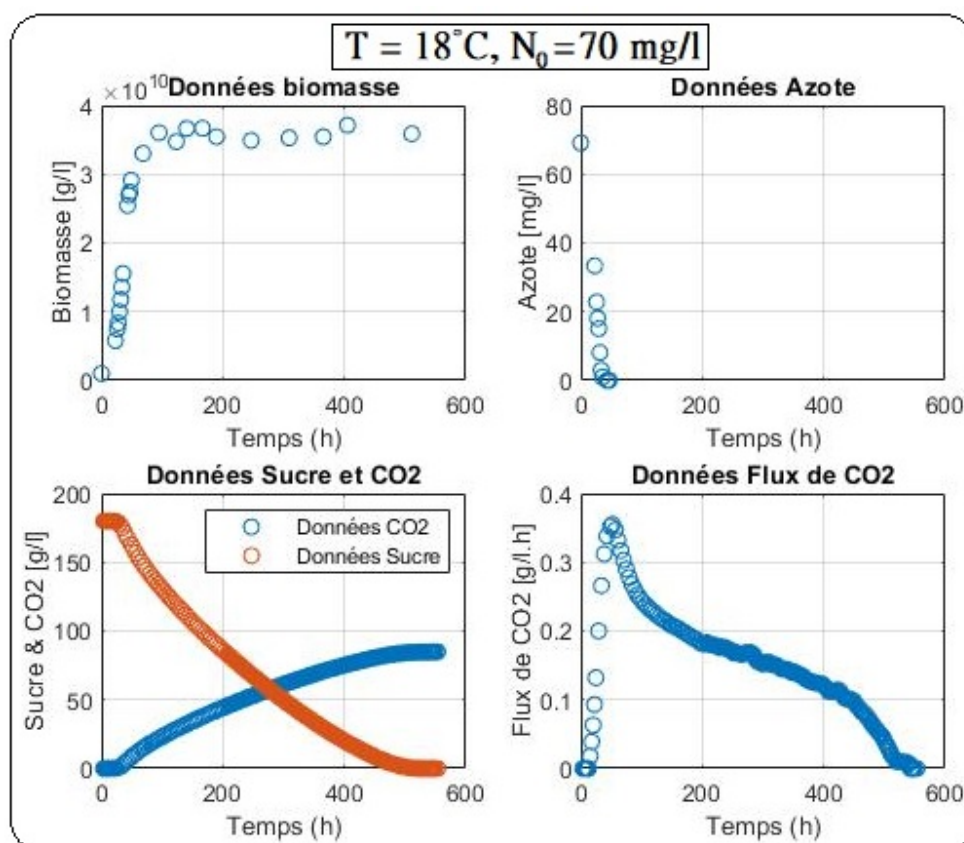
-
- [8] K. Raman and N. Chandra, “Flux balance analysis of biological systems : applications and challenges,” *Briefings in bioinformatics*, vol. 10, no. 4, pp. 435–449, 2009.
- [9] A. L. Koch, “The monod model and its alternatives,” in *Mathematical modeling in microbial ecology*, pp. 62–93, Springer, 1998.
- [10] O. Demin and I. Goryanin, *Kinetic modelling in systems biology*. CRC press, 2008.
- [11] N. r. Bonastre, C. de Mas, and C. Solà, “Vavilin equation in kinetic modeling of formaldehyde biodegradation,” *Biotechnol. Bioeng. ;(United States)*, vol. 28, no. 4, 1986.
- [12] K.-H. Bellgardt, “Bioprocess models,” in *Bioreaction Engineering*, pp. 44–105, Springer, 2000.
- [13] P. Li and Q. D. Vu, “Identification of parameter correlations for parameter estimation in dynamic biological models,” *BMC systems biology*, vol. 7, no. 1, pp. 1–12, 2013.
- [14] J. Schoukens and R. Pintelon, *Identification of linear systems : a practical guideline to accurate modeling*. Elsevier, 2014.
- [15] K. G. Gadkar, R. Gunawan, and F. J. Doyle, “Iterative approach to model identification of biological networks,” *BMC bioinformatics*, vol. 6, no. 1, pp. 1–20, 2005.
- [16] P. Vanrolleghem and D. Dochain, “Bioprocess model identification,” in *Advanced instrumentation, data interpretation, and control of biotechnological processes*, pp. 251–318, Springer, 1998.
- [17] D. Dochain, *Automatic control of bioprocesses*. John Wiley & Sons, 2013.
- [18] A. Munack, “Optimal feeding strategy for identification of monod-type models by fed-batch experiments,” in *Computer applications in fermentation technology : Modelling and control of biotechnological processes*, pp. 195–204, Springer, 1989.
- [19] K. Godfrey and J. DiStefano III, “Identifiability of model parameter,” *IFAC Proceedings Volumes*, vol. 18, no. 5, pp. 89–114, 1985.
- [20] E. Kyriakides and G. T. Heydt, “Calculating confidence intervals in parameter estimation : a case study,” *IEEE transactions on power delivery*, vol. 21, no. 1, pp. 508–509, 2005.

-
- [21] B. Petersen, *Calibration, identifiability and optimal experimental design of activated sludge models*. PhD thesis, Ghent University, 2000.
- [22] J. A. Robinson, “Determining microbial kinetic parameters using nonlinear regression analysis,” *Advances in microbial ecology*, pp. 61–114, 1985.
- [23] A. D. Rodman and D. I. Gerogiorgis, “Parameter estimation and sensitivity analysis for dynamic modelling and simulation of beer fermentation,” *Computers & Chemical Engineering*, vol. 136, p. 106665, 2020.
- [24] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom, “Systems biology : parameter estimation for biochemical models,” *The FEBS journal*, vol. 276, no. 4, pp. 886–902, 2009.
- [25] R. Hooke and T. A. Jeeves, ““direct search” solution of numerical and statistical problems,” *Journal of the ACM (JACM)*, vol. 8, no. 2, pp. 212–229, 1961.
- [26] D. F. Slezak, C. Suárez, G. A. Cecchi, G. Marshall, and G. Stolovitzky, “When the optimal is not the best : parameter estimation in complex biological models,” *PloS one*, vol. 5, no. 10, p. e13283, 2010.
- [27] D. Sornette, A. Davis, K. Ide, K. Vixie, V. Pisarenko, and J. Kamm, “Algorithm for model validation : Theory and applications,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 16, pp. 6562–6567, 2007.
- [28] F. Gross and M. MacLeod, “Prospects and problems for standardizing model validation in systems biology,” *Progress in biophysics and molecular biology*, vol. 129, pp. 3–12, 2017.
- [29] R. David, D. Dochain, J.-R. Mouret, A. V. Wouwer, and J.-M. Sablayrolles, “Nitrogen-backed modeling of wine-making in standard and nitrogen-added fermentations,” *Bioprocess and biosystems engineering*, vol. 37, no. 1, pp. 5–16, 2014.
- [30] R. Boulton, “The prediction of fermentation behavior by a kinetic model,” *American Journal of Enology and Viticulture*, vol. 31, no. 1, pp. 40–45, 1980.
- [31] J.-M. Salmon, O. Vincent, J. Mauricio, M. Bely, and P. Barre, “Sugar transport inhibition and apparent loss of activity in *saccharomyces cerevisiae* as a major limiting factor of enological fermentations,” *American Journal of Enology and Viticulture*, vol. 44, no. 1, pp. 56–64, 1993.
- [32] J. Slaughter, J. Smith, and W. Mitchell, “Glucose transport throughout fermentation by *saccharomyces cerevisiae* nyc 1108,” *Letters in applied microbiology*, vol. 12, no. 6, pp. 221–223, 1991.

