# Modeling the kinetic behaviour, mixing and local transfer pheonmena and biologicial population heterogeneity effects in industrial fermenters

Maxime Pigou

## To cite this version:

Maxime Pigou. Modeling the kinetic behaviour, mixing and local transfer pheonmena and biologicial population heterogeneity effects in industrial fermenters. Biotechnology. INSA de Toulouse, 2018. English. NNT : 2018ISAT0038 . tel-02917985

HAL Id: tel-02917985
https://theses.hal.science/tel-02917985v1

Submitted on 20 Aug 2020

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :**
*l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)*

---

**Présentée et soutenue le** *08/10/2018* **par :**
**Maxime PIGOU**

## Modélisation du comportement cinétique, des phénomènes de mélange, de transfert locaux et des effets d'hétérogénéité de population dans les fermenteurs industriels.

---

**JURY**

| | | |
|---|---|---|
| Frank DELVIGNE | Liège University | Rapporteur |
| Rodney O. FOX | Iowa State University | Rapporteur |
| Martine MEIRELES | Centre National de la Recherche Scientifique | Examinateur |
| Ralf TAKORS | University of Stuttgart | Examinateur |
| Angélique DELAFOSSE | Liège University | Examinateur |
| Marie-Isabelle PENET | Sanofi Chimie | Examinateur |
| Geoffrey LARONZE | Sanofi Chimie | Invité |
| Jérôme MORCHAIN | INSA de Toulouse | Co-directeur |
| Pascal FEDE | Université Paul Sabatier | Co-directeur |

---

**École doctorale et spécialité :**
*MEGEP : Génie des procédés et de l'Environnement*
**Unités de Recherche :**
*Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés*
*Institut de Mécanique des Fluides de Toulouse*
**Directeurs de Thèse :**
*Jérôme MORCHAIN* et *Pascal FEDE*
**Rapporteurs :**
*Frank DELVIGNE* et *Rodney O. FOX*

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :**
*l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)*

---

**Présentée et soutenue le** *08/10/2018* **par :**
**Maxime PIGOU**

## Modélisation du comportement cinétique, des phénomènes de mélange, de transfert locaux et des effets d'hétérogénéité de population dans les fermenteurs industriels.

---

### JURY

| | | |
|---|---|---|
| Frank DELVIGNE | Liège University | Rapporteur |
| Rodney O. FOX | Iowa State University | Rapporteur |
| Martine MEIRELES | Centre National de la Recherche Scientifique | Examinateur |
| Ralf TAKORS | University of Stuttgart | Examinateur |
| Angélique DELAFOSSE | Liège University | Examinateur |
| Marie-Isabelle PENET | Sanofi Chimie | Examinateur |
| Geoffrey LARONZE | Sanofi Chimie | Invité |
| Jérôme MORCHAIN | INSA de Toulouse | Co-directeur |
| Pascal FEDE | Université Paul Sabatier | Co-directeur |

---

**École doctorale et spécialité :**
*MEGEP : Génie des procédés et de l'Environnement*
**Unités de Recherche :**
*Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés*
*Institut de Mécanique des Fluides de Toulouse*
**Directeurs de Thèse :**
*Jérôme MORCHAIN* et *Pascal FEDE*
**Rapporteurs :**
*Frank DELVIGNE* et *Rodney O. FOX*

The work described in the present manuscript has been conducted at

Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés
UMR 5504 INSA, CNRS
UMR 792 INSA, INRA
135 Avenue de Rangueil
31077 Toulouse Cedex 4, France

and

Institut de Mécanique des Fluides de Toulouse
UMR 5502 CNRS, INPT, UPS
2 Allée du Professeur Camille Soula
31400 Toulouse, France

with the financial support and under the supervision of

Sanofi Chimie
9 Quai Jules Guesde
94400 Vitry-sur-Seine cedex, France

and

Sanofi Chimie
20 Avenue Raymond Aron
92165 Antony cedex, France

# RÉSUMÉ

Les biotechnologies sont un champ d'application majeur pour les travaux de recherche actuels et à venir [Noorman and Heijnen, 2017; Camarasa et al., 2018]. Elles sont actuellement appliquées aux productions industrielles de composés chimiques, de bio-carburants et bio-plastiques, ou de molécules à forte valeur ajoutée. La diversité naturelle des micro-organismes nous laisse espérer de nouvelles applications industrielles dans un futur proche, et pourrait aider à la lutte contre les crises environnementales et énergétiques.

De la découverte de nouvelles voies de production à leurs applications industrielles, plusieurs étapes d'ingénierie sont requises pour permettre le passage d'échelle. Ce travail peut être facilité par la modélisation et la simulation prédictive des bioréacteurs. Différentes approches ont été développées dans cet objectif, mais les modèles obtenus échouent généralement à reproduire les observations expérimentales à différentes échelles [Enfors et al., 2001]. Cela s'explique par l'utilisation d'hypothèses qui, bien que valides aux échelles de culture en laboratoire, ne s'appliquent pas aux échelles industrielles. Le défi principal abordé par ce travail est de s'affranchir de ces hypothèses afin d'améliorer les modèles tout en assurant un coût numérique plus faible que les approches usuelles.

Cette thèse se concentre sur le développement d'une structure de modèle pour les fermenteurs industriels, en visant un usage d'ingénierie. Ce travail se base sur deux postulats : (i) par leur taille, les fermenteurs industriels sont hétérogènes; (ii) par leur nature, les micro-organismes sont des systèmes dynamiques. Ces considérations induisent plusieurs conséquences. La principale est que les micro-organismes vont se différencier les uns des autres. La seconde est qu'ils seront en déséquilibre avec leur environnement, entraînant des baisses de performance à l'échelle du procédé. Pour intégrer l'ensemble de ces aspects dans une unique structure de modèle, tout en permettant des simulations rapides d'un fermenteur, l'approche proposée couple (i) un modèle métabolique dynamique pour décrire le comportement des cellules, (ii) un modèle de bilan de population pour suivre l'hétérogénéité biologique et (iii) un modèle de compartiments pour décrire l'hydrodynamique.

Un résultat majeur a été le développement d'une structure de modèle métabolique, qui a été confrontée avec succès à de nombreuses données expérimentales, tout en préservant un coût numérique significativement faible. Cette structure a été appliquée à deux micro-organismes, *Escherichia coli* et *Saccharomyces cerevisiae.* Une comparaison de plusieurs méthodes permettant de traiter les équations de bilans de population a été menée en termes de stabilité, précision et performances a conduit à la sélection de la méthode étendue de quadrature des moments. Une amélioration majeure de cette méthode a été identifiée, permettant un gain de stabilité et de performances, en réduisant son coût d'un facteur 10. L'hydrodynamique gaz-liquide d'un fermenteur industriel a été obtenue par Mécanique des Fluides Numérique (CFD) et des outils supplémentaires ont été développés pour extraire des modèles de compartiments à partir des simulations CFD. Finalement, le couplage des modèles métaboliques, de bilan de population et de compartiments a été illustré par la simulation d'une configuration industrielle.

*Mots clés : bioréacteur, modélisation, simulation, bilan de population, métabolisme, scale-up, CFD, modèle de compartiments, micro-mélange*

# SUMMARY

Biotechnologies form a significant field for current and upcoming research [Noorman and Heijnen, 2017; Camarasa et al., 2018]. They are currently applied to industrial production of chemicals, bio-fuels or high-value molecules. The natural –massive– diversity of micro-organisms is expected to unlock new industrial applications in a near future, and could help facing the environmental and energetic crises.

Between the lab discovery of new interesting pathways to their industrial application, multiple engineering stages are required to scale-up the process efficiently. In this context, the predictive modelling and simulation of bioreactors facilitates the scale-up procedure of new processes, and can be applied to existing systems to improve their performances. Attempts to produce such predictive models exist, but proposed approaches usually do not fit experimental results at different scales simultaneously [Enfors et al., 2001]. This can be explained by the use of modelling hypotheses that, while being valid at lab-scale, tend to be inaccurate in large-scale heterogeneous systems. Removing the need for these hypotheses while keeping the model complete, yet simple enough to allow fast simulations, is a challenge that this thesis attempts to tackle.

This PhD thesis focuses on developing a modelling framework for fermenter simulations, aiming at engineering applications. This work is based on two premises: (i) due to their size, industrial fermenters are heterogeneous; (ii) due to their nature, micro-organisms are dynamic systems. Coupling these considerations leads to a variety of consequences. The main one is that cells will tend to differentiate one from another, hence inducing biological heterogeneity at a population scale. The second one is that micro-organisms will be at disequilibrium with their environment, which induces loss in process performances. To embrace all of these aspects in a modelling framework, while allowing for fast simulations, we chose to couple (i) a dynamic metabolic model for cell behaviour, (ii) a population balance model to track biological heterogeneity and (iii) a compartment model to describe the fermenter hydrodynamics.

A significant achievement has been the development of a metabolic model structure which has shown to be accurate against numerous experimental data sets, while having a significantly low numerical cost. This structure has been applied to two micro-organisms: *Escherichia coli* and *Saccharomyces cerevisiae.* Multiple numerical methods exist to treat Population Balance Equations. They have been compared in terms of stability, accuracy and performances, which led to the selection of the Extended Quadrature Method of Moments (EQMOM). We happened to identify a major improvement to this method, which led to a further increase of its performances and stability, by reducing its cost by a factor 10. Computational Fluid Dynamics (CFD) has been used to access the gas-liquid hydrodynamics of an actual industrial fermenter. Tools have been developed to post-process CFD results and obtain compartment models. Finally, the coupling of compartment, population balance and metabolic models for engineering application has been illustrated on an industrial fermenter.

*Keywords: bioreactor, modelling, simulation, population balance, metabolism, scale-up, CFD, compartment model, micro-mixing*

# CONFIDENTIALITY NOTICE

For confidentiality reasons, the complete thesis cannot be made fully available. The current version is incomplete; this confidentially safe version comes with following modifications of the original manuscript:

- details about an industrial fermenter geometry and operating conditions were discarded;
- simulation results were normalized or changed to arbitrary units;
- content of Appendix D has been completely removed.

Despite these modifications, this manuscript version remains faithful to the original thesis. Observations, discussions and analysis were left untouched.

# REMERCIEMENTS

Le travail de thèse décrit dans ce manuscrit est l'aboutissement de plusieurs années passées au sein du Hall GPE sur le campus de l'INSA de Toulouse. J'ai rejoint cette école début 2011 en tant qu'étudiant Ingénieur et j'y ai rencontré d'une part celui qui allait ensuite être mon maître de stage puis mon directeur de thèse, et d'autre part celle avec qui je vis maintenant depuis 6 ans. Il est connu que le temps semble passer vite en bonne compagnie, et depuis 2011, le temps est passé vite, très vite! J'ai fait de très belles rencontres, et maintenant que je sais que je resterai encore de nombreuses années sur Toulouse, je sais que je ne perdrai pas de vue la majorité de celles et ceux que j'ai pu côtoyer, en particulier durant ma thèse.

Avant de commencer à travailler sur cette thèse, j'ai discuté avec plusieurs doctorants et leur principal conseil était "Choisis bien tes directeurs de thèse, davantage que le sujet". À ce niveau là, j'ai eu de la chance. La chance de rencontrer Jérôme Morchain, avec qui j'ai commencé à travailler lors d'un projet d'initiation à la recherche qui a ensuite débouché sur tout le début de ma carrière : stage, contrat d'ingénieur, puis thèse sur un sujet correspondant parfaitement à ma formation. Jérôme, tu as été pour moi un exemple, un mentor, et l'image de ce que devrait être tout directeur de thèse : impliqué tout en laissant beaucoup de libertés ; très humain là où certains ont bien moins de considération envers leurs doctorants ; et très actif pour faciliter l'insertion professionnelle de tes thésards en leur ouvrant ton réseau. Pour cela, pour les bonnes soirées passées au travers de nombreux congrès, pour ces journées de brainstorming intensif, et pour la place importante que tu m'as laissée au sein de ton projet de recherche, je te remercie très sincèrement. J'espère que nous continuerons à collaborer encore longtemps !

Le fait que ma thèse ait été co-encadrée par deux laboratoires et un partenaire industriel aurait pu être une source de complication, mais cela s'est au final extrêmement bien déroulé. Du côté IMFT, Pascal Fede m'aura apporté un regard de non-biologiste sur ce travail avec une expertise en mécanique des fluides multiphasique et une rigueur mathématique autrement manquante dans mes écrits. Pascal, merci de m'avoir ouvert les portes de l'IMFT ce qui aura –à terme– débouché sur mon poste actuel, et de m'avoir accueilli dans ton bureau à une période de ma thèse où j'avais besoin de changer d'environnement. Avec toi aussi j'espère continuer à collaborer longtemps.

Du côté industriel, j'ai également pu profiter d'un encadrement très humain, dont l'objectif était d'amener le projet aussi loin que possible et sans perdre de vue les thématiques industrielles ; sans jamais être contraignant, toujours en étant constructifs. J'ai bénéficié d'un véritable intérêt pour ce travail à toutes les échelles de la recherche chez Sanofi. Je tiens en particulier à remercier mes interlocuteurs directs, à savoir Geoffrey Laronze, Marie-Isabelle Penet et Marine Bertin. Vous avez porté ce projet de thèse et m'avez permis de travailler durant ces quelques années sur des thématiques nous intéressant tous profondément. Nos fréquents échanges, par téléconférence ou en face-à-face, ont toujours été productifs, constructifs, et se sont déroulés dans la bonne humeur. Travailler avec vous a été pour moi une très bonne expérience ; j'espère que vous partagez cet avis et que les collaborations vont pouvoir se poursuivre sur la lignée de cette première thèse.

Cette thèse s'est déroulé en parallèle de plusieurs projets partageant des thématiques communes. Cela a été la source de discussions intéressantes, et parfois critiques pour la poursuite de mon travail, ainsi que d'échanges qui ont permis à mon travail d'aller bien plus loin que ce que j'aurais été capable d'effectuer de manière isolée. Dans ce cadre, je tiens à remercier en particulier Bastien Polizzi, alors post-doctorant dans le cadre de la chaire d'attractivité BIREM, dont l'aide sur le développement d'un code de post-traitement de résultats de CFD a débloqué l'ensemble des résultats décrits dans les chapitres IV et V. Bastien, merci de m'avoir consacré du temps et de m'avoir débloqué dans une période où tu avais d'autres priorités. Merci également pour les nombreux retours que tu m'as fait sur mes développements et mes présentations. Je te souhaite une bonne continuation et j'espère que nos chemins professionnels se recroiseront. Je tiens également à remercier Guillaume Jambon qui, par le biais d'un projet qui n'a pas pu aller à son terme, m'a apporté la vision critique de quelqu'un n'étant pas de mes thématiques scientifiques. Cela m'a aidé à remettre en question des certitudes personnelles non justifiées et à donc limiter les biais dans ma démarche scientifique.

Bien que cela ne transparait que peu dans ce manuscrit, une partie non négligeable du temps consacré à ma thèse a porté sur la réalisation de simulations en mécanique des fluides numériques, en utilisant le logiciel NEPTUNE_CFD. Je tiens donc à remercier ceux qui m'ont aidé sur l'utilisation de cet outil, tout d'abord Lokman Bennani qui m'a formé à l'utilisation du logiciel, et ensuite Hervé Neau pour son aide tout au long de ma thèse. Hervé, non seulement tu as répondu à mes nombreuses demandes concernant Neptune, mais tu m'as également fait découvrir le métier d'Ingénieur en Calcul Scientifique. Tu m'as entendu lorsque dès ma deuxième année de thèse je t'ai dit être intéressé par ce métier, et voilà que deux ans plus tard j'ai le plaisir de travailler quotidiennement avec toi. À toi Hervé, mais également à l'ensemble du service CoSiNus, à savoir Alexei Stoukov, Annaïg Pedrono et Pierre Elyakime : merci de m'avoir accueilli dans votre équipe et de m'avoir aidé à préparer et à obtenir le concours grâce auquel je suis assuré de vous côtoyer ces prochaines années !

Pour arriver jusque là, j'ai eu de la chance. La chance de rencontrer d'une part ceux qui, tout au long de ma formation, m'ont laissé l'opportunité de faire mes preuves, m'ont formé et m'ont guidé, jusqu'à ma position actuelle. La chance d'autre part de rencontrer ceux qui ont fait de ces années de thèse une période si agréable de ma vie et qui auront créé cette ambiance chaleureuse propre au Hall GPE. La liste de ces personnes est longue et je vais en oublier quelques-un, qu'ils ne m'en veuillent pas. Merci à Sylvie Besses, le hasard aura voulu que d'Angers à Toulouse, je fasse une école d'ingénieurs et une thèse à quelques pas de là où vous avez fait vos propres études. Merci à Patrice Davodeau, je n'ai pas oublié notre conversation. Merci à Annie Leuridan et Laurent Bonaventure, cette année passée sur Nantes fût intense, mais vous m'avez donné les outils qui m'ont permis de poursuivre, merci de m'avoir fait confiance et d'avoir ouvert les portes de l'ATS à une formation jusque là inconnue ! Merci Aras Ahmadi pour l'ensemble de nos nombreuses discussions, l'une d'entre elle ayant eu un impact particulièrement important sur les développements que j'ai réalisés (POO), et pour ton expertise sur les questions d'optimisation qui m'aura fortement dépanné dans mes derniers mois de thèse. Merci Alain Liné, tu seras parvenu à me faire aimer la mécanique (des fluides, tout du moins), merci pour ton regard sur les aspects de simulation CFD que j'ai rencontrés dans ma thèse. Merci à Arnaud Cockx, pour l'organisation que tu as donné à l'équipe TIM, permettant aux doctorants de participer activement à la vie de l'équipe, tu m'as permis d'expérimenter à petite échelle un rôle de support informatique et cela se sera montré décisif pour mes choix d'orientation de carrière ! Finalement, merci à Maria Aurora Fernandez pour avoir été une directrice de département à l'écoute de ses étudiants, et pour m'avoir permis de réaliser mon stage de fin d'étude au Hall GPE –contre l'avis irrationnel et injustifié de la responsable des stages de 5A–, ce stage a été le commencement de ma carrière et m'a mené à mon poste actuel, et je n'aurais pas voulu qu'il en soit autrement. Pour cela et pour le reste, merci!

Il me reste à remercier chaleureusement tous ceux qui, par leur sourire, leur bonne humeur, leurs discussions, ont rendu le quotidien au Hall GPE si agréable! Tout d'abord, Daniel, Nour et Arezki : partager un bureau avec vous a été une formidable expérience à coups de luttes féministes (?...), d'entraide au branchement de câbles (par post-it intercalés), et de tentatives de réorganisation du mobilier (je maintiens qu'il y avait de bonnes idées ! Bien que le passage puisse être légèrement bloqué...). Ensuite, il y a l'ensemble de la vague des nouveaux doctorants de fin 2014 - début 2015 : David (mon co-gestionnaire de KFet'), Naïla, Elsa (responsable de l'événementiel), Angélica (qui aura relevé avec succès le défi "entre parent-thèse") et Paul. Ne pas oublier ceux qui étaient déjà dans les murs –Ana, Allan– et ceux qui allaient arriver peu après –Vincent G., Noémie (spécialiste de la fameuse Noémiade, mélange incertain de spontanéité et de raisonnements incompréhensiblement cohérents, mais elle ne serait pas d'accord avec cette définition...), Flavie, Ibrahima– ou beaucoup plus récemment –Dylan, Gaëlle, Ryma, Alexandre–. À ne citer que les (post-)doctorants, je risque de laisser l'image erronée d'un Hall GPE ou les ~~non-permanents~~ contractuels sont mis à part, alors que la secte des cruciverbistes est ouverte à tous ! Plus sérieusement, mes meilleurs souvenirs de ces dernières années seront ceux de la KFet' où, midis après midis, des liens se forgent entre deux équipes, entre (post-)doctorants, enseignants et/ou chercheurs et technicien(ne)s, autour de mots-fléchés, de café, de pâtisseries, de café, de discussions animées et de café. Je tiens d'ailleurs à immortaliser par écrit la première phrase de celui qui était alors mon directeur de stage lorsque j'ai rejoint le Hall GPE : "la recherche, c'est avant tout 50% de café, et 50% de cafet' " (Jérôme Morchain, Fev. 2014). Tout cela pour dire : Aurore, Nathalie, Manon, Colette, Claude, José, Nicolas, à tous ceux cités précédemment, et à tous ceux que j'ai malheureusement oublié de citer, *merci* pour ces moments ! Merci également de continuer à m'accueillir alors que j'ai maintenant quitté officiellement le LISBP depuis de nombreux mois, le sevrage risque de prendre encore quelques années, et pourrait ne jamais être complet !...

Si j'ai pu, après un BAC technologique et un BTS, poursuive mes études si loin, c'est pour beaucoup grâce au soutien de ma famille. D'une part de mon père, qui aura soutenu mes choix qui m'auront conduit à ce parcours atypique, et aura fait des sacrifices entre autre pour me permettre d'aller au bout de ces choix : je ne te le dis jamais assez, merci ! D'autre part à ma grand-mère, dont le hasard veut qu'elle ait travaillé sur le site industriel de Vitry-sur-Seine où se situe actuellement l'activité chez Sanofi en lien avec ma thèse, et que j'ai donc pu visiter à quelques reprises ces dernières années. Merci de m'avoir toujours soutenu et d'avoir éveillé en moi une fascination pour les sciences !

On dit de toujours garder le meilleur pour la fin, et le meilleur qui me soit arrivé depuis mon installation sur Toulouse porte le prénom de Marine. Merci de partager ma vie depuis 6 ans, à l'heure où j'écris ces lignes. Merci de supporter mes geekeries, mes divagations, mes expressions un tantinet désuettes et ma mémoire défaillante. Merci d'avoir supporté sans (trop) broncher le doctorant en fin de thèse, en rédaction, en préparation de soutenance (cumulé à un concours) que j'ai été. Je serai là pour te soutenir lorsque viendra ton tour, et pour ce qui viendra ensuite.

## NOTATIONS

### Roman symbols

| Symbol | Description | Unit |
|---|---|---|
| $a^\star$ | Specific gas-liquid transfer area | $\mathrm{m^2/m^3}$ |
| $a^\circ$ | Gas-liquid transfer area per gas volume unit | $\mathrm{m^2/m^3_G}$ |
| $\boldsymbol{C}$ | Vector of mass concentrations | $\mathrm{kg/m^3}$ |
| $d_{32}$ | Bubble Sauter diameter | m |
| $d_b$ | Bubble diameter | m |
| $D$ | Dilution rate | $\mathrm{s^{-1}}$ |
| $\mathcal{D}$ | Diffusion/dispersion coefficient | $\mathrm{m^2/s}$ |
| $f$ | Specific consumption or production molar transfer rate | $\mathrm{mol/kg_X \cdot s}$ |
| $\vec{g}$ | Gravity acceleration | $\mathrm{m/s^2}$ |
| $H$ | Gas solubility, Henry constant | $\mathrm{kg_L/kg_G}$ |
| $H$ | Shannon entropy | |
| $\boldsymbol{J}$ | Jacobi matrix | |
| $k_L$ | Vector of volumetric reaction rates | $\mathrm{kg/m^3 \cdot s}$ |
| $K$ | Biological affinity constant | $\mathrm{kg/m^3_L}$ |
| $K_i$ | Biological inhibition constant | $\mathrm{kg/m^3_L}$ |
| $m$ | Maintenance rate | $\mathrm{mol_G/kg_X \cdot s}$ |
| $m$ | Moment of a number density function | |
| $m_p$ | Particle mass | $\mathrm{kg_X}$ |
| $\vec{M}$ | Momentum exchange rate | $\mathrm{kg/m^2 \cdot s^2}$ |
| $M$ | Molar mass of chemical species | kg/mol |
| $n$ | Concentration density function of cell properties distribution | |
| $n_b$ | Number density function of bubble size distribution | |
| $N_{\mathrm{cell}}$ | Number of biological cells | |
| $\widehat{p}$ | Vector of biological properties | |
| $P$ | Pressure | Pa |
| $q$ | Specific reaction or metabolic-mode rate | $\mathrm{mol/kg_X \cdot s}$ |
| $\boldsymbol{r}$ | Vector of specific reaction rate | $\mathrm{kg/kg_X \cdot s}$ |
| $\boldsymbol{R}$ | Vector of volumetric reaction rates | $\mathrm{kg/m^3 \cdot s}$ |
| $t$ | Time | s |
| $T$ | Characteristic time-scale of phenomena | s |
| $\vec{u}$ | Velocity | m/s |
| $V$ | Volume | $\mathrm{m^3}$ |
| $\vec{x}$ | Vector of location | m |
| $Y$ | Conversion yield | mol/mol |

# Greek symbols

| Symbol | Description | Unit |
|---|---|---|
| $\alpha$ | Local phase fraction | $\mathrm{m_k^3/m^3}$ |
| $\beta$ | Probability Density Function of property redistribution | |
| $\kappa$ | Kernel Density Function | |
| $\mu$ | Dynamic fluid viscosity | $\mathrm{Pa \cdot s}$ |
| $\mu$ | Biological growth-rate | $\mathrm{kg_X/kg_X \cdot s}$ |
| $\boldsymbol{\varphi}$ | Specific consumption or production mass transfer rate | $\mathrm{kg/kg_X \cdot s}$ |
| $\boldsymbol{\Phi}$ | Volumetric mass transfer rate | $\mathrm{kg/m^3 \cdot s}$ |
| $\pi$ | Orthogonal polynomial | |
| $\rho$ | Phase density | $\mathrm{kg/m^3}$ |
| $\sigma$ | Glycolytic stress | |
| $\boldsymbol{\tau}$ | Stress tensor | $\mathrm{Pa}$ |
| $\zeta$ | Adaptation rate of biological properties | |

# Subscripts and superscripts

| Symbol | Description |
|---|---|
| $x_{@p}$ | Particle or individual attached variable |
| $x^*$ | Variable defined at equilibrium with environment |
| $x^{(a)}$ | Actual (realised) output of biological model |
| $x^{(b)}$ | Biological phase attached variable |
| $x^{(e)}$ | Environment attached variable |
| $x^{(g)}$ | Global variable accounting for biological and environmental conditions |
| $x^{(\max)}/x_{\max}$ | Maximum possible value |
| $\tilde{x}$ | Population mean value |
| $\langle x \rangle$ | Volumetric mean value |
| $\bar{x}$ | Time mean value |

# Tensors and operators

Tensors are denoted by a bold font in mathematical notations, *e.g.* $\boldsymbol{C}$. This manuscript refers to vectors and operators defined either in the geometrical space, *e.g.* $\vec{x}$ and $\vec{\nabla}$, or in a space of individual-scale properties as defined under the framework of Population Balance Models, *e.g.* $\widehat{x}$ and $\widehat{\nabla}$. On top of classical vectorial notations (dot-product $\vec{a} \cdot \vec{b}$, cross-product $\vec{a} \times \vec{b}$, tensor-product $\vec{a} \otimes \vec{b}$), this manuscript makes use of Hadamard operators for element-wise products ($\boldsymbol{A} \circ \boldsymbol{B}$) and element-wise divisions ($\boldsymbol{A} \oslash \boldsymbol{B}$).

# Abbreviations

| | |
|---|---|
| ATP | Adenosine TriPhosphate |
| BSD | Bubble Size Distribution |
| CDF | Concentration Density Function |
| CFD | Computational Fluid Dynamics |
| CFL | Courant-Friedrichs-Lewy |
| CFU | Colony-Forming Units |
| CMA | Compartment Model Approach |
| DNS | Direct Numerical Simulation |
| EQMOM | Extended Quadrature Method of Moments |
| FBP | Fructose-1,6-BiPhosphate |
| FLOP | Floating-Point Operations |
| GLIM | Generalized Large-Interface Model |
| G6P | Glucose-6-Phosphate |
| HMMC | High-order Moment conserving Method of Classes |
| KDEM | Kernel Density Element Method |
| KDF | Kernel Density Function |
| LES | Large-Eddies Simulation |
| LHS | Left-Hand Side |
| MRF | Multiple Reference Frames |
| NDF | Number Density Function |
| ODE | Ordinary Differential Equation |
| PBE | Population Balance Equation |
| PBM | Population Balance Model |
| PDE | Partial Differential Equation |
| PDF | Probability Density Function |
| PIV | Particle Image Velocimetry |
| POD | Proper Orthogonal Decomposition |
| PTS | PhosphoTransferase System |
| QMOM | Quadrature Method of Moments |
| RANS | Reynolds-Averaged Navier-Stokes |
| RHS | Right-Hand Side |
| RTD | Residential Time Distribution |
| TSM | Two-Size Moment |

# CONTENTS

# I

# CONTEXT, STATE OF THE ART AND OBJECTIVES

## I.1 Bioreactors in current economy

Biotechnologies refer to “any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use” [United Nations, 1992]. This broad definition encompasses a large-variety of applications in most fields of current industries, classified through so-called colors of biotechnologies. Green biotechnologies refer to agriculture use, blue to marine applications, red to health and pharmaceutical applications, white to industrial productions, and so on [Kafarski, 2012]. The global market size for biotechnologies was estimated at approximately USD 330 billion in 2015 and could reach up to USD 775 billion by 2024. This impressive expected growth can be attributed to recent breakthroughs in experimental capabilities to interact with biological systems, from fast and affordable genome sequencing to targeted genome edition [*e.g.* CRISPR-Cas9, Bolotin et al., 2005]. These experimental advances also benefit from a context of fossil fuels depletion to which biotechnologies are expected to provide an answer through biofuel, biogas or bio-material productions.

Whilst biotechnologies are not new and can be traced back to alcoholic beverage productions in all known civilisations, a major challenge for current industries is “to move from traditional methods towards more standardized industrial processes” [Noorman and Heijnen, 2017]. Most current productions occur in large-scale bioreactors: large mixed culture tanks, possibly aerated, in which a substrate is converted by a biological population into a product of interest. These products can be specific high value molecules (hormone, pharmaceuticals, ...), solvants, biopolymers, biofuels, biogas or even food ingredients. Considering the high volume or high value productions in each of these markets, any improvement to bioreactor processes will have significant financial impact. Doing so will require to understand local phenomenon occurring in such systems, and to couple these understandings of elementary phenomenon to predict the large-scale behaviour of the process. This, by itself, constitutes an exciting challenge for both academic and industrial communities.

## I.2 Challenges for operating robust large-scale bioprocesses

When designing a bioprocess, the first steps consist in identifying a micro-organism that converts an available substrate into the targeted molecule, either naturally, or after metabolic engineering. This organism is then tested in lab-scale cultures in flasks or litre-scale bioreactors, and optimal culture conditions are then determined (pH, oxygen/substrate concentrations, temperature, …). However, when scaling-up these cultures to industrial scales, conversion yield loss are usually observed.

Enfors et al. [2001] pointed out the existence of large-scale gradients as the main explanation for this difference of behaviour between lab-scale and industrial-scale cultures. Indeed, in litter-scale reactor, the mixing is fast enough to ensure homogeneous conditions in terms of pH, temperature, substrate or oxygen concentrations, … Therefore, when selecting micro-organisms for their high conversion performances in lab-scale cultures, the selection is biased in that chosen organisms perform well, but under tightly controlled conditions. To reach high production rates, industrial bioreactors are high volume systems. However, the higher the volume, the higher the mixing time. Therefore, it is usually not possible to replicate the perfectly controlled homogeneous conditions of lab-scale cultures into production-scale bioreactors, where gradients will then appear.

Cells travelling in an heterogeneous environment will experience fluctuations, of pH, temperature, concentration and so on. They will undergo periodic starvation and overfeeding [Ferenci, 2001], they will face aerobic and anaerobic conditions every few seconds or minutes, or switch between nitrogen and carbonaceous limitations [Löffler et al., 2017]. These conditions will be drastically different from tightly controlled lab-scale cultures in which the strain has been selected. This is known to cause undesired metabolic effect through the activation of unproductive biological reactions [Neubauer et al., 1995; Brand et al., 2018] as the natural objective of micro-organisms is survival, not industrial production.

## I.3 Using bioprocesses simulation to improve their robustness

As soon as large-scale bioreactors are considered, one has to admit that gradients will always appear in these systems; that cells will have to travel in these gradients and that their metabolism will be disturbed by these environmental fluctuations. *In silico* “experimentations” –or simulations– of bioreactors are useful to assess these nefarious effects. They can also become tools to design bioprocesses or optimise mixing, feeding strategy or other operating conditions in order to reduce unwanted large-scale effects and improve production yields. The use of simulation tools is expected to replace the use of empirical scale-up criteria that are, for now, used to design industrial bioprocesses [Takors, 2012].

A prerequisite to bioreactors simulations is the definition of a model, which sums up to a closed set of equations describing the dynamic evolution of a biological culture, and the use of suitable numerical methods to solve these equations. The basis of most models are local mass balance equations such as

$$\frac{\partial \boldsymbol{C}(\vec{x},t)}{\partial t} + \vec{u}(\vec{x},t)\cdot\vec{\nabla}\boldsymbol{C}(\vec{x},t) - \vec{\nabla}\cdot\left(\mathcal{D}\vec{\nabla}\boldsymbol{C}(\vec{x},t)\right) = \boldsymbol{R}(\boldsymbol{C}(\vec{x},t)) \tag{I.1}$$

with $\vec{x}$ and $t$ the location and time, $\vec{u}$ the local fluid velocity and $\boldsymbol{C}$ a vector of concentrations of dissolved species that are convected and dispersed/diffused by the fluid motion. This equation describes the evolution of local concentrations in an elementary volume of fluid in which some reactions occur.

If only chemical reactions were considered, the Right-Hand Side (RHS) term would be a sink or source term that could be attributed to these reactions and could be directly computed from thermodynamics and kinetic laws. In the case of bioreactors, these reaction terms must account for the contribution of each viable micro-organism to the overall observed local reaction rate.

Up to now, very few works attempted to perform large-scale simulations of bioreactors, and even less managed to use these simulations as decision assisting tools. Therefore, there is still room left for the emergence of engineering tools that will help the decision making process. One goal of this thesis is to provide a simulation framework that allows performing simulations of large-scale biological systems, in a reasonable amount of computation time, to pave the way for the development of such tools.

## I.4 Challenges in bioreactors modelling

To start, we shall explain what still limits the emergence of engineering tools based on bioreactor simulations. Bioreactors are large tanks, usually aerated through gas injection, mixed by one or multiple impellers, in which micro-organisms are cultivated. This implies that interactions occur continuously at multiples length and time scales. Even though Eq. (I.1) is an over-simplification of a bioreactor model, it still emphasize the fact that two aspects are core to the modelling: physical transport on the Left-Hand Side (LHS), and bioreactions on the RHS. Both aspects must therefore be dealt with and will bring complementary challenges to the modelling of bioprocesses.

### I.4.1 Describing large-scale hydrodynamics...

In terms of transport, a complete model must be able to describe the velocity field within the reactor. This can be done either by enforcing a circulation flow through so-called Compartment Models [Delafosse et al., 2014; Vrábel et al., 1999] or by solving equations of fluid mechanics, namely Navier-Stokes equations, for multiphase flows. Bioreactors are three-phases systems, with a liquid culture medium, gas bubbles for aeration, and a biological phase made of micrometric living particles.

Numerous works focus on the simulation of gas-liquid stirred systems using Computational Fluid Dynamics (CFD) [Schütze and Hengstler, 2006; Moilanen et al., 2008; Zhang et al., 2009; Elqotbi et al., 2013]. Due to the low Stokes number of micro-organisms, biological cells follow the same trajectories as fluid-particle which explains why two-phases simulations are sufficient to describe hydrodynamics in bioreactors [Delafosse, 2008; Linkès et al., 2014]. In these simulations, the liquid phase is actually a mixture representation of both liquid and biological phases, and one can integrate the rheological effects of biomass on the properties of that mixture phase [Bezzo et al., 2003; Moilanen et al., 2007; Laupsien, 2017].

### I.4.2 ... as well as single-cell functioning ...

Despite the modelling difficulties tied to CFD approaches, the hydrodynamic description might not be the most challenging aspect of bioreactors modelling. The simple, almost *naive*, reaction term $\boldsymbol{R}(\boldsymbol{C}(\vec{x},t))$ in Eq. (I.1) encompasses a wide variety of phenomena whose modelling is all but straightforward.

First, at the smallest scales, bioreactions occur within each viable micro-organism. These living systems uptake some substrates from their environment (*i.e.* sugars, oxygen, ammonium, organic molecules, ...) and process them through metabolic pathways to produce new cells, energy, and multiple by-products.

Internal metabolic –or reactive– pathways a cell is capable-of may differ between strains, but also between individuals of the same strain. Numerous modern experimental tools are used nowadays to investigate this wide variety of living systems, to understand and even predict the internal interactions, reactions and regulations that occur in each cell. These tools are usually gathered within the `omics family: fluxomics, metabolomics, proteomics, genomics, . . . [Winter and Krömer, 2013].

Thanks to recent progresses in experimental capabilities, databases and models start to emerge to describe the overall functioning of cells for strains commonly used in industrial bioprocesses. Orth et al. [2011] propose such a model for the bacteria *Escherichia coli* by describing 1366 genes, 2251 reactions between 1366 metabolites and numerous transports and interactions. Similarly, Heavner et al. [2012, 2013] provide incremental updates of a consensus yeast model for *Saccharomyces cerevisiae* that, in its version 6, described 900 genes, 1458 metabolites and 1888 possibles reactions between these.

For reactions to take place within a cell, substrates must first be up-taken from the culture medium. A particularity of biological system is that they manage to prevent thermodynamic equilibrium at the liquid-biological phases interface through the expense of energy. Transport between this two phases is regulated by the biological membrane and by the activation –or deactivation– of different transporters Ferenci [1996, 1999]; Gosset [2005]. The state of these transporters evolves along a cell's trajectory due to complex regulation mechanisms that respond to fluctuations of external (pH, concentrations, light, temperature, ...) and internal (metabolite concentrations, storages, ...) signals [*e.g.* Schlegel et al., 2002]. Tracking the state of these transporters appears to be required to predict the overall reaction terms in Eq. (I.1).

Overall, micro-organisms can be seen as reactive particles in heterogeneous catalysis systems. In order for the reaction to occur, substrate must first be brought near a cell membrane through external transport. Due to the size of micro-organisms, the description of this external transport cannot be done directly by solving fluid-dynamics equations. Indeed, this would require to solve all scales of turbulence up to the Kolmogorov and Batchelor length-scales (1 to 10 $\mu m$ under commonly observed mixing rates in bioreactors, Delafosse [2008]). This would roughly require discretizing the whole volume of a $100m^3$ bioreactor into $10^{17}$ to $10^{20}$ volume elements. To give an idea of the infeasibility of such a computation, as of November 2017, the super computer with the biggest memory could "only" hold approximately 1.7 $10^{15}$ bytes of information. But even if such a simulation was feasible, it would not be relevant if the same level of accuracy cannot be met when describing the behaviour of micro-organisms. In particular, no model is able to predict accurately the impact of short term fluctuations ($\approx 1\mu s$) in a cell environment on its metabolic behaviour. A more standard chemical engineering approach will make use of closure laws based on the comparison of characteristic times of micromixing and reaction. This approach will be sufficient for the overall modelling of bioreactions in a fluid particle.

### I.4.3 ... among billions of different cells.

One last source of complexity for the modelling of bioreactors is yet to be considered. As stated previously, micro-organisms respond to environmental fluctuations through regulation and dynamic evolution of internal concentrations of numerous metabolites. From one cell to an other, fluctuations will differ, and thus cells will differentiate in response to this so-called extrinsic noise [Delvigne and Goffin, 2014]. Moreover, cells may differentiate spontaneously due to stochastic noise in the occurrence of internal reactions through molecular crowding [Klumpp et al., 2013] or due to the uneven distribution of cellular content during cell division or budding. This stochastic differentiation is referred to as intrinsic noise, and ensures that biological heterogeneity exists, even in well controlled environments.

If this heterogeneity did not exist, the term $\boldsymbol{R}(\boldsymbol{C})$ in Eq. (I.1) could be written as simply as

$$\boldsymbol{R}(\boldsymbol{C}) = N_{\text{cell}}\ \boldsymbol{r}(\boldsymbol{C}) \tag{I.2}$$

with $N_{\text{cell}}$ the number of cell in a volume of liquid, and $\boldsymbol{r}$ the consumption and production rates of a single cell.