-
- [33] I. Caro, L. Perez, and D. Cantero, “Development of a kinetic model for the alcoholic fermentation of must,” *Biotechnology and bioengineering*, vol. 38, no. 7, pp. 742–748, 1991.
- [34] M. R. Marín, “Alcoholic fermentation modelling : Current state and perspectives,” *American journal of enology and viticulture*, vol. 50, no. 2, pp. 166–178, 1999.
- [35] A. C. Cramer, S. Vlassides, and D. E. Block, “Kinetic model for nitrogen-limited wine fermentations,” *Biotechnology and bioengineering*, vol. 77, no. 1, pp. 49–60, 2002.
- [36] M. C. Coleman, R. Fish, and D. E. Block, “Temperature-dependent kinetic model for nitrogen-limited wine fermentations,” *Applied and Environmental Microbiology*, vol. 73, no. 18, pp. 5875–5884, 2007.
- [37] S. Malherbe, V. Fromion, N. Hilgert, and J.-M. Sablayrolles, “Modeling the effects of assimilable nitrogen and temperature on fermentation kinetics in enological conditions,” *Biotechnology and bioengineering*, vol. 86, no. 3, pp. 261–272, 2004.
- [38] B. Sonnleitner and O. Käppli, “Growth of *saccharomyces cerevisiae* is controlled by its limited respiratory capacity : formulation and verification of a hypothesis,” *Biotechnology and bioengineering*, vol. 28, no. 6, pp. 927–937, 1986.
- [39] R. David, D. Dochain, J.-R. Mouret, A. V. Wouwer, and J.-M. Sablayrolles, “Dynamical modeling of alcoholic fermentation and its link with nitrogen consumption,” *IFAC Proceedings Volumes*, vol. 43, no. 6, pp. 496–501, 2010.
- [40] C. Leão and N. Van Uden, “Effects of ethanol and other alkanols on the glucose transport system of *saccharomyces cerevisiae*,” *Biotechnology and Bioengineering*, vol. 24, no. 11, pp. 2601–2604, 1982.
- [41] C. Leão and N. Van Uden, “Effects of ethanol and other alkanols on the ammonium transport system of *saccharomyces cerevisiae*,” *Biotechnology and bioengineering*, vol. 25, no. 8, pp. 2085–2089, 1983.
- [42] S. Malherbe, *Modélisation de la fermentation alcoolique en conditions oenologiques : élaboration du modèle à partir des principaux mécanismes limitants de la cinétique fermentaire, validation et application au contrôle des fermentations en cave (cuverie et besoins énergétiques)*. PhD thesis, Montpellier 2, 2003.

-
- [43] M. Bely, J.-M. Sablayrolles, and P. Barre, “Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions,” *Journal of fermentation and bioengineering*, vol. 70, no. 4, pp. 246–252, 1990.
- [44] R. David, D. Dochain, J.-R. Mouret, A. V. Wouwer, and J.-M. Sablayrolles, “Modeling of the aromatic profile in wine-making fermentation : the backbone equations,” *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 10597–10602, 2011.
- [45] D. B. Jones, *Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins*. No. 183, US Department of Agriculture, 1931.
- [46] M. Carpena, M. Fraga-Corral, P. Otero, R. A. Nogueira, P. Garcia-Oliveira, M. A. Prieto, and J. Simal-Gandara, “Secondary aroma : Influence of wine microorganisms in their aroma profile,” *Foods*, vol. 10, no. 1, p. 51, 2021.
- [47] P. M. Doran, *Bioprocess engineering principles*. Elsevier, 1995.
- [48] O. Ramirez, R. Zamora, G. Espinosa, E. Merino, F. Bolivar, and R. Quintero, “Kinetic study of penicillin acylase production by recombinant e. coli in batch cultures,” *Process Biochemistry*, vol. 29, no. 3, pp. 197–206, 1994.
- [49] J. W. Haefner, *Modeling Biological Systems : : Principles and Applications*. Springer Science & Business Media, 2005.
- [50] G. Lillacci and M. Khammash, “Parameter estimation and model selection in computational biology,” *PLoS Comput Biol*, vol. 6, no. 3, p. e1000696, 2010.
- [51] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, “Structural identifiability of systems biology models : a critical comparison of methods,” *PloS one*, vol. 6, no. 11, p. e27755, 2011.
- [52] A. Moser, “Bioprocess kinetics,” in *Bioprocess Technology*, pp. 197–306, Springer, 1988.
- [53] E. Walter and L. Pronzato, “On the identifiability and distinguishability of nonlinear parametric models,” *Mathematics and computers in simulation*, vol. 42, no. 2-3, pp. 125–134, 1996.
- [54] G. Styger, D. Jacobson, B. Prior, and F. Bauer, “Genetic analysis of the metabolic pathways responsible for aroma metabolite production by *saccharomyces cerevisiae*,” *Applied Microbiology and Biotechnology*, vol. 97, pp. 4429–4442, 2012.

ANNEXE I : Exemples de données

FIGURE A.1 – Données expérimentales pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$.

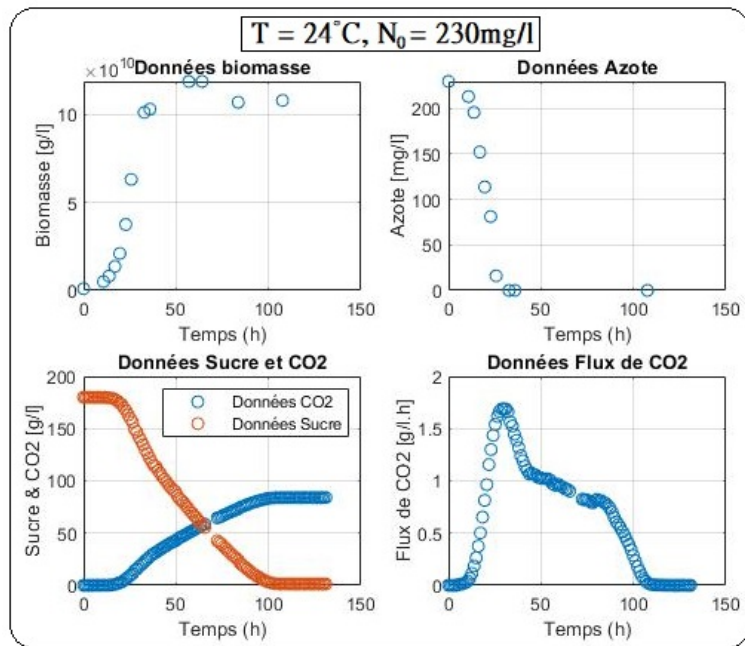


FIGURE A.2 – Données expérimentales pour $T = 24^{\circ}\text{C}$ et $N_0 = 230\text{mg/l}$.

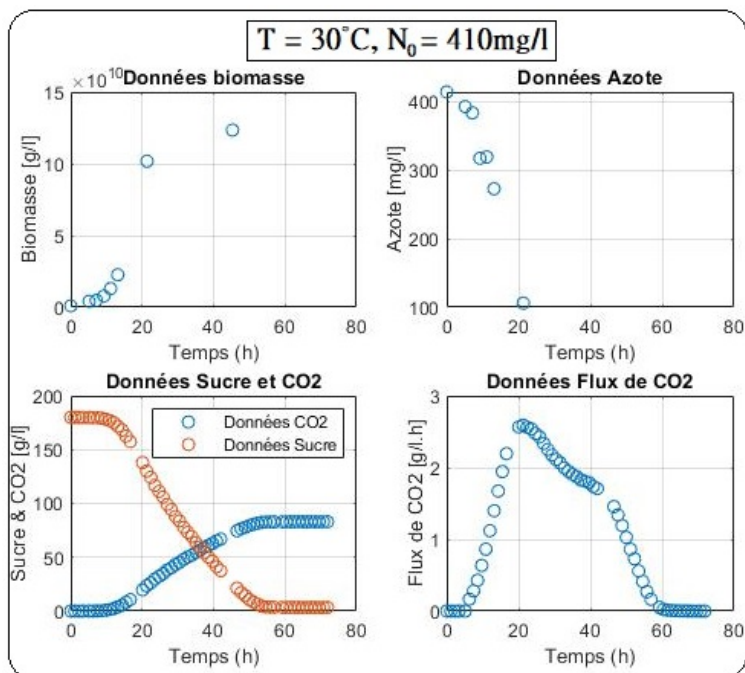


FIGURE A.3 – Données expérimentales pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$.

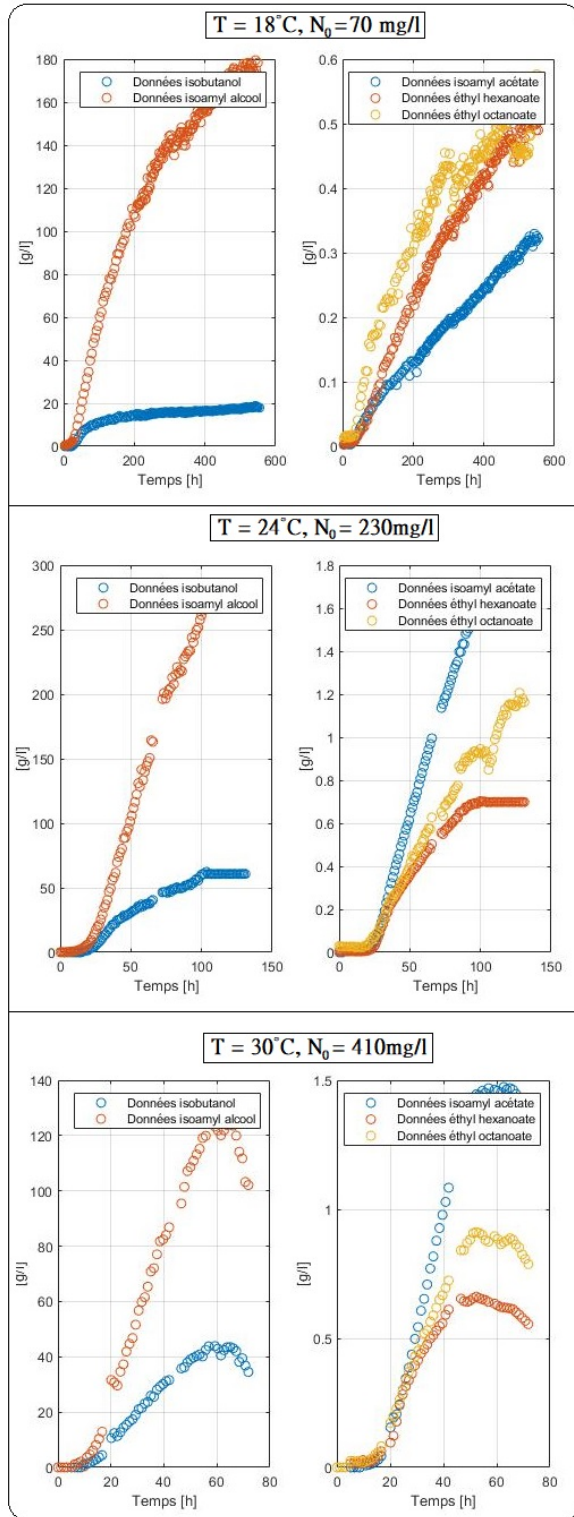


FIGURE A.4 – Quelques données aromatiques.
xviii

ANNEXE II : Comparatif ancien modèle/nouvelle implémentation

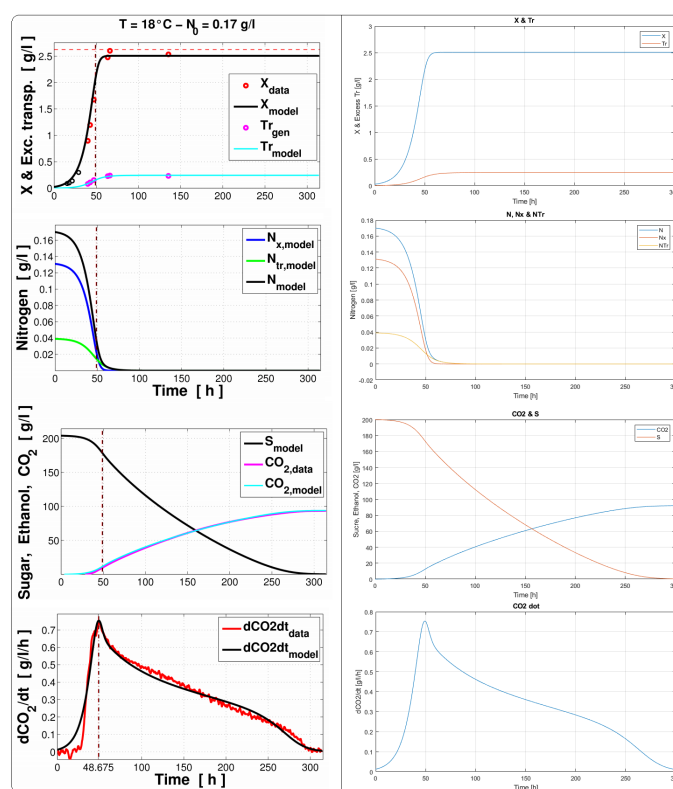


FIGURE B.1 – Comparaison des résultats de la nouvelle simulation, à droite, avec celle de référence, à gauche [29].



ANNEXE III : Quelques fonctions de sensibilité non-normées

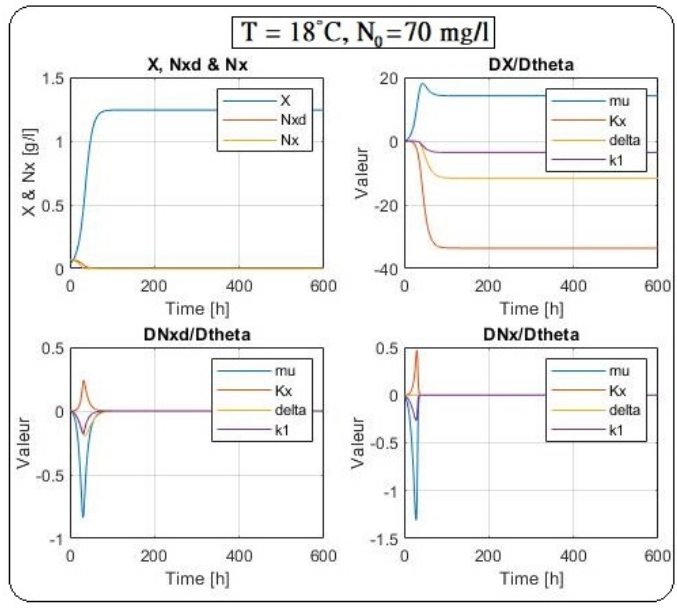


FIGURE C.1 – Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 18^\circ\text{C}$ et $N_0 = 70\text{mg/l}$.