Due to the biological heterogeneity, one must introduce a vector of cell properties, $\widehat{p}$, that will impact the metabolism and thus the rates of consumption and production. This vector may contain information such as its mass, size, age, internal concentrations of metabolites or other variables of interest. This vector is different between cells, and thus the source/sink term in Eq. (I.1) must be written

$$\boldsymbol{R}(\boldsymbol{C}) = \sum_{i=1}^{N_{\text{cell}}} \boldsymbol{r}(\boldsymbol{C}, \widehat{p}_i) \tag{I.3}$$

The introduction of this state vector comes with multiple modelling difficulties. Gathering experimental data at the scale of a biological population is still a difficult task. Few works ensure a stringent control of cells environment through microfluidics to access cell-attached properties [Yasuda, 2011; Nobs and Maerkl, 2014] and an other currently developped approach is the use of flow cytometry to access cell-scale information [Brognaux et al., 2013]. Modelling the evolution of these variables is then a challenge due to the limited accessibility to experimental data.

Moreover, even if one can predict the evolution of a micro-organism state over time, hardly no information is usually available about the distribution of biological state at the beginning of a biological culture.

Finally, a simulation tool based on a model that account for the varying biological state should be able to describe the variety of internal properties among a biological population. Note that a cell concentration of $10^9$ Colony-Forming Units (CFU) per millilitre is quite usual. If one aims at performing fast simulations of bioreactors while accounting for biological heterogeneity, a strong focus must be made on keeping track of the diversity among such a large population, without following each individual isolatedly.

## I.5 Bioreactors as three phases systems

At the most fundamental level of description, bioreactors are three-phases systems:

- an aqueous liquid phase, constituted by the culture medium which carries other phases;
- a gaseous phase, usually bubbles of air or pure oxygen, meant to transfer oxygen to the culture medium and possibly to strip carbon dioxide or volatile compounds from it;
- a biological phase which catalyses reactions of interest within a high number of micro-organisms.

Each of these phases carries dissolved matter, and they all exchange momentum , matter and heat. The description of these transport and transfer phenomenon are core to the modelling of bioreactors, in particular when considering the transfer of substrates and products between the culture medium and micro-organisms. Therefore, a complete modelling of these systems must account for all three phases.

### I.5.1 Euler-Euler treatment of gas and liquid phases

Large-scale gas-liquid systems are often modelled using the Euler-Euler description in which both phases are considered as inter-penetrating continua. This approach solves local volumetric phase fractions $\alpha_k$ and velocities velocity $\vec{u}_k$ by formulating continuity and momentum balance equations on each phase.

This Euler-Euler approach is especially suited to described homogeneous phases. The gas phase is actually dispersed, bubble size is not monodispersed and gas concentration may differ from one bubble to another. But overall, thanks to coalescence and breakage in stirred bioreactors, it can be considered that the gas phase is almost homogeneous so that one does not need to describe concentration heterogeneity between close bubbles.

Many authors have gathered experimental data about bubble-size distributions in agitated tank [Barigou and Greaves, 1992; Machon et al., 1997; Alves et al., 2002] and others have used this data to model breakage, coalescence and transport phenomena over bubble populations [Ribeiro and Lage, 2004; Laakkonen et al., 2007; Buffo et al., 2012; Yang and Xiao, 2017]. Despite the relevancy of these models for large-scale bioreactors, no particular focus was made during this work on the gas-liquid aspect of the overall modelling.

Overall, the Euler-Euler treatment of gas and liquid phases can be summed up to three equations for each phase.

Continuity equations are

$$\frac{\partial \alpha_G \rho_G}{\partial t} + \vec{\nabla} \cdot (\alpha_G \rho_G \vec{u}_G) = 0 \tag{I.4}$$

$$\frac{\partial \alpha_L \rho_L}{\partial t} + \vec{\nabla} \cdot (\alpha_L \rho_L \vec{u}_L) = 0 \tag{I.5}$$

where

- $k \in \{G, L, B\}$ designates respectively the gas, liquid and biological phase;
- $\alpha_k$ designates the volumetric phase fraction $(m_k^3/m^3)$, $\sum_k \alpha_k = 1$;
- $\rho_k$ is the phase density $(kg_k/m_k^3)$;

- $\vec{u}_k$ is the local velocity of phase $k$ $(m/s)$.

Momentum conservation equations are

$$\frac{\partial \alpha_G \rho_G \vec{u}_G}{\partial t} + \vec{\nabla} \cdot (\alpha_G \rho_G \vec{u}_G \otimes \vec{u}_G) = \alpha_G \left(\rho_G \vec{g} - \vec{\nabla} P\right) + \vec{\nabla} \cdot \boldsymbol{\tau}_G + \vec{M}_G \tag{I.6}$$

$$\frac{\partial \alpha_L \rho_L \vec{u}_L}{\partial t} + \vec{\nabla} \cdot (\alpha_L \rho_L \vec{u}_L \otimes \vec{u}_L) = \alpha_L \left(\rho_L \vec{g} - \vec{\nabla} P\right) + \vec{\nabla} \cdot \boldsymbol{\tau}_L + \vec{M}_L \tag{I.7}$$

where

- $\vec{g}$ is the body acceleration which only encompasses gravity in bioreactors ($\mathrm{m/s^2}$);
- $P$ is the local pressure, shared by all phases $(Pa)$;
- $\boldsymbol{\tau}_k$ is the stress tensor of phase $k$ $(Pa)$ ;
- $\vec{M}_k$ represents momentum exchange between phases $(kg_k/m^2.s^2)$.

Finally, both gas and liquid phases carry numerous dissolved species (substrate, oxygen, ...) whose concentrations are tracked by the following mass balance equations:

$$\frac{\partial \alpha_G \boldsymbol{C}_G}{\partial t} + \vec{\nabla} \cdot \left(\alpha_G \boldsymbol{C}_G \vec{u}_G - \vec{\nabla} \left(\alpha_G \boldsymbol{\mathcal{D}}_G \circ \boldsymbol{C}_G\right)\right) = \alpha_G \boldsymbol{R}_G + \boldsymbol{\Phi}_{LG} + \boldsymbol{\Phi}_{BG} \tag{I.8}$$

$$\frac{\partial \alpha_L \boldsymbol{C}_L}{\partial t} + \vec{\nabla} \cdot \left(\alpha_L \boldsymbol{C}_L \vec{u}_L - \vec{\nabla} \left(\alpha_L \boldsymbol{\mathcal{D}}_L \circ \boldsymbol{C}_L\right)\right) = \alpha_L \boldsymbol{R}_L + \boldsymbol{\Phi}_{GL} + \boldsymbol{\Phi}_{BL} \tag{I.9}$$

where

- $\boldsymbol{C}_k$ is a vector of concentrations carried by phase $k$ $(kg/m_k^3)$;
- $\boldsymbol{\mathcal{D}}_k$ is the vector of diffusion/dispersion rates of species carried by phase $k$ $(m^2/s)$;
- $\boldsymbol{R}_k$ is a source or sink term related to chemical reactions occurring in phase $k$ $(kg/m_k^3.s)$;
- $\boldsymbol{\Phi}_{kk'}$ is the rate of mass transfer from phase $k$ to phase $k'$ $(kg/m^3.s)$.

In most bioreactors, no significant chemical reaction occur in gas and liquid phases, except for acid-base reactions which play a role in the pH of the culture medium. We choose not to cover this last aspect, hence in this manuscript $\boldsymbol{R}_G = 0$ and $\boldsymbol{R}_L = 0$. Similarly, gaseous compounds usually dissolved into the liquid phase and do not transfer directly from the gas phase to micro-organisms which implies that $\boldsymbol{\Phi}_{BG} = 0$.

All reactions of interest occur within micro-organisms. Therefore, a strong focus must be made on mass transfer between liquid and biological phases. While mass transfer between gas and liquid phases can be deduced from thermodynamic laws (see I.6.1), biological systems have the unique capability of preventing thermodynamic equilibrium. Micro-organisms are indeed able to dynamically regulate the flow of matter through their membrane, at the expense of energy, in order to improve assimilation in poor environments [Ferenci, 1996], and to reduce assimilation in too rich environments. The modelling of $\boldsymbol{\Phi}_{BL}$ will then be a crucial aspect and will require to accurately describe the biological phase.

Opposite to the gas phase, micro-organism do not undergo coalescence or breakage and, as explained previously, they tend to differentiate from one another due to environmental fluctuations or to internal stochastic noise [Klumpp et al., 2013]. Thus, the biological phase must be modelled as a heterogeneous particulate phase. *i.e.* each particle has its own set of properties that impacts the computation of $\mathbf{\Phi}_{BL}$, and these properties differ between particles .

Note that Eqs. (I.4) and (I.5) are only valid under the assumption that the gas-liquid mass-transfer ($\mathbf{\Phi}_{GL}$) has a negligible impact of phase volumes. This mass transfer should be integrated, for instance if all oxygen is transferred from air to the liquid phase inducing a 21% loss of volume in the gas phase.

### I.5.2 Lagrangian description of micro-organisms

At its fundamental level, a Lagrangian treatment of a dispersed phase consists in a tracking of trajectories of numerous particles. In the present context of bioreactors modelling, micro-organisms can be modelled as Lagrangian particles whose trajectories follow liquid streamlines considering that they are non-inertial particles. This is verified by comparing their characteristic relaxation time, estimated around $\approx 10\mu$s by Delafosse [2008], against smallest turbulence time-scales, estimated by the same author to be around a few milliseconds.

Along with each particle trajectory, one solves a set of ordinary differential equations that defines the evolution of particle-attached properties. Therefore, the Lagrangian approach sums up to a trajectory equation, and a property evolution equation, attached to each particle:

$$\frac{\partial \vec{x}_{@p}}{\partial t} = \vec{u}_L(\vec{x}_{@p}) \tag{I.10}$$

$$\frac{\partial \widehat{p}_{@p}}{\partial t} = \widehat{\zeta}\left(\widehat{p}_{@p}, \boldsymbol{C}_L(\vec{x}_{@p})\right) \tag{I.11}$$

where

- subscript $@p$ designates variables attached to a Lagrangian particle;
- $\vec{x}_{@p}$ is the location of a particle;
- $\widehat{p}_{@p}$ is the vector of biological properties attached to a particle;
- $\widehat{\zeta}$ is the rate of evolution of $\widehat{p}_{@p}$ ([unit of $\widehat{p}$]$/s$).

Under this formalism, local transfer rates from liquid to cells is given by

$$\mathbf{\Phi}_{BL}(\vec{x}, t) = \sum_i D_p(\vec{x}, t) m_p \boldsymbol{\varphi}\left(\widehat{p}_{@p_i}, \boldsymbol{C}_L(\vec{x}, t)\right) \tag{I.12}$$

where

- the summation is performed over Lagrangian particle strictly located at $\vec{x}$;
- $\boldsymbol{\varphi}$ is the vector of specific consumption/production rates ($\mathrm{kg/kg_{cell}\cdot s}$);
- $D_p$ is the local particle density (particle$/m_L^3$).

- $m_p$ is the unitary particle mass ($\mathrm{kg_{cell}}$)

In order to reach high accuracies, this approach requires a high number of particles, so that high local particle density is obtained in every location of the considered bioreactor. Without this high density, no statistical convergence can be met and the evaluation of $\boldsymbol{\Phi}_{BL}$ will be skewed. This high density can be reached for small-scale reactors, but is yet to be seen in large-scale simulations.

Currently, the best work in terms of Lagrangian simulations for bioreactors made use of Lagrangian particles to record environmental perturbations seen by micro-organisms along their trajectory [Haringa et al., 2016]. However, this work did not consider the gas phase, and did not resolve the consumption of substrate using Lagrangian particles. Concentration fields where resolved using a kinetic model which did not account for biological heterogeneity. These simulations are then useful to assess the effect of environmental perturbation on biological populations in large-scale fermenters, but are not yet used for predictive modelling and simulation of bioreactors.

### I.5.3 Population Balance Equation for biological phase

Many scientific communities face similar modelling difficulties when dealing with heterogeneous large populations. An example among others is crystallizers in which numerous particles (crystals) have particle-attached properties such as size and morphology [Pagliolico et al., 1999; Jung et al., 2000]. The tracking of these properties at the particle scale is required to follow their evolution over time for each crystal.

A modelling framework has been developed for such situations: Population Balance Models (PBMs). These models describe the evolution of a statistical distribution of particle-attached properties. Let $n(\widehat{p})$ be a Number Density Function (NDF) describing this statistical distribution of properties:

$$N(\widehat{p} \in [\widehat{p}_0, \widehat{p}_0 + \Delta\widehat{p}]) = \int_{\widehat{p}_0}^{\widehat{p}_0 + \Delta\widehat{p}} n(\widehat{p}) \mathrm{d}\widehat{p} \tag{I.13}$$

$N$ is the amount (number, mass or local concentration) of particles whose state $\widehat{p}$ is included between $\widehat{p}_0$ and $\widehat{p}_0 + \Delta\widehat{p}$. Throughout this work, $n(\widehat{p})$ is defined as a Concentration Density Function (CDF), making $N$ homogeneous to a cell concentration ($\mathrm{kg_{cell}/m_L^3}$).

A PBM describes all phenomenon that modify $n(\widehat{p})$ under the formalism of a Population Balance Equation (PBE):

$$\frac{\partial \alpha_L n(\widehat{p})}{\partial t} + \vec{\nabla} \cdot \Big(\alpha_L \vec{u}_L n(\widehat{p}) - \mathcal{D}_B \vec{\nabla} \left(\alpha_L n(\widehat{p})\right)\Big) + \alpha_L \widehat{\nabla} \cdot \Big(\widehat{\zeta} \left(\widehat{p}, \boldsymbol{C}_L\right) n(\widehat{p})\Big) = B(\widehat{p}) - D(\widehat{p}) \tag{I.14}$$

where

- dependency over space $\vec{x}$ and $t$ is not detailed for sake of clarity but applies to $n(\widehat{p})$, $\alpha_L$, $\vec{u}_L$, $\boldsymbol{C}_L$, $B(\widehat{p})$ and $D(\widehat{p})$;
- $\mathcal{D}_B$ is the dispersion coefficient of cells ($\mathrm{m^2/s}$);
- $B(\widehat{p})$ is the rate of birth of new micro-organisms in the state $\widehat{p}$;
- $D(\widehat{p})$ is the rate of death of micro-organisms in the state $\widehat{p}$.

This PBE shares a similar structure with Eq. (I.9). The second term on the LHS accounts for the transport of cells thank to their carrying fluid, and is analogous to the trajectory equation under a Lagrangian treatment of the biological phase (Eq. (I.10)). Source and sink terms are on the RHS to describe new-cell formation through cell division, or their possible death, just like reaction or transfer terms describe added or removed dissolved species in Eq. (I.9).

The major difference between Eqs. (I.9) and (I.14) comes from the fact that the NDF does not only depends on location $\vec{x}$ and time $t$, but also on $\widehat{p}$. Under the PBM formalism, $\widehat{p}$ must be seen as a location, not in the physical space, but in the space of biological properties. Therefore, the evolution of cell's properties in the Lagrangian approach (Eq. (I.11)) is now described as a transport term in the internal properties space, with a velocity $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L)$, in the third term on the LHS of the PBE.

Provided some numerical methods to transform the PBE into a set of equation that exactly match the formalism of (I.9), this approach is directly compatible with a Eulerian multiphase modelling. In particular, under this formalism, the overall mass transfer-rate between the liquid and biological phases can be written

$$\boldsymbol{\Phi}_{BL}(\vec{x}, t) = \int_{\Omega_{\widehat{p}}} n(\widehat{p}, \vec{x}, t) \boldsymbol{\varphi} \left(\widehat{p}, \boldsymbol{C}_L(\vec{x}, t)\right) \mathrm{d}\widehat{p} \tag{I.15}$$

where $\boldsymbol{\varphi}$ is the specific rate of bioreactions ($\mathrm{kg/kg_{cell} \cdot s}$). The overall accuracy of this method depends on how accurate the NDF resolution is.

## I.6 Usual modelling closures for bioreactors

The description given in I.5 of elementary equation sets for bioreactor modelling is meant to be as generic as possible. All existing work about bioreactor models and simulations can be integrated under this formalism provided some closures or model reductions. Such usual closures or approach are detailed hereafter, to lay the foundations of discussions about the degree of accuracy that should be used on each aspect of the overall modelling.

### I.6.1 Gas-liquid mass transfer

Equations (I.8) and (I.9) both describe the transfer of matter between liquid and gas phases through the terms $\boldsymbol{\Phi}_{GL}$ and $\boldsymbol{\Phi}_{LG}$. This mass transfer occur at the gas-liquid interface. The transfer of oxygen or carbon dioxide is usually limited by a resistance in the liquid film. Therefore, the gas-liquid mass transfer of one of these compound is usually modelled as

$$\Phi_{GL} = k_L a^* (H C_G - C_L) \tag{I.16}$$

where

- $k_L$ is the liquid-film resistance to transfer (m/s);
- $a^*$ is the specific transfer area ($\mathrm{m^2/m^3}$);
- $H$ is the solubility ($\mathrm{kg_L/kg_G}$).

$H$ is a thermodynamic equilibrium constant illustating the linear relationship between liquid and gas phase concentrations (or partial pressure) of a compound at rest as described by Henry's law [Sander, 2015]. Its value mainly depends on temperature and on the composition of both phases.

The transfer resistance, $k_L$, is related to convection and diffusion phenomena in the liquid film. Numerous correlation are available in literature to estimate its value for bubbles in water. These correlations usually take the following form:

$$\mathrm{Sh} = a + b\mathrm{Re}^c + \mathrm{Sc}^d \tag{I.17}$$

with

$$\mathrm{Sh} = \frac{k_L\ d_b}{\mathcal{D}}, \qquad \mathrm{Re} = \frac{\rho_L\ d_b\ |\vec{u}_L|}{\mu_L}, \qquad \mathrm{Sc} = \frac{\mu_L}{\rho_L\ \mathcal{D}} \tag{I.18}$$

where

- Sh, Re and Sc respectively designate the dimensionless Sherwood, Reynolds and Schmidt numbers;
- $d_b$ is the bubble diameter (m);
- $\mathcal{D}$ is the dissolved gas molecular diffusivity ($\mathrm{m^2/s}$);
- $\mu_L$ is the liquid dynamic viscosity ($\mathrm{Pa \cdot s}$);
- $a$, $b$, $c$ and $d$ are the correlation fitting parameters.

As shown in the definition of dimensionless numbers, the resistance $k_L$ is dependent on the bubble diameter $d_b$. Numerous experimental analysis show that their exists a Bubble Size Distribution (BSD) in stirred aerated reactors. Let $n_b(d)$ be that distribution in a volume of reference $V$. It is possible to express the specific gas-liquid transfer area from that distribution:

$$[\text{specific area}] = \frac{[\text{total bubble area in V}]}{[\text{total bubble volume in V}]} \times \frac{[\text{total bubble volume in V}]}{V}$$

which numerically translates into

$$a^* = a^\circ \alpha_G \tag{I.19}$$

with $a^\circ$ the specific area expressed per unit of gas volume ($\mathrm{m^2/m^3_G}$):

$$a^\circ = \frac{\int_{\Omega d_b} \pi d_b^2 n_b(d_b)\mathrm{d}d_b}{\int_{\Omega d_b} \frac{\pi}{6} d_b^3 n_b(d_b)\mathrm{d}d_b} = \frac{6}{d_{32}} \tag{I.20}$$

$d_{32}$ is known as the Sauter diameter and is an integral property of the BSD.

If the BSD is resolved, the $k_L$ value used in Eq. (I.16) should be a mean value over all bubbles:

$$\tilde{k}_L = \frac{\int_{\Omega d_b} k_L(d_b) n_b(d_b)\mathrm{d}d_b}{\int_{\Omega d_b} n_b(d_b)\mathrm{d}d_b} \tag{I.21}$$

## I.6.2 Phase-fractions and velocity fields

All local equations (Eqs. (I.4-I.9), (I.10) and (I.14)) require the knowledge of local phase fractions and velocities. These informations could be accessed locally by experimental measurements based, for instance, on Particle Image Velocimetry (PIV) and ombroscopy methods. However, these methods are limited to small set-ups and would be impractical on large-scale bioreactors. In particular, they require transparent vessels and fluids which is not applicable to industrial fermenters.

For now, experimental measurements are applied to small-scale systems using model fluids [Laupsien, 2017]. These measures are then used to improve existing hydrodynamic model and to perform large-scale simulation of hydrodynamics, by solving numerically the fluid mechanics equations given in Eqs. (I.4-I.7).

### CFD-RANS approach

The Euler-Euler modelling of gas-liquid systems has been presented previously (see Eqs. (I.4)-(I.9)) and consists in considering that both phases are interpenetrated continuous phases each characterised by its own velocity $\vec{u}_k$ and local phase fraction $\alpha_k$. To solve these variables over time and space, CFD software apply some numerical methods, such as the finite-volume or finite-element methods, to degenerate Navier-Stokes equations into a set of discrete integrable equations. However, due to their strong non-linearities, these equations can not be directly resolved as soon as large turbulent systems are considered. Bioreactors will then be modelled using the Reynolds-Averaged Navier-Stokes (RANS) approach in which velocity and local fractions are split into a mean and a fluctuating component:

$$\vec{u}_k(\vec{x},t) = \vec{U}_k(\vec{x}) + \vec{u}'_k(\vec{x},t) \tag{I.22}$$

$$\alpha_k(\vec{x},t) = \bar{\alpha}_k(\vec{x},t) + \alpha'_k(\vec{x},t) \tag{I.23}$$

Injecting these decompositions into momentum conservation equations (Eqs. (I.6)-(I.7)), leads to following equations for conservation of the mean local momentum:

$$\frac{\partial \bar{\alpha}_G \rho_G \vec{u}_G}{\partial t} + \vec{\nabla} \cdot (\bar{\alpha}_G \rho_G \vec{u}_G \otimes \vec{u}_G) = \bar{\alpha}_G \left(\rho_G \vec{g} - \vec{\nabla} P\right) + \vec{\nabla} \cdot (\boldsymbol{\tau}_G + \boldsymbol{\tau}'_G) + \vec{M}_G \tag{I.24}$$

$$\frac{\partial \bar{\alpha}_L \rho_L \vec{u}_L}{\partial t} + \vec{\nabla} \cdot (\bar{\alpha}_L \rho_L \vec{u}_L \otimes \vec{u}_L) = \bar{\alpha}_L \left(\rho_L \vec{g} - \vec{\nabla} P\right) + \vec{\nabla} \cdot (\boldsymbol{\tau}_L + \boldsymbol{\tau}'_L) + \vec{M}_L \tag{I.25}$$

where $\boldsymbol{\tau}'_k$ are Reynolds stress tensors, defined as

$$\tau'_{kij} = \rho_k \overline{u'_{ki} u'_{kj}} \tag{I.26}$$

which represent the mean effect of velocity fluctuations on the dispersion of momentum. Under RANS approaches, these fluctuations are not resolved. Therefore, the Reynolds stress tensors will need to be modelled by introducing conserved turbulence characterising variables. Multiple models have been developed for that purpose [Couderc et al., 2008] and the most commonly applied, though not the most accurate, is the two-equations $k-\epsilon$ model first proposed by Launder and Spalding [1972]. This model resolves the turbulent kinetic energy $k$ (J/kg) and the rate of viscous dissipation of that energy $\epsilon$ (W/kg).

### Closure for momentum transfer between liquid and gas phase

In Eqs. (I.24)-(I.25), terms $\vec{M}_k$ represent the interfacial transfer of momentum between gas and liquid phases. Here, we only consider momentum exchange between a dispersed gas phase (*i.e.* bubbles) and a surrounding liquid. When transferred from phase A, momentum will either be received by phase B, or will be used to extend the interfacial area. When this area is left unchanged, one can model the momentum transfer as

$$\vec{M}_G = -\vec{M}_L = \vec{f}_B + \vec{f}_D + \vec{f}_{VM} + \vec{f}_L \tag{I.27}$$

with

- $\vec{f}_B$: buoyancy force resulting from body and gravitational forces,

  $$\vec{f}_B = (\rho_G - \rho_L)\vec{g} \tag{I.28}$$

- $\vec{f}_D$: drag force resulting from pressure and viscous effects on the gas-liquid interface, usually modelled as

  $$\vec{f}_D = F_D(\vec{u}_L - \vec{u}_G) \quad \text{with } F_D = \frac{18\mu_L}{d_b^2}\frac{C_D Re_b}{24} \tag{I.29}$$

  for spherical bubbles of diameter $d_b$ of Reynolds number $Re = \frac{\rho_L d_b |\vec{u}_G - \vec{u}_L|}{\mu_L}$. $C_D$ is the bubble drag coefficient.

- $\vec{f}_{VM}$: virtual (or added) mass force. This is the force required to accelerate the fluid surrounding a bubble, modelled as

  $$\vec{f}_{VM} = C_{VM}\rho_L\left(\frac{\partial \vec{u}_L}{\partial t} - \frac{\partial \vec{u}_G}{\partial t}\right) \tag{I.30}$$

  with $C_{VM} = 0.5$ the virtual mass coefficient.

- $\vec{f}_L$: lift force due to the unbalanced distribution of pressure and viscous constraints on the gas-liquid interface modelled as:

  $$\vec{f}_L = C_L\rho_L\alpha_G(\vec{u}_L - \vec{u}_G) \times (\vec{\nabla} \times \vec{u}_L) \tag{I.31}$$

  with $C_L$ the lift force coefficient.

See Couderc et al. [2008], Laupsien [2017] and ANSYS Fluent [2015] for further references on these models and for closures for drag, virtual mass and lift force coefficients. These topic are vast and the development of models suiting the properties of culture broth (rheology, gas-liquid interface contamination, ...) are still on-going. However, this does not constitute a core topic of this thesis.

### Reducing the spatial resolution: compartment approaches

While CFD simulations form the current trend for hydrodynamic description of reactors, they have not always been as widely available as today. Chemical engineers then developed other approaches to describe the transport in heterogeneous systems that are still relevant nowadays including mainly the Compartment Model Approach (CMA).

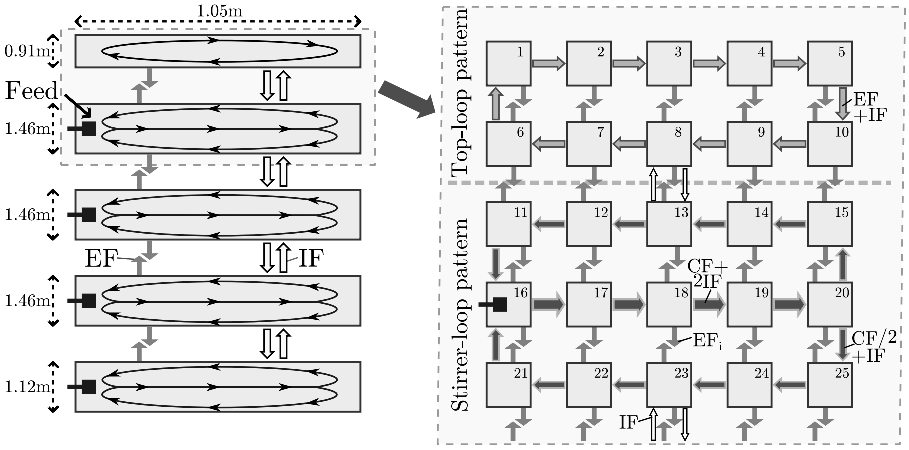


Figure I.1: Compartmental representation of macroscopic flow patterns in a 22 m$^3$ fed-batch bioreactor [Pigou and Morchain, 2015]. Model from Vrábel et al. [1999].

In these models, the reactor volume is split into a low number of sub-volumes (*i.e.* the compartments, or zones) assumed to be homogeneous. The transport is then described by defining the volume flow-rates of each phase between adjacent compartments. These models can be deduced from experimental observations [Reuss, 1991; Mayr et al., 1993; Vrábel et al., 1999; Zahradník et al., 2001] or from CFD simulations [Bezzo et al., 2003; Le Moullec et al., 2010; Delafosse et al., 2014; Zhao et al., 2016]. An example of such a compartment-based model is illustrated in Fig. I.1.

### I.6.3 Biological behaviour

Once biological heterogeneity is tackled through Lagrangian or PBE approach, one still needs to define how a cell, in the state $\widehat{p}$ and in an environment characterized by concentrations $\boldsymbol{C}_L$, will behave. This question encompasses two aspects:

- What will be the rates of consumption of substrates and production of (by-)products? *i.e.* How to compute $\boldsymbol{\varphi}(\widehat{p}, \boldsymbol{C}_L)$ in Eqs. (I.12) and (I.15)?
- How will evolve the cell's properties? *i.e.* How to compute $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L)$ in Eqs. (I.11) and (I.14)?

Both aspects must be considered together, and answers will strongly depend on which cell properties are tracked in $\widehat{p}$. Thereafter are listed some examples of approaches to model this overall biological behaviour. In terms of metabolism, two distinctions are possible:

- kinetic/metabolic model
- structured/unstructured model

Kinetic models will be constant conversion yields model and can usually be summed up to a simple pseudo-reaction, for instance:

$$Y_{SX} S + Y_{OX} O_2 \xrightarrow{X} X + Y_{PX} P \tag{I.32}$$

with $S$: substrate, $O_2$: oxygen, $X$: biomass and $P$: product. $Y_{AB}$ is the conversion yield of $A$ into $B$. These simple models do not account for the actual complexity of metabolic networks. They rely on the hypothesis of balanced growth with constant biomass composition over time and act as black box models of the internal cell functioning.

On the opposite side, metabolic models rely on a description of the metabolic network and let substrates flow through different pathways depending on current internal and external conditions. Therefore, depending on which pathways are activated and on their respective efficiency for substrate conversion, a metabolic model will exhibit varying global conversion yields.

The structured aspect of metabolism modelling relates to whether the model tracks cell attached quantities. These can be composition/internal concentrations, physiology parameters (mass, size, age, ...), or variable without physically defined definition (cybernetic variables, Tartakovsky et al. [1997]). These are the quantities that can be tracked through Population Balance approaches detailed previously.

If a model is structured and tracks some cell-attached properties, these information can be used to determine the rates of reactions described through kinetic or metabolic approaches. If the model is unstructured, only external information (*i.e.* environmental concentrations) are used to model these rates. Hereafter are four examples illustrating each association of the kinetic/metabolic and structured/unstructured modelling.

UNSTRUCTURED KINETIC MODEL A simple unstructured kinetic model can be based on the previous pseudo-reaction (Eq. (I.32)). The growth rate is often defined using Monod-type law [Monod, 1952]:

$$\mu = \mu_{\max} \frac{S}{S + K_S} \frac{O}{O + K_O} \tag{I.33}$$

with $S$ and $O$ the liquid phase concentrations of substrate and oxygen, $K_S$ and $K_O$ the respective affinities of the cell toward these compounds, and $\mu_{\max}$ the maximum growth rate. Chapter II discusses the applicability of this law.

Once the growth-rate $\mu$ ($\mathrm{g_X/g_X \cdot h}$) is determined from environmental conditions (hence the unstructured nature of the model), the volumetric consumption or production rates of $S$, $O_2$, $X$ and $P$ ($R_i$ expressed in $g_i/L.h$) can be computed by considering stoichiometry from Eq. (I.32):

$$R_S = -Y_{SX}\mu X \tag{I.34}$$

$$R_{O_2} = -Y_{OX}\mu X \tag{I.35}$$

$$R_X = \mu X \tag{I.36}$$

$$R_P = Y_{PX}\mu X \tag{I.37}$$

UNSTRUCTURED METABOLIC MODEL Xu et al. [1999] propose a metabolic model describing the behaviour of *Escherichia coli* using either glucose or acetate as a carbon source, and able to switch between oxidative and fermentative energy production modes. This model will be further detailed in Chapter II and is illustrated in Fig. I.2.

In an environment characterised by moderate glucose and oxygen availability, this model will predict a behaviour summed-up by the pseudo-reaction:

$$Y_{SX}S + Y_{OX}O \rightarrow X \tag{I.38}$$

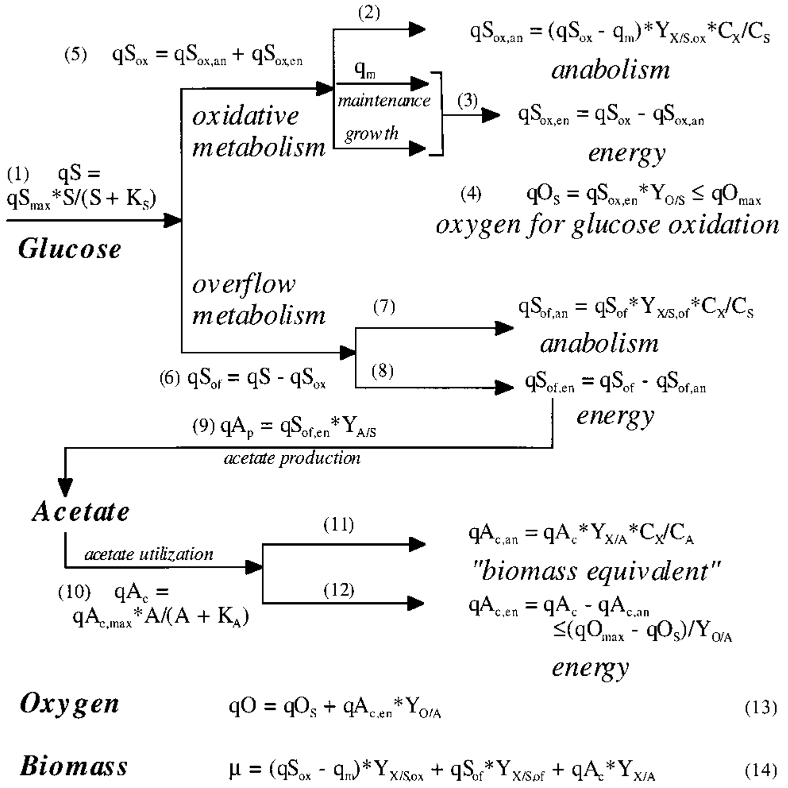


Figure I.2: Structure of model for *E. coli* from Xu et al. [1999]. Reprinted by permission from John Wiley and Sons.

In the absence of oxygen, the observed pseudo-reaction is:

$$Y_{SX}S \rightarrow X + Y_{AX}A \tag{I.39}$$

with $A$ the acetate, here produced as a fermentation by-product. In the absence of glucose and presence of both acetate and oxygen, acetate remplaces glucose as the main carbon source:

$$Y_{AX}A + Y_{OX}O \rightarrow X \tag{I.40}$$

Depending on environmental conditions, the model will actually yield any combination of this three behaviours.

The switch between these modes is only defined from biological constants and environmental conditions, hence the unstructured aspect of this model. The model capability to switch between multiple behaviour and global conversion yields is typical from metabolic models.

STRUCTURED KINETIC MODEL Structured kinetic models are unusual. Their idea would be to track internal cell properties but not to use them to predict how matter flows in the metabolic network. The use of these properties would then be limited to predicting the kinetics of bioreactions. Note however that kinetic models have a narrow range of application (stationnary or pseudo-stationnary cultures) in which laws such a Monod laws exist to predict these reaction rates.

To formulate a structured kinetic model, one could simply track an internal cell glucose concentration $S_{in}$ whose evolution could be dictated by

$$\frac{\partial S_{in}}{\partial t} = \frac{1}{\tau_S}(S - S_{in}) \tag{I.41}$$

with $\tau_S$ the characteristic time of substrate assimilation, and then base the Monod law on this internal concentration

$$\mu = \mu_{\max} \frac{S_{in}}{S_{in} + K_S} \frac{O}{O + K_O} \tag{I.42}$$

Doing so would induce a delay in the answer of a cell in terms of growth rates following a sudden change in its environment glucose concentration.

STRUCTURED METABOLIC MODEL Structured metabolic models form the vast majority of current cell modelling. They describe internal properties of cells and use that information to predict which metabolic pathways are activated and at which rates. They can account for internal metabolic regulations, for lack or excess of metabolites, and for cell-cycle effects.

Quantities that can be tracked are

- Internal concentrations [Reuss, 1991; Rizzi et al., 1997; Lei et al., 2001; Lapin et al., 2004; Matsuoka and Shimizu, 2013]
- Physiology quantities such as age, mass, size, ... [Hatzis et al., 1995; Mantzaris et al., 1999; Robert et al., 2014]
- Cybernetic variables [Sweere et al., 1988; Abulesz and Lyberatos, 1989; Morchain et al., 2013]
- Sub-population fractions [Le Moullec et al., 2010]

In particular, internal concentration based metabolic models can benefit from recent advances in metabolic flux analysis [Nöh et al., 2007; Kajihata et al., 2015].

Bioreaction rates in structured metabolic models can be computed by different ways, the most common being:

- Use of kinetic expressions based on internal concentrations.
- Optimization algorithm aiming a maximizing one rate or yield, usually the growth rate [Varma and Palsson, 1994; Orth et al., 2011]
- Decision-tree based closure [Sweere et al., 1988]

The evolution of biological properties directly depends on their definitions. Age increases naturally over-time and is reset to 0 at cell-division. Cell mass or internal concentrations evolution can be modelled through mass balances. The evolution of cybernetic variables is usually more tricky as their definition and physical meaning is not direct. Their evolution will be modelled on a per-variable manner.

## I.7 Handling the PBE in a Eulerian transport scheme

Whereas many authors already attempted using Population Balance Modelling for the simulation of bioreactors, few works actually made use of PBE in conjunction with a transport scheme for fermenters. Example of such attempts are Heins et al. [2015] who transported a PBM developed by Lencastre Fernandes et al. [2012] in a two-compartments set-up. An other example actually preceded the current PhD thesis and was presented in Pigou and Morchain [2015]. In this work, a PBM was applied to the simulation of a 22m$^3$ fermenter using a metabolic model for *E. coli* derived from the work of Xu et al. [1999].

One of the main reason this coupling is hardly ever considered is because of the need for numerical methods to make the PBE (Eq. (I.14)) compatible with the generic local transport equation (Eq. (I.9)). Such methods already exist and were applied by numerous scientific communities for whom the coupling of PBEs with a transport scheme is especially required.

Overall, all methods will consist in forming a vector $\boldsymbol{P}(\vec{x},t)$ from which one can deduce an approximation $\tilde{n}(\widehat{p},\vec{x},t)$ of $n(\widehat{p},\vec{x},t)$. $\boldsymbol{P}(\vec{x},t)$ will be referred to as a vector of population-scale biological properties, opposed to $\widehat{p}$ which is the vector of cell-scale biological properties. By construction, $\boldsymbol{P}(\vec{x},t)$ will be transportable in the Euler-Euler framework of gas-liquid description:

$$\frac{\partial \alpha_L \boldsymbol{P}(\vec{x},t)}{\partial t} + \vec{\nabla} \cdot \left(\alpha_L \vec{u}_L \boldsymbol{P} - \mathcal{D}_B \vec{\nabla}(\alpha_L \boldsymbol{P})\right) = \boldsymbol{S}(\boldsymbol{P}, \boldsymbol{C}_L) \tag{I.43}$$

where $\boldsymbol{S}(\boldsymbol{P}, \boldsymbol{C}_L)$ is the overall source or sink term which accounts for all phenomenon that modify the distribution in the space of internal properties: transport in the internal space, birth and death of cells in a specific state $\widehat{p}$.

Two main approaches will be detailed hereafter, and illustrated with the example of a one dimensional PBE whose cell-scale biological property is the cell age, $\widehat{p} = a$:

$$\frac{\partial \alpha_L n(a)}{\partial t} + \vec{\nabla} \cdot \left(\alpha_L \vec{u}_L n(a) - \mathcal{D}_B \vec{\nabla}(\alpha_L n(a))\right) + \alpha_L \frac{\partial(\zeta_a(a, \boldsymbol{C}_L) n(a))}{\partial a} = B(a) - D(a) \tag{I.44}$$

where

- $\zeta_a$ is the rate of age increase over time, therefore $\zeta_a = 1$ s/s;
- $B(a)$ is the rate of formation of new cells whose age is $a$. All cells are born with a age $a = 0$ so that $B(a) = \Gamma_b \delta(a)$ with $\Gamma_b$ the rate of formation of new cells ($\mathrm{kg_{cell}/m_L^3 \cdot s}$) and $\delta$ the Dirac delta function;
- $D(a)$ is the rate of death of cells of age $a$ with $D(a) = \Gamma_d(a) n(a)$ with $\Gamma_d(a)$ the specific rate of death ($\mathrm{kg_{cell}/kg_{cell} \cdot s}$).

### I.7.1 Sectional methods

This first category of approach to solve the PBE in a Eulerian transport scheme has multiple designation: sectional methods [Nguyen et al., 2016], pivot methods [Kumar and Ramkrishna, 1996a,b] or class methods [Alopaeus et al., 2006, 2007, 2008; Morchain et al., 2013].

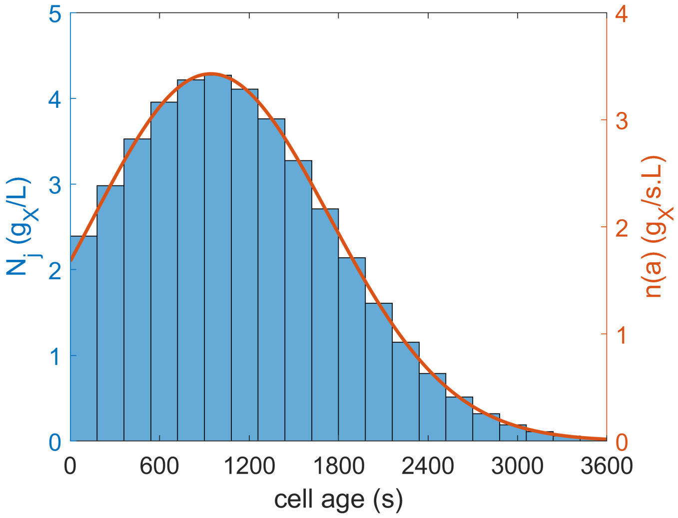


Figure I.3: Illustration of a continuous Number Density Function of age distribution and associated 20-classes discretization.

This approach consists in discretizing the space of biological properties into numerous “sections” or “classes”. Under this formalism, $\boldsymbol{P}$ is a vector whose values are the concentrations of cells which belong to each section. For each section, one attaches a “pivot” $\widehat{p}_i$ and all cells within that section are assumed to be actually located at that pivot in the space of internal properties. For the cell-age distribution example, Fig. I.3 illustrates the application of this discretization to an arbitrary distribution.

Let $\boldsymbol{P} = [N_1, N_2, \ldots, N_{20}]^T$. In this 1D example, one must perform integrations of the PBE on the support of each class to obtain the law of evolution of $\boldsymbol{P}$. For instance, the evolution of $N_{10}$ is given by integrating each term of the PBE from Eq. (I.44) over $a \in \left[a_{10-1/2}; a_{10+1/2}\right]$:

$$\frac{\partial \alpha_L N_{10}}{\partial t} + \vec{\nabla}\cdot\left(\alpha_L \vec{u}_L N_{10} - \mathcal{D}_B \vec{\nabla}(\alpha_L N_{10})\right) = \int_{a_{10-1/2}}^{a_{10+1/2}} (B(a) - D(a))\mathrm{d}a - \left[\zeta_a(a, \boldsymbol{C}_L) n(a)\right]_{a_{10-1/2}}^{a_{10+1/2}} \tag{I.45}$$

The third term of the LHS is the flux of cells between the $10-th$ class and its surrounding classes.

$n(a)$ can be approximated from $\boldsymbol{P}$ as a discrete distribution where all cells are considered to be located at the pivot of their class, or as a uniform by part distribution. By increasing the number of classes, the error between $n(a)$ and its approximation decreases.

Variants of this method aims at higher accuracy by solving multiple values in each section such as the Two-Size Moment (TSM) sectional method [Nguyen et al., 2016] and High-order Moment conserving Method of Classes (HMMC) [Alopaeus et al., 2006, 2007, 2008].

### I.7.2 Moment methods

Moments are integral properties of a NDF. The $k-th$ order integer moment of a mono-variable NDF such as $n(a)$ is defined as

$$m_k = \int_{\Omega_a} a^k n(a) \mathrm{d}a \tag{I.46}$$

where

- $m_0$ is the total amount of individual; here the total concentration of cells whatever their age;
- $m_1$ is related to the population mean age $\tilde{a}$ by $m_1 = \tilde{a} m_0$;
- $m_2$ holds information about the standard deviation $\sigma_a$ of the age distribution: $\sigma_a = \sqrt{\frac{m_2 m_0 - 2m_1^2}{m_0}}$

Higher order moments are related to age distribution skewness, kurtosis, flatness, and other descriptors of the distribution shape.

Methods of moments, applied to a PBE, will track moments of the NDF whose evolution is described by that PBE. Let $\boldsymbol{P} = [m_0, m_1, \ldots, m_{N-1}]^T$ be the vector of the first $N$ integer moments of $n(a)$. The moment equations, which describes the evolution of each element of $\boldsymbol{P}$, is obtained by multiplying all terms of the PBE (here Eq. (I.44)) by $a^k$ and integrating over the space of definition of $a$, $\Omega_a$:

$$\begin{aligned} \frac{\partial \alpha_L m_k}{\partial t} + \vec{\nabla} \cdot \left( \alpha_L \vec{u}_L m_k - \mathcal{D}_B \vec{\nabla} \left( \alpha_L m_k \right) \right) + \left[ a^k \zeta_a n(a) \right]_{\partial \Omega_a} \\ - \int_{\Omega_a} k a^{k-1} \zeta_a n(a) \mathrm{d}a = \int_{\Omega_a} a^k (B(a) - D(a)) \mathrm{d}a \end{aligned} \tag{I.47}$$

where the third term on the LHS represent the flux of cells, weighted by $a^k$, at the limits of the space of age definition. As the ageing process will not allow cells to go out out the space of age, this term is necessarily null.

After a reorganization of terms, and an expansion of $B(a)$ and $D(a)$ definitions, one obtains the following moment equations:

$$\frac{\partial \alpha_L m_0}{\partial t} + \vec{\nabla} \cdot \left( \alpha_L \vec{u}_L m_0 - \mathcal{D}_B \vec{\nabla} \left( \alpha_L m_0 \right) \right) = \Gamma_b - \int_{\Omega_a} \Gamma_d(a) n(a) \mathrm{d}a \tag{I.48}$$

$$\begin{aligned} \frac{\partial \alpha_L m_k}{\partial t} + \vec{\nabla} \cdot \left( \alpha_L \vec{u}_L m_k - \mathcal{D}_B \vec{\nabla} \left( \alpha_L m_k \right) \right) = k \zeta_a m_{k-1} - \int_{\Omega_a} a^k \Gamma_d(a) n(a) \mathrm{d}a, \\ k \in \{1, \ldots, N-1\} \end{aligned} \tag{I.49}$$

If $\Gamma_d(a)$ is not defined as a polynomial of $a$, the integral term on the RHS of Eqs. (I.48) and (I.49) will not be closed in terms of moments of $n(a)$. Moreover, even if $\Gamma_d(a)$ is indeed a polynomial but is not constant, the volution of $m_k$ will depend on higher order moments such as $m_{k+1}$ which leads to an unclosed set of equations. For these reasons, numerical methods to estimate unknown integral properties of $n(a)$ from its resolved moments are usually required. Examples of such methods are the Quadrature Method of Moments (QMOM) [Marchisio et al., 2003a], the Extended Quadrature Method of Moments (EQMOM) [Yuan et al., 2012] and the Maximum Entropy method [Mead and

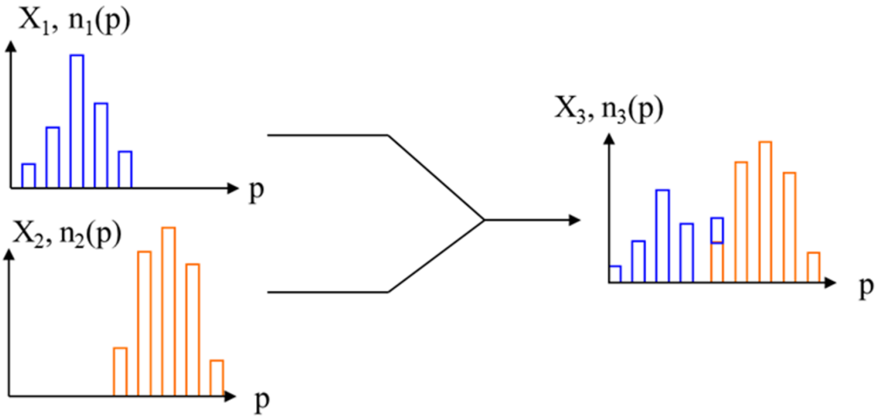


Figure I.4: Segregated description of the mixing of two different populations. Reproduced with permission from Elsevier [Morchain, 2017].

Papanicolaou, 1984] which are fully discussed in chapter III. New numerical methods to close moment equations are still developed, and multiple recent works –not in the field of bioprocesses– couple them along with CFD simulations [Marchisio et al., 2003b; Passalacqua et al., 2017; Askari et al., 2018].

Multi-variate population balance models can also be tackled through methods of moments through numerical methods that are not explored in this manuscript such as the Conditional Quadrature Method of Moments (CQMOM, Yuan and Fox [2011]) and the Bivariate EQMOM [Pollack et al., 2016].

### I.7.3 Unsegregated models

Previously described approaches are said to be segregated. They consider the biological phase as discrete and attach sets of properties at the individual scale. Therefore, they allow to correctly describe the mixing of two different populations as illustrated in Figure I.4. Properties remain attached to cells and therefore do not mix, only their distributions are mixed. For instance, a mix of 10 cells of $1\mu$m and 10 cells of $2\mu$m does not yield a population of 20 cells of 1.5 $\mu$m, but yields a population made of two sub-populations of each cell size.

Unsegregated approaches do not track distributions during mixing, they consider that all individuals in the same location are identical. For the previous example, the unsegregated approach would predict that indeed, the population mix yields a population of cells that share the same size 1.5 $\mu$m. Though the unsegregated approach are obviously inaccurate to describes the discrete aspect of the biological phase, it has the benefit of a significantly low numerical cost and is therefore often used in place of segregated approaches.

## I.8 Overview of current state-of-the-art

Section I.5 developed a generic set of equations to describe the three phases constituting a bioreactor, and section I.6 lists multiple closures or approaches of different accuracies for these equations. These foundations for bioreactors modelling being laid down, it will be interesting to see how multiple works about bioreactor simulation integrate under this generic modelling framework.

Table I.1 is a summary of which approaches were used by multiple authors in terms of

- Approach for hydrodynamic description
  - CFD: Numerical resolution of fluid mechanics equations.
  - CMA: Formulation of a compartment model for transport.
  - 0D-model: Use of spatially homogeneous equations.
- Bioreactions
  - Kinetic: Use of kinetic (constant yield) model.
  - Metabolic: Use of varying yield model.
  - Structured: Use of at least one internal dynamic property.
- Biological heterogeneity (applicable only to structured models)
  - Segregated: multiple biological states are considered in the same spatial location $\vec{x}$
    - Lagrange: Lagrangian treatment of the biological phase.
    - Section: Sectional treatment of a PBE (pivot/class methods).
    - Moment: Moment treatment of a PBE.
  - Unsegregated: All cells are locally assumed to be identical even though their properties may evolve in time and space.

This list is not exhaustive, and focuses only on works in which biological cultures are simulated by coupling an hydrodynamic model and a reactive biological phase. Therefore, works focusing solely on bioreactor hydrodynamics or on the pure analysis of experimental results were discarded.

Most of these articles were classified within four categories, depending on their approach and goals:

I Simulations based on 0D hydrodynamics using structured unsegregated metabolic models.

II Simulations based on CMA coupled with structured unsegregated metabolic models.

III Simulations based on CFD coupled with Lagrangian modelling of the biological phase.

IV Simulations based on CMA coupled with structured metabolic model and population balance approaches.

The category I makes the vast majority of listed works, and mainly corresponds to the development of metabolic models for the purpose of fitting a few sets of experimental data obtained in lab-scale –homogeneous– cultures. These models make use of structured approach and therefore could require taking into account the segregated aspect, however, neither works from categories I and II do implement any approach to account for biological heterogeneity. The fact is that, even if these models are supposedly compatible with

Table I.1: Overview of existing work on bioreactor simulation.

| Reference | Hydrodynamics | | | Bioreactions | | | Heterogeneity | | | | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0D | CMA | CFD | Kinetic | Metabolic | Structured | Segregated | | | Un-segregated | |
| | | | | | | | Lagrange | Section | Moment | | |
| Young et al. [1970] | X | - | - | X | - | - | - | - | - | - | |
| Sweere et al. [1988] | X | - | - | X | - | - | - | - | - | - | |
| Reuss [1991] | X | - | - | - | X | X | - | - | - | X | I |
| Nielsen and Villadsen [1992] | X | - | - | - | X | X | - | - | - | X | I |
| Bellgardt [1994] | X | - | - | X | - | X | - | - | - | - | |
| Varma and Palsson [1994] | X | - | - | - | X | X | - | - | - | X | I |
| Hatzis et al. [1995] | X | - | - | - | - | X | X | - | - | - | |
| Rizzi et al. [1997] | X | - | - | - | X | X | - | - | - | X | |
| Xu et al. [1999] | X | - | - | - | X | - | - | - | - | - | |
| Vrábel et al. [1999] | X | X | - | - | X | - | - | - | - | - | II |
| Lei et al. [2001] | X | - | - | - | X | X | - | - | - | X | I |
| Mantzaris et al. [2001] | X | - | - | X | - | - | - | X | - | - | |
| Vrábel et al. [2001] | X | X | - | - | X | - | - | - | - | X | II |
| Chassagnole et al. [2002] | X | - | - | - | X | - | - | - | - | X | |
| Bezzo et al. [2003] | - | X | X | - | X | - | - | - | - | - | II |
| Gerdtzen et al. [2004] | X | - | - | - | X | X | - | - | - | X | I |
| Rodríguez et al. [2004] | X | - | - | - | X | X | - | - | - | X | I |
| Lapin et al. [2004] | - | - | X | X | X | X | X | - | - | - | III |
| Hristov et al. [2004] | - | X | - | X | - | - | - | - | - | - | |
| Hjersted and Henson [2006] | X | - | - | - | X | X | - | - | - | X | I |
| Laakkonen et al. [2006] | - | X | X | X | - | - | - | - | - | - | |
| Lapin et al. [2006] | - | - | X | - | X | X | X | - | - | - | III |
| Morchain and Fonade [2009] | X | - | - | - | X | X | - | - | - | X | I |
| Meadows et al. [2010] | X | - | - | - | X | X | - | - | - | X | I |
| Le Moullec et al. [2010] | - | X | X | - | X | X | - | - | - | X | II |
| Lencastre Fernandes et al. [2012] | X | - | - | X | - | X | - | - | - | - | |
| Elqotbi et al. [2013] | X | - | X | X | - | - | - | - | - | - | |
| Matsuoka and Shimizu [2013] | X | - | - | - | X | X | - | - | - | X | I |
| Morchain et al. [2013] | X | - | - | X | - | X | - | X | - | - | |
| Morchain et al. [2014] | - | - | X | X | - | X | - | X | - | - | |
| Goldrick et al. [2015] | X | - | - | - | X | X | - | - | - | X | I |
| Heins et al. [2015] | - | X | - | X | - | X | - | X | - | - | IV |
| Pigou and Morchain [2015] | X | X | - | - | X | X | - | X | - | - | IV |
| Haringa et al. [2016] | - | - | X | - | - | X | X | - | - | - | III |
| Ordoñez et al. [2016] | X | - | - | - | X | X | - | - | - | X | I |
| Farzan and Lerapetritou [2017] | - | - | X | X | X | X | - | - | - | X | III |
| Kuschel et al. [2017] | - | - | X | - | - | X | X | - | - | - | III |
| Brand et al. [2018] | X | - | - | - | X | X | - | - | - | X | I |

segregation by their structured nature, they are usually not designed with this consideration in mind. They may therefore have a prohibitive complexity through the use of a high number of internal variables, or the parameter identification might have been flawed by assuming that all cells are identical in a lab-scale system while they may actually differentiate through stochastic processes [Delvigne et al., 2014].

First attempts to account for cell's history and exposure to different environmental fluctuations were carried out through the Euler-Lagrange framework using CFD simulations (see category III). Developments related to this approach are still ongoing and for now the Lagrangian method is used to characterise environmental fluctuations along cell's trajectories rather than to predict their effect of biological heterogeneity and reverse coupling on concentration field gradients. For instance, Haringa et al. [2016] tracked the signal of substrate concentration fluctuations in a large-scale fermenter using Lagrangian particles, but for the "prediction" of concentration fields evolutions, they relied on an unstructured kinetic model.

Finally, some attempts are being made to couple the biological heterogeneity with intermediate hydrodynamic resolutions (category IV). On that regard, developments of CMA models from CFD is an interesting topic that has been considered by numerous authors classified in category II (all those for which both CMA and CFD approaches are ticked), and is also investigated in papers not referenced here (*i.e.* Delafosse et al. [2015]).

Bailey [1998]'s classification is at the origin of (un)segregated and (un)structured classifications of biological models. Table I.2, excerpted from Zieringer and Takors [2018] summarizes pros and cons of these categories. From these, it is easily understood why the majority of developments on bioreactor simulation relies on unsegregated descriptions, due to the mathematical complexity on segregated approaches. However, it is also clear that structured segregated models will be required for the simulation of industrial fermenters.

Table I.2: Partial reproduction of the summary of pros and cons of (un)segregated and (un)structured approaches for biological models listed by Zieringer and Takors [2018].

| Model class | Application | Advantages | Disadvantages |
|---|---|---|---|
| Structured | Systems in transient state | Cellular compartmentation | Biological knowledge |
| Unstructured | Steady-state systems | Easy to build | Only penomenological cell description |
| Segregated | Heterogeneous, individual cell systems | More representative and informative | Difficult to handle mathematically |
| Unsegregated | Systems with average cell description | Easy to build for a large number of cells | Average cell description |

## I.9 Proposed approach, thesis goals and outline

The main goal of this PhD thesis is to develop a model structure that encompasses all aspects of bioreactors modelling while keeping a low numerical cost. This model structure is applied to the simulation of an industrial fermenter in which the baker yeast, *Saccharomyces cerevisiae*, is cultivated.

In order to reach all-coupling simulations of bioreactors, it is necessary to develop a modelling strategy and to focus on each aspect of the overall model to ensure that lowest possible numerical costs are achieved while not sacrificing accuracy. Indeed, aiming at low numerical cost will be useless if each part of the model is not accurate enough to obtain an overall predictive model. Therefore, along this thesis, discussions about the trade-off between numerical cost and accuracy will be a common thread.

Considering all previously detailed aspects, the proposed modelling strategy is as follow:

1. First, define a structured metabolic for the considered micro-organism. This definition should answer the following questions:
    - What are chemical species that are used as source of matter/energy, and/or are produced by the considered strain? This step consists in defining $\boldsymbol{C_L}$.
    - What is the minimal list of biological properties that should be tracked for each cell to determine its behaviour? Examples would be age, size, internal concentrations, storage level, or others... This step consists in defining $\widehat{p}$
    - What mathematical model allows predicting the rates of consumption or production of chemical species for a cell defined by $\widehat{p}$ in an environment characterized by concentrations $\boldsymbol{C_L}$? This steps consists in defining closure for $\boldsymbol{\varphi}(\widehat{p}, \boldsymbol{C_L})$.
    - Along with the definition of $\boldsymbol{\varphi}(\widehat{p}, \boldsymbol{C_L})$, what mathematical model gives the evolution of cell-attached properties for a cell defined by $\widehat{p}$ in an environment

characterized by concentrations $\boldsymbol{C}_L$? This step consists in defining closure for $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L)$.

2. Define how to model biological heterogeneity
   - Based on $\widehat{p}$ definition chose the most suited numerical method by weighting accuracy against numerical cost.
3. Define how to describe large-scale hydrodynamics
   - Arbitrate about the implementation of biological models within CFD or CMA frameworks by considering hydrodynamic and biological time-scales as well as numerical cost.

These steps actually draw the structure of this PhD thesis. Chapter II focuses on the metabolism modelling. In particular, this chapter details a modular and numerically efficient model structure. This structure is first applied to a bacteria, *E. coli*, to illustrate its application to a simple micro-organism. Then, the same structure is applied to *S. cerevisiae* to prepare large-scale simulations of an industrial fermenter cultivating this specie. Both models are compared to, and validated against, experimental data available throughout the literature.

In this work, the PBM approach is considered to be the key element that allows large-scale simulations of bioreactors. Chapter III makes a strong focus on this aspect by comparing both sectional and moment methods to solve 1D population balance models. Studying the core principles of moment methods led to a major improvements of a numerical method used to close equations in monovariate moment approaches. These improvements of the numerical method are described too in this chapter.

In order to illustrate the coupling of metabolic models and population balances with hydrodynamics, Chapter IV focuses on performing numerical simulations of the gas-liquid hydrodynamics of an actual industrial fermenter. To further exploit these CFD simulation results, Chapter IV details how to produce a compartment model from highly resolved CFD results.

Finally, Chapter V couples developments from Chapter II to IV to perform the simulation of industrial biological cultures and show the capabilities and applications of the proposed modelling framework.

Table I.3: Approaches used in following Chapters.

| Reference | Hydrodynamics | | | Bioreactions | | | Heterogeneity | | | | Category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Segregated | | | Un-segregated | |
| | 0D | CMA | CFD | Kinetic | Metabolic | Structured | Lagrange | Section | Moment | | |
| Chapter II | X | - | - | - | X | X | - | - | - | X | I |
| Chapter III | X | X | - | - | X | X | - | X | X | - | IV |
| Chapter IV | - | X | X | - | - | - | - | - | - | - | |
| Chapter V | - | X | - | - | X | X | - | - | X | - | IV |

Table I.3 is based on the same structure than Table I.1, but summarizes which approaches are developed in the current work. First, Chapter II focuses on developing a low-cost metabolic model structure and relies on homogeneous non-segregated simulations. Note that this model structure is designed for upcoming coupling with segregated approaches. These are developed in Chapter III were numerical methods to solve Population Balance Equation are investigated. This chapter implements both homogeneous and compartment model approaches to assess numerical methods in terms of accuracy, stability and performances. In order to not be bound to Compartment Models available from the literature, and to make possible the use of this work for any bioreactor, Chapter IV focuses on the use of CFD simulations to formulate Compartment Models. Finally, Chapter V illustrates the coupling of developments from all three previous chapters to perform the predictive simulation of a large-scale fedbatch culture, and underlines how this tool can be used for engineering purposes.

# II

# MODÉLISATION DU MÉTABOLISME

## Résumé

La principale complexité associée à la modélisation d'un bioréacteur industriel provient de la description de la phase biologique. Les micro-organismes ne sont pas des particules inertes ou catalysant de simples réactions chimiques, mais sont des systèmes vivants et dynamiques dotés par leur évolution Darwinienne de capacité de régulation et d'adaptation aux perturbations dans leur environnement. Ils sont donc capables de sélectionner de manière préférentielle leurs substrats, de réguler leur assimilation en s'affranchissant des équilibres thermodynamiques au prix d'une dépense énergétique, et d'orienter la matière assimilée vers de multiples voies métaboliques par le biais de régulations internes fortement couplées.

Ce premier chapitre de développements se focalise sur la modélisation de ces aspects, dans la continuité de très nombreux travaux sur ce sujet enrichissant la littérature depuis plusieurs décennies. Ces travaux tombent généralement soit dans un excès de simplicité en négligeant de nombreuses dynamiques cellulaires, soit dans un excès de complexité en cherchant à décrire les échelles les plus fines. Il va de soi qu'aucune de ces deux modélisations limites ne peuvent être envisagées pour la simulation prédictive de bioréacteurs industriels. Quelques modèles se situant à un niveau intermédiaire de complexité existent, mais ils n'ont en général pas été pensés ni construits en vue d'un couplage avec le transport de la phase biologique dans un système hétérogène.

Nous chercherons à intégrer les dynamiques biologiques, les régulations des flux et l'orientation de la matière entre différentes voies métaboliques dans un modèle de faible coût numérique. Cela permettra le couplage avec des approches aussi bien Lagrangienne que de Bilan de Population pour prendre en considération la diversité biologique, c'est-à-dire le fait que les micro-organismes peuvent se différencier suite à des perturbations environnementales ou stochastiques. Il sera alors possible d'envisager des simulations de cultures biologiques aux échelles industrielles pour un coût numérique raisonnable.

Ce chapitre propose essentiellement une structure de modèle métabolique. Cette structure, dérivée du travail de Xu et al. [1999] et des approches de réduction de réseaux de réactions métaboliques par la définition de modes métaboliques, permettra effectivement d'intégrer les dynamiques, les régulations et les différentes orientations de la matière, sans faire appel à des procédures coûteuses d'optimisation. La mise en œuvre de cette structure est illustrée d'une part pour la modélisation d'*Escherichia coli*, et d'autre part pour décrire les comportements dynamiques de *Saccharomyces cerevisiae*. En particulier ce dernier modèle, qui sera utilisé par la suite dans le Chapitre V, est testé sur plusieurs jeux de données expérimentales obtenus par différentes équipes de chercheurs, pour plusieurs souches sauvages de la levure, et sur des modes de culture distincts. La

confrontation massive à différents jeux de données expérimentales constitue un test sévère pour un modèle. À notre connaissance, c'est la première fois qu'un modèle métabolique de si faible coût numérique est confronté à autant de données expérimentales.

# II

# METABOLISM MODELLING

## Summary

The main challenge in the modelling of industrial bioreactors comes from the depiction of the biological phase. Micro-organisms are not inert particles or catalysts for simple chemical reactions. These are living and dynamic systems which, through Darwinian evolution, obtained regulation and adaptation capabilities to face environmental perturbations. They are then able to select preferential substrates, regulate their assimilation by over-coming thermodynamic equilibrium at the expense of energy, and to channel assimilated matter toward multiple metabolic pathways through internal and tightly coupled regulations.

This first development chapter focuses on modelling these aspects, in the continuation of numerous works which covered this topic over the last decades. These works tended to fall either in an oversimplification excess, by neglecting most of cell dynamics, or in a complexity excess by attempting to describe smallest scales. It is obvious that both of these limit modelling approaches may not be considered to perform predictive simulations of large-scale fermenters. Some models lie at an intermediary level of complexity, but these were not designed by anticipating their coupling with the transport of the biological phase in an heterogeneous system.

This chapter describe how to integrate biological dynamics, flux regulation and the channelling of matter through multiple metabolic pathways in a low numerical cost model. This low-cost permits the integration of this model under Lagrangian or Population Balance modellings to account for biological diversity, *i.e.* the fact that micro-organisms will differentiate due to environmental or stochastic perturbations. This shall allow the simulation of industrial-scale biological cultures with a reasonable and tractable numerical cost.

This chapter mainly describes a structure proposal for metabolic models. This structure, derived from the work of Xu et al. [1999] and from metabolic-mode approaches for metabolic networks reduction, describes indeed dynamics, regulations and matter channelling with an especially low numerical cost. This structure is applied to the modelling of *Escherichia coli* and to the description of dynamic behaviours of *Saccharomyces cerevisiae*. In particular the model developed for this yeast, which will be used in Chapter V, is challenged against numerous experimental data sets produced by multiple research teams with different wild strains and culture modes. Challenging a model against such different data sets is quite a severe test. To the best of our knowledge, this is the first attempt to produce such a low-cost metabolic model that fits so many experimental data sets.

# II

# METABOLISM MODELLING

## II.1 Objectives and constraints of the biological modelling

In the proposed modelling framework, the metabolic model serves to predict the mass transfer rates between the liquid and the biotic phase. The starting point is the knowledge of the cell state and the concentrations in the surrounding liquid. To develop metabolic models, two questions must be tackled. (i) What limits the reaction rates? (ii) Through which pathways will the matter be transformed? This notion of pathways encompasses the need to define reaction stoichiometry, but also energetic cost of the conversion.

At the most fundamental levels, biological reactions depend on oxidative reactions of carbonaceous reduced matter (sugars, fat). Just like in combustion, these reactions release energy, part of it as heat, but most of this chemical energy will be used by the living system to produce highly complex biomolecules, new cells, to interact with the environment by controlling flux at the membrane level, or even by using this energy to move toward more favourable environments.

In combustion, energy is only released as heat in fast reactions, whose time-scales can be as low as nanoseconds [Caudal et al., 2013]. The need for biological system to control the release of energy implies that they must be able to limit and control the rate of bioreactions. Some of these limitations occur naturally, for instance due to limiting diffusive transport within a cell cytoplasm. Cells also actively regulate their internal reaction rates by decreasing the concentrations of some enzymes, or by limiting fluxes at the membrane level at the expense of energy. For these reasons, even in rich environment, bioreactions will never be as fast as combustion.

If limitations do not come from internal reaction rates, or from limiting cross-membrane transfer, the limitation will be external to the cell. It is indeed straightforward to understand that cells without internal reserves cannot perform any reaction in a depleted environment where no substrate is available.

Both internal and external limitations will have to be considered when designing a biological model for large-scale simulation. Indeed, in such systems, the environment will be rich in substrate near the feeding point but might be depleted in other parts of the reactor. Similarly, oxygen will be highly available near the bubble spargers where gas is still rich, but may be limited elsewhere if the oxygen transfer rate is insufficient compared to the overall biological needs. Consequently, due to this possibly cyclic environmental availabilities, the proposed model will need to handle both sources of limitations.

If the biological reaction could be summed up to

$$\text{Glucose} + \text{Oxygen} \rightarrow \text{Cells}$$

the rate of this reaction would simply be the limiting value between the possible assimilation rates of glucose and oxygen, and the maximum production rate of new cells. Unfortunately, the chemistry of living systems is much more complicated than that.

This chapter will detail proposed models for a bacteria, *Escherichia coli*, and for a yeast, *Saccharomyces cerevisiae*. Both of these micro-organisms are said to be optional aero-anaerobic. This implies that they are both able to grow either with, or without, oxygen. If oxygen is present, cells produce energy from oxidative catabolism, *i.e.* energy producing pathway, whose final electron receiver is oxygen. These pathways are analogous to full combustion of organic matter in presence of oxygen. On the opposite, if no oxygen is available, fermentative catabolism still allows producing energy. These pathways lead to the formation of some by-products (alcohol or acid compounds) just like partial combustion.

The proposed models will then have to be able to switch between these two modes of energy production depending on oxygen availability, but complexity goes on. While glucose is the preferential source of carbon and energy for these micro-organisms, they are also able to grow on other carbonaceous substrate. In particular, by-products from fermentative catabolism are still reduced carbon matter and can be assimilated again in oxygen-rich environments, both for energy and new cell productions.

Moreover, the model will be used in large-scale where the biomass undergoes both (i) short-term perturbations such as sudden expositions to high substrate concentrations when cells pass by the feeding point, as well as (ii) long term modifications of the overall environment, for instance if the oxygen supply becomes limiting due to increasing biomass concentration over time.

Due to the last aspect, the metabolic model must be able to switch between multiple modes and must behave well both in long-term steady situations and in dynamic environments. The model construction will have to account for these aspects, and must be challenged against both steady and dynamically changing experimental conditions.

Finally, one must keep in mind that this model will be used for long-term large-scale simulations. This model will be called in each volume of reference solved by the hydrodynamic part of the overall model, and for each biological state considered in the biological heterogeneity part of the whole model. The numerical cost of the biological model must then be maintained as low as possible and can not rely on expensive computations.

## II.2 Proposed metabolic model structure

### II.2.1 Inspiration from a decision-tree model by Xu et al. [1999]

Foundations of the proposed structure come from Xu et al. [1999] who developed a simple model for *E. coli*. This model describes growth of this bacteria in anaerobic conditions, where glucose could be fermented producing energy as well as acetate as a by-product; and in aerobic conditions where both glucose and acetate could be used as main carbon source along with an oxidative energy production. This model is illustrated on Fig. II.1 and relies on two simple yet powerful ideas:

1. Growth is limited by some maximum fluxes of matter (glucose, acetate, oxygen).
2. Available matter will first be directed toward high-yield and/or high-affinity metabolic modes, only leftovers will be directed toward other modes.

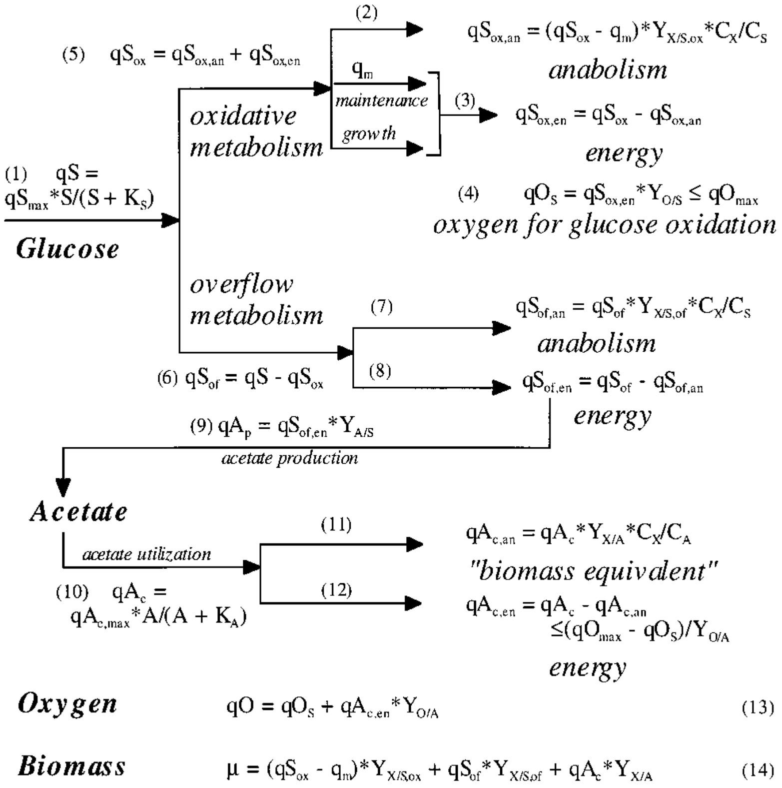


Figure II.1: Structure of model for *E. coli* from Xu et al. [1999]. Reprinted by permission from John Wiley and Sons.

The first idea allows describing the overall metabolism only in terms of flux of matter and energy. As energy is produced from biochemical reactions, all fluxes are coupled through mass and energy balances. For instance, if the flux of glucose and oxygen that is processed by a cell is known, one can deduce the associated flux of energy-carrying metabolites produced through the oxidative pathway.

The second idea comes as a replacement for a commonly used approach when solving metabolic models. Most models based on an extensive description of a metabolic network rely on constraint-based flux optimization to deduce an internal cell composition, and matter orientation between metabolic pathways, from environmental composition. Examples are the YEAST model for *S. cerevisiae* [Heavner et al., 2012, 2013], and the *i*JO1366 model for *E. coli* [Orth et al., 2011]. The flux optimization usually aims at identifying the cell state that maximizes growth.

Instead of performing expensive optimization computations, the *E. coli* model from Xu et al. [1999] enforces that matter will be used in priority within efficient growth modes, sorted in the following order:

1. Growth on glucose with oxidative energy production;
2. Growth on glucose with fermentative energy production;
3. Growth on acetate with oxidative energy production.

Obviously, if no oxygen is available, only the mode 2 will be activated. Similarly, a lack of glucose in presence of acetate will only activate the mode 3. Overall, this approach ensures that the most efficient metabolic mode is used when associated substrates are available. This indirectly allows the model to maximize growth, without actually relying on expensive optimization algorithms. Note that if a strain has been engineered so that it tends to maximize an other process than growth, this can still be accounted for by simply sorting metabolic modes differently. This approach is not limited to strains maximizing growth.

Therefore, the model from Xu et al. [1999] does integrate two of the three constraints that we are interested in. The computation of reaction rates is fast as no specific expensive algorithm is involved. Moreover, by construction, it allows describing multiple cell functioning modes.

This model structure removes the need for solving numerous intracellular concentrations and for computing reaction rates using over-parametrized kinetic expressions based on all these internal concentrations. Such models would be the one from Rizzi et al. [1997], or those applied to metabolic flux analysis [Matsuoka and Shimizu, 2013]. The high number of dynamics tracked by these models, *i.e.* number of internal concentrations, make them hardly compatible with heterogeneous bioreactor modelling. While simplifying the computation of bioreaction rates, it is crucial to observe that this computation depends on the prior knowledge of some key fluxes. This approach then only reduces the number of degrees of liberty of the system to be solved. In the present case of Xu et al. [1999], the glucose uptake rate and the maximum respiration rate have to be known. The first one is modelled as a function of two constant cell properties $qS_{\max}$ and $K_S$ along with the glucose concentration (see Eq. (1) in Fig. II.1) and a similar modelling is proposed by the authors for the oxygen consumption rate. It is reasonable to question these closure laws with at least two ideas in mind: do all cells share the same state and will uptake at the same rates in the same environment? Also, some closure laws have to be defined for some fluxes, but which ones are the most relevant?

## II.2.2 Brief description of the model structure

The model from Xu et al. [1999] lacks the integration of dynamics variables. A first attempt of integrating this aspect was done in work preceding this PhD thesis [Pigou and Morchain, 2015]. The proposal was to associate a growth capability to micro-organisms which would evolve over time. Metabolic modes aiming at growth, *i.e.* new cell production, could be limited by this growth capability even if sufficient nutrients were available. If a strong disequilibrium occurred between glucose consumption, and glucose requirements for growth, the excess glucose would be diverted into the well-known overflow metabolism which induces acetate production when *E. coli* is suddenly exposed to high glucose concentrations [Xu et al., 1999; Matsuoka and Shimizu, 2013; Yamanè and Shimizu, 1984]. This initial model was fully detailed in Pigou and Morchain [2015] and kept a structure similar to Xu et al. [1999]. Here, we further improve upon this previous work by defining a procedure to define metabolic models. Overall, the design of these models will rely on:

1. Identifying self-sufficient metabolic modes that can be observed in experimental cultures.

2. Sorting metabolic modes by order of priority. A general rule would be to prioritize growth-mode with high yield efficiency (substrate affinity, more efficient energetic production, ...).

3. Identifying what can limit the rate of each mode.

The model structure will be more easily explained through application examples to illustrate each of these three basis aspects, and how their combination makes the model structure suitable for dynamic simulations of industrial fermenters. Nevertheless, we start with a short generic explanation of the proposed structure.

### Metabolic modes

The concept of metabolic modes is usually associated to the model reduction of complex metabolic networks through linear analysis by identifying limits of a convex hyper-volume of admissible metabolic flux. This approach is extensively detailed by Trinh et al. whose summary figure is reproduced in Fig. II.2. Readers interested in these model reduction approaches should refer to their article [Trinh et al., 2008].