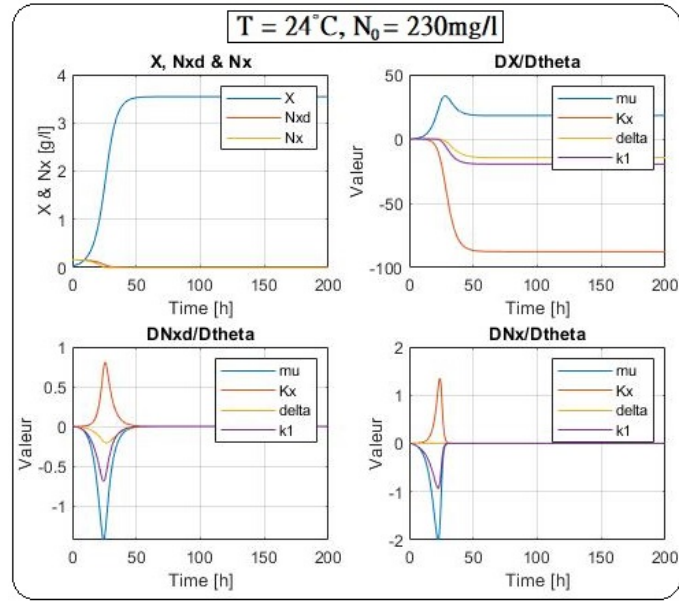


FIGURE C.2 – Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

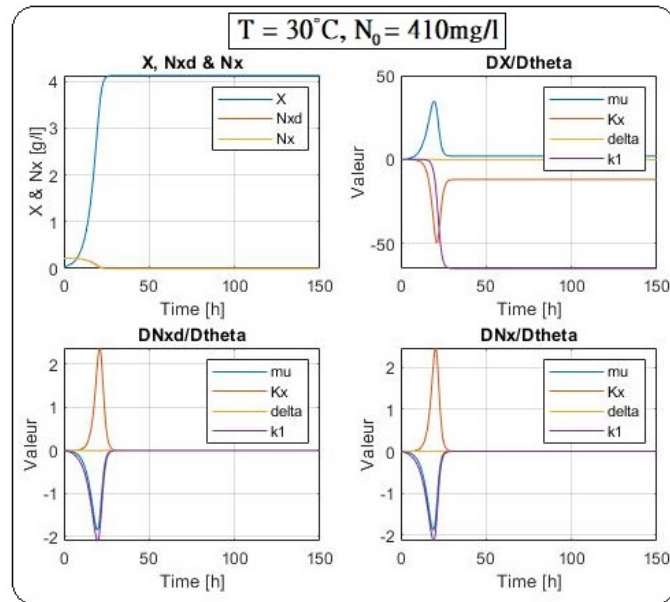


FIGURE C.3 – Fonctions de sensibilité de μ_{max} , K_x , k_1 et δ pour $T = 30^\circ\text{C}$ et $N_0 = 410\text{mg/l}$.

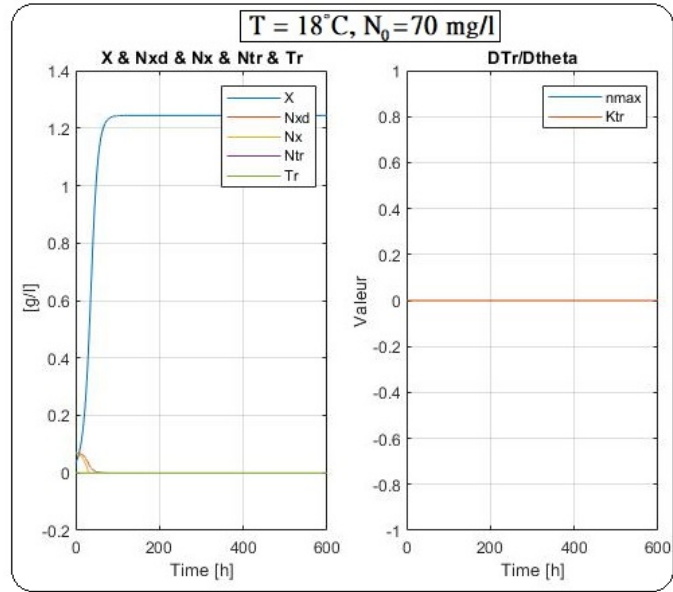


FIGURE C.4 – Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 18^\circ\text{C}$ et $N_0 = 70\text{mg/l}$.

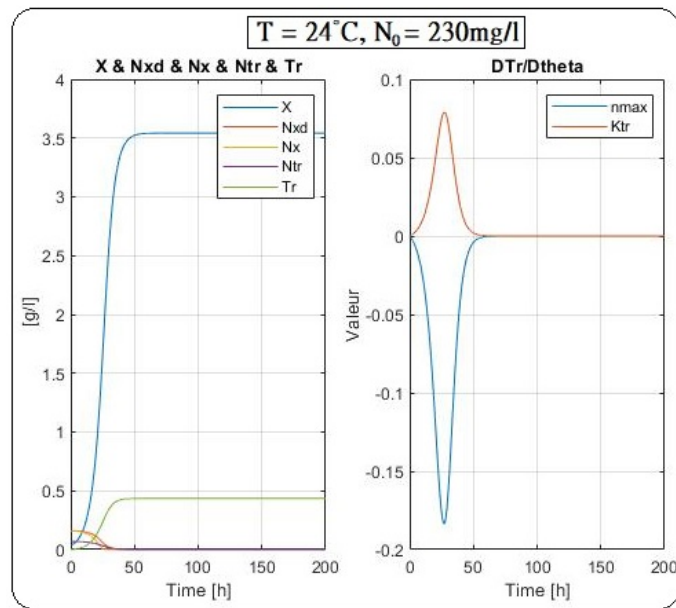


FIGURE C.5 – Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

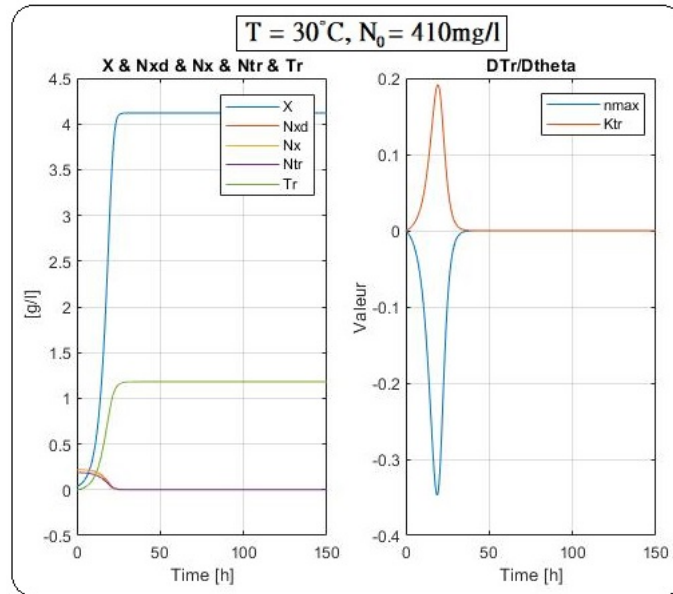


FIGURE C.6 – Fonctions de sensibilité de η_{max} et K_{tr} pour $T = 30^{\circ}\text{C}$ et $N_0 = 410\text{mg/l}$.

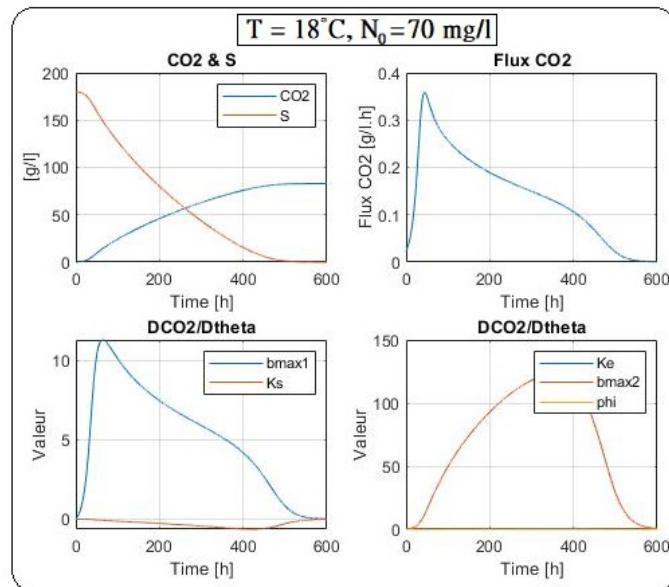


FIGURE C.7 – Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 18^{\circ}\text{C}$ et $N_0 = 70\text{mg/l}$.