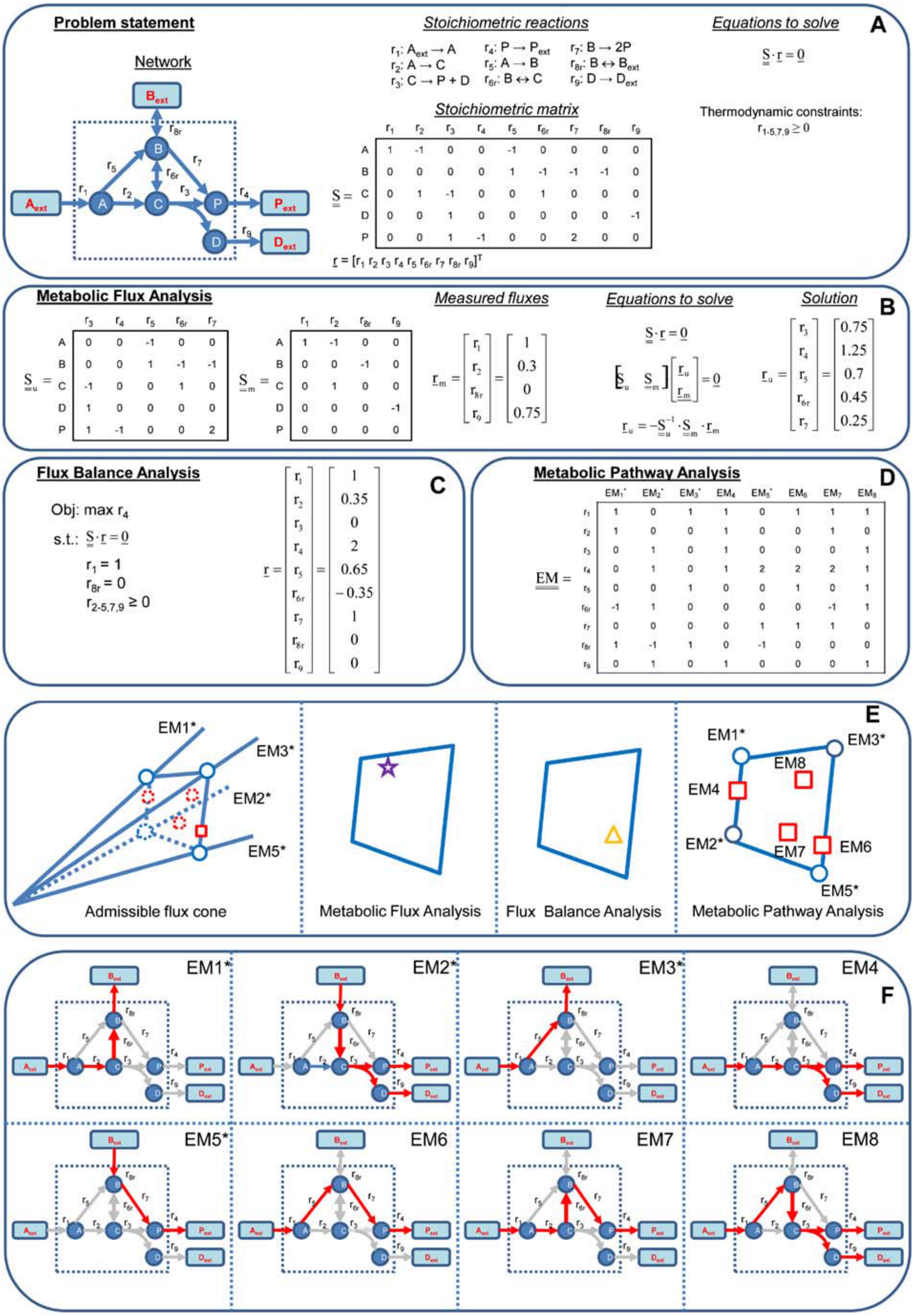
In short, metabolic modes are elementary functionings of a cell and each modes describes the transformation of some input matter (substrates, oxygen, . . . ) into some output matter (products, by-products, new cells, . . . ). One property of these modes is that, taken individually, they appear as constant stoichiometry bioreactions. All possible thermodynamically admissible metabolic behaviours can be expressed as a linear combination of these elementary modes.

The way metabolic modes are combined evolves over time due to (i) the evolution of environmental concentrations and (ii) the evolution of biological capabilities to process matter through each possible pathway. These changing combinations of elementary modes explain how the overall conversion yields of matter by cells may change from one experimental condition to another, even though underlying stoichiometry does not actually change.

The usual approach consists in deducing metabolic elementary modes from a detailed metabolic network, coupled to thermodynamic constraints over this network [Trinh et al., 2008]. Highly detailed metabolic network models exist for *E. coli*[Orth et al., 2011] and *S. cerevisiae*[Heavner et al., 2012, 2013] and model reduction methods could be applied to them [Gerdtzen et al., 2004]. However, we choose to manually design simplified metabolic models instead of reducing complex ones. This was decided in order to ensure a low-cost final model, and to only account for selected metabolic features. Doing so allows for keeping the model complexity low which improves the identifiability of the model parameters.

Therefore, we propose to manually define metabolic modes to list, in an exhaustive manner, all metabolic features of interest of the considered strain. Such features could be:

- capability to grow using glucose and oxygen only, *i.e.* oxidative glucose-based growth;
- capability to grow using only glucose and producing acetate as a by-product, *i.e.* fermentative glucose-based growth;
- . . .



$r_1$: $A_{ext} \rightarrow A$
$r_2$: $A \rightarrow C$
$r_3$: $C \rightarrow P + D$
$r_4$: $P \rightarrow P_{ext}$
$r_5$: $A \rightarrow B$
$r_{6r}$: $B \leftrightarrow C$
$r_7$: $B \rightarrow 2P$
$r_{8r}$: $B \leftrightarrow B_{ext}$
$r_9$: $D \rightarrow D_{ext}$

$$\underline{\underline{S}} = \begin{array}{c|ccccccccc|} & r_1 & r_2 & r_3 & r_4 & r_5 & r_{6r} & r_7 & r_{8r} & r_9 \\ \hline A & 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ B & 0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & 0 \\ C & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ D & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ P & 0 & 0 & 1 & -1 & 0 & 0 & 2 & 0 & 0 \\ \hline \end{array}$$

$\underline{r} = [r_1\ r_2\ r_3\ r_4\ r_5\ r_{6r}\ r_7\ r_{8r}\ r_9]^T$

$$\underline{\underline{S}} \cdot \underline{r} = \underline{0}$$

Thermodynamic constraints:
$r_{1\text{-}5,7,9} \geq 0$

$$\underline{\underline{S}}_u = \begin{array}{c|ccccc|} & r_3 & r_4 & r_5 & r_{6r} & r_7 \\ \hline A & 0 & 0 & -1 & 0 & 0 \\ B & 0 & 0 & 1 & -1 & -1 \\ C & -1 & 0 & 0 & 1 & 0 \\ D & 1 & 0 & 0 & 0 & 0 \\ P & 1 & -1 & 0 & 0 & 2 \\ \hline \end{array} \quad \underline{\underline{S}}_m = \begin{array}{c|cccc|} & r_1 & r_2 & r_{8r} & r_9 \\ \hline A & 1 & -1 & 0 & 0 \\ B & 0 & 0 & -1 & 0 \\ C & 0 & 1 & 0 & 0 \\ D & 0 & 0 & 0 & -1 \\ P & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

$$\underline{r}_m = \begin{bmatrix} r_1 \\ r_2 \\ r_{8r} \\ r_9 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.3 \\ 0 \\ 0.75 \end{bmatrix}$$

$$\underline{\underline{S}} \cdot \underline{r} = \underline{0}$$

$$\begin{bmatrix} \underline{\underline{S}}_u & \underline{\underline{S}}_m \end{bmatrix} \begin{bmatrix} \underline{r}_u \\ \underline{r}_m \end{bmatrix} = \underline{0}$$

$$\underline{r}_u = -\underline{\underline{S}}_u^{-1} \cdot \underline{\underline{S}}_m \cdot \underline{r}_m$$

$$\underline{r}_u = \begin{bmatrix} r_3 \\ r_4 \\ r_5 \\ r_{6r} \\ r_7 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 1.25 \\ 0.7 \\ 0.45 \\ 0.25 \end{bmatrix}$$

Obj: max $r_4$
s.t.: $\underline{\underline{S}} \cdot \underline{r} = \underline{0}$
$r_1 = 1$
$r_{8r} = 0$
$r_{2\text{-}5,7,9} \geq 0$

$$\underline{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_{6r} \\ r_7 \\ r_{8r} \\ r_9 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.35 \\ 0 \\ 2 \\ 0.65 \\ -0.35 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{\underline{EM}} = \begin{array}{c|cccccccc|} & EM_1^* & EM_2^* & EM_3^* & EM_4 & EM_5^* & EM_6 & EM_7 & EM_8 \\ \hline r_1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ r_2 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ r_3 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ r_4 & 0 & 1 & 0 & 1 & 2 & 2 & 2 & 1 \\ r_5 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ r_{6r} & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 1 \\ r_7 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ r_{8r} & 1 & -1 & 1 & 0 & -1 & 0 & 0 & 0 \\ r_9 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hline \end{array}$$

Figure II.2: Illustration of elementary mode analysis of metabolic networks. Reprinted by permission from Springer Customer Service Centre GmbH: Springer Nature, Applied Microbiology and Biotechnology, *Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism*, Trinh, Wlaschin and Srienc, COPYRIGHT (2008). License number 4355340054177.

Selected metabolic modes must be non-colinear, meaning that no mode should be a linear combination of others.

### Modes prioritization

Once metabolic modes are defined, the challenge is to determine how much each elementary mode is activated, based on environmental substrate availabilities, and internal biological limits. On that regard, we keep the approach from Xu et al. [1999] by giving a higher priority to metabolic modes that make use of high-affinity substrates and high-efficiency energetic pathways.

Each mode makes use of resources, and the reaction rate associated to a mode will be limited by resources availability. These may be substrate resources, reactive resources, and/or transport resources. For instance, an elementary mode that describes growth using glucose and oxidative catabolism may be limited by:

- External transport limitation due to turbulence, mixing, external concentrations, ...
    - Maximum glucose uptake rate allowed by the environment $\varphi_G^{(e)}$ $(\mathrm{kg_G/kg_X \cdot h})$.
    - Maximum oxygen uptake rate allowed by the environment $\varphi_O^{(e)}$ $(\mathrm{kg_O/kg_X \cdot h})$.
- Internal bioreaction rate limitations due to enzyme content, induction or repression of pathways, ...
    - Limit on glucose utilization (maximum rate of glycolysis for instance).
    - Maximum rate of respiratory chain for oxidative energy production.
    - Maximum rate of any sub-reaction required to produce new cell building blocks.
    - ...
- Mass transfer limitations due to insufficient membrane transporters, repression of these transporters, ...
    - Maximum transport rate of glucose through the biological membrane
    - Possible transports between cytoplasm and mitochondria in eukaryotic cells.
    - ...

These resource limits are shared by all metabolic modes. Hence, the resulting linear combination of metabolic modes cannot use more resources than those available. Therefore, in order to respect this constraint, and to account for the mode prioritization, the resolution of metabolic model will consist in

- Computing available resources
    - Maximum mass transfer/transport rates of $S$: $\varphi_S^{(\max)}$ $(\mathrm{kg_S/kg_X \cdot h})$
    - Maximum rate of $i$-th reaction in the metabolic network: $q_i^{(\max)}$ $(\mathrm{mol_i/kg_X \cdot h})$
- Maximizing the rate of each metabolic mode, one by one, in their priority order, while always respecting global constraints
    1. Identify limiting resource for mode A;

2. Set rate of mode A such that its limiting resource is depleted;
3. Compute remaining resources for mode B;
4. Identify limiting resource for mode B;
5. Set rate of mode B such that its limiting resource is depleted;
6. Compute remaining resources for mode C;
7. ...

The main benefit of this structure is its low numerical cost. The dynamic aspect of this model comes from how one computes available resources. If these resources are only defined in terms of external concentrations, the output of the metabolic model will be entirely controlled by the environment (unstructured modelling). The preferred approach will be to track some dynamic biological properties and use them to compute these resources, then the model will behave in a dynamic way and the "memory" of cells will affect their metabolic behaviour. Some dynamic properties could be

- Internal concentrations, or storage quantities;
- Number and/or activity of membrane transporters (porines, PTS, ...);
- Non-biochemical properties of cells such as age, mass, length, number of bud scars for yeasts, ...
- Reactive capabilities, either expressed in terms of enzyme concentrations, or in terms of maximum achievable reaction rates.

Instead of tracking all possible dynamic properties, following model developments will focus on (i) identifying what dynamic behaviours are of interest, (ii) listing which metabolic pathways and elementary modes are necessary to describe them, and (iii) defining which biological properties are required to account for these dynamics. The goal will be to obtain models that exhibit correct behaviours in lab-scale dynamic cultures, while keeping the model size as low as possible in terms of elementary modes and dynamic variables.

## II.3 Application to a simple micro-organism: *Escherichia coli*

### II.3.1 General presentation

*Escherichia coli* is a well-known bacteria whose genome sequence is known since 1997 [Blattner et al.]. This micro-organism has multiple industrial uses through the introduction of heterologous metabolism pathways. It is mainly known for its historic use in therapeutic insulin production, but is also used for the production of amino acids, monomers (1, 3-Propanediol, 1, 4-Butanediol), other protein therapeutics, organics acids, and so on [Theisen and Liao, 2016].

As stated in introduction (II.1), this bacteria is optional aero-anaerobic, meaning that it can produce energy using either oxidative of fermentative pathways. It is able to use glucose as a primary source of carbon and energy, and produces multiple organic acids, the main one being acetate, when using fermentative pathways. These acids can also serve as carbon source when sugars are depleted and oxygen is available. Finally, *E. coli* is known to produce acetate when exposed to significantly high glucose concentrations, this is known as the overflow metabolism. This phenomenon has been extensively studied over the last decades as this acetate production inhibits growth, but also diverts glucose from metabolic pathways of industrial interest and therefore affects the conversion yield of the process [Xu et al., 1999; Anane et al., 2017]. All of these metabolic aspects are considered in the following model.

## II.3.2 Elementary metabolic modes, limitations and global rates

### Metabolic network

The basis of metabolic models is the considered metabolic network which sums up all bioreactions. Here, the considered network is kept close to the one illustrated in Fig. II.1 and is illustrated in Fig. II.3. This network consists in following bioreactions:

$$G + Y_{EG}E \xrightarrow{q_{r_1}^{(G)}} Y_{XG}X \qquad (r_1)$$

$$A + Y_{EA}E \xrightarrow{q_{r_{1'}}^{(A)}} Y_{XA}X \qquad (r_1')$$

$$G + Y_{OG}O \xrightarrow{q_{r_2}^{(G)}} Y_{EG}^{(o)}E \qquad (r_2)$$

$$A + Y_{OA}O \xrightarrow{q_{r_{2'}}^{(A)}} Y_{EA}^{(o)}E \qquad (r_2')$$

$$G + \xrightarrow{q_{r_3}^{(G)}} Y_{EG}^{(f)}E + Y_{AG}A \qquad (r_3)$$

$$G \xrightarrow{q_{r_4}^{(G)}} Y_{AG}A \qquad (r_4)$$

$$G \xrightarrow{q_{r_5}^{(G)}} \emptyset \qquad (r_5)$$

with G: Glucose, A: Acetate, O: Oxygen, X: Cells/Biomass and E: Adenosine TriPhosphate, an energy-carrying molecule. $Y_{ij}$ are molar conversion yields ($\mathrm{mol_i/mol_j}$), and $q_i^{(j)}$ the specific reaction rates of reaction $i$, expressed in terms of processed quantity of $j$ ($\mathrm{mol_j/kg_X \cdot h}$).

$r_1$ represents the production of new cells using glucose and energy. $r_2$ and $r_3$ are reactions for glucose-based production of energy, respectively through oxidative and fermentative catabolism. $r_4$ is the dissimilation of excess glucose into acetate through the overflow metabolism. $r_1'$ and $r_2'$ are acetate-based reactions for cell production and energy production through an oxidative pathway.

Reaction $r_5$ was not explicitly described in the model from Pigou and Morchain [2015]. It corresponds to a use of glucose for unproductive processes generally referred to as maintenance processes [Hempfling and Mainzer, 1975]. Usually, maintenance is associated to a use of energy through ATP consumption [Meadows et al., 2010]. However, the model described in Pigou and Morchain [2015] (Appendix A) made use of Pirt's law [1965] to account for the effect of maintenance on the conversion yield of glucose into biomass, $Y_{XG}$. This was done by varying this yield depending on the performed growth rate. By separating maintenance into a different reaction, $Y_{XG}$ is now a model constant. The fact that reaction $r_5$ uses glucose in place of ATP is related to the definition of the maintenance

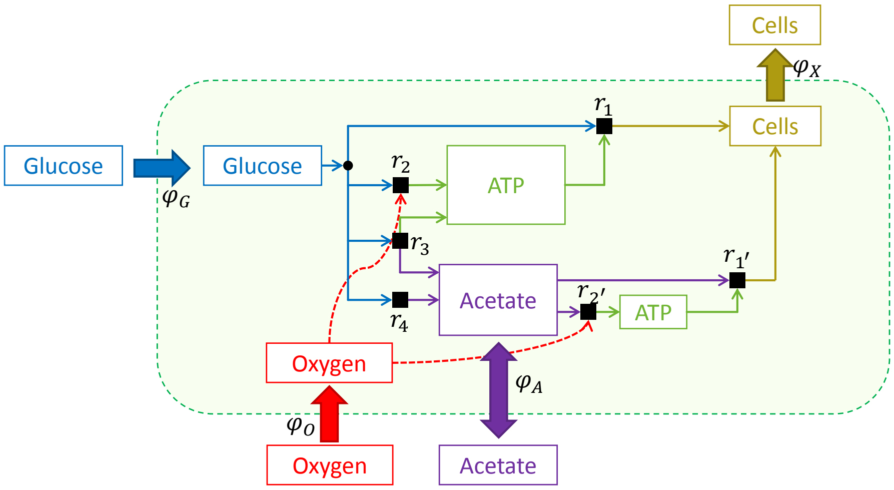


Figure II.3: Simplified metabolic network considered for *E. coli* modelling.

rate in Pirt's law, expressed as a consumption of substrate ($\mathrm{mol_G/g_X \cdot h}$). This was kept here to prevent the current model and the one described in Appendix A from differing.

### Elementary modes

The *E. coli* model will associate previously defined bioreactions into the five following elementary modes:

$\mathcal{A}$. Maintenance

$\mathcal{B}$. Oxidative growth on glucose

$\mathcal{C}$. Fermentative growth on glucose

$\mathcal{D}$. Overflow

$\mathcal{E}$. Oxidative growth on acetate

These modes are here sorted by their order of priority. Indeed, maintenance corresponds to the highest needs of a cell to ensure its minimal functioning. Glucose will then be oriented toward this mode in priority. Then come modes related to growth using the highest affinity substrate: glucose. Priority is given to the oxidative growth as its efficient energy-producing pathway allows for highest yields of conversion. *If not all glucose is used for growth, leftovers are directed into the overflow metabolism. This is the non-equilibrium metabolism that was not present in Xu et al. [1999] model.* Finally, a mode enables growth on acetate using an oxidative energy producing pathway.

All five modes are illustrated in Fig. II.4. Let $q_i^{(j)}$ be the rate of mode $i \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$ ($\mathrm{mol_j/kg_X \cdot h}$), $\varphi_j^{(i)}$ be the mass rate of production of $j \in \{G, A, O, X\}$ by the mode $i$ ($\mathrm{kg_j/kg_X \cdot h}$), and $f_j^{(i)}$ be the molar rate of production of $j$ ($\mathrm{mol_j/kg_X \cdot h}$). Note that

$$\varphi_j^{(i)} = M_j f_j^{(i)} \tag{II.1}$$

with $M_j$ the molar mass of $j$ ($\mathrm{kg_j/mol_j}$). Biomass is represented by the typical chemical formula $C_5H_7NO_2$.

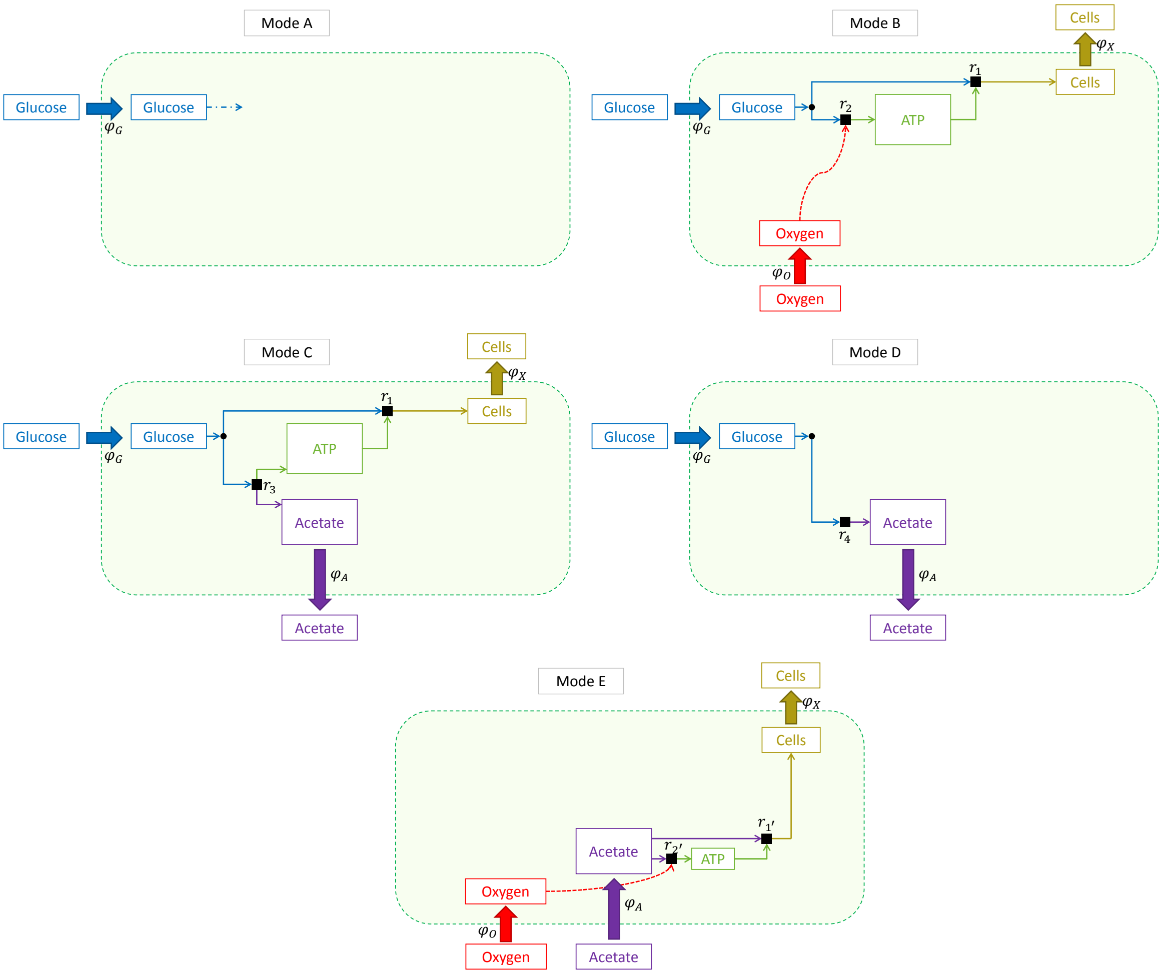


Figure II.4: Elementary modes for *E. coli* metabolic modelling.

Table II.1: Stoichiometry of elementary modes for *E. coli* metabolic model: consumption and production rates.

| Mode, $i$ | Mode rate, $q_i^{(j)}$ ($\mathrm{mol_j/kg_X \cdot h}$) | $f_G^{(i)}/q_i^{(j)}$ ($\mathrm{mol_G/mol_j}$) | $f_A^{(i)}/q_i^{(j)}$ ($\mathrm{mol_A/mol_j}$) | $f_O^{(i)}/q_i^{(j)}$ ($\mathrm{mol_O/mol_j}$) | $f_X^{(i)}/q_i^{(j)}$ ($\mathrm{mol_X/mol_j}$) |
|---|---|---|---|---|---|
| $\mathcal{A}$ | $q_{\mathcal{A}}^{(G)}$ | $-1$ | $0$ | $0$ | $0$ |
| $\mathcal{B}$ | $q_{\mathcal{B}}^{(G)}$ | $-1$ | $0$ | $-\frac{Y_{EG}Y_{OG}}{Y_{EG}+Y_{EG}^{(o)}}$ | $\frac{Y_{XG}Y_{EG}^{(o)}}{Y_{EG}+Y_{EG}^{(o)}}$ |
| $\mathcal{C}$ | $q_{\mathcal{C}}^{(G)}$ | $-1$ | $\frac{Y_{AG}Y_{EG}}{Y_{EG}+Y_{EG}^{(f)}}$ | $0$ | $\frac{Y_{XG}Y_{EG}^{(f)}}{Y_{EG}+Y_{EG}^{(f)}}$ |
| $\mathcal{D}$ | $q_{\mathcal{D}}^{(G)}$ | $-1$ | $Y_{AG}$ | $0$ | $0$ |
| $\mathcal{E}$ | $q_{\mathcal{E}}^{(A)}$ | $0$ | $-1$ | $-\frac{Y_{EA}Y_{OA}}{Y_{EA}+Y_{EA}^{(o)}}$ | $\frac{Y_{XA}Y_{EA}^{(o)}}{Y_{EA}+Y_{EA}^{(o)}}$ |

Table II.2: Stoichiometry of elementary modes for *E. coli* metabolic model: reaction rates. Values are ratios between reaction rates and mode rates (mol/mol).

| | | Reaction rate, $q_i^{(j)}$ ($\mathrm{mol_j/kg_X \cdot h}$) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $q_{r_1}^{(G)}$ | $q_{r_2}^{(G)}$ | $q_{r_3}^{(G)}$ | $q_{r_4}^{(G)}$ | $q_{r_5}^{(G)}$ | $q_{r_{1'}}^{(A)}$ | $q_{r_{2'}}^{(A)}$ |
| Mode rate, $q_i^{(j)}$ ($\mathrm{mol_j/kg_X \cdot h}$) | $q_{\mathcal{A}}^{(G)}$ | $0$ | $0$ | $0$ | $0$ | $1$ | $0$ | $0$ |
| | $q_{\mathcal{B}}^{(G)}$ | $\dfrac{Y_{EG}^{(o)}}{Y_{EG}+Y_{EG}^{(o)}}$ | $\dfrac{Y_{EG}}{Y_{EG}+Y_{EG}^{(o)}}$ | $0$ | $0$ | $0$ | $0$ | $0$ |
| | $q_{\mathcal{C}}^{(G)}$ | $\dfrac{Y_{EG}^{(f)}}{Y_{EG}+Y_{EG}^{(f)}}$ | $0$ | $\dfrac{Y_{EG}}{Y_{EG}+Y_{EG}^{(f)}}$ | $0$ | $0$ | $0$ | $0$ |
| | $q_{\mathcal{D}}^{(G)}$ | $0$ | $0$ | $0$ | $1$ | $0$ | $0$ | $0$ |
| | $q_{\mathcal{E}}^{(A)}$ | $0$ | $0$ | $0$ | $0$ | $0$ | $\dfrac{Y_{EA}^{(o)}}{Y_{EA}+Y_{EA}^{(o)}}$ | $\dfrac{Y_{EA}}{Y_{EA}+Y_{EA}^{(o)}}$ |

Though this would be feasible through the use of dynamic variables, we will not describe the accumulation of intracellular matter. Therefore, one can easily access the overall stoichiometry of each mode by formulating mass and energy balances over intracellular content. This leads to the stoichiometry table in Tab. II.1 where the consumption and/or production rates $\varphi_j^{(i)}$ are expressed in terms of mode rates $q_i^{(j)}$, and to Tab. II.2 which details the activity level of each reaction for all modes.

It is crucial to observe that the strength of this modelling approach lies in its ability to detect if some substrate is available for side reactions. This ability comes from the a priori knowledge of both the uptake rate and the maximum glucose utilization rate for growth. It is worth explaining this feature in some more details because it is also tightly related to the dynamic capabilities of the metabolic model.

### Growth-rate dynamics

The model proposed in Pigou and Morchain [2015] (article available in Appendix A) is based on the idea that growth-rate is a dynamic property of cells. By tracking the growth capability of a cell, one is able to define its needs in terms of glucose consumption. On the other hand, it has been observed that glucose uptake is not directly correlated to growth-rate [Natarajan and Srienc, 2000] as their “reactivity” [Leegwater et al., 1982] allows them to dynamically adapt their uptake capacity in response to environmental fluctuations in substrate availability [Ferenci, 1996]. Therefore, by (i) tracking growth capability and (ii) modelling glucose uptake, one will be able to measure a disequilibrium between glucose needs and uptake, which will condition the activation of the overflow metabolism if glucose is in excess.

This notion of a dynamic growth-rate is often not considered for the modelling of bioreactors. Instead, one often assume a local equilibrium between a cell and its environment by making use of equilibrium kinetics that usually follow the [Monod, 1952], already presented in I.6.3:

$$\mu^* = \mu_{\max}\frac{G^*}{G^* + K_G} \tag{II.2}$$

where $\mu^*$ and $G^*$ are steady-state –at equilibrium– growth-rate and glucose concentration, $\mu_{\max}$ is the maximum mean growth-rate observed at the population level, and $K_G$ is the affinity for glucose ($kg_G/m_L^3$). Other terms can be added to this law to account for affinity toward other substrates and/or for inhibition by some compounds. These laws are deduced from experimental observations as detailed hereafter.

Using Monod kinetics law to model steady-state conditions from the point of view of a cell is a valid approach. However, large-scale bioreactors are heterogeneous, and a cell travelling in concentration gradients will experience dynamic fluctuations in its surrounding environment. Using Monod kinetics in these systems would then be an over-simplification “by allowing instantaneous adaptation of the cell to the abiotic environment” [Silveston et al., 2008].

This is perfectly illustrated by the experimental response of a chemostat system to a dilution-rate shift-up. Chemostats are continuous homogeneous lab-scale cultures characterized by their dilution rate $D = Q/V$ ($\mathrm{h}^{-1}$) with $V$ their volume ($\mathrm{m}^3$) and $Q$ the flow rate passing through the reactor ($\mathrm{m}^3/\mathrm{h}$). A simple mass balance on such a system shows that at steady state, the dilution rate of a chemostat matches the mean growth rate of a biological population ($\widetilde{\mu} = D$ at steady-state).

As illustrated in Fig. II.5, if one suddenly increases the dilution rate from $D$ to $D'$ in a chemostat culture, the biomass concentration first decreases (meaning that $\widetilde{\mu} < D'$), then increases ($\widetilde{\mu} > D'$) to finally reach a new equilibrium ($\widetilde{\mu} = D'$). This experimentally proves that the growth-rate is indeed a dynamic variable.

This problematic of whether Monod models can be applied to dynamic chemostat cultures was actually fully investigated half a century ago [Young et al., 1970]. The conclusion of these authors was that inducing a delay in the evolution of growth-rate was necessary to achieve a reasonable fitting with experimental data.

Let $\mu_{c@p}$ be the glucose-based growth-rate of a cell, whose evolution is described as

$$\frac{\partial \mu_{c@p}}{\partial t} = \zeta_\mu(\mu_{c@p}, \boldsymbol{C}_L) \tag{II.3}$$

Chapter III will detail how such a biological property can be tracked over an entire biological population. $\zeta_\mu(\mu_c, \boldsymbol{C}_L)$ will be defined in such a way that $\mu_c$ will evolve over-

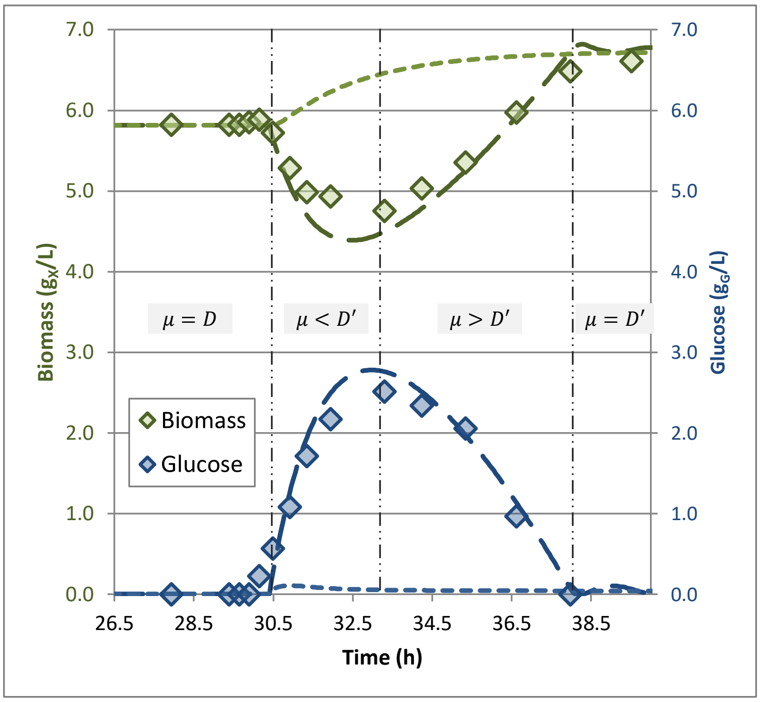


Figure II.5: Experimental measurements of biomass and glucose concentrations in a chemostat culture after a dilution rate shift-up [Kätterer et al., 1986]. Dashed-lines: modelling with dynamic growth-rate. Dotted-lines: modelling with equilibrium growth-rate.

time toward $\mu^*$ (computed from $\boldsymbol{C}_L$). The following model was proposed by Morchain and Fonade [2009]:

$$\zeta_\mu(\mu_{c@p}, \boldsymbol{C}_L) = \left(\frac{1}{T} + \mu_{c@p}\right)\left(\mu^*(\boldsymbol{C}_L) - \mu_{c@p}\right) \tag{II.4}$$

This formulation implies that all cells will adapt their growth rate to eventually reach $\mu^*(\boldsymbol{C}_L)$ but that fastest growing cells will be able to adapt their growth machinery faster. The question is then to identify the time-scale $T$ and the equilibrium law $\mu^*(\boldsymbol{C}_L)$.

This delay between an environmental change and the biological answer in terms of growth rate has been observed on *Candida tropicalis* and *Trichosporon cutaneum* by Kätterer et al. [1986], on *Azobacter vinelandi* and *Saccharomyces cerevisiae* [Young et al., 1970], and on *Lactococcus lactis* [Adamberg et al., 2009]. Each time, the time-scale of that delay was about an hour. This seems to imply that producing the growth-related cellular machinery is as complex and slow as producing new cells, considering that the characteristic time of growth is about an hour too. Therefore, $T \approx 1$h and can be identified through dilution-rate shift-up experiments [Kätterer et al., 1986] or in accelerostat cultures [Adamberg et al., 2009].

A subtlety implemented in Pigou and Morchain [2015] was the use of different time-scales $T_u$ and $T_d$ depending on whether the growth-rate adaptation was performed upward ($\mu_c < \mu^*$) or downward ($\mu_c > \mu^*$). This was necessary for fitting the modelling to experimental data in large-scale bioreactors and implied that upward adaptation is a faster phenomenon than downward adaptation.

The notion of equilibrium, and thus the Monod law (Eq. (II.2)), can be analysed experimentally in chemostat cultures. A simple mass balance on the biomass over a chemostat will be

$$\frac{\partial X_T}{\partial t} = \left(\widetilde{\mu} - \frac{Q}{V}\right) X_T \tag{II.5}$$

where $X_T = \int_{\Omega_\mu} n(\mu_c)\mathrm{d}\mu_c$ is the total biomass concentration ($kg_X/m^3$), $\widetilde{\mu}$ is the population mean growth-rate and $V$ the reactor volume. At steady-state, one can easily observe that $\widetilde{\mu} = Q/V = D$ with $D$ being the chemostat dilution rate. Therefore, by controlling the feeding flow rate $Q$, one can enforce a growth-rate at steady-state.

By doing so, it is possible to experimentally observe an equilibrium law between a biological population mean growth-rate and the residual substrate concentration. An example is given in Fig. II.6 from Schmideder et al. [2015] where lab-scale bioreactors were used to access steady-state kinetics of *E. coli* on glucose. Parameters of the Monod law can be extracted from such experiments. $\mu_{\max}$ is the maximum growth rate achievable by the population. For high dilution rates, $D > \mu_{\max}$, the population can not growth fast enough and is flushed out of the reactor. This allows identifying $\mu_{\max}$ as the maximum dilution rate at which biomass is still present ($\approx 0.52\mathrm{h}^{-1}$ in Fig. II.6).

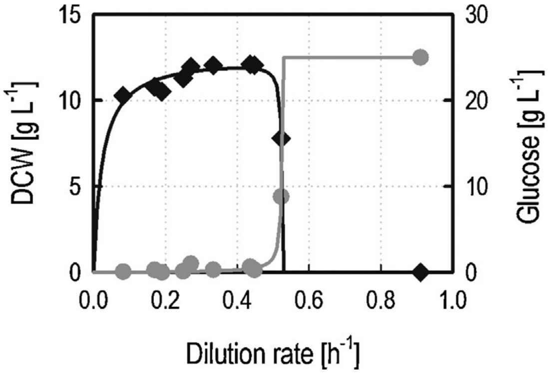


Figure II.6: State of equilibrium between a *E. coli* population and its environment in milliliter-scale chemostat cultures [Schmideder et al., 2015].

### Mode limitations

Pigou and Morchain [2015] detailed a procedure to compute reaction rates $q_{r_1}^{(G)}$, $q_{r_2}^{(G)}$, ... from environment concentrations $G_L$, $A_L$ and $O_L$ and the glucose-based growth capability $\mu_{c@p}$ (see Appendix A of that article for full details of the initial procedure). Here, the goal is to define the maximum consumption rates or reaction rates, *i.e.* available resources, such that the metabolic mode-based model will behave in a similar manner than the decision-tree model detailed in Pigou and Morchain [2015]. As far as consumption rate are concerned, they can be limited either by biological limits, *e.g.* insufficient membrane transporters, or by environmental limits, *e.g.* depleted environment or insufficient micromixing. The model from Pigou and Morchain [2015] did not explicitly consider both aspects. In terms of notations, environmental limits are noted $\varphi_i^{(e)}$, biological limits $\varphi_i^{(b)}$ and global limits $\varphi_i^{(g)}$ ($\mathrm{kg_i}/\mathrm{kg_X}\cdot\mathrm{h}$).

First, the maintenance elementary mode will be limited by the maintenance rate $m$ initially defined in Pirt's law as the quantity of glucose required for maintenance operations ($\mathrm{mol_G/kg_X \cdot h}$). This will simply translate into a limit reaction rate for reaction $r_5$

$$q_{r_5}^{(g)} = m \tag{II.6}$$

Models from Xu et al. [1999] and Pigou and Morchain [2015] both used the same formulation for a maximal oxygen consumption rate:

$$\varphi_O^{(g)} = \varphi_O^{(\max)} \frac{O}{O + K_O} \frac{K_{i,A}^{(O)}}{A + K_{i,A}^{(O)}} \tag{II.7}$$

with $\varphi_O^{(\max)}$ the maximum achievable oxygen consumption rate, $K_O$ the affinity constant of *E. coli* toward oxygen, and $K_{i,A}^{(O)}$ the inhibition constant of oxygen uptake by acetate.

Equilibrium growth rates on glucose and acetate are defined by following Monod laws:

$$\mu_G^* = \mu_G^{(\max)} \frac{G}{G + K_G} \frac{O}{O + K_O} \frac{K_{i,A}}{A + K_{i,A}} \tag{II.8}$$

$$\mu_A^* = \mu_A^{(\max)} \frac{A}{A + K_A} \frac{O}{O + K_O} \frac{K_{i,G}}{G + K_{i,G}} \tag{II.9}$$

Elementary modes leading to cell-production might be limited by growth capabilities, defined as maximum reaction rates for reactions $r_1$ and $r_1'$. In accordance with the model from Pigou and Morchain [2015], these maximum growth reaction rates are defined from cell-attached growth capability $\mu_{c@p}$ and from equilibrium growth rates $\mu_G^*$ and $\mu_A^*$:

$$q_{r_1}^{(g)} = \frac{\min\left(\mu_{c@p}, \mu_G^*\right)}{M_X Y_{XG}} \tag{II.10}$$

$$q_{r_1'}^{(g)} = \frac{\min\left(\mu_{c@p} - \min\left(\mu_{c@p}, \mu_G^*\right), \mu_A^*\right)}{M_X Y_{XA}} \tag{II.11}$$

The underlying idea is simply that cells grow as fast as their internal capabilities allow them to, unless their environment is limiting them.

The last limitation introduced in Pigou and Morchain [2015] is related to glucose consumption. Considering that glucose uptake systems are quickly regulated compared to the adaptation of growth capabilities [Ferenci, 1996; Quedeville et al., 2018], we assumed that all cells will assimilate as much glucose as a cell at equilibrium with its environment. Therefore, $\varphi_G^{(g)}$ is defined as the total glucose consumption rate of a cell characterized by $\mu_{c@p} = \mu_G^*$ with null mode rates for overflow and acetate-based growth elementary metabolic modes ($q_{\mathcal{D}}^{(G)} = 0$ and $q_{\mathcal{E}}^{(A)} = 0$).

## Mode rates computation

Once metabolic modes are defined in terms of stoichiometry and priority ranks, along with maximum available resources, computing actual mode rates $q_i^{(j)}$ is straightforward. To summarize, available resources are:

- Consumption rates
    - of oxygen: $\varphi_O^{(g)}$ ($\mathrm{kg_O/kg_X \cdot h}$) or $f_O^{(g)} = \varphi_O^{(g)}/M_O$ ($\mathrm{mol_O/kg_X \cdot h}$);

  - of glucose: $\varphi_G^{(g)}$ $(\mathrm{kg_G/kg_X \cdot h})$ or $f_G^{(g)} = \varphi_G^{(g)}/M_G$ $(\mathrm{mol_G/kg_X \cdot h})$.

- Reaction rates
  - of glucose-based growth: $q_{r_1}^{(g)}$ $(\mathrm{mol_G/kg_X \cdot h})$;
  - of acetate-based growth: $q_{r_1'}^{(g)}$ $(\mathrm{mol_A/kg_X \cdot h})$;
  - of maintenance: $q_{r_5}^{(g)}$ $(\mathrm{mol_G/kg_X \cdot h})$.

This process of computing mode rates simply consists in following operations

1. Compute mode $\mathcal{A}$ rate by considering limits in maintenance rate $(q_{r_5}^{(g)})$ and glucose availability $f_G^{(g)}$,

$$q_{\mathcal{A}}^{(G)} = \min\left(q_{r_5}^{(g)}; f_G^{(g)}\right)$$

2. Compute resources left for modes $\mathcal{B}$ to $\mathcal{E}$ by decreasing glucose availability,

$$f_G^{(g)} \leftarrow f_G^{(g)} - q_{\mathcal{A}}^{(G)}$$

3. Compute mode $\mathcal{B}$ rate by considering limits in glucose-based growth reaction rate $q_{r_1}^{(g)}$, glucose availability $f_G^{(g)}$ and oxygen availability $f_O^{(g)}$,

$$q_{\mathcal{B}}^{(G)} = \min\left(\frac{Y_{EG} + Y_{EG}^{(o)}}{Y_{EG}^{(o)}} q_{r_1}^{(g)}; f_G^{(g)}; \frac{Y_{EG} + Y_{EG}^{(o)}}{Y_{EG} Y_{OG}} f_O^{(g)}\right)$$

4. Compute resources left for modes $\mathcal{C}$ to $\mathcal{E}$ by decreasing glucose and oxygen availabilities, as well as glucose-based growth capability,

$$\begin{aligned} f_G^{(g)} &\leftarrow f_G^{(g)} - q_{\mathcal{B}}^{(G)} \\ f_O^{(g)} &\leftarrow f_O^{(g)} - \frac{Y_{EG} Y_{OG}}{Y_{EG} + Y_{EG}^{(o)}} q_{\mathcal{B}}^{(G)} \\ q_{r_1}^{(g)} &\leftarrow q_{r_1}^{(g)} - \frac{Y_{EG}^{(o)}}{Y_{EG} + Y_{EG}^{(o)}} q_{\mathcal{B}}^{(G)} \end{aligned}$$

5. Compute mode $\mathcal{C}$ rate by considering limits in glucose based growth reaction rate $q_{r_1}^{(g)}$ and glucose availability $f_G^{(g)}$;

$$q_{\mathcal{C}}^{(G)} = \min\left(\frac{Y_{EG} + Y_{EG}^{(f)}}{Y_{EG}^{(f)}} q_{r_1}^{(g)}; f_G^{(g)}\right)$$

6. Compute resources left for modes $\mathcal{D}$ and $\mathcal{E}$ by decreasing glucose availability,

$$f_G^{(g)} \leftarrow f_G^{(g)} - q_{\mathcal{C}}^{(G)}$$

7. Compute mode $\mathcal{D}$ rate by considering only glucose availability $f_G^{(g)}$,

$$q_{\mathcal{D}}^{(G)} = f_G^{(g)} \tag{II.12}$$

8. Compute mode $\mathcal{E}$ rate by considering limits in acetate-based growth reaction rate $q_{r1'}^{(g)}$ and oxygen availability $f_O^{(g)}$,

$$q_{\mathcal{E}}^{(A)} = \min\left(\frac{Y_{EA}+Y_{EA}^{(o)}}{Y_{EA}^{(o)}}q_{r1'}^{(g)}; \frac{Y_{EA}+Y_{EA}^{(o)}}{Y_{EA}Y_{OA}}f_O^{(g)}\right)$$

**Overall bioreaction rates**

Once mode rates are known, the actual growth-rate, substrate consumption rates, and (by-)product production rates are simply given by a linear combination of mode rates based on stoichiometry coefficients from Tab. II.1. These actual rates are denoted by the upper-script $^{(a)}$:

$$\begin{aligned}
\mu^{(a)} &= \left(\frac{Y_{XG}Y_{EG}^{(o)}}{Y_{EG}+Y_{EG}^{(o)}}q_{\mathcal{B}}^{(G)} + \frac{Y_{XG}Y_{EG}^{(f)}}{Y_{EG}+Y_{EG}^{(f)}}q_{\mathcal{C}}^{(G)} + \frac{Y_{XA}Y_{EA}^{(o)}}{Y_{EA}+Y_{EA}^{(o)}}q_{\mathcal{E}}^{(A)}\right)M_X \\
\varphi_G^{(a)} &= -\left(q_{\mathcal{A}}^{(G)} + q_{\mathcal{B}}^{(G)} + q_{\mathcal{C}}^{(G)} + q_{\mathcal{D}}^{(G)}\right)M_G \\
\varphi_A^{(a)} &= \left(\frac{Y_{AG}Y_{EG}}{Y_{EG}+Y_{EG}^{(f)}}q_{\mathcal{C}}^{(G)} + Y_{AG}q_{\mathcal{D}}^{(G)} - q_{\mathcal{E}}^{(A)}\right)M_A \\
\varphi_O^{(a)} &= -\left(\frac{Y_{EG}Y_{OG}}{Y_{EG}+Y_{EG}^{(o)}}q_{\mathcal{B}}^{(G)} + \frac{Y_{EA}Y_{OA}}{Y_{EA}+Y_{EA}^{(o)}}q_{\mathcal{E}}^{(A)}\right)M_O
\end{aligned}$$

### II.3.3 Discussion

This *E. coli* model mainly serves as an illustration of the metabolic mode based structure, by reusing and restructuring developments published prior to this PhD thesis where the model predictions were challenged against experimental data [Pigou and Morchain, 2015]. This model was voluntarily kept simple by only describing a slow cell dynamic, the growth-rate adaptation, to show the interest of coupling such simple metabolic models with biological properties tracking. It will be shown hereafter, in Section II.4, that adding a few dynamics, while preserving this model structure, allows representing more advanced metabolic behaviour, here in the case of the baker's yeast *Saccharomyces cerevisiae*.

The most important aspect of the proposed model is its low numerical cost. Many quantities can be precomputed, the numerical cost is then mainly related to (i) the computation of available resources, (ii) the computation of the maximum mode rate for each mode, which was shown to simply consist in the identification of the most limiting rate and (iii) the computation of resources left for ulterior modes.

This simple order of computations, consisting only in multiplications, addition, and calls to the min function, allows for fast computations. In particular, if this model must be called numerous times –which will be the cased in Chapter V for large-scale simulations– it is important to note that such a model structure is compatible with vectorization. This can be done either in [MATLAB, 2016] using vectorization notations, or more broadly using instruction sets available in last generation processors (SSE, AVX, ...), to further accelerate computations of the metabolic behaviour.

In addition to its low cost and possible computation speed-ups allowed by the computation procedure simplicity, a major feature of this model structure is its facilitated modularity. For instance, one could easily implement a mortality phenomenon that would be triggered if maintenance requirements are not fulfilled. Doing so would simply require a supplementary reaction that would describe cell death:

$$X \xrightarrow{q_{r_6}^{(X)}} \emptyset \tag{$r_6$}$$

A new mode, say $\mathcal{F}$, would trigger death to a rate proportional to the unsatisfied maintenance rate, $q_{r_5}^{(g)}$. If maintenance requirements are fulfilled by mode $\mathcal{A}$, $q_{r_5}^{(g)} = 0$ for modes $\mathcal{B}$ to $\mathcal{F}$ and therefore no death occurs. Otherwise, cell death is triggered.

This modularity comes from the fluency of adding or removing metabolic modes without influencing much other aspects of the model, but also from the easiness of adding dynamics of interest. For instance, we defined the glucose uptake limit, $\varphi_G^{(g)}$, directly from a cell equilibrium state and thus we assumed that glucose uptake is instantaneously adapted to environmental fluctuations. It can be experimentally shown that the uptake rate of glucose after a sudden exposure to high concentrations induces a quick regulation of membrane transporters to limit the influx of glucose [Sunya et al., 2012]. This is explained by Ferenci [1996] through the coexistence of two membrane transporter families, (i) porines which are passive transport systems, based on natural diffusivity through protein channel and (ii) the PhosphoTransferase System (PTS) which is an active (*i.e.* energy consuming) transport system [Chassagnole et al., 2002]. Activities and number of both transporters can be regulated with a time-scale of a few seconds to a few minutes, which induces a short-term dynamic response of glucose uptake capabilities after environmental perturbations. If one models dynamics of these transporter systems, it is possible to derive $\varphi_G^{(g)}$ from this cell-attached information and therefore improve the overall model to short-term perturbations. Actually, this work has been started in a project parallel to this PhD thesis, but needs few refinements before coupling it with current metabolic model [Quedeville et al., 2018].

Summing-up steps required for the definition of a mode-based metabolic model, one needs to

1. list reactions defining the metabolic network and their respective stoichiometry;
2. define metabolic modes as self-sufficient combinations of elementary reactions;
3. define the order of priority of elementary modes;
4. identify and quantify substrate uptake, and reactive, resources that may limit the rate associated to each mode.

Once the networks and modes are defined in terms of stoichiometry matrices, and closures are provided for limiting resource, the computation of mode rates is straightforward and numerically efficient.

A final aspect of interest lies in the choice of cell-attached dynamic variables. Usual approaches consist in tracking cell-size, mass or age [Fredrickson and Tsuchiya, 1963; Hatzis et al., 1995; Heins et al., 2015] while we propose to track the growth-rate. The rational behind this choice is that from this growth-rate, it is easy to deduce reactive and transfer flux in the metabolic network. Here, the growth-rate is associated to the capability of cells to process substrate to form new cells, which is sufficient information to deduce the activation rate of growth-related metabolic modes. If we tracked physical cell properties (age, mass, . . . ), more disputable closures would be required to deduce the rates of metabolic modes from these quantities.

## II.4 Application to *Saccharomyces cerevisiae*

### II.4.1 General presentation

*Saccharomyces cerevisiae* is one of the most employed strain for industrial production purposes [Theisen and Liao, 2016]. Thanks to its versatility, this micro-organism is employed both for (i) high-quantity low-value productions such as xylose-based biofuel [Kim et al., 2013] and for (ii) low-quantity high-value productions such as pharmaceutical molecules [Paddon and Keasling, 2014].

Industrial cultures of *S. cerevisiae* are subjected to the Glucose –or Crabtree– effect. Indeed, like *E. coli*, *S. cerevisiae* is an optional aero-anaerobic micro-organism. Therefore, it is expected to produce energy through oxidative pathways in presence of oxygen, and through fermentative pathways in anaerobic conditions. However, the Crabtree effect leads to the production of ethanol, showing that fermentative metabolism is active, even in presence of dissolved oxygen. This phenomenon is known to occur when the culture medium is rich in glucose, but with triggering concentrations varying between experimentations. This ethanol production is often undesired and detrimental due to the inhibitory effect of this by-product on growth, and the yield-loss resulting from the diversion of the main substrate toward ethanol producing pathways.

Just as *E. coli* which grows on both glucose and acetate, *S. cerevisiae* can use multiple carbon sources. It will be shown hereafter that glucose, ethanol and acetate are all possible carbon sources for this yeast.

Finally, the growth rate dynamics that were observed on *E. coli* are actually generalizable to most micro-organisms, including the baker's yeast. Therefore, these dynamics will be considered too for the modelling of this strain.

Many experimental data for *S. cerevisiae* cultures are available in the literature. The proposed model will rely on these observations and will be designed to fit, at least qualitatively, with observed tendencies, as well as quantitatively when possible.

### II.4.2 Metabolic network

The metabolic modelling will be dedicated to wild strains of *S. cerevisiae* and will focus on the central carbon metabolism. The specificity of this metabolism in *S. cerevisiae* compared to *E. coli* is the fact that the oxidative metabolism occurs in mitochondria whose main purpose is to produce energy, mainly Adenosine TriPhosphate (ATP) molecules, through the oxidation of co-enzymes that are reduced by other parts of the metabolism.

The central carbon metabolism has been modelled by Rizzi et al. [1997] who considered the metabolic network illustrated in Fig. II.7. However, closure for their model was obtained by tracking up to 15 internal concentrations, while we attempt to significantly limit the number of biological properties.

Therefore, we based our metabolic network on a more simple one, from Pronk et al. [1996], where main pathways are simplified into a few global reactions. Their minimalistic metabolic network is given in Fig. II.8, and the version we used, with associated reactions and trans-membrane mass transfers is illustrated in Fig. II.9.

Considered reactions are listed in Tab. II.3. Most reactions and isolated pathways have well defined stoichiometry. In these cases, coefficients directly appear in reaction formulas. Cell production (reaction $r_{10}$) is not an actual reaction so stoichiometry coefficients in

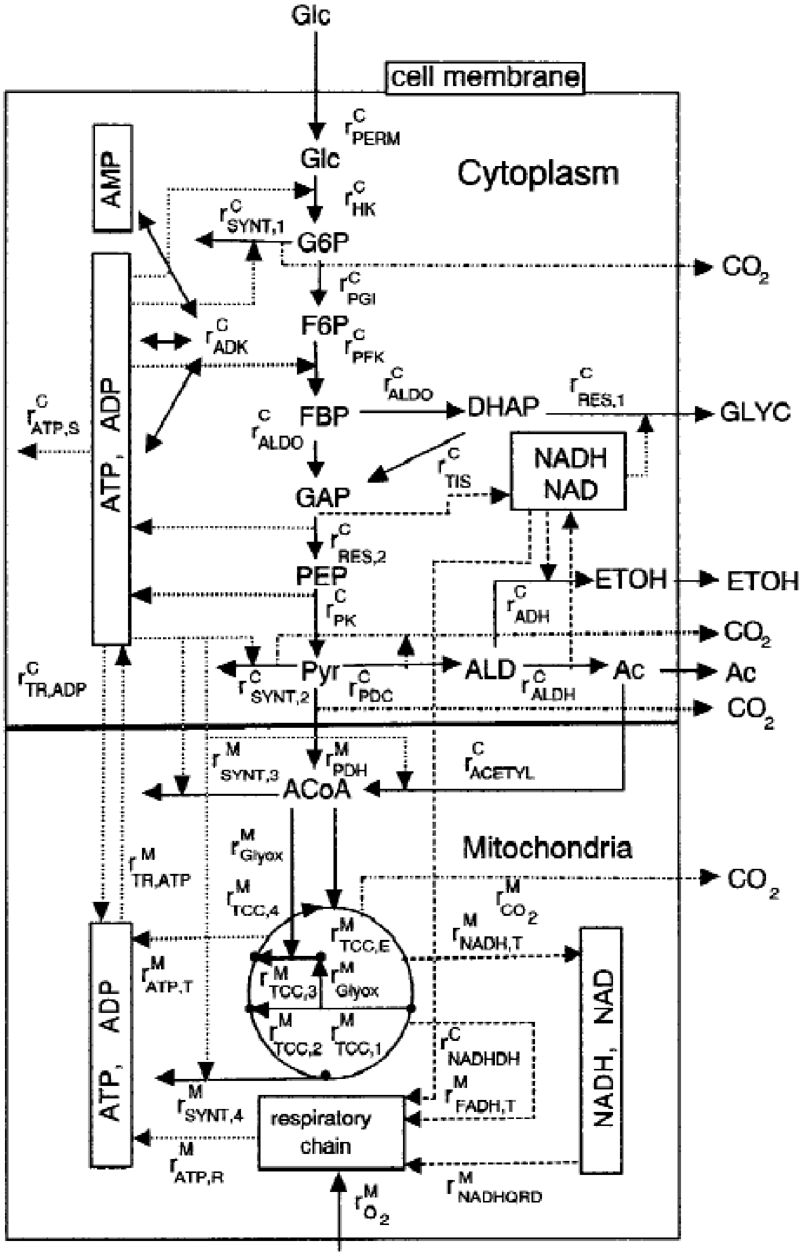


Figure II.7: Structure of the metabolic network considered for *S. cerevisiae* modelling by Rizzi et al. [1997]. Reprinted by permission from John Wiley and Sons: Biotechnology & Bioengineering, *In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model*, Rizzi, Baltes, Theobald and Reuss, COPYRIGHT (1997).

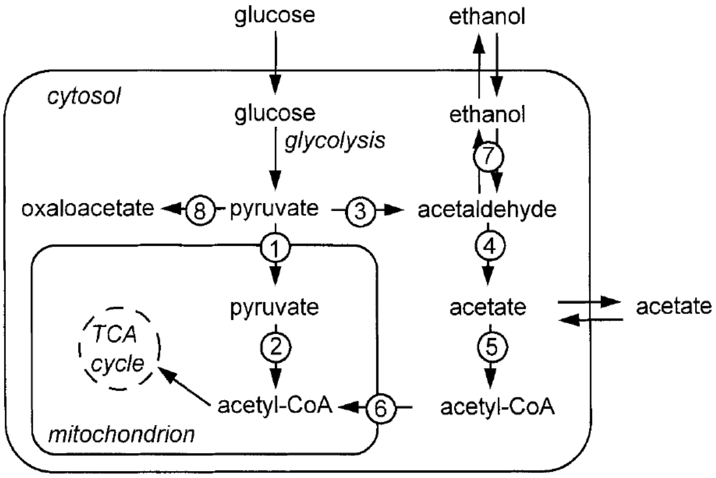


Figure II.8: Illustration of the central carbon metabolism of *S. cerevisiae* as detailed by Pronk et al. [1996]. Reprinted by permission from John Wiley and Sons: YEAST, *Pyruvate Metabolism in Saccharomyces cerevisiae*, Pronk, Yde Steensma and Van Dijken, COPYRIGHT (1996).

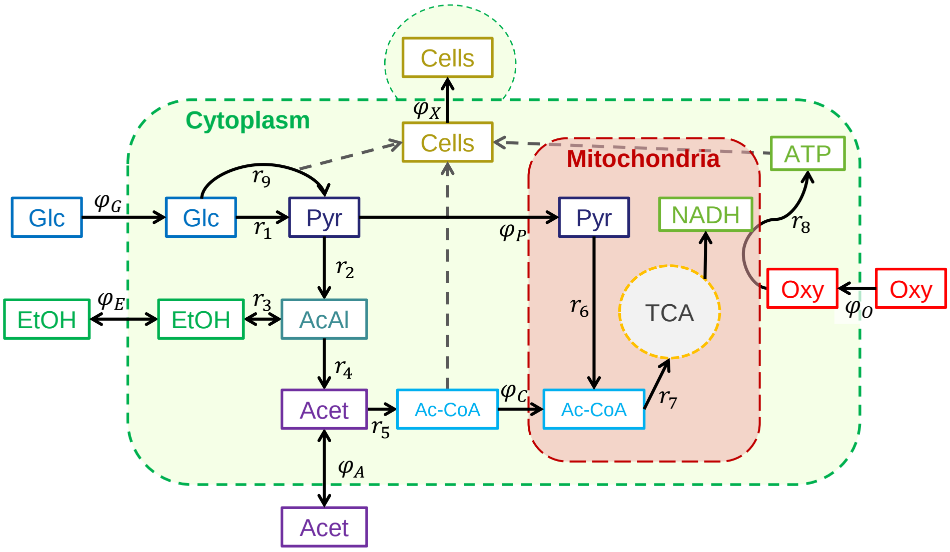


Figure II.9: Simplified metabolic network considered for *S. cerevisiae* modelling.

Table II.3: Detail of reactions used in the metabolic network from Fig. II.9.

| Reaction | Formula | Description |
|---|---|---|
| $r_{1a}$ | $\text{GLC} + \text{ATP} \xrightarrow{q_{r_{1a}}} \text{F6P}$ | Glycolysis I |
| $r_{1b}$ | $\text{F6P} + \text{ATP} \xrightarrow{q_{r_{1b}}} 2\text{G3P}$ | Glycolysis II |
| $r_{1c}$ | $\text{G3P} \xrightarrow{q_{r_{1c}}} \text{Pyr} + \text{NADH} + 2\text{ATP}$ | Glycolysis III |
| $r_2$ | $\text{Pyr} \xrightarrow{q_{r_2}} \text{AcAl}$ | Pyruvate decarboxylation |
| $r_3$ | $\text{EtOH} \xrightarrow{q_{r_3}} \text{AcAl} + \text{NADH}$ | Ethanol dehydrogenation |
| $r_3'$ | $\text{EtOH} \xrightarrow{q_{r_3'}} \text{AcAl} + \text{NADPH}$ | Ethanol dehydrogenation |
| $r_4$ | $\text{AcAl} \xrightarrow{q_{r_4}} \text{Acet} + \text{NADH}$ | Acetaldehyde dehydrogenation |
| $r_5$ | $\text{Acet} + \text{ATP} \xrightarrow{q_{r_5}} \text{AcCoA}$ | Acetate utilization |
| $r_6$ | $\text{Pyr} \xrightarrow{q_{r_6}} \text{AcCoA} + \text{NADH}$ | Pyruvate dehydrogenation |
| $r_7$ | $\text{AcCoA} \xrightarrow{q_{r_7}} 3.6\text{NADH} + \text{ATP}$ | TriCarboxylic Acid cycle |
| $r_8$ | $2\text{NADH} + \text{Oxy} \xrightarrow{q_{r_8}} 5\text{ATP}$ | Respiratory chain |
| $r_9$ | $3\text{GLC} \xrightarrow{q_{r_9}} 6\text{NADPH} + \text{G3P} + 2\text{F6P}$ | Pentose-Phosphate Pathway |
| $r_{10}$ | $Y_{CX}\text{AcCoA} + Y_{EX}\text{ATP} + Y_{NX}\text{NADPH} \xrightarrow{q_{r_{10}}} X$ | Cell production |

this reaction will be fitting parameters, and are allowed to differ from one elementary mode to another. Note that reactions formulas given in this table are simplifications of actual reactions, and only describe elements on which we will formulate mass balances to determine each mode global stoichiometry. For instance, we only describe reduced coenzymes (NADH and NADPH) and not their oxidized counterpart, as their mass balances are symmetric.

## II.4.3 Elementary modes

### Mode definition

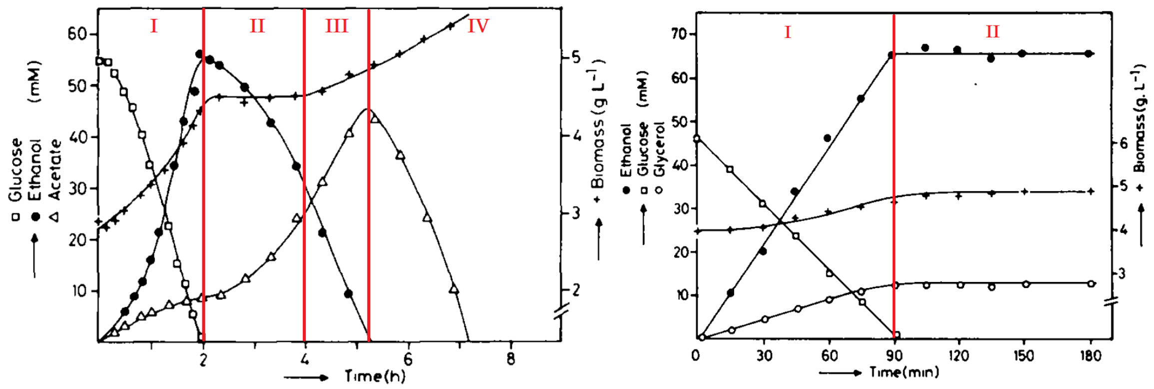
The choice of elementary modes comes from the analysis of experimental data available in the literature. What defines a mode is the list of its "reactant" (substrates, oxygen, ...), and the list of its "products" (cells, by-product or product of industrial interest). Many data sets were used to define the following list of metabolic modes, but those represented in Fig. II.10 are sufficient to exhibit most of the following modes:

$\mathcal{G}_1$. Oxidative growth on glucose.

$$Y_{GX}^{(\mathcal{G}_1)}\text{GLC} + Y_{OX}^{(\mathcal{G}_1)}\text{Oxy} \xrightarrow{q_{\mathcal{G}_1}^{(X)}} X$$

$\mathcal{G}_2$. Fermentative growth on glucose.

$$Y_{GX}^{(\mathcal{G}_2)}\text{GLC} \xrightarrow{q_{\mathcal{G}_2}^{(X)}} X + Y_{EX}^{(\mathcal{G}_2)}\text{EtOH}$$



(a) Aerobic batch culture of *S. cerevisiae* CBS8066 after addition of excess glucose in the culture medium.

(b) Anaerobic batch culture of *S. cerevisiae* CBS8066 after addition of excess glucose in the culture medium.

Figure II.10: van Dijken and Scheffers, *Redox balances in the metabolism of sugars by yeasts*, FEMS Microbiology Reviews, 1986, volume 32, issue 3, p. 199-224, by permission of Oxford University Press.

$\mathcal{E}_1$. Oxidative growth on ethanol.

$$Y_{EX}^{(\mathcal{E}_1)}\mathrm{EtOH} + Y_{OX}^{(\mathcal{E}_1)}\mathrm{Oxy} \xrightarrow{q_{\mathcal{E}_1}^{(X)}} X$$

$\mathcal{E}_2$. Oxidation of ethanol.

$$\mathrm{EtOH} + Y_{OE}^{(\mathcal{E}_2)}\mathrm{Oxy} \xrightarrow{q_{\mathcal{E}_2}^{(E)}} Y_{AE}^{(\mathcal{E}_2)}\mathrm{Acet}$$

$\mathcal{A}_1$. Oxidative growth on acetate.

$$Y_{AX}^{(\mathcal{A}_1)}\mathrm{Acet} + Y_{OX}^{(\mathcal{A}_1)}\mathrm{Oxy} \xrightarrow{q_{\mathcal{A}_1}^{(X)}} X$$

The culture illustrated in Fig. II.10a is split into four phases, in which different modes are, or can be, activated. In phase I, the fact that ethanol and cells are produced along with glucose consumption indicates that mode $\mathcal{G}_2$ is activated, moreover, the production of acetate indicated that mode $\mathcal{E}_2$ is activated too. The fact that no cell is produced in phase II once glucose is depleted tends to indicate that growth only happens through glucose consumption in phase I. In absence of data about oxygen consumption, the possibility that mode $\mathcal{G}_1$ is activated during this first culture phase cannot be confirmed nor discarded. In phase II, glucose is depleted and no growth is observed but ethanol is converted into acetate, which implies that only mode $\mathcal{E}_2$ is activated. In phase III, acetate keeps being oxidized by mode $\mathcal{E}_2$ and growth is observed but it is not clear whether it should be attributed to either mode $\mathcal{E}_1$, $\mathcal{A}_1$ or a combination of both.

In Fig. II.10b, the anaerobic conditions ensure that modes $\mathcal{G}_1$, $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{A}_1$ are disabled. Therefore, only mode $\mathcal{G}_2$ can explain the observed growth and ethanol production in phase I. The steady conditions in phase II confirm that the metabolism of *S. cerevisiae* cannot process ethanol in the absence of oxygen.

None of these observed phases of culture prove the existence of mode $\mathcal{G}_1$, but it is known that in aerobic continuous cultures of *S. cerevisiae*, and at low dilution rate, growth is possible on glucose by consuming oxygen, and without the production of ethanol or other fermentation by-products [Postma et al., 1989].

Stoichiometry coefficients for each mode are computed from mass balance equations on all species that appear in reactions described in Tab. II.3. These mass balance equations differ from one mode to another depending on which reactions are activated per mode as illustrated in Fig. II.11.

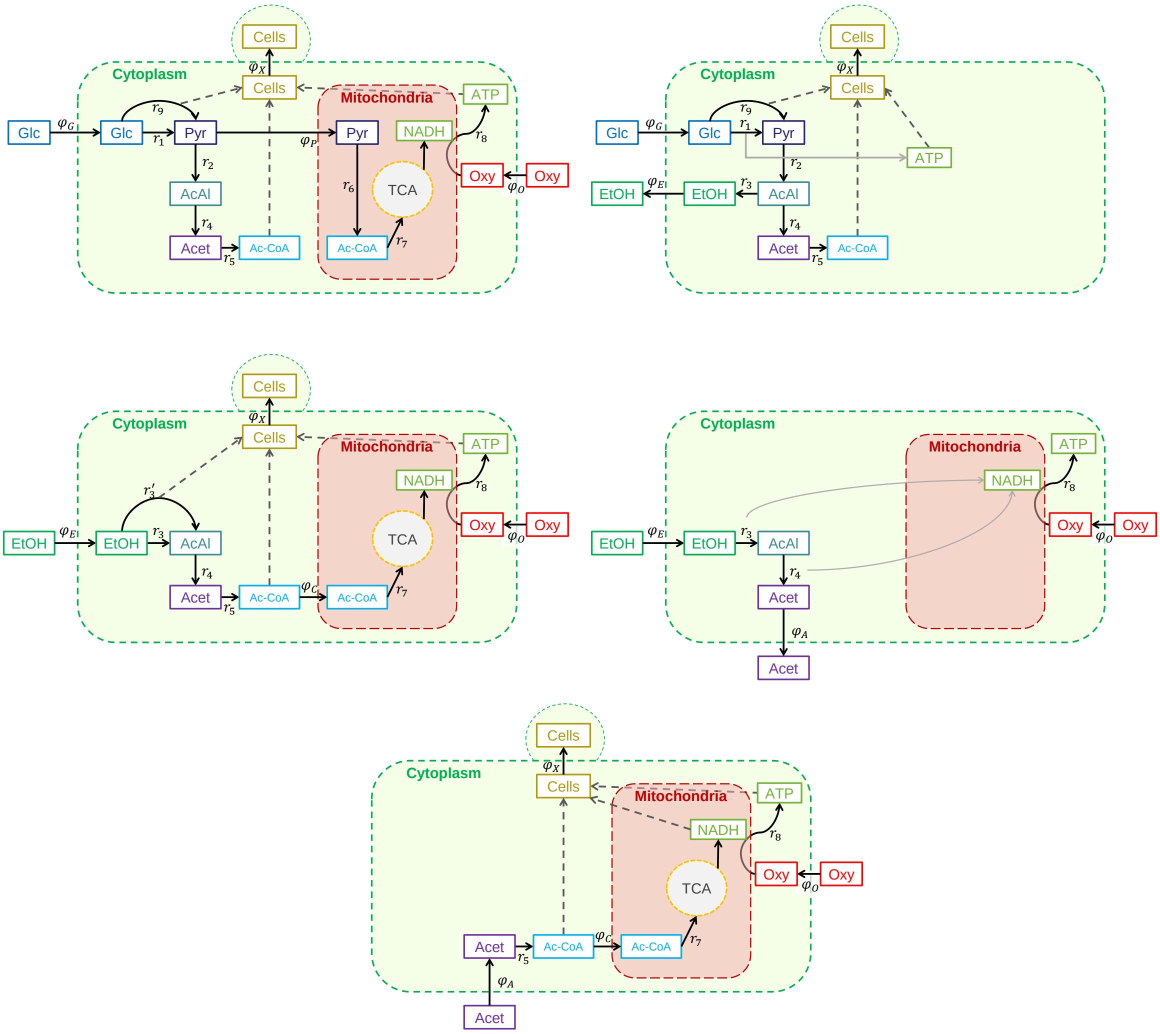


Figure II.11: Elementary modes for *S. cerevisiae* metabolic modelling.

As stoichiometry coefficients for anabolism (reaction $r_{10}$) are undefined, the global stoichiometry of modes will be parametrized by yields $Y_{CX}$, $Y_{EX}$ and $Y_{NX}$. We detail hereafter the computation of stoichiometry parameters for mode $\mathcal{G}_1$ by showing the underlying linear system associated to this mode. The same approach has been applied to all modes but only final results are shown for sake of clarity.

### Stoichiometry of mode $\mathcal{G}_1$

As shown in Fig. II.11, the mode corresponding to oxidative growth on glucose makes use of reactions $r_1$, $r_2$, $r_3$ $r_4$, $r_5$, $r_6$, $r_7$, $r_8$, $r_9$ and $r_{10}$, and requires the transfer of glucose and oxygen from the environment to the cell, and of pyruvate between the cytoplasm and the mitochondria. Therefore, we aim at identifying the rate of each of these reactions and transfers from the overall rate of the mode, $q_{\mathcal{G}_1}^{(X)}$ ($\mathrm{mol_X/g_X \cdot h}$). By assuming that metabolites do not accumulate in cells, and considering stoichiometry of reactions (see II.3), one can formulate mass balances on GLC, ATP, F6P, G3P, NADH, AcAl, Acet, Oxy, NADPH, biomass ($X$), Pyr and AcCoA. Note that these last two metabolites are located both in cytoplasm and in mitochondria. These mass balances are summed up in the following linear system

$$\begin{pmatrix}
-1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -3 & \cdot & -1 & \cdot & \cdot & \cdot \\
-1 & -1 & 2 & \cdot & \cdot & -1 & \cdot & 1 & 5 & \cdot & -Y_{EX}^{(\mathcal{G}_1)} & \cdot & \cdot & \cdot & \cdot \\
1 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 2 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & 1 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot \\
\cdot & \cdot & 1 & \cdot & 1 & \cdot & 1 & 3.6 & -2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & 1 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & -Y_{CX}^{(\mathcal{G}_1)} & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot & \cdot & \cdot & -1 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 6 & -Y_{NX}^{(\mathcal{G}_1)} & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & -1 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1
\end{pmatrix}
\times
\begin{pmatrix}
q_{r_{1a}}^{(\mathcal{G}_1)} \\ q_{r_{1b}}^{(\mathcal{G}_1)} \\ q_{r_{1c}}^{(\mathcal{G}_1)} \\ q_{r_2}^{(\mathcal{G}_1)} \\ q_{r_4}^{(\mathcal{G}_1)} \\ q_{r_5}^{(\mathcal{G}_1)} \\ q_{r_6}^{(\mathcal{G}_1)} \\ q_{r_7}^{(\mathcal{G}_1)} \\ q_{r_8}^{(\mathcal{G}_1)} \\ q_{r_9}^{(\mathcal{G}_1)} \\ q_{r_{10}}^{(\mathcal{G}_1)} \\ f_G^{(\mathcal{G}_1)} \\ f_O^{(\mathcal{G}_1)} \\ f_P^{(\mathcal{G}_1)} \\ f_X^{(\mathcal{G}_1)}
\end{pmatrix}
=
\begin{pmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ q_{\mathcal{G}_1}^{(X)}
\end{pmatrix}$$

Mass transfer rates $f_i$ ($\mathrm{mol_i/g_X \cdot h}$) are defined such that a positive value corresponds to matter leaving the cytoplasm. This linear system has been inverted using MATLAB [2016] symbolic toolbox, which leads to following reaction and mass transfer rates:

$$\begin{pmatrix}
q_{r_{1a}}^{(\mathcal{G}_1)} \\ q_{r_{1b}}^{(\mathcal{G}_1)} \\ q_{r_{1c}}^{(\mathcal{G}_1)} \\ q_{r_2}^{(\mathcal{G}_1)} \\ q_{r_4}^{(\mathcal{G}_1)} \\ q_{r_5}^{(\mathcal{G}_1)} \\ q_{r_6}^{(\mathcal{G}_1)} \\ q_{r_7}^{(\mathcal{G}_1)} \\ q_{r_8}^{(\mathcal{G}_1)} \\ q_{r_9}^{(\mathcal{G}_1)} \\ f_G^{(\mathcal{G}_1)} \\ f_O^{(\mathcal{G}_1)} \\ f_P^{(\mathcal{G}_1)}
\end{pmatrix}
= Y_{EX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)}
\begin{pmatrix}
1/32 \\ 1/32 \\ 1/16 \\ 0 \\ 0 \\ 0 \\ 1/16 \\ 1/16 \\ 7/40 \\ 0 \\ -1/32 \\ -7/40 \\ 1/16
\end{pmatrix}
+ Y_{CX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)}
\begin{pmatrix}
11/32 \\ 11/32 \\ 11/16 \\ 1 \\ 1 \\ 1 \\ -5/16 \\ -5/16 \\ 1/8 \\ 0 \\ -11/32 \\ -1/8 \\ -5/16
\end{pmatrix}
+ Y_{NX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)}
\begin{pmatrix}
-83/192 \\ -19/192 \\ -1/32 \\ 0 \\ 0 \\ 0 \\ -1/32 \\ -1/32 \\ -7/80 \\ 1/6 \\ -13/192 \\ 7/80 \\ -1/32
\end{pmatrix}$$

and by definition $q_{r_{10}}^{(\mathcal{G}_1)} = f_X^{(\mathcal{G}_1)} = q_{\mathcal{G}_1}^{(X)}$. Overall yields of this modes are simply given by $Y_{GX}^{(\mathcal{G}_1)} = -f_G^{(\mathcal{G}_1)}/f_X^{(\mathcal{G}_1)}$ and $Y_{OX}^{(\mathcal{G}_1)} = f_O^{(\mathcal{G}_1)}/f_X^{(\mathcal{G}_1)}$

The model will describe limits over some of these reaction and mass transfer rates and these stoichiometry coefficients will allow determining the limiting factor for mode $\mathcal{G}_1$. The same approach, based on linear systems, has been applied to determine the stoichiometry of other modes.

### Stoichiometry of mode $\mathcal{G}_2$

The mode $\mathcal{G}_2$ which describes fermentative growth on glucose activates reactions $r_1$, $r_2$, $r_3$, $r_4$, $r_5$, $r_9$ and $r_{10}$. It consumes a single substrate, glucose, and produces ethanol as a fermentation by-product. Rates of these reactions and mass transfers can be expressed from the rate of cell production by this mode, $q_{\mathcal{G}_2}^{(X)}$ ($\mathrm{mol_X/g_X \cdot h}$). The procedure is identical to the one used for mode $\mathcal{G}_1$ and considers mass balances on GLC, ATP, F6P, G3P AcAl, Acet, NADPH, biomass ($X$), Pyr, EtOH and AcCoA. Note that a mass balance on NADH would be a linear combination of mass balances on Pyr and AcAl which explains why it is not considered. Overall, we obtain following reaction and transfer rates:

$$\begin{pmatrix} q_{r_{1a}}^{(\mathcal{G}_2)} \\ q_{r_{1b}}^{(\mathcal{G}_2)} \\ q_{r_{1c}}^{(\mathcal{G}_2)} \\ q_{r_2}^{(\mathcal{G}_2)} \\ q_{r_3}^{(\mathcal{G}_2)} \\ q_{r_4}^{(\mathcal{G}_2)} \\ q_{r_5}^{(\mathcal{G}_2)} \\ q_{r_9}^{(\mathcal{G}_2)} \\ f_G^{(\mathcal{G}_2)} \\ f_E^{(\mathcal{G}_2)} \end{pmatrix} = Y_{EX}^{(\mathcal{G}_2)} q_{\mathcal{G}_2}^{(X)} \begin{pmatrix} 1/2 \\ 1/2 \\ 1 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ -1/2 \\ -1 \end{pmatrix} + Y_{CX}^{(\mathcal{G}_2)} q_{\mathcal{G}_2}^{(X)} \begin{pmatrix} 1/2 \\ 1/2 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ -1/2 \\ 0 \end{pmatrix} + Y_{NX}^{(\mathcal{G}_2)} q_{\mathcal{G}_2}^{(X)} \begin{pmatrix} -2/3 \\ -1/3 \\ -1/2 \\ -1/2 \\ 1/2 \\ 0 \\ 0 \\ 1/6 \\ -1/6 \\ 1/2 \end{pmatrix}$$

and $q_{r_{10}}^{(\mathcal{G}_2)} = f_X^{(\mathcal{G}_2)} = q_{\mathcal{G}_2}^{(X)}$.