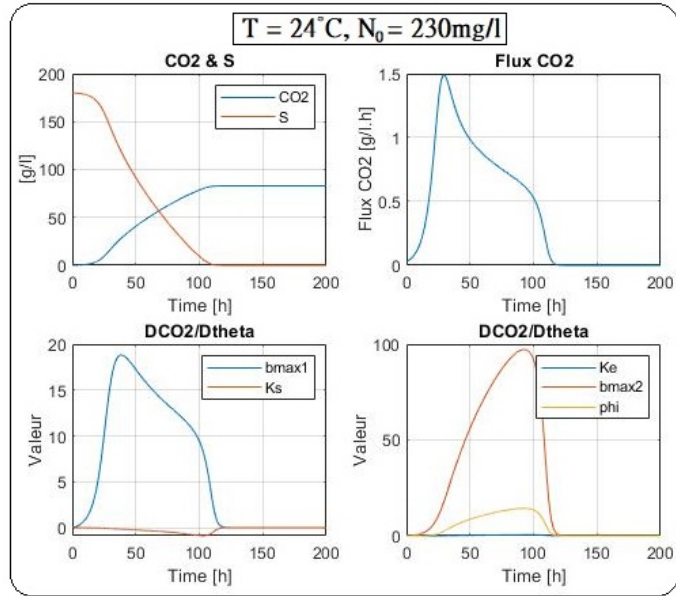


FIGURE C.8 – Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 24^\circ\text{C}$ et $N_0 = 230\text{mg/l}$.

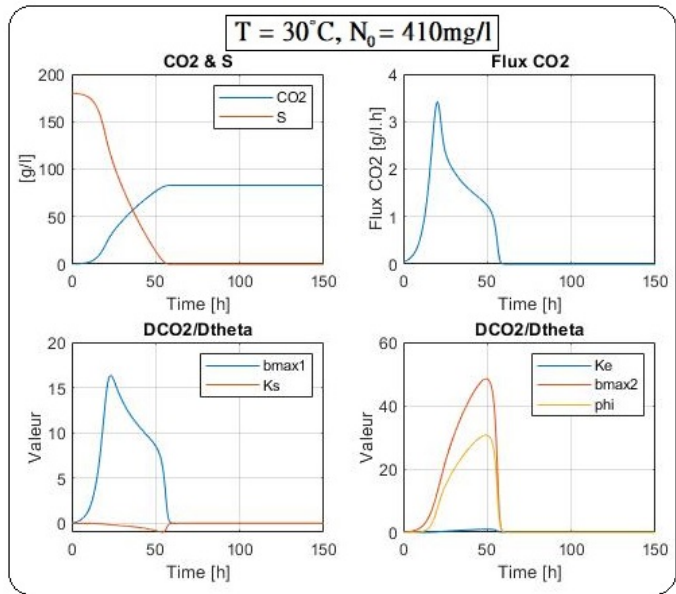


FIGURE C.9 – Fonctions de sensibilité de $\beta_{max,1}$, $\beta_{max,2}$, K_E , K_S et ϕ pour $T = 30^\circ\text{C}$ et $N_0 = 410\text{mg/l}$.



ANNEXE IV : Coefficients des polynômes des lois d'évolutions des paramètres

$$\begin{aligned}K_S(T) &: [0.04444, -2.583, 39.3] \\K_E(T) &: [-0.1805, 7.083, -24] \\\beta_{max,1}(T) &: [-0.0003652, 0.032108, -0.105799] \\\beta_{max,2}(T) &: [0.004375, -0.16792, 1.94] \\\eta_{max}(T) &: [0.000282, -0.001791, 0.0281] \\\delta(N_0) &: [220.0681, -131.24697, 27.7978, -2.18474, 0.1618] \\\delta(T) &: [0.0007305, -0.019392, 0.2392] \\\mu_{max}(T) &: [3.0555e - 05, 0.0117, -0.134] \\\phi(T) &: [-0.006944, 0.3416, -3.5] \\k_1(N_0) &: [0.469771, -0.578431, 0.218188]\end{aligned}\tag{D.1}$$

ANNEXE V : Etude informative de la cinétique aromatique

Cette tentative de modélisation de la cinétique de production des arômes se base sur une simple réflexion phénoménologique. Puisqu'il a été observé que les courbes de production des molécules organoleptiques sont semblables à celle de production du CO_2 , utiliser une expression cinétique proche de celle décrivant la production du CO_2 semble être un bon candidat pour une première approximation.

Ainsi, l'expression choisie est

$$\dot{Ar\grave{o}me} = \mu \frac{K_1}{K_1 + CO_2} \frac{S}{K_2 + S} X$$

Cette expression est empruntée d'une ancienne modélisation de la production du CO_2 [39] et convient bien à une première approximation grâce à son faible nombre de paramètres.

Les paramètres μ , K_1 et K_2 ont donc été identifiés pour chaque arôme. μ fait office de vitesse spécifique de réaction tandis que K_1 et K_2 jouent le rôle de constantes de saturation. Les valeurs identifiées de K_1 et K_2 sont indiquées à la [Table E.1](#). Les valeurs de μ correspondantes sont visibles aux [Figures E.1, E.2, E.3, E.4 et E.5](#). On observe que des profils d'évolution de μ se démarquent, fonctions de N_0 et T . Les [Figures E.6, E.7, E.8, E.9 et E.10](#) montrent le comparatif entre les données expérimentales et les prédictions de ce premier modèle informatif.

Arôme	K_1	K_2
Isobutanol	40	20
Isoamyl alcool	20	5
Isoamyl acétate	100	0.5
Ethyl hexanoate	40	10
Ethyl octanoate	45	10

TABLE E.1 – Valeurs identifiées de K_1 et K_2 pour chaque arôme.

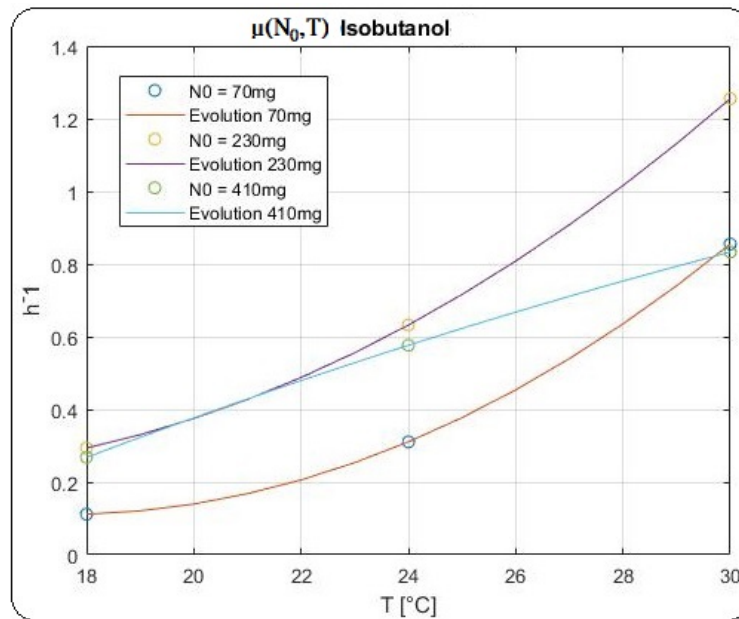


FIGURE E.1 – Valeurs identifiées de $\mu(N_0, T)$ pour l'isobutanol.