### Stoichiometry of mode $\mathcal{E}_1$

Mode $\mathcal{E}_1$ describes the oxidative growth on ethanol. It requires reactions $r_3$, $r_3'$, $r_4$, $r_5$, $r_7$, $r_8$, $r_{10}$, consumes both ethanol and oxygen, and AcCoA produced in cytosol is transferred to mitochondria to enter the TriCarboxylic Acid Cycle. By formulating mass balances as previously, we obtain following rates parametrized by the mode rate $q_{\mathcal{E}_1}^{(X)}$ ($\mathrm{mol_X/g_X \cdot h}$):

$$\begin{pmatrix} q_{r_3}^{(\mathcal{E}_1)} \\ q_{r_3'}^{(\mathcal{E}_1)} \\ q_{r_4}^{(\mathcal{E}_1)} \\ q_{r_5}^{(\mathcal{E}_1)} \\ q_{r_7}^{(\mathcal{E}_1)} \\ q_{r_8}^{(\mathcal{E}_1)} \\ f_E^{(\mathcal{E}_1)} \\ f_O^{(\mathcal{E}_1)} \\ f_C^{(\mathcal{E}_1)} \end{pmatrix} = Y_{EX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} \begin{pmatrix} 1/14 \\ 0 \\ 1/14 \\ 1/14 \\ 1/14 \\ 1/5 \\ -1/14 \\ -1/5 \\ -1/14 \end{pmatrix} + Y_{CX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} \begin{pmatrix} 5/7 \\ 0 \\ 5/7 \\ 5/7 \\ -2/7 \\ 1/5 \\ -5/7 \\ -1/5 \\ 2/7 \end{pmatrix} + Y_{NX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} \begin{pmatrix} -23/28 \\ 1 \\ 5/28 \\ 5/28 \\ 5/28 \\ 0 \\ -5/28 \\ 0 \\ -5/28 \end{pmatrix}$$

and $q_{r_{10}}^{(\mathcal{E}_1)} = f_X^{(\mathcal{E}_1)} = q_{\mathcal{E}_1}^{(X)}$.

### Stoichiometry of mode $\mathcal{E}_2$

Mode $\mathcal{E}_2$ is the only one which does not describe a growth phenomenon but the oxidation of ethanol into acetate. It is based on reactions $r_3$, $r_4$ and $r_8$ which produce acetate from ethanol and oxygen. Rates associated to this mode, parametrized by the mode rate $q^{(E)}_{\mathcal{E}_2}$ ($\mathrm{mol_E/g_X \cdot h}$), are simply:

$$\begin{pmatrix} q^{(\mathcal{E}_2)}_{r_3} \\ q^{(\mathcal{E}_2)}_{r_4} \\ q^{(\mathcal{E}_2)}_{r_8} \\ f^{(\mathcal{E}_2)}_{A} \\ f^{(\mathcal{E}_2)}_{O} \end{pmatrix} = q^{(E)}_{\mathcal{E}_2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

and $f^{(\mathcal{E}_2)}_{E} = q^{(E)}_{\mathcal{E}_2}$.

### Stoichiometry of mode $\mathcal{A}_1$

The mode $\mathcal{A}_1$ describes the oxidative growth on acetate. It is based on reactions $r_5$, $r_7$, $r_8$ and $r_{10}$, it consumes acetate and oxygen and, similarly to mode $\mathcal{E}_1$, transfers AcCoA from cytosol to mitochondria. Rate associated to this mode are parametrized by the mode rate $q^{(X)}_{\mathcal{A}_1}$ ($\mathrm{mol_X/g_X \cdot h}$) as following:

$$\begin{pmatrix} q^{(\mathcal{A}_1)}_{r_5} \\ q^{(\mathcal{A}_1)}_{r_7} \\ q^{(\mathcal{A}_1)}_{r_8} \\ f^{(\mathcal{A}_1)}_{A} \\ f^{(\mathcal{A}_1)}_{O} \\ f^{(\mathcal{A}_1)}_{C} \end{pmatrix} = Y^{(\mathcal{A}_1)}_{EX} q^{(X)}_{\mathcal{A}_1} \begin{pmatrix} 1/9 \\ 1/9 \\ 1/5 \\ -1/9 \\ -1/5 \\ -1/9 \end{pmatrix} + Y^{(\mathcal{A}_1)}_{CX} q^{(X)}_{\mathcal{A}_1} \begin{pmatrix} 10/9 \\ 1/9 \\ 1/5 \\ -10/9 \\ -1/5 \\ -1/9 \end{pmatrix} + Y^{(\mathcal{A}_1)}_{NX} q^{(X)}_{\mathcal{A}_1} \begin{pmatrix} 5/18 \\ 5/18 \\ 0 \\ -5/18 \\ 0 \\ -5/18 \end{pmatrix}$$

and $q^{(\mathcal{A}_1)}_{r_{10}} = f^{(\mathcal{A}_1)}_{X} = q^{(X)}_{\mathcal{A}_1}$.

### Fitting parameters for metabolic modes

In order for the model to quantitatively fit experimental data, parameters $Y_{EX}$, $Y_{CX}$ and $Y_{NX}$ will be adjusted. However, actual pathways underlying the anabolism reaction ($r_{10}$) may differ depending on used substrates or sources of energy. Therefore, one can expect that conversion yields $Y_{EX}$, $Y_{CX}$ and $Y_{NX}$ may slightly differ from one mode to an other. Consequently, the fitting of experimental data will be done by allowing a small variance for these parameters between modes.

## II.4.4 Cell dynamics

### Growth rate dynamics on multiple substrates

Elementary modes forming the metabolic network allow for growth on three different carbonaceous substrates: glucose, ethanol and acetate. Similarly to what has been considered in the model developed in Section II.3.2 for *E. coli*, Young et al. [1970] refer to observed dynamics on growth rate adaptation to environmental changes in cultures of

*S. cerevisiae.* Therefore, we will attach three biological properties to cells to account for their growth capability on each substrate:

- $\mu_{G@p}$: growth capability on glucose ($\mathrm{g_X/g_X \cdot h}$);
- $\mu_{E@p}$: growth capability on ethanol ($\mathrm{g_X/g_X \cdot h}$);
- $\mu_{A@p}$: growth capability on acetate ($\mathrm{g_X/g_X \cdot h}$).

We model the adaptation rate of all three growth rate capabilities with a similar adaptation law:

$$\frac{\partial \mu_{k@p}}{\partial t} = \zeta_{\mu_k}\left(\mu_{k@p}, \boldsymbol{C}_L\right) = \frac{1}{T_{\mu_k}}\left(\mu_k^*(\boldsymbol{C}_L) - \mu_{k@p}\right), \qquad k \in \{G, E, A\} \tag{II.13}$$

This first-order adaptation law was chosen, instead of the second-order law used for *E. coli* (see Eq. (II.4)), to limit the number of parameters in the model, and facilitate parameter identification. It will be shown in Section II.5, that this law allows for a satisfactory enough fitting of the model against experimental data. These adaptation laws still require the definition of an equilibrium growth rate for each substrate. We defined these, using Monod-like kinetic laws, by considering that growth on glucose is slightly inhibited by both ethanol and acetate, and that growth based on these by-products is strongly inhibited by glucose which is the preferential carbon source. Equilibrium laws are then given by

$$\begin{aligned}
\mu_G^*(\boldsymbol{C}_L) &= \mu_G^{(\max)} \frac{G}{G + K_G} \frac{K_{i,E}^{(G)}}{E + K_{i,E}^{(G)}} \frac{K_{i,A}^{(G)}}{A + K_{i,A}^{(G)}} \\
\mu_E^*(\boldsymbol{C}_L) &= \mu_E^{(\max)} \frac{E}{E + K_E} \frac{K_{i,G}^{(G)}}{G + K_{i,G}^{(G)}} \\
\mu_A^*(\boldsymbol{C}_L) &= \mu_A^{(\max)} \frac{A}{A + K_A} \frac{K_{i,G}^{(A)}}{G + K_{i,G}^{(A)}} \frac{K_{i,E}^{(A)}}{E + K_{i,E}^{(A)}}
\end{aligned}$$

Time-scales associated to the adaptation of each growth rate, $T_{\mu_G}$, $T_{\mu_E}$ and $T_{\mu_A}$, will have a similar order of magnitude than the time-scale used in the *E. coli* model, about an hour. These time-scales are allowed to slightly differ between substrates.

### Crabtree effect: adapting the oxidative capacity

Once growth-rate dynamics are tackled, the main dynamic effect to be considered about *S. cerevisiae* is the Crabtree effect. As already stated, this effect consists in the production of ethanol, which is a fermentative pathway by-product, under aerobic conditions. Pronk et al. [1996] actually differentiate two phenomenon, the long-term Crabtree effect and the short-term Crabtree effect, each having distinct origins.

The long-term effect corresponds to the production of ethanol in culture at high growth-rate, no matter the residual glucose concentration. It can be observed by tracking oxygen consumption, carbon dioxide production, and sugar to cell conversion yield against increasing dilution rates in chemostat cultures as shown in Fig. II.12.

At low growth rates ($\mu < 0.3\mathrm{h}^{-1}$), oxygen consumption increases almost linearly with the growth-rate. The respiratory quotient (molar ratio between carbon dioxide production and oxygen consumption) is near 1.0 which denotes an exclusive use of oxidative

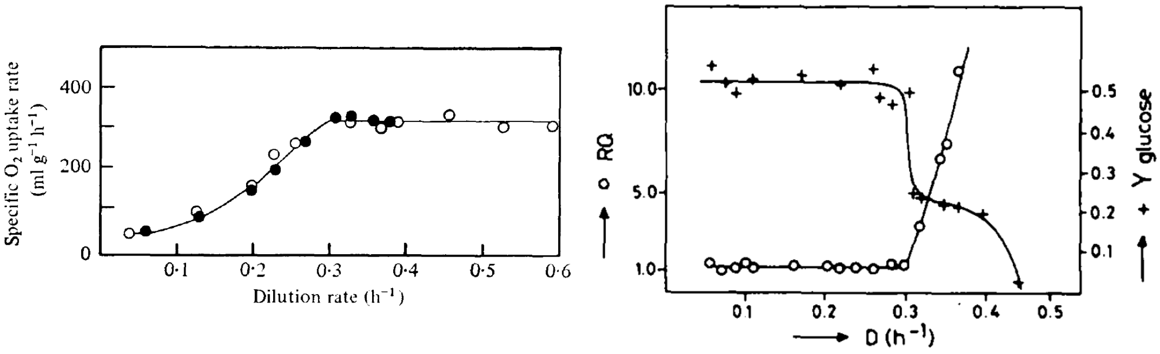


Figure II.12: Illustration of long-term Crabtree effect in chemostat cultures. (left) Specific oxygen consumption rate. Reproduced with permission of Microbiology Society via Copyright Clearance Center [Barford and Hall, 1979]. (right) Respiratory quotient and conversion yield of glucose [van Dijken and Scheffers, 1986].

pathways. Once the dilution –and growth– rates reach the critical value of $0.3\text{h}^{-1}$, oxygen consumption reaches a plateau, and the respiratory quotient starts increasing linearly with the growth-rate. This RQ increase can be attributed to the use of fermentative pathways. This also explains the associated drop in glucose conversion yield, as fermentative pathways are less efficient than oxidative ones. Similar results have been observed by Petrik et al. [1983] and Rieger et al. [1983].

Overall, the long-term Crabtree effect can be attributed to a limiting oxidative capacity which cannot exceed the capacity required to grow at a critical rate $\mu_{\text{crit}} \approx 0.3\text{h}^{-1}$. The steady-state data shown in Fig. II.12 are however not sufficient to determine whether the oxidative capacity is a dynamic property of cells. Indeed, the capacity could always be at its peak value, and would be under-solicited at low growth-rate. Another hypothesis would be that cells dynamically adapt their capacity at low growth-rates.

It has been assumed that the oxidative capacity is indeed a dynamic property of cells. While growth-rate dynamics denotes the adaptation of cells to varying substrate availability, the adaptation of oxidative capacity will denote the adaptation of cells to varying energy production requirements. This assumption is in-line with the modelling of Scerevisiae by Sweere et al. [1988] where a dynamic oxidative capacity was considered too.

Assessing whether this assumption is correct is difficult due to a lack of experimental evidence. Indeed, due to the short-term Crabtree effect which prevents an efficient use of oxidative pathways, one cannot observe whether the oxidative capacity is a limiting factor after a sudden increase of dilution-rate in chemostat culture. A possibility would be to perform anaerobic chemostat culture at moderate growth-rate and to observe whether switching to aerobic conditions allows an instantaneous use of oxygen resources.

For now, we will describe the evolution of a cell-attached property $\varphi^{(b)}_{O@p}$ ($\text{g}_O/\text{g}_X \cdot \text{h}$) which is the maximum amount of oxygen that can be processed by cells in the respiratory chain to produce energy. This variable will evolve toward an equilibrium capacity which depends on the equilibrium growth-rate associated to the environment. This equilibrium capacity is modelled against experimental data illustrated in Fig. II.13 by

$$\varphi^*_O = \begin{cases} \varphi^{(\max)}_O \dfrac{\mu^*_G}{\mu^{O_{\lim}}_G} & \text{if } \mu^*_G \leq \mu^{O_{\lim}}_G \\ \varphi^{(\max)}_O & \text{otherwise} \end{cases} \tag{II.14}$$

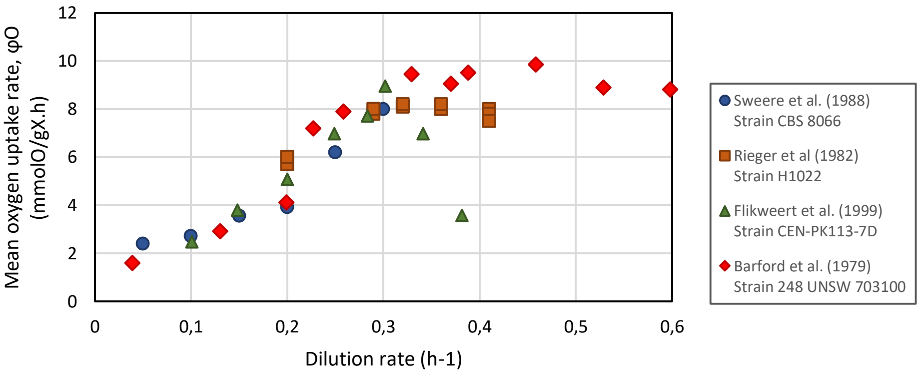


Figure II.13: Experimentally measured specific oxygen consumption rate in chemostat cultures at equilibrium.

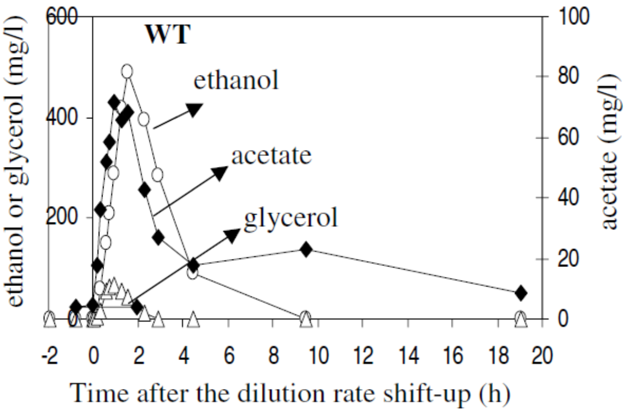


Figure II.14: Concentrations of ethanol (○), acetate (♦) and glycerol (△) after a dilution-rate shift-up (0.05 to 0.15 $\text{h}^{-1}$) in a glucose-fed chemostat culture of *S. cerevisiae*. Reproduced by permission of Oxford University Press [Guillou et al., 2004].

with $\mu_G^{O\text{lim}}$ the minimal growth-rate at which the long-term Crabtree effect occurs and $\varphi_O^{(\text{max})}$ the corresponding oxidative capacity. The cell capacity $\varphi_{O@p}^{(b)}$ will evolve toward this equilibrium value at the following rate:

$$\frac{\partial \varphi_{O@p}^{(b)}}{\partial t} = \zeta_{\varphi_O} = \frac{1}{T_{\varphi_O}} \left( \varphi_O^* - \varphi_{O@p}^{(b)} \right) \tag{II.15}$$

Sweere et al. [1988] also modelled *S. cerevisiae* by introducing a dynamic oxidative capacity. This property evolved with a time-scale of 1.8h.

On top of the long-term Crabtree effect, which is simply described by the observed plateau in specific oxygen consumption in Fig. II.12, the critical dynamic behaviour for *S. cerevisiae* modelling is the short-term Crabtree effect. This effect corresponds to the production of ethanol by *S. cerevisiae*, in aerobic cultures, after a sudden exposition to high glucose concentrations. This is illustrated in Fig. II.14 where ethanol is produced in a chemostat culture after a dilution-rate shift-up.

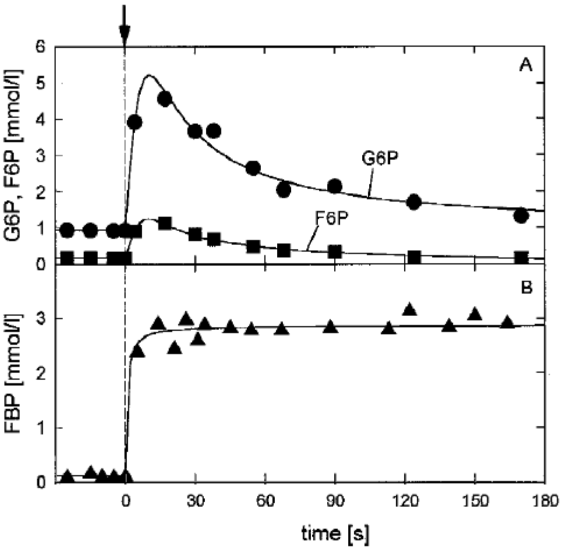


Figure II.15: Intracellular concentration in Glucose-6-Phosphate (G6P), Fructose-6-Phosphate (F6P) and Fructose-1,6-BiPhosphate (FBP), after a glucose pulse in chemostat culture of *S. cerevisiae*. Reproduced by permission of John Wiley and Sons [Theobald et al., 1997].

The short-term Crabtree effect is usually attributed to a saturation of the respiratory chain, which induces an overflow at the pyruvate node in the metabolic network, and results in a diversion of carbon toward ethanol production [Pronk et al., 1996]. More recently, Rosas-Lemus et al. [2014] isolated mitochondria from both Crabtree-positive and Crabtree-negative yeasts and could measure their activity when exposed to different conditions. In particular, they observed that one metabolite of the glycolysis pathway, the Fructose-1,6-BiPhosphate (FBP), closes mitochondrial unspecific channels which are responsible for most mass-transfers between cytoplasm and mitochondria.

Theobald et al. [1997] measured intracellular concentrations of multiple metabolites, including FBP, in *S. cerevisiae* right after a glucose pulse in chemostat culture. Their most interesting result to understand the short-term Crabtree effect is shown in Fig. II.15 where one can observe that after the pulse, glucose is assimilated and consequently converted into $G6P$, $F6P$ and then FBP. As long as the environment remains rich in glucose, the intracellular concentration of FBP remains at high levels, which shows that the conversion of FBP by the fructose-bisphosphate aldolase is a limiting reaction in the glycolysis pathway.

Coupling measurements of intracellular concentrations after a glucose pulse [Theobald et al., 1997] with the observation of isolated mitochondria by Rosas-Lemus et al. [2014] allows inferring a probable cause for the short-term Crabtree effect: if a cell is suddenly exposed to high glucose concentrations, this substrate will enter the cytoplasm but the cellular machinery will not be sufficient to process efficiently this high substrate quantity. Therefore, FBP will accumulate which will close mitochondria unspecific channels and prevent part of the pyruvate to enter the TriCarboxylic Acid cycle which reduces the energetic efficiency of the cell by under-using the respiratory chain. As pyruvate accumulates, and the use of oxidative pathways is diminished, some of that pyruvate is diverted toward the fermentative pathway which explains the ethanol production.

Overall, this short-term Crabtree dynamics can be described by considering that (i) in glucose-rich environments, some metabolites accumulate quickly –within a few seconds (Fig. II.15)– in cells which reduces the use of oxidative metabolic modes and (ii) in environments with non-excessive glucose concentrations, these metabolites are processed and the respiratory capacity can be fully exploited once again. The time-scale associated to the return to standard cell functioning is not clearly identifiable from previous results and will be a fitting parameters of the model.

Instead of quantitatively track intracellular concentrations to determine whether the short-term Crabtree effect occurs, we chose to track an adimensional quantity, $\sigma_{@p} \in [0,1]$ which we refer to as the glycolytic stress. If $\sigma_{@p} = 0$, cells are unstressed and the oxidative capacity can be fully exploited. Otherwise, oxidative modes will be inhibited.

Stress evolution is modelled by the following law:

$$\frac{\partial \sigma_{@p}}{\partial t} = \zeta_\sigma = \frac{\kappa_\sigma}{T_\sigma^{(u)}} (1 - \sigma) - \frac{(1 - \kappa_\sigma)}{T_\sigma^{(d)}} \sigma \tag{II.16}$$

$\kappa_\sigma$ is a variable close to 1 in stressful environment and close to 0 in unstressing environments. Therefore, the first term on the Right-Hand Side (RHS) describes the stress increase with a time-scale $T_\sigma^{(u)} \approx 1\text{s}$ and the second term describe stress recovery over a longer time-scale $T_\sigma^{(d)} \approx 1\text{min}$. It was decided to characterise the stressfulness of an environment by comparing the difference between a cell's growth capacity on glucose $\mu_{G@p}$ with the equilibrium growth-rate $\mu_G^*(\boldsymbol{C}_L)$ against a threshold value $\delta\mu_\sigma$:

$$\kappa_\sigma = \left( \frac{1}{1 + \exp\left(-50 \left(\mu_G^* - \mu_{G@p} - \delta\mu_\sigma\right)\right)} \right)^2 \approx \begin{cases} 0 & \text{if } \mu_G^* - \mu_{G@p} < \delta\mu_\sigma \\ 1 & \text{otherwise} \end{cases} \tag{II.17}$$

$\mu_{G@p}$ is here taken as an estimate of a cell's capability to process glucose, and $\mu_G^*$ a rough estimate of the glucose uptake rate. Therefore, the difference $\mu_G^* - \mu_{G@p}$ is a measure of the disequilibrium between a cell and its environments. The introduced threshold value $\delta\mu_\sigma$ reflects the fact that small scale disequilibriums are indeed dumped by minor metabolic adjustments whereas larger disequilibrium are stressful and will turn into major metabolic perturbations.

Observing conversion yields and respiration rates under active Crabtree-effect, we could notice that oxidative modes are not entirely inhibited by this effect. Overall, after a significant stressing event (glucose pulse, dilution rate shift-up), respiration is reduced by approximately 60% compared to the oxygen consumption rate preceding the event. Therefore, we compute an inhibition factor $k_{i,\sigma}$ by

$$k_{i,\sigma} = \frac{k_{i,\sigma}^{(\min)}}{k_{i,\sigma}^{(\min)} + \left(1 - k_{i,\sigma}^{(\min)}\right) \sigma_{@p}} \tag{II.18}$$

with $k_{i,\sigma}^{(\min)} = 0.4$. The computation of rates associated to each mode will still be based on the identification of the limiting factor, but afterwards, the rate of oxidative modes will be multiplied by $k_{i,\sigma}$. Doing so, if a cell is fully stressed ($\sigma_{@p} \approx 1$), oxidative modes will be slowed down by 60%. Otherwise, these mode rates will be left almost unchanged.

### II.4.5 Mode limitations

Similarly to the *E. coli* model developed in Section II.3, the rates of metabolic modes in the *S. cerevisiae* model will be computed by defining available resources which, once depleted, will limit metabolic mode rates. We recall here that generally speaking, the uptake rates $\varphi^{(g)}$ ($\mathrm{g}/\mathrm{g_X}\cdot\mathrm{h}$) is either limited by the availability in the environment $\varphi^{(e)}$ or by the cell capability to import the substrates $\varphi^{(b)}$. None of these limits is actually constant in industrial bioprocessed but in this section, it will be assumed the $\varphi^{(g)} = \varphi^{(b)}$, *i.e.* that the environment is not limiting. The modelling of $\varphi^{(e)}$ and its combination with $\varphi^{(b)}$ to compute $\varphi^{(g)}$ is an aspect that will be developed in Chapter V.

Biological limits on carbonaceous substrate uptake-rates will simply be defined as biological constants, $\varphi^{(b)}_{\{G,E,A\}} = \varphi^{(\max)}_{\{G,E,A\}}$, which will be fitting parameters for the model. Limits on oxygen consumption rate is defined by the cell-attached biological property $\varphi^{(g)}_{O} = \varphi^{(b)}_{O@p}$.

Cell-attached growth capabilities $\mu_{G@p}$, $\mu_{E@p}$ and $\mu_{A@p}$ form limits on cell production rates for growth modes. It is known that fermentative growth is slower than oxidative growth on glucose, $\mu_G^{(\max,o)} > \mu_G^{(\max,f)}$. In order to account for this, while only attaching a single glucose-based growth capability $\mu_{G@p}$ to cells, the maximum growth rate on glucose through fermentation is set to a fraction of the maximum value for oxidative growth. This aspect will clearly appear hereafter in the computation of the rate for mode $\mathcal{G}_2$. Note that this approach makes the implicit assumption that a cell fully adapted to oxidative growth on glucose will be instantaneously able to grow a full efficiency in a fermentative mode is oxygen is suddenly depleted. This is not an accurate description of the switch between oxidative and fermentative growths, but it allows limiting the number of cell-attached dynamic properties.

### II.4.6 Mode rates computation

Available resources available for *S. cerevisiae* metabolic modes are:

- Consumption/uptake rates
  - of glucose: $\varphi_G^{(g)}$ ($\mathrm{kg_G}/\mathrm{kg_X}\cdot\mathrm{h}$) or $f_G^{(g)} = \varphi_G^{(g)}/M_G$ ($\mathrm{mol_G}/\mathrm{kg_X}\cdot\mathrm{h}$);
  - of ethanol: $\varphi_E^{(g)}$ ($\mathrm{kg_E}/\mathrm{kg_X}\cdot\mathrm{h}$) or $f_E^{(g)} = \varphi_E^{(g)}/M_E$ ($\mathrm{mol_E}/\mathrm{kg_X}\cdot\mathrm{h}$);
  - of acetate: $\varphi_A^{(g)}$ ($\mathrm{kg_A}/\mathrm{kg_X}\cdot\mathrm{h}$) or $f_A^{(g)} = \varphi_A^{(g)}/M_A$ ($\mathrm{mol_A}/\mathrm{kg_X}\cdot\mathrm{h}$);
  - of oxygen: $\varphi_O^{(g)}$ ($\mathrm{kg_O}/\mathrm{kg_X}\cdot\mathrm{h}$) or $f_O^{(g)} = \varphi_O^{(g)}/M_O$ ($\mathrm{mol_O}/\mathrm{kg_X}\cdot\mathrm{h}$).
- Growth capabilities
  - on glucose $\mu_G^{(g)} = \mu_{G@p}$ ($\mathrm{kg_X}/\mathrm{kg_X}\cdot\mathrm{h}$);
  - on ethanol $\mu_E^{(g)} = \mu_{E@p}$ ($\mathrm{kg_X}/\mathrm{kg_X}\cdot\mathrm{h}$);
  - on acetate $\mu_A^{(g)} = \mu_{A@p}$ ($\mathrm{kg_X}/\mathrm{kg_X}\cdot\mathrm{h}$);

Stoichiometry of modes has been defined previously (see Section II.4.3) and here, we will only make use of global conversion yields associated to each mode. Note however that these yields actually depend on stoichiometric coefficients associated to the anabolism reaction which will be fitting parameters in Section II.5.

Steps for the computation of mode rates are

1. Compute the inhibition factor of oxidative modes by glycolytic stress $k_{i,\sigma}$.

2. Compute mode $\mathcal{G}_1$ rate considering inhibition by stress and limits in oxygen and glucose uptake rates ($f_O^{(g)}$ and $f_G^{(g)}$) and glucose-based growth capacity $\mu_G^{(g)}$,

$$q_{\mathcal{G}_1}^{(X)} = k_{i,\sigma} \min \left( \frac{f_O^{(g)}}{Y_{OX}^{(\mathcal{G}_1)}}; \frac{f_G^{(g)}}{Y_{GX}^{(\mathcal{G}_1)}}; \frac{\mu_G^{(g)}}{M_X} \right)$$

3. Compute remaining resources for modes $\mathcal{G}_2$, $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{A}_1$,

$$\begin{aligned} f_O^{(g)} &\leftarrow f_O^{(g)} - Y_{OX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)} \\ f_G^{(g)} &\leftarrow f_G^{(g)} - Y_{GX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)} \\ \mu_G^{(g)} &\leftarrow \mu_G^{(g)} - M_X q_{\mathcal{G}_1}^{(X)} \end{aligned}$$

4. Compute mode $\mathcal{G}_2$ rate considering limits in glucose uptake rate $f_G^{(g)}$ and glucose-base growth capacity $\mu_G^{(g)}$ corrected for fermentative growth,

$$q_{\mathcal{G}_2}^{(X)} = \min \left( \frac{f_G^{(g)}}{Y_{GX}^{(\mathcal{G}_2)}}; \frac{\mu_G^{(\max,f)}}{\mu_G^{(\max,o)}} \frac{\mu_G^{(g)}}{M_X} \right)$$

5. Compute mode $\mathcal{E}_1$ rate considering the stress level, upper bounds on uptake rates for oxygen and ethanol ($f_O^{(g)}$ and $f_E^{(g)}$) and the limited growth capacity on ethanol $\mu_E^{(g)}$,

$$q_{\mathcal{E}_1}^{(X)} = k_{i,\sigma} \min \left( \frac{f_O^{(g)}}{Y_{OX}^{(\mathcal{E}_1)}}; \frac{f_E^{(g)}}{Y_{EtX}^{(\mathcal{E}_1)}}; \frac{\mu_E^{(g)}}{M_X} \right)$$

6. Compute resources available for modes $\mathcal{E}_2$ and $\mathcal{A}_1$,

$$\begin{aligned} f_O^{(g)} &\leftarrow f_O^{(g)} - Y_{OX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} \\ f_E^{(g)} &\leftarrow f_E^{(g)} - Y_{EX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} \end{aligned}$$

7. Compute mode $\mathcal{E}_2$ rate considering limits in oxygen and ethanol uptake rates ($f_O^{(g)}$ and $f_E^{(g)}$),

$$q_{\mathcal{E}_2}^{(E)} = \min \left( \frac{f_O^{(g)}}{Y_{OE}^{(\mathcal{E}_2)}}; f_E^{(g)} \right)$$

8. Compute resources available for mode $\mathcal{A}_1$,

$$f_O^{(g)} \leftarrow f_O^{(g)} - Y_{OE}^{(\mathcal{E}_2)} q_{\mathcal{E}_2}^{(E)}$$

9. Compute mode $\mathcal{A}_1$ rate considering the stress level, upper bounds on uptakes rates for oxygen and acetate ($f_O^{(g)}$ and $f_A^{(g)}$) and the limited growth capacity on acetate $\mu_A^{(g)}$,

$$q_{\mathcal{A}_1}^{(X)} = k_{i,\sigma} \min\left(\frac{f_O^{(g)}}{Y_{OX}^{(\mathcal{A}_1)}}; \frac{f_A^{(g)}}{Y_{AX}^{(\mathcal{A}_1)}}; \frac{\mu_A^{(g)}}{M_X}\right)$$

Once mode rates are computed through this straightforward procedure, actual consumption/production and growth rates are computed by

$$\begin{aligned}
\mu^{(a)} &= \left(q_{\mathcal{G}_1}^{(X)} + q_{\mathcal{G}_2}^{(X)} + q_{\mathcal{E}_1}^{(X)} + q_{\mathcal{A}_1}^{(X)}\right) M_X \\
\varphi_G^{(a)} &= -\left(Y_{GX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)} + Y_{GX}^{(\mathcal{G}_2)} q_{\mathcal{G}_2}^{(X)}\right) M_G \\
\varphi_E^{(a)} &= \left(Y_{EX}^{(\mathcal{G}_2)} q_{\mathcal{G}_2}^{(X)} - Y_{EX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} - q_{\mathcal{E}_2}^{(E)}\right) M_E \\
\varphi_A^{(a)} &= \left(Y_{AE}^{(\mathcal{E}_2)} q_{\mathcal{E}_2}^{(E)} - Y_{AX}^{(\mathcal{A}_1)} q_{\mathcal{A}_1}^{(X)}\right) M_A \\
\varphi_O^{(a)} &= -\left(Y_{OX}^{(\mathcal{G}_1)} q_{\mathcal{G}_1}^{(X)} + Y_{OX}^{(\mathcal{E}_1)} q_{\mathcal{E}_1}^{(X)} + Y_{OX}^{(\mathcal{A}_1)} q_{\mathcal{A}_1}^{(X)}\right) M_G
\end{aligned}$$

## II.5 Challenging the model developed for *S. cerevisiae*

### II.5.1 Comparison data

The model described previously is meant to reproduce the behaviour of a wild *S. cerevisiae* strain in steady environment but also in dynamic situations. Therefore, we attempted to fit the model against multiple experimental data sets coming from different culture modes. To do so, we extracted data points from the literature for experiments using different wild strains of *S. cerevisiae*. A summary of these data sets is given in Table II.4.

Table II.4: References for data sets used to fit the *S. cerevisiae* model.

| Source | Strain | Culture mode |
|---|---|---|
| Flikweert et al. [1999] | CEN-PK113-7D | Steady-state chemostat<br>Glucose pulse in chemostat |
| Barford and Hall [1979] | 248 UNSW703100 | Steady-state chemostat |
| van Dijken and Scheffers [1986] | CBS 8066 | Aerobic batch cultures<br>Anaerobic batch culture |
| Sweere et al. [1988] | CBS 8066 | Glucose pulse in chemostat |

### II.5.2 Model fitting strategy

For now, the model has only been described qualitatively and only order of magnitudes were given for the model parameters. All model parameters were fitted manually against previously listed data sets. This was done by iterating over two fitting steps:

1. Fit parameters that only have an impact on steady-state behaviour (conversion yields, affinity/inhibition constants, maximum kinetic rates, ...) against steady-state chemostat data.

2. Fit parameters that have an impact on dynamic behaviour (adaptation rates of biological properties, order of magnitude of short-term Crabtree effect, . . . ).

Final model parameters are listed in Table II.5.

Table II.5: List of fitted parameters for *S. cerevisiae* model.

| Kinetic parameters | |
|---|---|
| Symbol | Value |
| $\mu_G^{(\max,o)}$ | $0.49\ \mathrm{h}^{-1}$ |
| $\mu_G^{(\max,f)}$ | $0.37\ \mathrm{h}^{-1}$ |
| $\mu_E^{(\max)}$ | $0.12\ \mathrm{h}^{-1}$ |
| $\mu_A^{(\max)}$ | $0.20\ \mathrm{h}^{-1}$ |
| $\varphi_O^{(\max)}$ | $8.90\ \mathrm{mol_O/kg_X \cdot h}$ |
| $\varphi_G^{(\max)}$ | $19.86\ \mathrm{mol_G/kg_X \cdot h}$ |
| $\varphi_E^{(\max)}$ | $3.87\ \mathrm{mol_E/kg_X \cdot h}$ |
| $\varphi_A^{(\max)}$ | $4.13\ \mathrm{mol_A/kg_X \cdot h}$ |

| Affinity/Inhibition constants | |
|---|---|
| Symbol | Value |
| $K_G$ | $41.37\ \mathrm{g_G/m_L^3}$ |
| $K_{i,E}^{(G)}$ | $40.0\ \mathrm{kg_E/m_L^3}$ |
| $K_{i,A}^{(G)}$ | $6.0\ \mathrm{kg_A/m_L^3}$ |
| $K_E$ | $561.8\ \mathrm{g_E/m_L^3}$ |
| $K_{i,G}^{(E)}$ | $10\ \mathrm{g_G/m_L^3}$ |
| $K_A$ | $450.6\ \mathrm{g_E/m_L^3}$ |
| $K_{i,G}^{(A)}$ | $2.0\ \mathrm{g_G/m_L^3}$ |
| $K_{i,E}^{(A)}$ | $2.0\ \mathrm{g_E/m_L^3}$ |

| Dynamics time-scales | |
|---|---|
| Symbol | Value |
| $T_{\mu_G}$ | 2.5 h |
| $T_{\mu_E}$ | 3.5 h |
| $T_{\mu_A}$ | 0.4 h |
| $T_{\varphi_O}$ | 1.6 h |
| $T_\sigma^{(u)}$ | 3 s |
| $T_\sigma^{(d)}$ | 120 s |

| Mode yields | |
|---|---|
| Symbol | Value |
| $Y_{EX}^{(\mathcal{G}_1)}$ | $16.02\ \mathrm{mol_{ATP}/mol_X}$ |
| $Y_{EX}^{(\mathcal{G}_2)}$ | $12.02\ \mathrm{mol_{ATP}/mol_X}$ |
| $Y_{EX}^{(\mathcal{E}_1)}$ | $16.02\ \mathrm{mol_{ATP}/mol_X}$ |
| $Y_{EX}^{(\mathcal{A}_1)}$ | $18.02\ \mathrm{mol_{ATP}/mol_X}$ |
| $Y_{CX}^{(\mathcal{G}_1)}$ | $2.77\ \mathrm{mol_{AcCoA}/mol_X}$ |
| $Y_{CX}^{(\mathcal{G}_2)}$ | $3.02\ \mathrm{mol_{AcCoA}/mol_X}$ |
| $Y_{CX}^{(\mathcal{E}_1)}$ | $2.52\ \mathrm{mol_{AcCoA}/mol_X}$ |
| $Y_{CX}^{(\mathcal{A}_1)}$ | $3.02\ \mathrm{mol_{AcCoA}/mol_X}$ |
| $Y_{NX}^{(\mathcal{G}_1)}$ | $29.4\ \mathrm{mmol_{NADPH}/mol_X}$ |
| $Y_{NX}^{(\mathcal{G}_2)}$ | $29.4\ \mathrm{mmol_{NADPH}/mol_X}$ |
| $Y_{NX}^{(\mathcal{E}_1)}$ | $29.4\ \mathrm{mmol_{NADPH}/mol_X}$ |
| $Y_{NX}^{(\mathcal{A}_1)}$ | $29.4\ \mathrm{mmol_{NADPH}/mol_X}$ |
| $Y_{AE}^{(\mathcal{E}_2)}$ | $0.5\ \mathrm{mol_A/mol_E}$ |

| Constants | |
|---|---|
| Symbol | Value |
| $M_G$ | $180.2\ \mathrm{g_G/mol_G}$ |
| $M_E$ | $46.0\ \mathrm{g_E/mol_E}$ |
| $M_A$ | $59.0\ \mathrm{g_A/mol_A}$ |
| $M_O$ | $32.0\ \mathrm{g_O/mol_O}$ |
| $M_X$ | $113.1\ \mathrm{g_X/mol_X}$ |

| Diverse | |
|---|---|
| Symbol | Value |
| $\mu_G^{O_{lim}}$ | $0.31\ \mathrm{h}^{-1}$ |
| $\delta\mu_\sigma$ | $0.2\ \mathrm{h}^{-1}$ |
| $k_{i,\sigma}^{(\min)}$ | 0.4 |

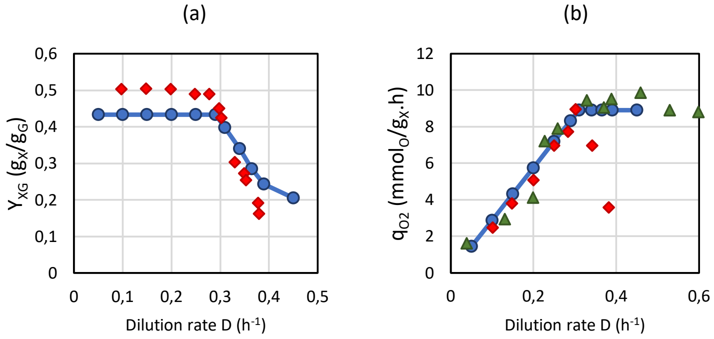


Figure II.16: Comparison of model predictions against experimental steady-state chemostat measurements. ♦ Data from Flikweert et al. [1999]. ▲ Data from Barford and Hall [1979]. (a) Conversion yield of glucose to cell. (b) Specific oxygen consumption rate. (c) Respiratory quotient.

## II.5.3 Comparison results

### Steady-state comparison

First, the model has been assessed by simulating glucose-fed chemostat culture at multiple dilution rates. The culture state was considered steady with the stopping criterion

$$\left\| \left( \frac{\partial \boldsymbol{C}}{\partial t} \right) \oslash \boldsymbol{C} \right\| < 10^{-3}\text{h}^{-1} \tag{II.19}$$

*i.e.* when concentrations would vary by less than 0.1% over an hour. At this steady-state, we extracted from the mode rates the global conversion yield of glucose into cell, the specific oxygen consumption rate and the respiratory quotient RQ. The latter is the molar ratio between rates of produced carbon dioxide and consumed oxygen. This is a metrics often used to assess the well functioning of a culture as this ratio is close to 1 if the metabolic is fully oxidative, and significantly increases when fermentative pathways are active. Results are illustrated in Fig. II.16 and clearly show correct tendencies as well as the switch of metabolism occurring around $D = 0.31\text{h}^{-1} = \mu_G^{O_\text{lim}}$ due to the long-term Crabtree effect.

### Dynamic simulations

van Dijken and Scheffers [1986] have performed lab-scale cultures of *S. cerevisiae* which are of particular interest for the development of our model. They performed batch cultures, but with the particularity that these batch cultures directly follow pre-culture in chemostat. As explained in Section II.3, chemostat cultures allow for enforcing a specific growth-rate over a biological population. As batch culture are highly sensitive to initial condition, having performed these chemostat pre-culture helps us defining the initial state of the biological population. Note that this knowledge of initial state is often missing. Therefore, we will focus the fitting of the model on data sets from these authors having a great confidence in their exploitability.

Three data sets corresponding to three batch culture are actually available. The experimental data and corresponding simulations are shown in Fig. II.17. Sub-figures (a)

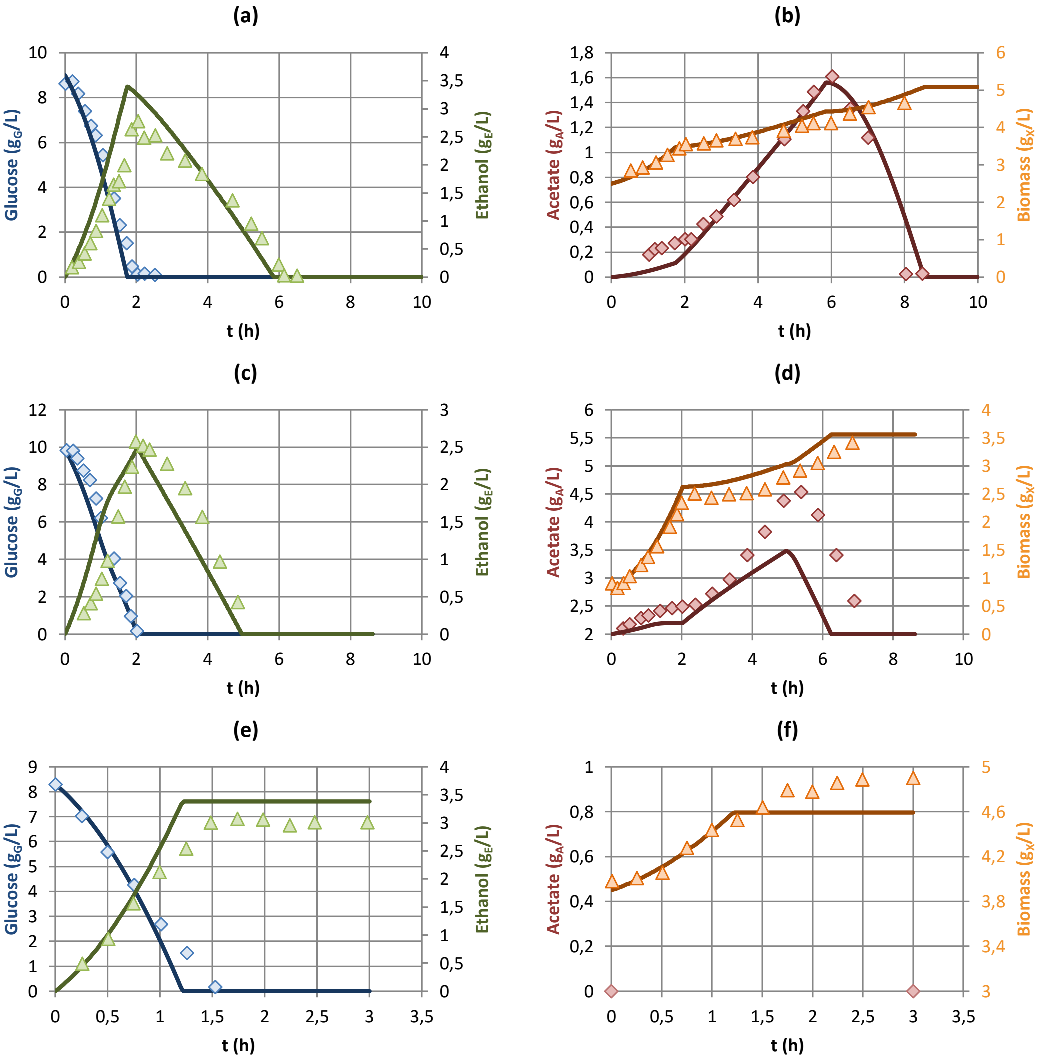


Figure II.17: Comparison of model predictions with batch culture following precultivation in chemostat from van Dijken et al. [2000]. (a) and (b): preculture at dilution rate $0.1\mathrm{h}^{-1}$. Aerobic batch culture. (c) and (d): preculture at dilution rate $0.2\mathrm{h}^{-1}$. Aerobic batch culture. (e) and (f) preculture at dilution rate $0.1\mathrm{h}^{-1}$. Anaerobic batch culture.

and (b) correspond to an aerobic batch culture with a pre-culture dilution rate set to $D = 0.1\mathrm{h}^{-1}$. Sub-figures (c) and (d) are similar but with an higher pre-culture dilution rate set to $D = 0.2\mathrm{h}^{-1}$. Finally, sub-figures (e) and (f) correspond to an anaerobic batch culture with a pre-culture dilution rate of $D = 0.1\mathrm{h}^{-1}$.

Overall, all metabolic modes play their role in different phases of theses cultures as shown previously in Fig. II.10. Therefore, we could adjust conversion yield but also time-scales for biological properties adaptation over these data sets. Following this, we could attempt reproducing chemostat cultures disturbed by glucose pulses by Flikweert et al. [1999] (see Fig. II.18) and Sweere et al. [1988] (see Fig. II.19). Due to the use of different strains of *S. cerevisiae*, the imperfect fitting is not surprising. However, main tendencies seem to be well reproduced and the use of automated fitting tool in place of a manual adjustment of model parameters could certainly yield even better agreement.

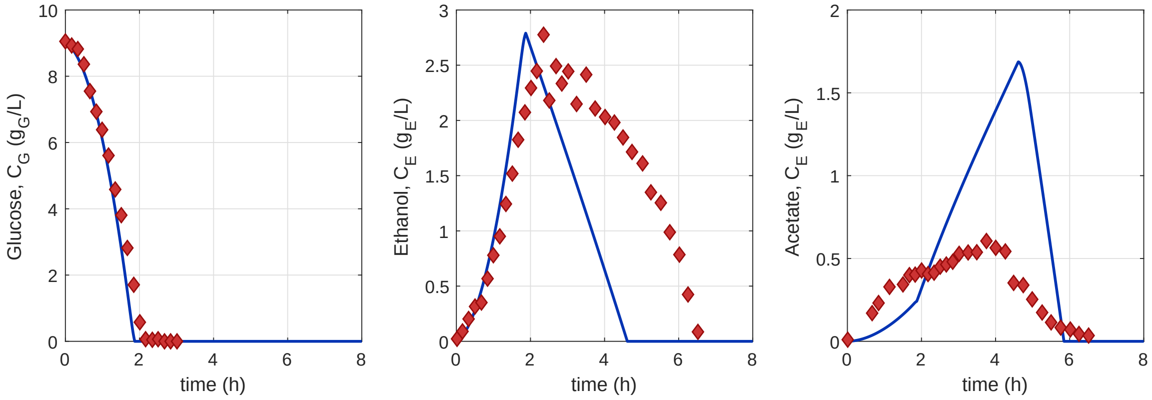


Figure II.18: Comparison of model predictions with glucose pulse experiment in chemostat culture by Flikweert et al. [1999]. Dilution rate: $0.10\mathrm{h}^{-1}$.

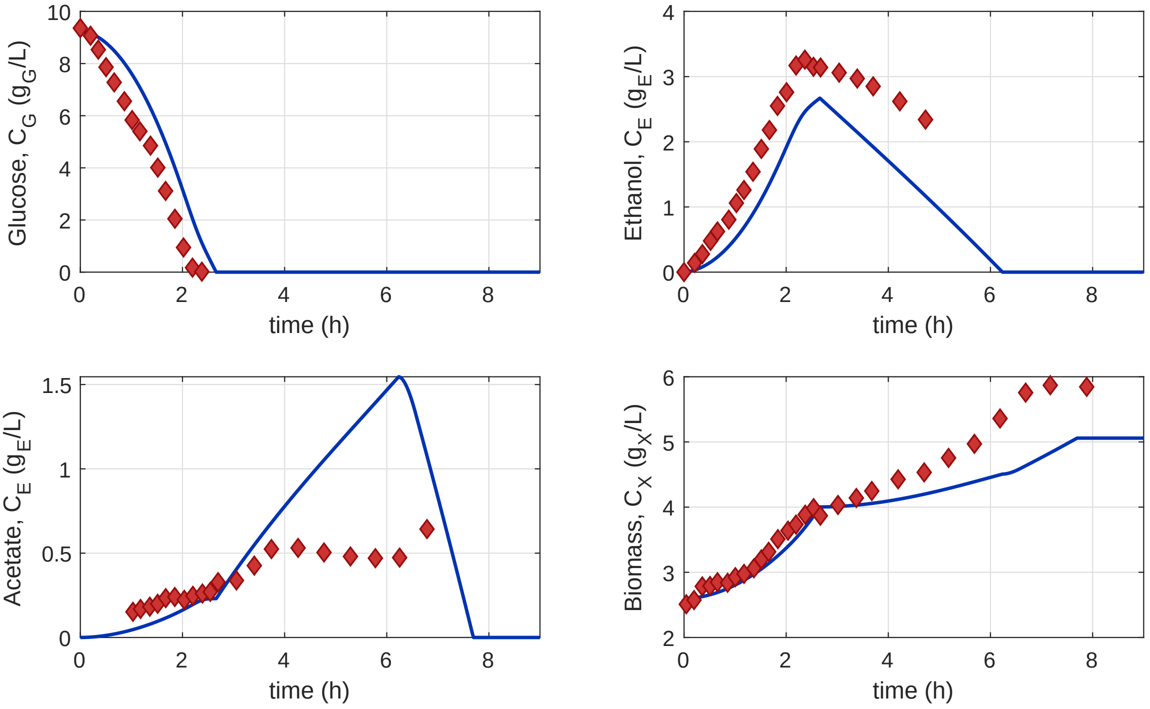


Figure II.19: Comparison of model predictions with glucose pulse experiment in chemostat culture by Sweere et al. [1988]. Dilution rate: $0.08\mathrm{h}^{-1}$.

## II.6 Conclusion

This chapter focused on developing a low-cost structure for metabolic models which can account for dynamic biological properties and multiple metabolic modes. This work is actually the continuation of a Master thesis where a simple metabolic model for *E. coli* by Xu et al. [1999] was derived to account for a dynamic growth capability and disequilibrium between a cell and its environment [Pigou and Morchain, 2015]. Principles underlying these first models have been reapplied under a better organized formalism, based on a metabolic mode definition and a simple methodology to compute rates of these modes. This was mainly done by selecting key metabolic fluxes and by providing simple closure laws to estimate them. This in turn allows computing all fluxes and thus consumption and production rates.

The proposed structure has been illustrated on two standard micro-organisms in industrial bio-processes: *Escherichia coli* and *Saccharomyces cerevisiae*. The first one mainly served as an illustration of the structure. The second one will be used in Chapter V to perform large-scale simulations of an industrial bioreactor.

The *S. cerevisiae* model has been challenged against multiple experimental data sets, originating from multiple labs and research teams. A preliminary fitting has been performed manually which yields a satisfactory match between simulations and experimental results. An ulterior step would be to improve numerically the fitting while also assessing the model sensitivity to its parameters through an optimization based procedure [Sánchez et al., 2014]. For now, the proposed model is a proof-of-concept which illustrates that the proposed structure can indeed reproduce observed tendencies over a wide range of culture modes even in dynamic conditions. Note that this is done using a low number of floating-point operations (flop), about 140 flop, and a limited number of parameters too (35 different parameters in Table II.5). To give a sense of comparison, Rizzi et al. [1997] proposed a intracellular concentrations-based model tracking 15 metabolites concentrations which required fitting 87 model parameters and relied on at least 330 floating-point operations. This model did not handle long-term cultures as the growth-rate was a constant model parameter.

# III

# BILANS DE POPULATION APPLIQUÉS AUX SYSTÈMES BIOLOGIQUES

## Résumé

Afin de prédire le comportement des bioréacteurs aux échelles de pilotes de laboratoire et de procédés industriels, tout en préservant un même modèle, l'élément clé est de considérer les micro-organismes comme des systèmes dynamiques. Suivre leurs propriétés dynamiques est alors un prérequis pour une modélisation précise des fermenteurs.

Dans le Chapitre II, des modèles métaboliques sont présentés pour *Escherichia coli* et *Saccharomyces cerevisiae* et sont basés sur la définition de variables dynamiques directement liées aux vitesses de réactions internes. Les variables alors choisies sont les capacités de croissance, de mobilisation des voies oxydatives, et le stress glycolitique. Ces variables ont été sélectionnées car directement utilisables pour prédire l'orientation des flux de matières entre les modes métaboliques élémentaires du modèle. Ces choix de variables étaient donc justifiés par leur couplage avec la structure de modèle métabolique proposée. Cependant, de nombreux auteurs ont considéré d'autres quantités attachées aux cellules telles que leur âge, masse, taille, ... Au delà du choix de ces variables, le principal défi est de suivre ces quantités à l'échelle d'une population biologique.

Pour résoudre la diversité de ces variables sur l'ensemble de la population de micro-organismes, et grâce au nombre important d'individu, il va être possible d'appliquer des approches statistiques. Au lieu de suivre les propriétés attachées à chaque cellule prise individuellement, nous proposons d'appliquer la modélisation par bilan de population, dont le principe est de résoudre l'évolution spatiale et temporelle d'une distribution statistique. Le formalisme des bilans de population se couple parfaitement avec les modèles métaboliques présentés en Chapitre II, mais aussi avec les descriptions de l'hydrodynamique qui seront détaillées en Chapitre IV. Cependant, cette approche peut se révéler numériquement coûteuse selon la méthode choisie pour résoudre l'Équation de Bilan de Population.

Ce chapitre vise principalement à identifier la méthode la plus adaptée pour résoudre les équations de bilan de population appliquées à la modélisation des bioréacteurs. Deux catégories de méthodes sont considérées : méthodes des classes et méthodes des moments. La méthode des classes a déjà été appliquée à la modélisation des bioréacteurs avant ce travail de thèse [Pigou and Morchain, 2015]. Appliquer les méthodes aux moments à la modélisation des fermenteurs est cependant un élément nouveau proposé au cours de ce travail.

Une première partie de ce chapitre présente (i) une variante des méthodes des classes, dite de pivots-fixes et (ii) des méthodes de quadratures des moments afin de fermer le jeu d'équations obtenu dans la formulation aux moments : la méthode de Quadra-

ture des Moments (QMOM), sa version étendue (EQMOM) et l'approche du Maximum d'Entropie. Une seconde partie se concentre sur la comparaison de ces méthodes en termes de précision, de stabilité et de performance. Cette comparaison a été publiée dans *Chemical Engineering Science* [Pigou et al., 2017]. Au cours de ce travail, EQMOM a été identifiée comme répondant aux besoins associés à la simulation des bioréacteurs. Le travail d'implémentation de méthodes pour leur comparaison nous a permis d'identifier une amélioration majeure qu'il était possible d'apporter à EQMOM. Une troisième partie de ce chapitre décrit brièvement cette amélioration, et se concentre sur la mesure des gains qu'elle apporte. En particulier, il est montré que le coût numérique d'EQMOM a été réduit d'un facteur 5 à 30. L'ensemble des détails relatifs à cette nouvelle version d'EQMOM ont été publiés dans le *Journal of Computational Physics* [Pigou et al., 2018a].

# III

# POPULATION BALANCES FOR BIOLOGICAL SYSTEMS

## Summary

In order to use similar models to predict the behaviour of bioreactors at both lab-scale and industrial scale, the key consideration is that micro-organisms are dynamic systems. Tracking their dynamic properties is therefore a prerequisite for an accurate modelling of fermenters.

In Chapter II, metabolic models are developed for *Escherichia coli* and *Saccharomyces cerevisiae* and are based on the definition of dynamic variables directly related to internal reaction rates. Chosen variables were growth capabilities, oxidative capacity and glycolytic stress. These variables were selected as they directly contain information useful for orienting matter between elementary metabolic modes. These variable choices were then motivated by the proposed metabolic model structure. However, many authors have considered other quantities such as cell age, mass, size, budding index, ... Beyond this variable choice, the main numerical challenge is to track these quantities over a large biological population.

To track the diversity of these variables among whole biological populations, one can benefit from the high number of cells to apply statistical approaches. Instead of tracking these properties cell-by-cell, we propose to make use of the Population Balance Modelling (PBM) approach which tracks statistical distributions over time and space. The PBM formalism can be perfectly coupled with metabolic models as proposed in Chapter II, and with hydrodynamic framework that will be discussed in Chapter IV. However, this approach can be computationally expensive depending on the method used to solve the Population Balance Equation.

This chapter aims at identifying the best suited method to handle Population Balances for the simulation of large-scale bioreactors. Two categories of methods are considered: methods of classes and methods of moments. The class method has already been used in the context of bioreactor modelling before this PhD thesis [Pigou and Morchain, 2015] however this problem is an original application case for method of moments that have been mainly developed in other chemical engineering fields.

The first part of this chapter introduces (i) the fixed-pivot method which is one variant of methods of classes and (ii) quadrature methods used to close the set of equations in moment formulation, namely the Quadrature Method of Moments (QMOM), its Extended version (EQMOM) and the Maximum-Entropy approach. The second part focuses on comparing these methods in terms of accuracy, stability and performance. This comparison has been published in *Chemical Engineering Science* [Pigou et al., 2017]. Finally, over the course of this work, EQMOM was found to be a method suiting the goal of bioreactor simulations, and we happened to identify a major improvement in the basis of

this method. These improvements are quickly presented in a third part, and are shown to significantly reduce the numerical cost of EQMOM by a factor ranging between 5 and 30. Full details of this new EQMOM version were published in *Journal of Computational Physics* [Pigou et al., 2018a].

# III

# POPULATION BALANCES FOR BIOLOGICAL SYSTEMS

## III.1 Introduction

Bioreactors are heterogeneous systems and so are biological populations. This fact is well documented [Delvigne and Goffin, 2014; Heins et al., 2015; Simen et al., 2017] and constitutes the main challenge in bioreactor modelling. The proposed modelling framework accounts for this by associating a vector of biological properties, $\widehat{p}$, to cells. These properties evolve along cell's trajectories, and influence their metabolic behaviour.

In chapter II, two biological models have been detailed. The first one describing *E. coli* considers a single biological property, the cell's growth capability $\mu_c$, hence $\widehat{p} = \mu_c$. The second model describes the behaviour of *S. cerevisiae* and requires multiple dynamic variables, namely the specific growth capability associated to each carbon substrate (glucose, ethanol and acetate), the unstressed oxidative capacity, and the glycolytic stress, hence $\widehat{p} = \left[\mu_G,\ \mu_E,\ \mu_A,\ \varphi_O^{(b)},\ \sigma\right]^T$. These models are defined at the cell-scale. They take as input the vector of biological properties $\widehat{p}$ and the vector of concentrations in the liquid environment of a cell, $\boldsymbol{C}_L$. From these inputs, both models yield (i) a vector of specific consumption/production rates $\boldsymbol{r}(\widehat{p}, \boldsymbol{C}_L)$ which will be source/sink terms in mass balance equations of dissolved species, as well as (ii) rates of evolution of biological properties $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L)$.

The cell-scale behaviour being modelled, the remaining challenge is to describe this biological behaviour at the scale of the whole population, in a numerically tractable way. As stated in chapter I, these questions are not new and arise in many engineering fields. Suitable modelling frameworks have already been developed to tackle these problems, and the challenge is mainly to transpose and adapt them to the modelling of heterogeneous bioreactors.

The modelling of heterogeneous population will be handled by Population Balance Models (PBMs). They are based on Population Balance Equations (PBEs) which track a statistical distribution $n(\widehat{p})$ of heterogeneous properties in space and time. PBEs allow taking into account all phenomena that influence the statistical distribution. The cell-scale evolution of biological properties, modelled through the term $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L)$, will be a transport term in the space of biological properties, which is the support of the Number Density Function (NDF) $n(\widehat{p})$. The fact that new cells are formed by cell-division (for bacteria) or budding (for yeasts) and that these new cells can have properties differing from their mothers cells can also be taken into account, as soon as one describes the probability that a mother of properties $\widehat{p}'$ will produce a daughter cell of properties $\widehat{p}$. This will be referred to as the redistribution probability law $\beta(\widehat{p}, \widehat{p}')$.

This chapter focuses on how to handle the PBE in a way compatible with other aspects of a bioreactor modelling. The core question is: how to track $n(\widehat{p})$ in a numerically efficient way? For the modelling of bioreactors, having access to this distribution is of interest as it allows estimating local rates of consumption or production of dissolved species (glucose, acetate, oxygen, ...):

$$\boldsymbol{R}\left(n(\mu), \boldsymbol{C}_L\right) = \int_{\Omega_{\widehat{p}}} n(\widehat{p}) \boldsymbol{r}(\widehat{p}, \boldsymbol{C}_L) \mathrm{d}\widehat{p} \tag{III.1}$$

Approaches developed hereafter will not directly track $n(\widehat{p})$ but will yield approximations of this reaction term.

A generic PBE compatible with models developed in chapter II is

$$\frac{\partial n(\widehat{p})}{\partial t} + \widehat{\nabla} \cdot \left(\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L) n(\widehat{p})\right) = \int_{\Omega_{\widehat{p}}} \mu(\widehat{p}', \boldsymbol{C}_L) \beta(\widehat{p}, \widehat{p}') n(\widehat{p}') \mathrm{d}\widehat{p}' \tag{III.2}$$

Spatial convection and dispersion terms must be added on the Left-Hand Side (LHS), refer to Eq. (I.14) for the complete PBE. A supplementary term on the Right-Hand Side (RHS) could account for cell death, but this has not been considered in this work. Specific PBEs for *E. coli* and *S. cerevisiae*, based on models from Chapter II, will be detailed in Section III.2.

As explained in Chapter I, PBEs are not directly compatible with mass balance equations as described in classical Euler-Euler gas-liquid modelling and simulation frameworks (see Eq.(I.9)), due to the transport term in the space of cell's properties. Both sectional and moment methods can be applied to transform a PBE into a set a equations having the expected structure. This is presented in Section III.3 in the case of developed PBEs.

When moment methods are applied to a PBE, some integral terms may not be formulated directly in terms of moments. For bioreactors, and in particular if one uses the metabolic model structure developed in Chapter II, this situation will occur. Therefore, supplementary numerical methods will be required to provide closure for these integral terms. Some of these methods were assessed by John et al. [2007] and Lebaz et al. [2016]. Supplementary comparisons have been performed, aiming in particular toward the simulation of bioreactors. Methods first detailed in Section III.4 are then compared to the sectional method as presented in Section III.5 which sums up an article published in *Chemical Engineering Science* [Pigou et al., 2017, accessible in Appendix B].

Numerous readings while implementing these methods, in particular of Chihara [1978], Dette and Studden [1997], and Nguyen et al. [2016], allowed identifying –almost by serendipity– a way to improve an already efficient method, the Extended Quadrature Method of Moments (EQMOM). This led to an article [Pigou et al., 2018a, accessible in Appendix C] whose main developments and results are summed up in Section III.6. Note that these improvements are not specific to the study-case of a bioreactor modelling, and can benefit all communities using moment methods for one-dimension PBMs.

Most developments presented in this chapter are related to one-dimensional PBMs and are thus applied to the PBE related to *E. coli* for sake of illustration. However, the modelling of *S. cerevisiae* as presented in Chapter II requires five dynamic biological properties. Therefore, Section III.7 describes the trade-off that is applied to consider these five variables while using only numerical methods for 1D PBEs.

# III.2 Considered Population Balance Equations

## III.2.1 A generic 1D population balance model on growth-rate

As stated in Chapter II, this PhD thesis is the continuation of a worked started with the modelling of large-scale cultures of *E. coli*. This preliminary work integrated the use of a Population Balance Model based on the growth capability $\widehat{p} = \mu_c$ of a cell, *i.e.* the growth rate a cell can achieve if substrates are sufficiently available in the environment. This PBM was first proposed by Morchain et al. [2013, 2014] and takes the following form:

$$\frac{\partial n(\mu_c)}{\partial t} + \frac{\partial}{\partial \mu_c}\left(n(\mu_c)\zeta_\mu(\mu_c, \boldsymbol{C}_L)\right) = \int_{\Omega\mu} \mu(\mu_c', \boldsymbol{C}_L) n(\mu_c') \beta(\mu_c, \mu_c') \mathrm{d}\mu_c' \qquad \text{(III.3)}$$

with $\mu(\mu_c', \boldsymbol{C}_L)$ being the growth rate achieved by a cell based on its growth capabilities and on the substrate availability in its environment. $\zeta_\mu(\mu_c, \boldsymbol{C}_L)$ is the rate of evolution of the growth rate, and $\beta(\mu, \mu')$ is the probability density function that a mother cell of growth capability $\mu'$ will yield a daughter cell of growth capability $\mu$.

The computation of $\mu(\mu_c, \boldsymbol{C}_L)$ was detailed in Chapter II.3. Over the course of this work, various proposal were made for the terms $\zeta_\mu(\mu_c, \boldsymbol{C}_L)$ and $\beta(\mu, \mu')$. They are detailed hereafter.

### Dynamics of growth-rate adaptation

Growth-rate adaptation corresponds to the second LHS term in Eq. (III.3) where $\zeta_\mu(\mu_c, \boldsymbol{C}_L)$ is the rate of adaptation of a cell's growth rate. This notion of a dynamic growth-rate has been fully detailed in Chapter II for the specific case of *E. coli*, but is actually applicable to a variety of other micro-organisms (see Section II.3).

The proposed model for growth-rate adaptation to an environment is based on the fact that, in constant environments, the mean growth-rate of a population reaches an equilibrium value that usually follows the Monod [1952] law, already presented in I.6.3:

$$\mu^* = \mu_{\max} \frac{G^*}{G^* + K_G} \qquad \text{(III.4)}$$

where $\mu^*$ and $G^*$ are steady-state –at equilibrium– growth-rate and glucose concentration, $\mu_{\max}$ is the maximum mean growth-rate achievable by the population, and $K_G$ is the affinity of the cell with glucose $(kg_G/m_L^3)$. Other terms can be added to this law to account for affinity toward other substrates and/or for inhibition by some compounds.

In unsteady environments, it can be observed that growth-rate evolves dynamically and requires a few hours to reach a new equilibrium. Therefore, the following model was suggested by Morchain and Fonade [2009]:

$$\zeta_\mu(\mu_c, \boldsymbol{C}_L) = \left(\frac{1}{T} + \mu_c\right)(\mu^*(\boldsymbol{C}_L) - \mu_c) \qquad \text{(III.5)}$$

with $T$ the time-scale of adaptation ($\approx$ 1h) and $\mu^*(\boldsymbol{C}_L)$ the Monod law adapted to the considered micro-organism.

### Redistribution of growth rate at cell-division

The redistribution of growth-rate when new cells are formed, let it be by cell division or budding, is modelled through the probability density function $\beta(\mu_c, \mu_c')$ which represents the probability that a new cell will have a growth-rate $\mu_c$ if its mother cell is characterised by a growth-rate $\mu_c'$.

The simplest approach is to assume that the growth-rate is inherited at cell division, therefore

$$\beta(\mu_c, \mu_c') = \delta(\mu_c, \mu_c') \tag{III.6}$$

where $\delta$ is the Dirac delta distribution, that can be defined by its sifting property

$$\int f(x)\delta(x, x')\mathrm{d}x = f(x') \tag{III.7}$$

Though being a simple approach, which in particular prevents the need for integral term on the RHS of Eq. (III.3), this model is not supported by experimental measurements.

The choice of the growth rate as a discriminating factor between individuals facilitates the computation of reaction rates as shown in Chapter II. However, the growth rate is a variable that is more easily attached to a population rather than to individuals. This can reveal to be problematic if one must produce a Probability Density Function (PDF) defining the growth-rate of a daughter cell based on the growth rate of its mother cell, as one requires the knowledge of cell-attached growth-rates.

The growth-rate has two definitions depending on whether one considers the total mass of the biological population, or the total number of cells. In terms of mass, it is defined as the rate of population mass increase, hence $\mu$ being defined with a unit $\mathrm{kg_{biomass}/kg_{biomass} \cdot h}$. It is also well-known that the growth rate can be defined by the population doubling time $\widetilde{\mu} = \ln(2)/\tau$ where $\tau$ is the time required to double the number of cells in the population. As thoroughly analysed by Quedeville et al. [2018], these two definitions only match when cells mean-mass remains constant. This only happens in steady-state culture (chemostat) or during exponential growth in batch cultures.

The number based definition of $\mu$ may give access to cell-attached growth-rate if one considers that the population doubling time is relatable to the cell division-time –if one consider bacteria exhibiting cell division– or budding-time if one considers budding yeast. The modelling of the probability density function $\beta(\mu_c, \mu_c')$ can then be based on experimental data that follow the division time along cells lineage in a steady environment.

Such experimental data are not widely available but Yasuda [2011] did perform such lineage analysis on *E. coli* cultures. Similar analysis were also conducted by Nobs and Maerkl [2014] on *Schyzosaccharomyces pombe*. This yeast exhibits a rod-shape and cell-division similar to *E. coli*. These authors developed microfluidic devices allowing for the visually monitored culture of micro-organisms in chemostat conditions, using perfused feeding. Camera and image processing tool were then used to track individual cells over-time which gives access to statistically significant measures of doubling times along lineages. As demonstrated in Morchain et al. [2016], we could post-process these data to model the redistribution $\beta(\mu_c, \mu_c')$. The proposed model is given by the following skew-normal distribution:

$$\beta(\mu_c, \mu_c') = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mu_c - l)^2}{2\sigma^2}\right)\left(1 + \operatorname{erf}\left(\frac{\alpha(\mu_c - l)}{\sigma\sqrt{2\pi}}\right)\right) \tag{III.8}$$

with $l = k_l\ \tilde{\mu}$ and $\sigma = k_\sigma\ \tilde{\mu}$. Numerical parameters were deduced from data from Yasuda [2011] and are $k_l \approx 0.65$, $k_\sigma \approx 0.46$ and $\alpha \approx 3.65$.

Note that this model is still to be improved. Indeed, one can notice that $\beta(\mu_c, \mu_c')$ is actually not dependent on $\mu_c'$ meaning that in this model, the growth-rate of a daughter cell is not related to that of its mother. Instead, the redistribution law refers to the population mean growth rate $\tilde{\mu}$ and $\beta(\mu_c, \mu_c')$ becomes $\beta_{\tilde{\mu}}(\mu_c)$.

Attaching mean population properties to the individual behaviour can be considered as a major flaw of this proposed model. This is due to the fact that the only available experimental data was obtained in non-limiting environment where $\tilde{\mu} = \mu_{\max}$. Then it was necessary to extend the model to less favourable environments which is done by assuming a linear dependency of the PDF parameters with $\tilde{\mu}$. The redistribution probability function becomes narrower in slowly growing populations as shown in Figure III.1.

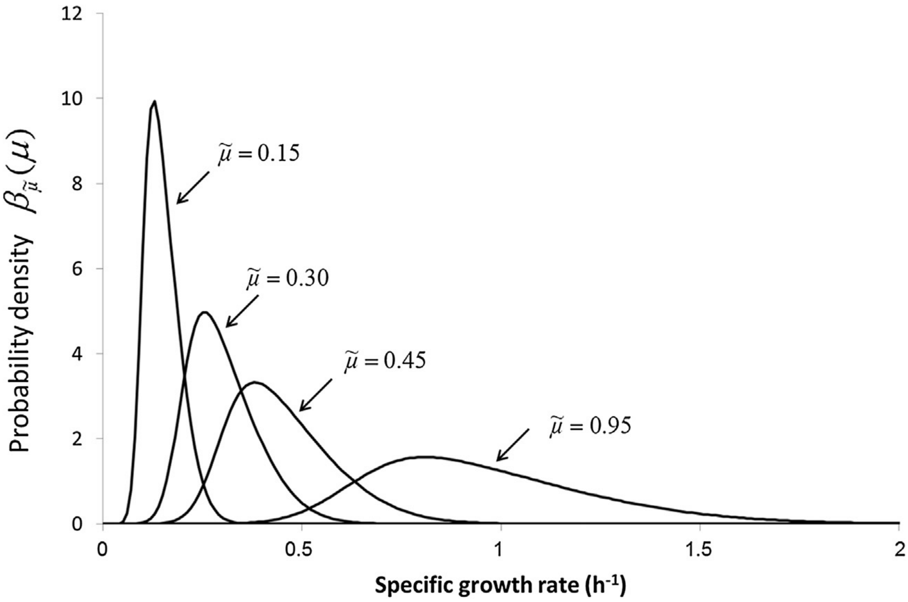


Figure III.1: Illustration of the extrapolation of $\beta(\mu_c, \mu_c')$ to slow or fast growing populations [Morchain et al., 2016].

Further work will be required to improve this model by removing the dependency of cell-attached behaviour to population-attached mean properties. For instance, the skewed distribution observed by Yasuda [2011] tends to indicate that the higher the mother cell's growth-rate, the higher the difference between the mother's and the daughter's cell growth-rate. A proposition would then be that $\beta(\mu_c, \mu_c')$ is a standard PDF (Gaussian or Log-normal for instance) dependent on $\mu_c$, whose mean value is $\mu_c'$ and whose spreading parameter is proportional to $\mu_c'$. This improved modelling has not been extensively looked into and is, for now, limited by experimental data availability.

### III.2.2 Modelling *S. cerevisiae* dynamics

**Biological properties**

As detailed previously in Chapter II, we model the metabolic behaviour of *S. cerevisiae* by considering five dynamic biological properties attached to each cell:

- $\mu_G^{(b)}$: maximum rate of glucose-based growth in non-limiting aerobic environment $(\mathrm{g_X}/\mathrm{g_X}\cdot\mathrm{h})$.
- $\mu_E^{(b)}$: maximum rate of ethanol-based growth in non-limiting environment $(\mathrm{g_X}/\mathrm{g_X}\cdot\mathrm{h})$.
- $\mu_A^{(b)}$: maximum rate of acetate-based growth in non-limiting environment $(\mathrm{g_X}/\mathrm{g_X}\cdot\mathrm{h})$.
- $\varphi_O^{(b)}$: maximum oxygen consumption rate $(\mathrm{g_O}/\mathrm{g_X}\cdot\mathrm{h})$.
- $\sigma$: cell glycolytic stress (no unit, normalized between 0 and 1).

The three maximum growth rates and the maximum oxygen consumption rate are properties that slowly change over time in order to reach values of equilibrium with the environment:

$$\frac{\partial \mu_{\{G,E,A\}}^{(b)}}{\partial t} = \zeta_{\mu_{\{G,E,A\}}} = \frac{1}{T_{\mu_{\{G,E,A\}}}}\left(\mu_{\{G,E,A\}}^{*} - \mu_{\{G,E,A\}}\right) \tag{III.9}$$

$$\frac{\partial \varphi_O^{(b)}}{\partial t} = \zeta_{\varphi_O} = \frac{1}{T_{\varphi_O}}\left(\varphi_O^{*} - \varphi_O^{(b)}\right) \tag{III.10}$$

where star values are population mean rates that are experimentally observed in a steady environment. Time constants $T_{\mu_G,\mu_E,\mu_A,\varphi_O}$ have an order or magnitude of an hour.

The stress $\sigma$ is a variable which keeps track of whether a cell has been exposed to high glucose concentrations recently. If not, $\sigma \approx 0$ which will imply in the metabolic model that the short-term Crabtree effect is not active. If the cell has indeed been recently exposed to high glucose concentrations (stressing environment), $\sigma \approx 1$ and the short-term Crabtree effect is triggered. Therefore, the evolution of this variable is described by two mutually exclusive terms:

$$\frac{\partial \sigma}{\partial t} = \zeta_\sigma = \frac{\kappa_\sigma}{T_\sigma^{(u)}}(1-\sigma) - \frac{(1-\kappa_\sigma)}{T_\sigma^{(d)}}\sigma \tag{III.11}$$

where $\kappa_\sigma \in [0,1]$ qualifies the stressfulness of the environment, $T_\sigma^{(u)}$ is the time constant of stress increase in stressful environments ($\approx$ 1s) and $T_\sigma^{(d)}$ is the characteristic time for a return to unstressed metabolic behaviour ($\approx$ 100s).

Closures for $\mu_{\{G,E,A\}}^{*}$ and $\varphi_O^{*}$ were defined in Chapter II and are based on the vector of liquid-phase concentrations $\boldsymbol{C_L} = [C_G,\ C_E,\ C_A,\ C_O]$. $\kappa_\sigma$ computation also relies on $\boldsymbol{C_L}$ but also on the value of $\mu_G$.

**Population Balance Equation**

In order to track these properties and their respective evolution, not at a cell-scale, but at the population scale, a multivariate population balance model can be formulated. Let $\widehat{p} = [\mu_G^{(b)},\ \mu_E^{(b)},\ \mu_A^{(b)},\ \varphi_O^{(b)},\ \sigma]$ be the vector of biological properties of a cell, and $n(\widehat{p})$ the

Number Density Function which represents the statistical distribution of these properties over the biological population. A Population Balance Equation tracking the evolution of $n(\widehat{p})$ is

$$\frac{\partial n(\widehat{p})}{\partial t} + \sum_{i\in\{G,E,A\}} \left( \frac{\partial}{\partial \mu_i^{(b)}} \left(n(\widehat{p})\zeta_{\mu_i}(\widehat{p}, \boldsymbol{C}_L)\right)\right) + \frac{\partial}{\partial \varphi_O^{(b)}}(n(\widehat{p})\zeta_{\varphi O}(\widehat{p}, \boldsymbol{C}_L) \\ + \frac{\partial}{\partial \sigma}(n(\widehat{p})\zeta_{\sigma}(\widehat{p}, \boldsymbol{C}_L) = \int_{\Omega_{\widehat{p}}} \mu^{(a)}(\widehat{p}', \boldsymbol{C}_L) n(\widehat{p}')\beta(\widehat{p}, \widehat{p}')\mathrm{d}\widehat{p}' \quad \text{(III.12)}$$

or, more simply by defining $\widehat{\zeta} = [\zeta_{\mu_G},\ \zeta_{\mu_E},\ \zeta_{\mu_A},\ ,\zeta_{\varphi_O},\ \zeta_\sigma]^T$:

$$\frac{\partial n(\widehat{p})}{\partial t} + \widehat{\nabla} \cdot \left(\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_L) n(\widehat{p})\right) = \int_{\Omega_{\widehat{p}}} \mu^{(a)}(\widehat{p}', \boldsymbol{C}_L) n(\widehat{p}')\beta(\widehat{p}, \widehat{p}')\mathrm{d}\widehat{p}' \quad \text{(III.13)}$$

### Redistribution term

The RHS of (III.12) is analogous to the RHS of (III.3) and represents the distribution of biological properties when new cell are formed. We did not attempt to provide a fine description for this term as the difficulty of finding relevant experimental data observed for *E. coli* is worsened by the multi-dimensionality of the current *S. cerevisiae* modelling.