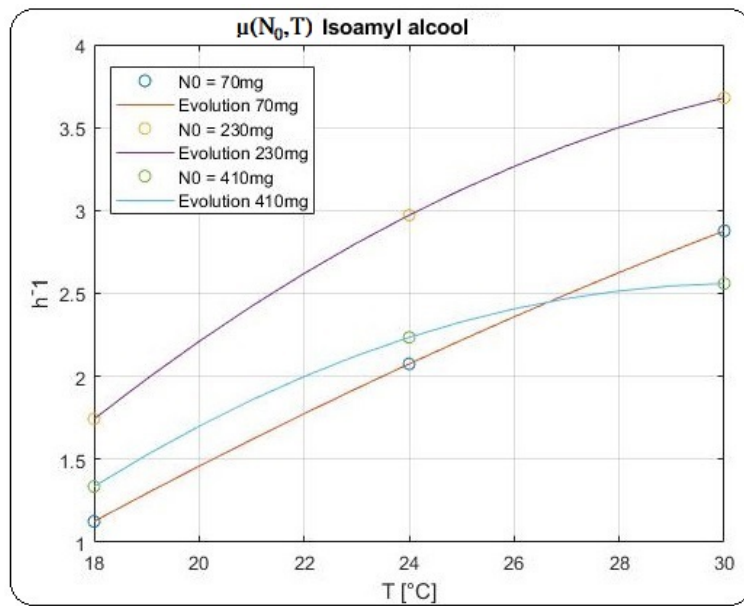


FIGURE E.2 – Valeurs identifiées de $\mu(N_0, T)$ pour l'isoamyl alcool.

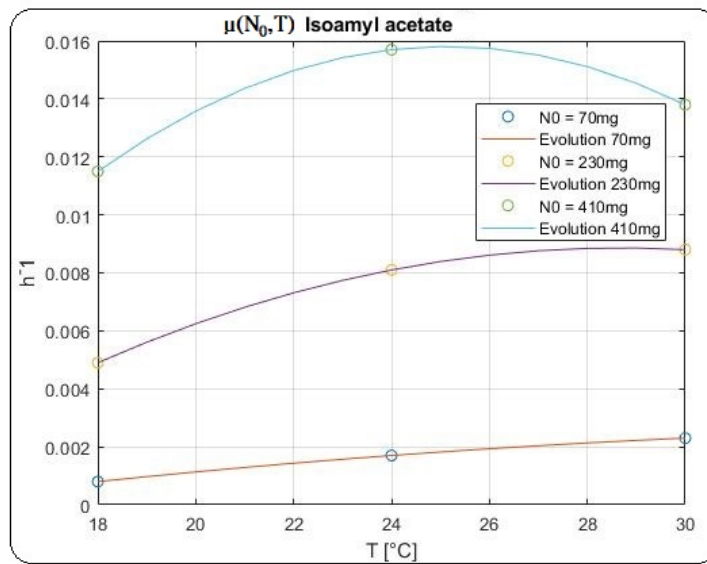


FIGURE E.3 – Valeurs identifiées de $\mu(N_0, T)$ pour l'isoamyl acétate.

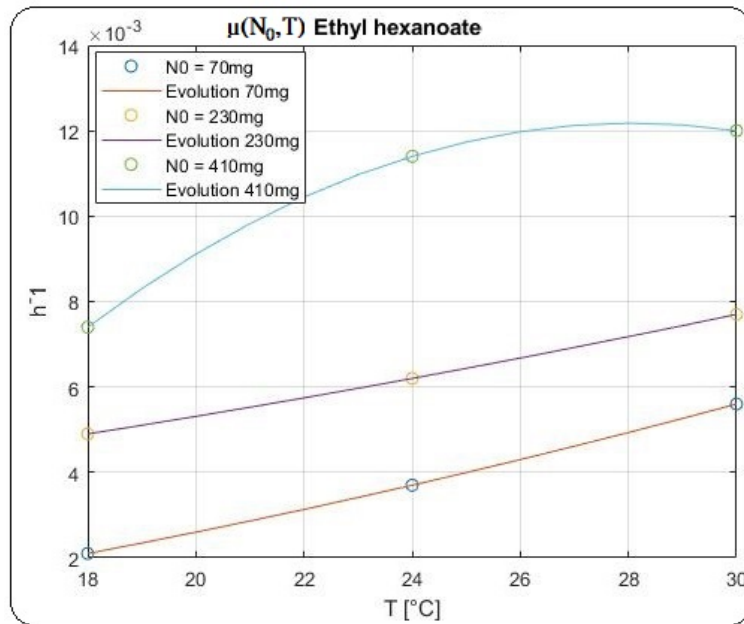


FIGURE E.4 – Valeurs identifiées de $\mu(N_0, T)$ pour l'éthyl acétate.

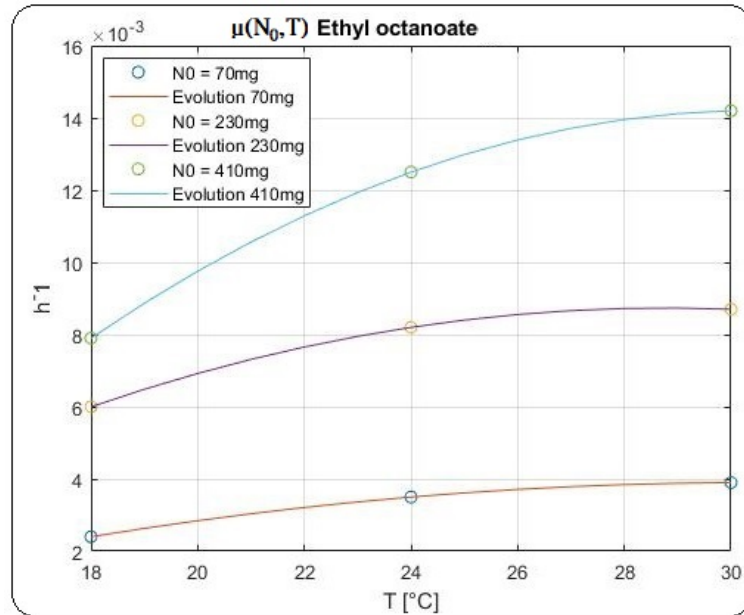


FIGURE E.5 – Valeurs identifiées de $\mu(N_0, T)$ pour l'éthyl octanoate.

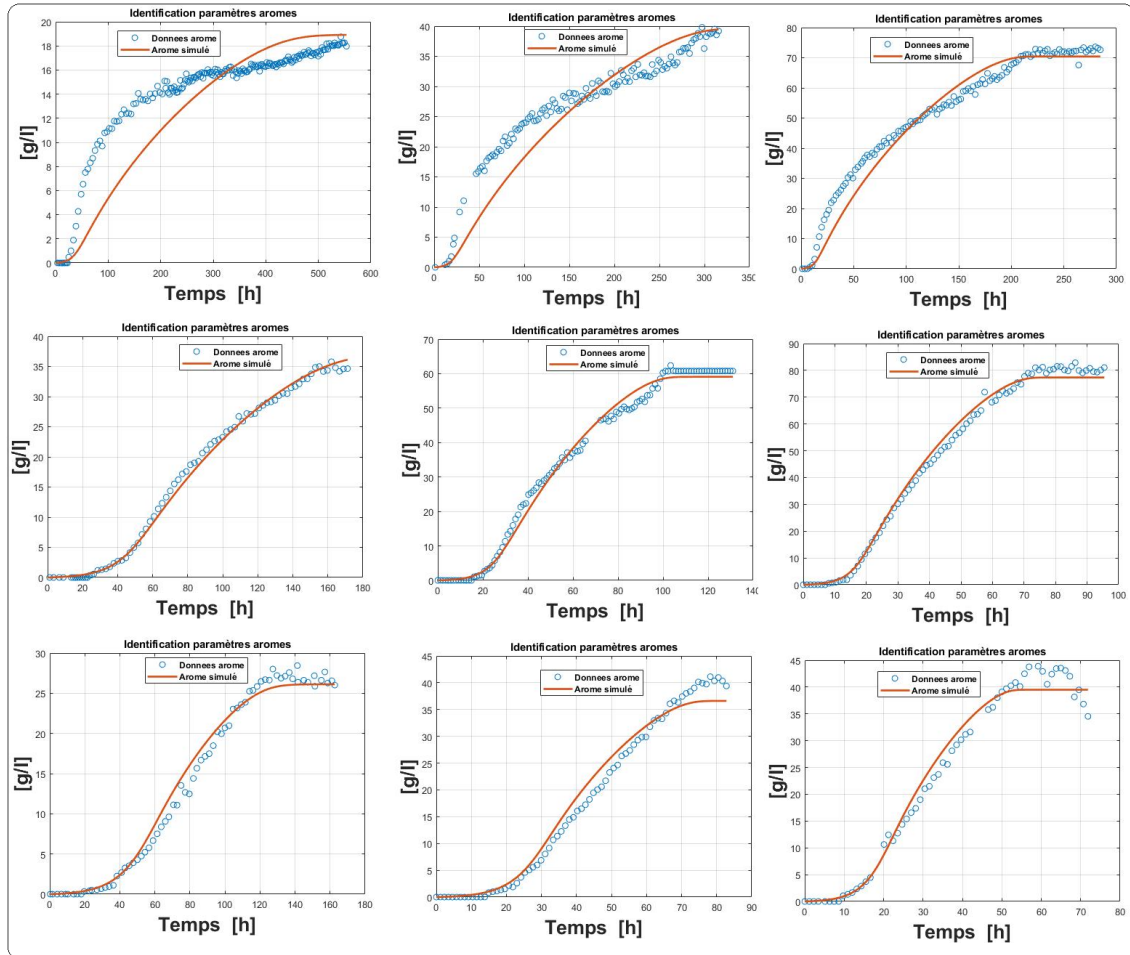


FIGURE E.6 – Comparatif entre le modèle et les données expérimentales pour l’isobutanol. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l.

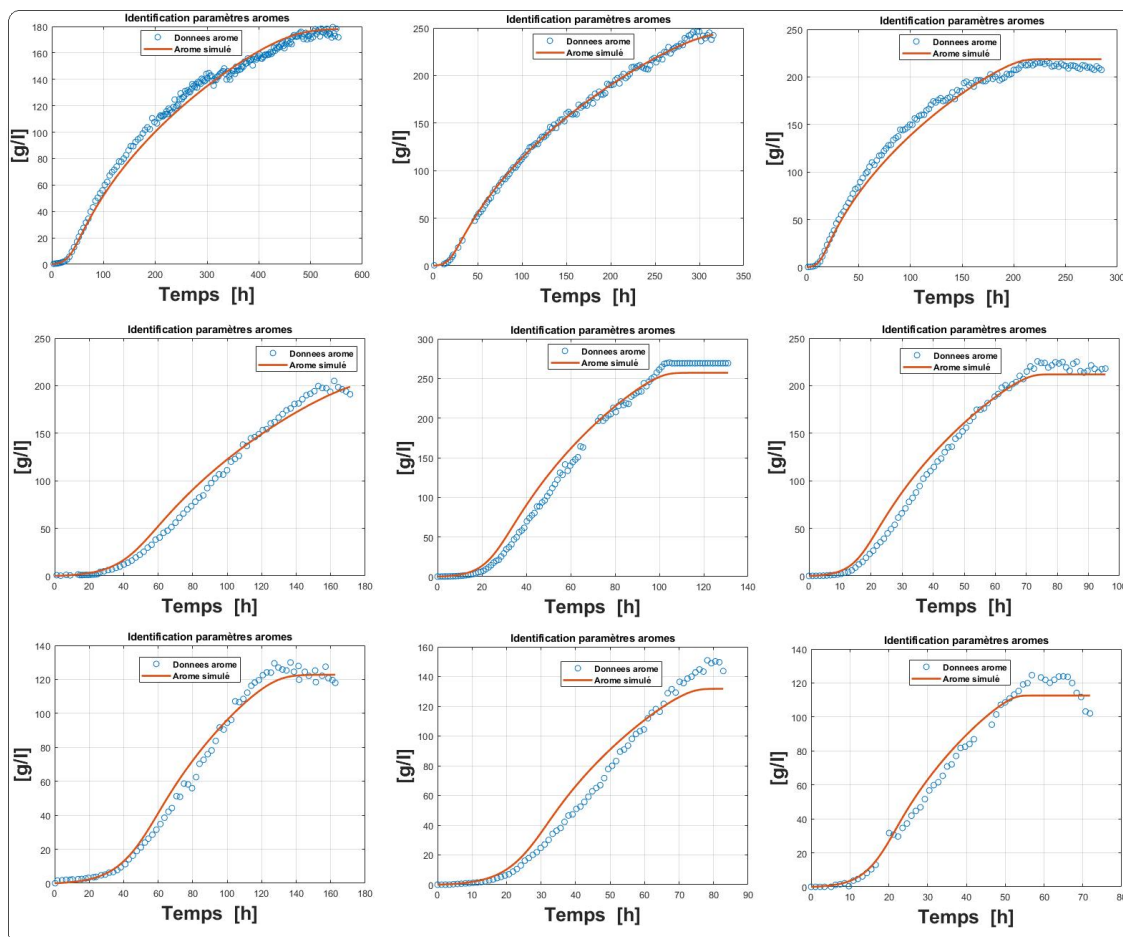


FIGURE E.7 – Comparatif entre le modèle et les données expérimentales pour l'isoamyl alcool. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l.

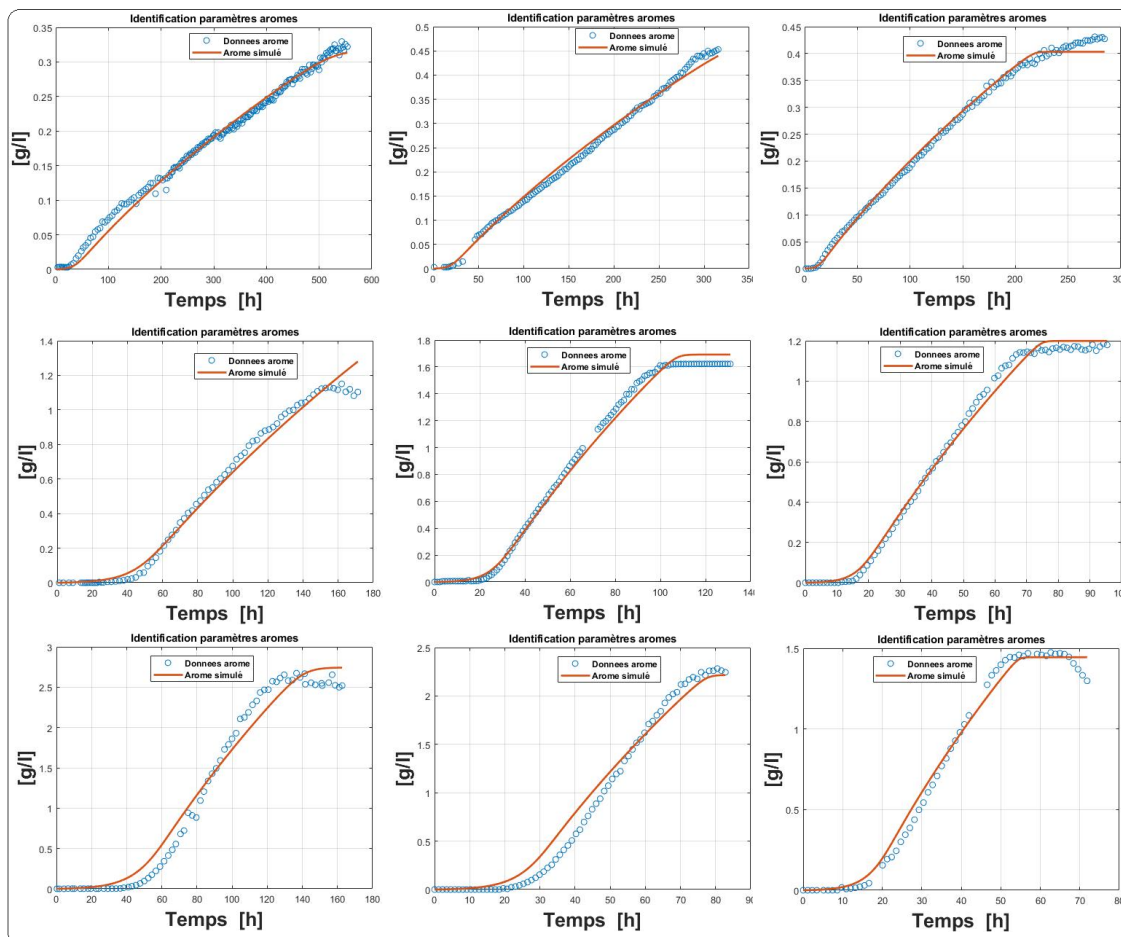


FIGURE E.8 – Comparatif entre le modèle et les données expérimentales pour l'isoamyl acétate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l.