Therefore we applied an inheritance assumption by making daughter cells identical to their mother. This is modelled by defining the transition distribution $\beta(\widehat{p}, \widehat{p}')$ as a 5D Dirac distribution:

$$\int_{\Omega_{\widehat{p}}} f(\widehat{p}')\beta(\widehat{p}, \widehat{p}')\mathrm{d}\widehat{p}' = f(\widehat{p}) \quad \text{(III.14)}$$

which yields

$$\int_{\Omega_{\widehat{p}}} \mu^{(a)}(\widehat{p}', \boldsymbol{C}_L) n(\widehat{p}')\beta(\widehat{p}, \widehat{p}')\mathrm{d}\widehat{p}' = \mu^{(a)}(\widehat{p}, \boldsymbol{C}_L) n(\widehat{p}) \quad \text{(III.15)}$$

## III.3 Applying sectional and moment methods to *E. coli* PBE

As stated previously, Population Balance Equations have a structure that is not directly compatible with mass balance equations in Eulerian modelling framework. To allow compatibility between the Population Balance Model framework with simulation software based on local mass balances, both sectional and moment methods can be applied to replace the PBE by a set of local mass balance equations.

Here, the focus is made on applying these methods on one-dimensional PBEs. Methods for higher order PBEs were not investigated in this PhD and are being tackled in an other ongoing project. Therefore, following sections illustrate the use of both sectional and moment methods on the 1D PBE associated to the modelling of *E. coli* (see Eq. (III.3)). This PBE is here recalled with an explicit convection term:

$$\frac{\partial n(\mu_c)}{\partial t} + \vec{\nabla}\cdot(\vec{u}_L n(\mu_c)) + \frac{\partial}{\partial \mu_c}\left(n(\mu_c)\zeta_\mu(\mu_c, \boldsymbol{C}_L)\right) = \int_{\Omega_\mu} \mu(\mu_c', \boldsymbol{C}_L) n(\mu_c')\beta(\mu_c, \mu_c')\mathrm{d}\mu_c' \quad \text{(III.16)}$$

### III.3.1 Sectional methods

Sectional methods, also referred to as class methods or pivot methods, consist in a discretization of the support of a continuous NDF.

In the considered case of the PBE describing the evolution of the growth-rate of *E. coli*, the considered support is $\mu_c \in [0, \mu_{\text{lim}}[$ where $\mu_{\text{lim}}$ is the maximum growth-rate achievable by a single cell. $\mu_{\text{lim}}$ must not be confused with the maximum growth-rate $\mu_{\text{max}}$ as seen in the Monod kinetic law (see Eq. (III.4)). $\mu_{\text{max}}$ is the maximum growth-rate that can be observed at the population scale. This value is only a mean value, meaning that there exists cells that grow at a higher rate than $\mu_{\text{max}}$. Experimental results have indeed shown that at the cell-scale, some cells grow and divide at a higher rate indeed, hence $\mu_{\text{lim}} > \mu_{\text{max}}$. In the proposed model, such individuals whose growth rate is higher than $\mu_{\text{max}}$ appear thank to the modelling of the cell-division redistribution function $\beta(\mu_c, \mu_c')$ which can theoretically generate cells with infinitely high growth-rate but with low probability. We set $\mu_{\text{lim}} = 2.2\mu_{\text{max}}$ considering that the probability to generate individuals such that $\mu_c > 2.2\mu_{\text{lim}}$ is negligible ($< 0.07\%$).

We considered the simplest of sectional methods consistent with the bounded support, that is the fixed pivot method using regular grid [Kumar and Ramkrishna, 1996a]. In this approach, the support $\Omega_\mu = [0, \mu_{\text{lim}}[$ is discretized into $\mathcal{N}_{\text{class}}$ intervals of same lengths. Bounds of the $i$-th interval, or class, are defined as

$$\mu_{i-1/2} = (i-1)\frac{\mu_{\text{lim}}}{\mathcal{N}_{\text{class}}}$$
$$\mu_{i+1/2} = i\frac{\mu_{\text{lim}}}{\mathcal{N}_{\text{class}}}$$

with $i \in \{1, \ldots, \mathcal{N}_{\text{class}}\}$. All cells whose property lies within the bounds of the $i$-th class, for which $\mu_c \in [\mu_{i-1/2}, \mu_{i+1/2}[$, are assumed to share the same property $\mu_i$ defined as the mean class property:

$$\mu_i = (i-1/2)\frac{\mu_{\text{lim}}}{\mathcal{N}_{\text{class}}}$$

An illustration of this discretization is given in Fig. III.2 for $\mathcal{N}_{\text{class}} = 10$.

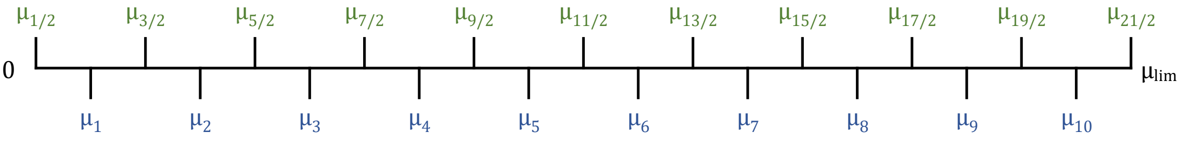


Figure III.2: Example of NDF support discretization in regular grid fixed pivot technique.

Let $N_i$ be the amount of cells in the $i$-th class ($kg_X/m_L^3$). The mass balance equation on $N_i$ is obtained by integrating Eq. (III.16) over $[\mu_{i-1/2}, \mu_{i+1/2}[$ which leads to

$$\frac{\partial N_i}{\partial t} + \vec{\nabla}\cdot(\vec{u}_L N_i) = -\left[n(\mu_c)\zeta_\mu(\mu_c, \boldsymbol{C}_L)\right]_{\mu_{i-1/2}}^{\mu_{i+1/2}} + \sum_{j=1}^{\mathcal{N}_{\text{class}}}\left(\mu(\mu_j, \boldsymbol{C}_L)N_j\int_{\mu_{i-1/2}}^{\mu_{i+1/2}}\beta(\mu_c, \mu_j)\mathrm{d}\mu_c\right) \tag{III.17}$$

The first term on the RHS is the net flux of cell moving between adjacent classes due to growth-rate adaptation. The second term on the RHS is the rate of production of new cells whose growth-rate is located in the $i$-th interval. This second term is already closed and only requires the analytical or numerical (pre)computation of integrals $\int_{\mu_{i-1/2}}^{\mu_{i+1/2}} \beta(\mu_c, \mu_j) \mathrm{d}\mu_c$ for each pair of classes.

The first term is not directly accessible from $\boldsymbol{N} = [N_1, \ldots, N_{\mathcal{N}_{\text{class}}}]^T$ as is requires the knowledge of the values $n(\mu_c)$ not at the pivot locations $\mu_i$, but at the interval boundaries $\mu_{i\pm 1/2}$. The following transport scheme was then used by Morchain et al. [2013] as it allows preserving the positivity of $\boldsymbol{N}$ elements [Kumar and Ramkrishna, 1997]:

$$- \left[n(\mu_c)\zeta_\mu(\mu_c, \boldsymbol{C}_L)\right]_{\mu_{i-1/2}}^{\mu_{i+1/2}} \approx \frac{1}{\Delta\mu} \left[N_{i-1}\zeta_\mu^u(\mu_{i-1}, \boldsymbol{C}_L) + N_{i+1}\zeta_\mu^d(\mu_{i+1}, \boldsymbol{C}_L) - N_i|\zeta_\mu(\mu_i, \boldsymbol{C}_L)|\right] \tag{III.18}$$

where

$$\zeta_\mu^u(\mu_c, \boldsymbol{C}_L) = \begin{cases} \zeta_\mu(\mu_c, \boldsymbol{C}_L) & \text{if } \mu_c < \mu^*(\boldsymbol{C}_L) \\ 0 & \text{otherwise} \end{cases}$$

$$\zeta_\mu^d(\mu_c, \boldsymbol{C}_L) = \begin{cases} -\zeta_\mu(\mu_c, \boldsymbol{C}_L) & \text{if } \mu_c > \mu^*(\boldsymbol{C}_L) \\ 0 & \text{otherwise} \end{cases}$$

This first scheme is applicable here because all cells move toward the same $\mu^*(\boldsymbol{C}_L)$ value.

The fixed pivot method lies in the family of finite volume methods to solve partial derivative equations. It is then also possible to apply the well-known Rusanov scheme which is suited for transport terms in finite-volume approaches and ensures an accurate, conservative and non-oscillatory solution[1].

Under that formalism, the computation of bioreaction rates over the population is given by a sum over all classes:

$$\boldsymbol{R}(\boldsymbol{N}, \boldsymbol{C}_L) = \sum_{i=1}^{\mathcal{N}_{\text{class}}} N_i \boldsymbol{r}(\mu_i, \boldsymbol{C}_L) \tag{III.19}$$

### III.3.2 Methods of moments

Methods of moments do not track directly local values of the NDF but resolve a finite set of its moments $\boldsymbol{m}_N = [m_0, m_1, \ldots, m_N]^T$ with

$$m_k = \int_{\Omega_\mu} \mu_c^k n(\mu_c) \mathrm{d}\mu_c \tag{III.20}$$

While defining a cut-off value $\mu_{\text{lim}}$ was required for the sectional method, this is not required any more for moment methods, therefore $\Omega_\mu = [0, +\infty[$.

---

[1] The application of Rusanov scheme for the closure of mass-balance equations when applying the sectional method was suggested by Dr. Bastien POLIZZI.

Multiplying each term of Eq. (III.16) by $\mu_c^k, k \in \{0, \dots, N\}$ and integrating over $\Omega_\mu$ yields

$$\frac{\partial m_0}{\partial t} + \vec{\nabla} \cdot (\vec{u}_L m_0) = \int_{\Omega_\mu} \mu(\mu'_c, \boldsymbol{C}_L) n(\mu'_c) \beta(\mu_c, \mu'_c) \mathrm{d}\mu'_c \tag{III.21}$$

$$\frac{\partial m_k}{\partial t} + \vec{\nabla} \cdot (\vec{u}_L m_k) = k \int_{\Omega_\mu} \mu_c^{k-1} n(\mu_c) \zeta_\mu(\mu_c, \boldsymbol{C}_L) \mathrm{d}\mu_c + \iint_{\Omega_\mu} \mu_c^k \mu(\mu'_c, \boldsymbol{C}_L) n(\mu'_c) \beta(\mu_c, \mu'_c) \mathrm{d}\mu'_c \mathrm{d}\mu_c \tag{III.22}$$

The total amount of cells $m_0$ is not affected by the adaptation of growth-rate, but only by the total number of newly created cells (RHS of Eq (III.21)). Both the adaptation and the formation of new cells will affect higher order moments, and therefore the shape of the growth-rate distribution. The first term on the RHS of Eq. (III.22) can be expressed in terms of moments if one considers the model for $\zeta_\mu(\mu_c, \boldsymbol{C}_L)$ given in Eq. (III.5):

$$\int_{\Omega_\mu} \mu_c^{k-1} n(\mu_c) \zeta_\mu(\mu_c, \boldsymbol{C}_L) \mathrm{d}\mu_c = \frac{\mu^*(\boldsymbol{C}_L)}{T} m_{k-1} + \left(\mu^*(\boldsymbol{C}_L) - \frac{1}{T}\right) m_k - m_{k+1} \tag{III.23}$$

Note however that this closure is not possible if $T$ depends on $\mu_c$ and $\boldsymbol{C}$. This was the case in the modelling from Pigou and Morchain [2015] where two different characteristic times $T_u$ and $T_d$ are used as explained previously (see Section III.2.1).

For the modelling of bioreactors, one should expect unclosed integral terms when methods of moments are applied. For the currently considered PBE, these unclosed terms are mainly related to the fact that biological reaction rates cannot be directly expressed in terms of moments of $n(\mu_c)$. This is the case for the cell-division integral terms in moment equations (see Eqs. (III.21) and (III.22)), but also for the reaction rates in dissolved species mass balance equations

$$\boldsymbol{R}(n, \boldsymbol{C}_L) = \int_{\Omega_\mu} n(\mu_c) \boldsymbol{r}(\mu_c, \boldsymbol{C}_L) \mathrm{d}\mu_c \tag{III.24}$$

which cannot be expressed directly in terms of moments of the distribution $n(\mu_c)$, at least because, under the formalism proposed in Chapter II, elements of $\boldsymbol{r}(\mu_c, \boldsymbol{C}_L)$ are not continuously differentiable with respect to $\mu_c$.

## III.4 Numerical methods providing closure for moment formulations

As shown previously, moment equations do not form a closed system for the considered case of bioreactors modelling. Numerous integral terms appear in this formulation, and their closure requires the knowledge of the raw distribution $n(\mu_c)$, which is not accessible from only a finite set of its moments. Therefore, a numerical method is required to provide closure by approximating integral properties of the distribution when only few moments of that distribution are known.

Multiple numerical methods exist to tackle this issue. All of these methods propose an approximation $\tilde{n}(\mu_c)$ of $n(\mu_c)$ as well as a quadrature rule such that

$$\int_{\Omega_\mu} f(\mu_c) n(\mu_c) \mathrm{d}\mu_c \approx \sum_{i=1}^{\mathcal{N}_{\text{node}}} w_i f(\xi_i) \tag{III.25}$$

where $w_i$ are the weights of the quadrature, $\xi_i$ its nodes, and $\mathcal{N}_{\text{node}}$ the number of quadrature nodes. These weights and nodes are computed from moments $\boldsymbol{m}_N = [m_0, \ldots, m_N]^T$ of $n(\mu_c)$ in such a way that the quadrature rule (Eq. (III.25)) correctly reproduces these moments, *i.e.* yields exact results if $f(\mu_c) = \mu_c^k, k \in \{0, \ldots, N\}$.

John et al. [2007] and Lebaz et al. [2016] assessed multiple moment-based closure methods and noticed that the following ones were either inaccurate or unstable for their respective test-cases:

- The Kernel Density Element Method [Athanassoulis and Gavriliadis, 2002]
- The Spline Method [John et al., 2007]

For these reasons, these two methods were not furthermore considered for the simulation of bioreactors. Hereafter are detailed other existing methods that have been considered during this work to hopefully provide low-cost, stable and accurate closures. These methods will be compared in section III.5 to the class method.

### III.4.1 Quadrature Method Of Moments

The Quadrature Method of Moments was first proposed by McGraw [1997] to provide closure in moment-evolution equations in the modelling of aerosol size distributions. It is derived from the theory of orthogonal polynomial [Chihara, 1978] which serves as a basis for Gauss-quadratures whose main purpose is to provide approximations of integral terms in general.

Using an even number $2P$ of moments, Quadrature Method of Moments (QMOM) yields a $P$ nodes quadratures. Its weights $\boldsymbol{w}_P$ and nodes $\boldsymbol{\xi}_P$ are obtained by computing eigenvalues and normalized eigenvectors of a $P \times P$ tridiagonal symmetric matrix whose elements are deduced from $\boldsymbol{m}_{2P-1}$. Exact details of this computation are given in Section III.6.

In a sense, one can consider that QMOM yields an approximation of $n(\mu_c)$ defined by

$$\tilde{n}(\mu_c) = \sum_{i=1}^{P} w_i \delta(\mu_c, \xi_i) \tag{III.26}$$

where $\delta$ is the Dirac delta distribution defined in Eq. (III.7).

This approximation is discontinuous whereas $n(\mu_c)$ is expected to be a continuous distribution. Therefore, QMOM cannot be used to estimate point-wise values of $n(\mu_c)$ and must be restricted to situations where only integral properties are required. When applying QMOM, the generic quadrature rule from Eq. (III.25) becomes

$$\int_{\Omega_\mu} f(\mu_c) n(\mu_c) \mathrm{d}\mu_c \approx \sum_{i=1}^{P} w_i f(\xi_i) \tag{III.27}$$

### III.4.2 Extended Quadrature Method of Moments

[Yuan et al., 2012] introduced the EQMOM as a generalisation of bi-Gaussian approximations of NDFs proposed by Chalons et al. [2010]. EQMOM approximates a NDF as a convex mixture of Kernel Density Functions (KDFs) of the same parametric family:

$$\tilde{n}(\mu_c) = \sum_{i=1}^{P} w_i \kappa(\mu_c, \xi_i, \sigma) \tag{III.28}$$

where $\kappa(\mu, \xi, \sigma)$ is a probability density function, referred to as the EQMOM kernel, whose location parameter is $\xi$ and shape parameter is $\sigma$. Over the past few years, multiple kernels were found to be compatible with the method:

- Gaussian kernel [Chalons et al., 2010; Marchisio and Fox, 2013];

$$\kappa_G(\mu, \xi, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\mu-\xi)^2}{2\sigma^2}\right) \tag{III.29}$$

- Laplace kernel [Pigou et al., 2018a]

$$\kappa_\lambda(\mu, \xi, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|\mu-\xi|}{\sigma}\right) \tag{III.30}$$

- Log-Normal kernel [Madadi-Kandjani and Passalacqua, 2015]:

$$\kappa_L(\mu, \xi, \sigma) = \frac{1}{\sigma\mu\sqrt{2\pi}} \exp\left(-\frac{(\log(\mu)-\log(\xi))^2}{2\sigma^2}\right) \tag{III.31}$$

- Gamma kernel [Yuan et al., 2012]:

$$\kappa_\Gamma(\mu, \xi, \sigma) = \frac{\mu^{(l-1)}\exp(-\mu/\sigma)}{\Gamma(l)\sigma^l} \text{ with } l = \frac{\xi}{\sigma} \text{ and } \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t \tag{III.32}$$

- Weibull kernel [Pigou et al., 2018a]:

$$\kappa_W(\mu, \xi, \sigma) = \frac{1}{\sigma\xi}\left(\frac{\mu}{\xi}\right)^{\frac{1-\sigma}{\sigma}} \exp\left(-\left(\frac{\mu}{\xi}\right)^{1/\sigma}\right) \tag{III.33}$$

- Beta kernel [Yuan et al., 2012]:

$$\kappa_\beta(\mu, \xi, \sigma) = \frac{\mu^{(l-1)}(1-\mu)^{(m-1)}}{B(l,m)} \text{ with } l = \frac{\xi}{\sigma}, m = \frac{1-\xi}{\sigma}$$
$$\text{and } B(l,m) = \int_0^1 x^{(l-1)}(1-x)^{(m-1)}\mathrm{d}x \tag{III.34}$$

The core of this method is the moment-inversion procedure which is an iterative process whose goal is to identify values of the weights $\boldsymbol{w}_P$, nodes $\boldsymbol{\xi}_P$ and shape parameter $\sigma$ from an odd number of moments $2P+1$. This procedure was first developed by Yuan et al.

[2012] and has been further improved by Nguyen et al. [2016]. An even more efficient moment-inversion procedure has been identified during the course of this work and is detailed in section III.6.

For all kernels, it is possible to establish a quadrature rule such that

$$\int_{\Omega_\mu} f(\mu)\kappa(\mu,\xi,\sigma)\mathrm{d}\mu \approx \sum_{j=1}^{Q} \omega_j(\xi,\sigma) f(\lambda_j(\xi,\sigma)) \tag{III.35}$$

with $\boldsymbol{\omega}_Q(\xi,\sigma) = [\omega_1(\xi,\sigma),\dots,\omega(\xi,\sigma)_Q]$ and $\boldsymbol{\lambda}_Q(\xi,\sigma) = [\lambda_1(\xi,\sigma),\dots,\lambda_Q(\xi,\sigma)]$ the weights and nodes of this $Q$-nodes quadrature rule. Therefore, when using EQMOM, integral terms over the population $n(\mu_c)$ can be approximated by

$$\int_{\Omega_\mu} f(\mu_c) n(\mu_c)\mathrm{d}\mu \approx \sum_{i=1}^{P}\sum_{j=1}^{Q} w_i\omega_j(\xi_i,\sigma) f\left(\lambda_j(\xi_i,\sigma)\right) \tag{III.36}$$

More importantly, EQMOM yields a continuous approximation of $n(\mu_c)$ from its moments which allows approximating point-wise values of this distribution. This is not possible with QMOM which only yields discrete distributions.

### III.4.3 Maximum-Entropy Method

The Maximum-Entropy method was assessed by Lebaz et al. [2016] and favorably selected when compared to Kernel Density Element Method (KDEM) and Spline Methods for the modelling of a depolymerization process. This method was introduced by Mead and Papanicolaou [1984] and has been further improved by Tagliani [1999, 2001]. Its core idea is to select, between all NDFs whose first moments are given by $\boldsymbol{m}_N$, the one whose Shannon entropy is the highest.

The Shannon entropy of a NDF $p(x)$ defined on $\Omega_x = [0,\ 1]$ is defined as

$$H[p] = -\int_0^1 p(x)\log p(x)\mathrm{d}x \tag{III.37}$$

and is a measure of the amount of information required to define $p(x)$.

Solving the Moment Maximum-Entropy problem requires identifying the NDF $p(x)$ of highest Shannon Entropy under the constraints

$$\int_0^1 x^k p(x)\mathrm{d}x = m'_k \quad k \in \{0,\dots,N\}, m'_0 = 1 \tag{III.38}$$

with $\boldsymbol{m}'_N$ moments of $p(x)$.

Mead and Papanicolaou [1984] and Tagliani [1999] show that the solution to this problem is a NDF whose expression takes the following form:

$$p(x) = \exp\left(-\sum_{i=0}^{N} \lambda_i x^i\right) \tag{III.39}$$

which is the exponential of a polynomial function. The core procedure of the Maximum-Entropy approach then lies in the identification of polynomial coefficients $\boldsymbol{\lambda}_N = [\lambda_0,\dots,\lambda_N]$.

It is shown that the researched $\boldsymbol{\lambda}_N$ values minimize the convex function $\Gamma(\boldsymbol{\lambda}_N)$ defined as [Mead and Papanicolaou, 1984]

$$\Gamma(\boldsymbol{\lambda}_N) = \sum_{i=1}^{N} \lambda_i m'_i + \log\left(\int_0^1 \exp\left(-\sum_{i=1}^{N} \lambda_i x^i\right) \mathrm{d}x\right) \tag{III.40}$$

and whose Jacobian and Hessian matrices are easily expressed in terms of the following integrals [Abramov, 2007]

$$\hat{m}'_k = \int_0^1 x^k \exp\left(-\sum_{i=0}^{N} \lambda_i x^i\right) \mathrm{d}x \quad k \in \{0, \dots, 2N\} \tag{III.41}$$

Finding the Maximum Entropy reconstruction from a moment set $\boldsymbol{m}_N$ is then equivalent to finding the minimum of $\Gamma(\boldsymbol{\lambda}_N)$. This is feasible by applying the Newton-Raphson iterative procedure to identify the root of the Jacobian matrix of this function. Note however that each iteration of this procedure requires $2N$ numerical integrations (see Eq. (III.41)) and that the accuracy and stability of the method depends on the accuracy of these numerical integrations. Vié et al. [2013] then suggested the use of an adaptive quadrature rule. The basics of this adaptive quadrature is to split the interval [0, 1] into multiple sub-intervals depending on the values of intermediary reconstructions, and then use a Gauss-Legendre quadrature on each on these sub-intervals.

Hitherto, the method has only been presented for a NDF $p(x)$ defined on $\Omega_x = [0, 1]$. We must then transform our problem of reconstructing the NDF $n(\mu_c)$ on $\Omega_\mu = [0, \mu_{\mathrm{lim}}]$ to reach the expected formalism. This is done by a simple variable change

$$p(x) = n(x * \mu_{\mathrm{lim}}) \tag{III.42}$$

which implies the following relation between moments $\boldsymbol{m}_N$ of $n(\mu_c)$ and moments $\boldsymbol{m}'_N$ of $p(x)$:

$$m'_k = \mu_{\mathrm{lim}}^{-(k+1)} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} m_{k-j} \tag{III.43}$$

Among methods compared in this work (QMOM, EQMOM and ME), the Maximum Entropy approach can be seen as the most –mathematically speaking– elegant. Its core principle is to “extract” as much information as possible from the known moments, and does not requires any hypothesis about an expected shape of the raw NDF.

## III.5 Comparison of numerical methods

As detailed in an article published before the current PhD work [Pigou and Morchain, 2015], we performed simulations of a large-scale bioreactor, using the model for *E. coli* which inspired the formulation of metabolic models proposed in Chapter II, the Population Balance Model presented previously (see Section III.2.1), and the Compartment Model from Vrábel et al. [1999]. These simulations relied on the use of the class method to solve the Population Balance Equation.

Accuracy was obtained through the use of 100 classes while 80 of them contained virtually no cell. This is a usual drawback of sectional methods which implies that most of the memory allocated to the tracking of the biological-population state holds no useful information. This explains why we considered the use of methods of moments, in order to make a better use of allocated memory. Less tracked variables implies both less used memory, and also less mass balance equations to be solved, which can be of significant interest in particular when the biological modelling (metabolic model coupled with the PBM) is embedded in Computational Fluid Dynamics (CFD) simulations.

These gains expected from a switch toward methods of moments come with the necessity to use quadrature methods as presented previously. This implies an expected loss of accuracy difficult to estimate a priori. Therefore, estimating the accuracy of quadrature methods compared to the class method is a prerequisite to their use in large-scale simulations. As far as the class method is concerned, increasing the number of classes to assess the sensibility of results to that number, and to tightly control the method accuracy, is straightforward.

An other key element of comparison is the numerical cost associated to each method. As already stated, the class method requires solving coupled mass balance equations over all classes, as well as one call to the metabolic model for each class that holds non negligible quantity of micro-organisms. Moreover, the discretization of the biological properties space induces a stringent constraint on the simulation time-step analogous to the Courant-Friedrichs-Lewy (CFL) condition, well known to CFD-software users [Courant et al., 1928]. On the other side, each quadrature method presented in section III.4 comes with different numerical costs related to

- varying number of resolved moments;
- the computation of the quadrature rule;
- a varying number of calls to the metabolic model depending on the size of the produced quadrature rule.

Finally, stability will also be a concern. For the class method, stability is ensured by using low-order transport schemes as detailed previously (see III.3.1). Stability is more difficult to achieve for methods of moments where time-integration can produce non-realisable moment sets, *i.e.* vector of moments that cannot be associated to any well-defined NDF [Vikas et al., 2011], or if a quadrature method fails to produce an accurate quadrature rule.

Over the course of this work, a strong focus has been made on comparing the class method with all three considered quadrature methods (QMOM, EQMOM and the Maximum entropy approach latter referred to as MaxEnt) on each of these aspects: memory use, numerical cost and stability. This is detailed hereafter by first explaining the comparison methodology and then the observed results. Full details of this comparison can be found in Pigou et al. [2017].

### III.5.1 Comparison methodology

In order to assess methods of moment against the class method, two biological cultures have been simulated. They are both shortly described hereafter and have been fully detailed in Pigou et al. [2017].

The first simulation set-up allowed assessing the stability of all methods for the simulation of a stressed homogeneous system, and to measure the dependency of results on the number of resolved moments.

The second simulation set-up is that of a large-scale fedbatch culture already simulated in Pigou and Morchain [2015] using the class method and whose results are reproduced using moment methods. This allows to assess the accuracy and stability of these methods in heterogeneous systems.

#### Stressed chemostat culture

The first simulation case corresponds to experimental results shown previously in Chapter II, Fig. II.5, from Kätterer et al. [1986]. It is a chemostat culture started with a dilution rate $D = 0.10h^{-1}$ which is suddenly shifted toward a higher dilution rate $D' = 0.42h^{-1}$. In this set-up, the bioreactor can be considered as homogeneous, and the biological population only experiences a single environmental perturbation. This allows for a comparison of methods in a tightly controlled set-up. The time integration was performed by a first-order explicit Euler method with a fixed time-step, shared by all methods, chosen so that all results are independent from that time-step.

In order for these simulations to quantitatively compare against experimental results from Kätterer et al. [1986], constants of the *E. coli* metabolic model from Pigou and Morchain [2015] have been changed to match the conversion yields of *Candida tropicalis*.

#### Large-scale fedbatch culture

Vrábel et al. [2001] performed simulations of a large-scale (22m$^3$) fedbatch culture of *E. coli* using the metabolic model from Xu et al. [1999] and the hydrodynamic Compartment Model Approach developed by Vrábel et al. [1999]. We reproduced these simulations as detailed in Pigou and Morchain [2015] by integrating the Population Balance Model for the dynamic adaptation of *E. coli* growth-rate as detailed previously (see III.2.1) using the class method to solve the PBE. Please refer to Pigou and Morchain [2015]; Pigou et al. [2017] for a full presentation of the hydrodynamic modelling by Compartment Model Approach (CMA).

We reproduced these simulations by using moment methods (QMOM, EQMOM and MaxEnt) with a number of moment based on results from the simulation of the stressed chemostat culture.

### III.5.2 Comparison results

#### Stressed chemostat cuture

Fig. III.3 sums-up simulation results for biomass and glucose concentrations for each method and different orders of resolution. The first observation is that all methods manage to correctly reproduce the overall dynamics of the system. This dynamics is mainly related to the accuracy of the PBE treatment which describes the growth-rate dynamic evolution (see Eq. (III.16)). In particular, the moment equation (Eq. (III.22)) of the last resolved moment $m_{N-1}$ requires the knowledge of the unknown moment $m_N$. The predicted dynamics will then depend on the approximation of this unknown moment from the quadrature rule yield by each method. An analysis conducted on this aspect has shown that all methods and order correctly predicted this unknown moment with less than 0.2% of relative error. This implies that all moment method correctly predict the dynamics indeed.

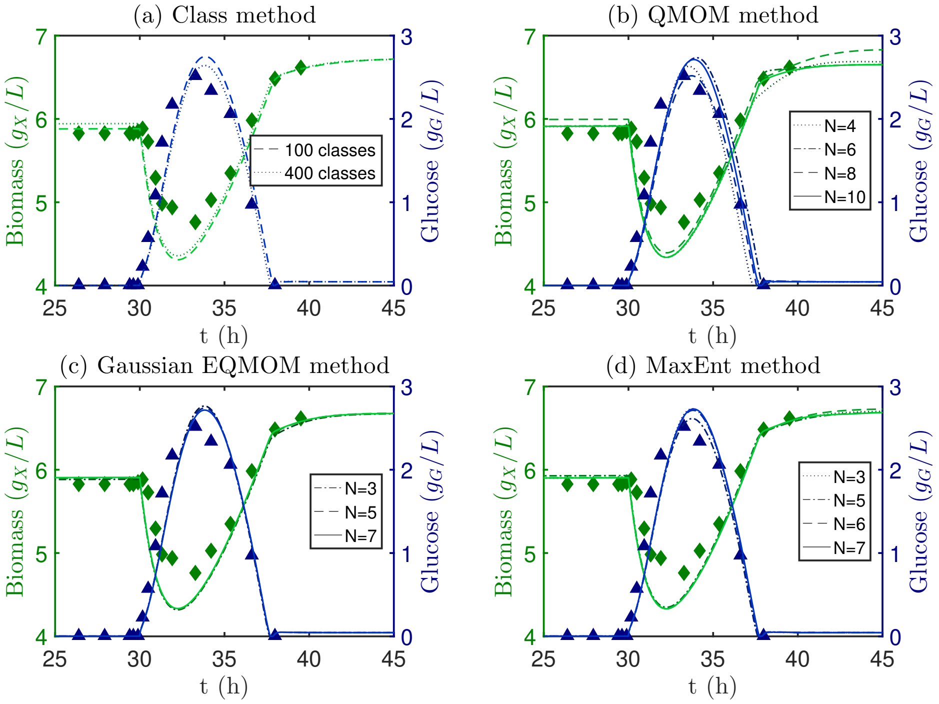


Figure III.3: Comparison of experimental data from Kätterer et al. [1986] against simulation results from each method. For methods of moments, $N$ is the number of resolved moments.

Clearly, all methods yield accurate approximate of unknown moment. A more sensitive variable will be the acetate production. As stated in Chapter II, Section II.3, the metabolic model for *E. coli* predicts that cells whose growth-rate $\mu_c$ is lower than the equilibrium growth rate $\mu^*$ will produce acetate through the overflow metabolism. Here, we reused this model by only changing some conversion yields to fit experimental data in terms of glucose and biomass concentration (see Fig. III.3) but this overflow metabolism is still enabled. Since acetate production is closely related to the shape of the distribution $n(\mu_c)$, the acetate concentration should be sensitive to methods order and accuracy. Predicted evolutions of these concentrations are given in Fig. III.4.

Let $A_M(t)$ be the evolution of acetate concentration as predicted by moment methods, and $A_C(t)$ the same evolution as predicted by the class method with 400 classes. Fig. III.5 summarizes the error between moment methods against the class method with

$$\varepsilon = \frac{\int_{t_0}^{t_f} |A_M(t) - A_C(t)| \mathrm{d}t}{\int_{t_0}^{t_f} A_C(t) \mathrm{d}t} \tag{III.44}$$

with $t_0 = 25$h and $t_f = 45$h. This figure shows that in terms of accuracy, MaxEnt is the more accurate method, even with as low as 3 moments, for the prediction of acetate production in this stressed chemostat culture. EQMOM and MaxEnt share similar accuracy as soon as 5 moments are tracked. Finally, QMOM requires twice as many moments as other methods to reach a similar accuracy of 5%.

While performing these simulations, we noticed stability issues with MaxEnt as soon as 7 moments were used. Stability could still be maintained at the cost of extra operations as detailed in Pigou et al. [2017] but no more than 7 moments could be tracked. In terms of

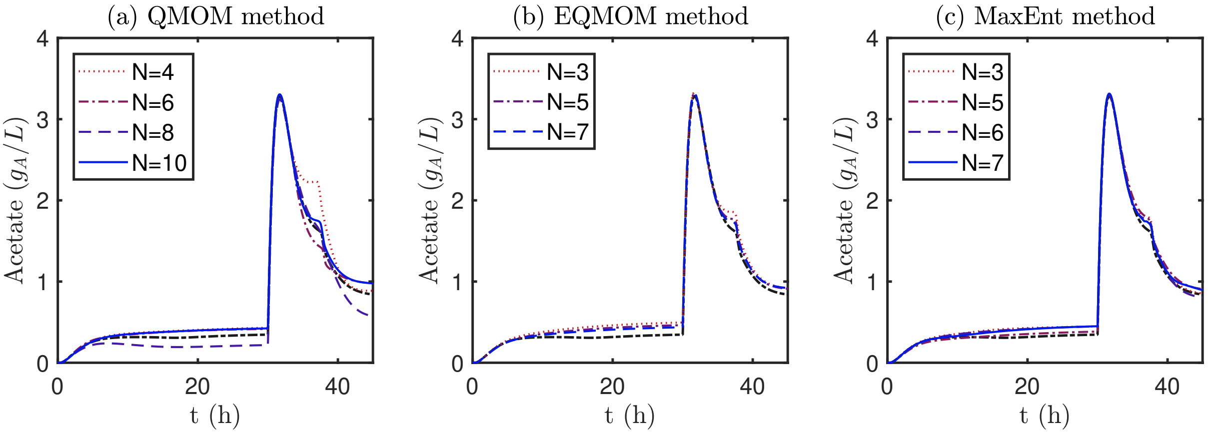


Figure III.4: Evolution of acetate concentrations as predicted by each moment method. Black dash-dotted line: results from class method using 400 classes.

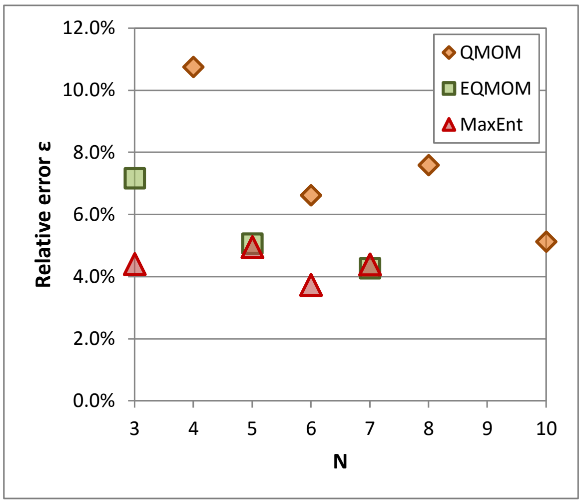


Figure III.5: Relative error on acetate concentration predictions for all methods and orders. The 400 classes method serves as a reference.

numerical costs, Fig. III.6 details the run-time associated with each method. The QMOM method appears twice in this figure because two different algorithms were tested to perform the quadrature computation (PDA: Product Difference Algorithm, WA: Wheeler Algorithm, also referred to as Chebyshev Algorithm is Section III.6). The “Blank” measurement corresponds to simulations were reaction terms were not computed in order to isolate bioreaction and population balance related computations from other tasks performed by the simulation software. A complete analysis of these run-times is provided in Pigou et al. [2017], main observations are:

- Overall, all method have a similar order of magnitude in term of associated run-time even though this numerical cost has different sources for each method.

- The cost associated to the class method mainly comes from calling the metabolic model for each class, as well as computing the redistribution term by integrating a skew-normal distribution over each class (see Eq. (III.8)).

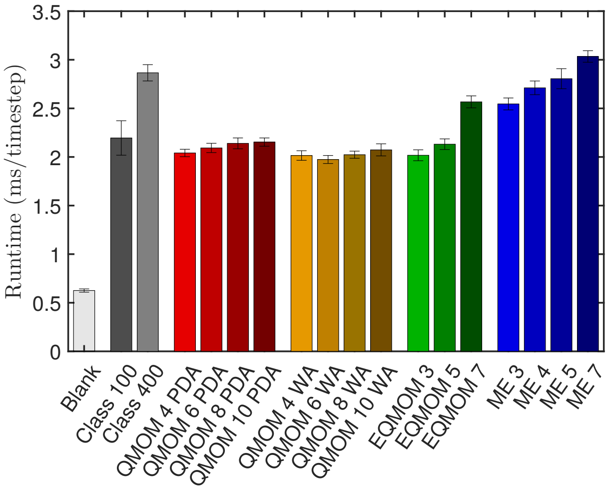


Figure III.6: Mean run-time per time-step for each tested method and order (ms/ts) (± standard-deviation measured on 20 simulations).

- The iterative process underlying the MaxEnt approach often converges in a single iteration, but this increases up to 10 iteration by time-step right after the dilution-rate shift-up. Simulating the next few hours following this shift significantly increases the mean observed run-time for this method, but it should be quite fast if used in systems with minor perturbations.

- QMOM has a computational cost hardly dependent on the method order and is the fastest method as it yield the lower number of quadrature nodes.

- Analytical solutions were used for 3 and 5 moments with EQMOM which explains the significant increase in numerical cost for $N = 7$.

One surprising aspect in this comparison of run-time is the low difference between the class method and QMOM. One would expect a significantly lower cost for QMOM compared to the class method which is not visible here. This observation has been latter linked to the use of a simulation software based on Oriented Object Programming in MATLAB [2016] which happened to be badly optimized. The transfer of data between the time integration routine and modelling blocks (code for quadrature computations, for metabolic model computations, ...) significantly slowed down the simulations. Therefore, all previous comparisons only have a qualitative value but not a quantitative one. A better approach would have been to compare the number of required Floating-Point Operations per time-step but this is a less accessible metrics. The key element of comparison will then be stability and accuracy of quadrature methods.

### Large-scale fedbatch culture

Initial simulations of a homogeneous system led to the selection of EQMOM and MaxEnt with $N = 5$ as well as QMOM with $N = 10$ to perform simulations of a $22\text{m}^3$ heterogeneous fedbatch culture. Figure III.7 details the volumetric means of glucose, biomass and acetate concentrations, as well as local acetate production/consumption rates, at three different heights referred to as *top*, *middle* and *bottom*. Once again, all moment methods