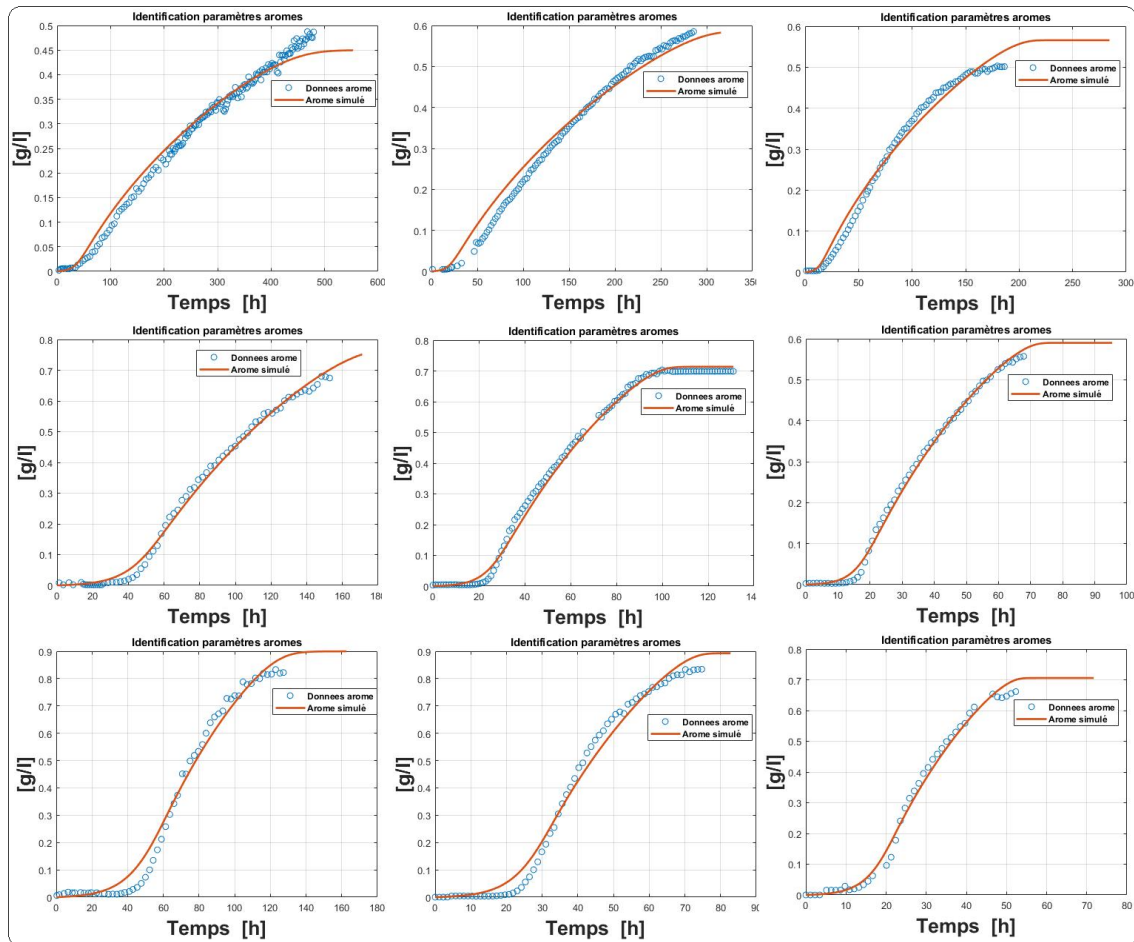


FIGURE E.9 – Comparatif entre le modèle et les données expérimentales pour l'éthyl hexanoate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l.

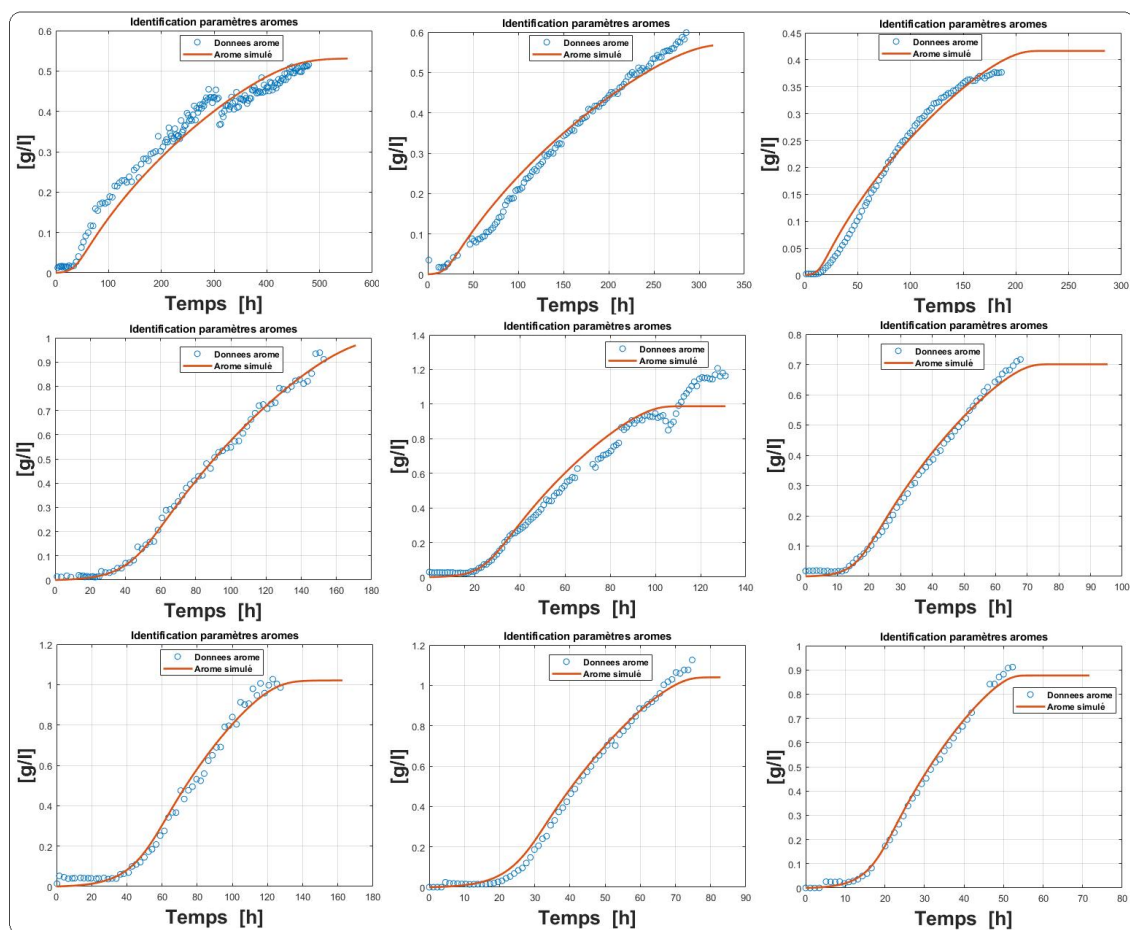


FIGURE E.10 – Comparatif entre le modèle et les données expérimentales pour l'éthyl octanoate. De gauche à droite, T varie entre 18, 24 et 30°C. De haut en bas, N_0 varie entre 70, 230 et 410mg/l.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl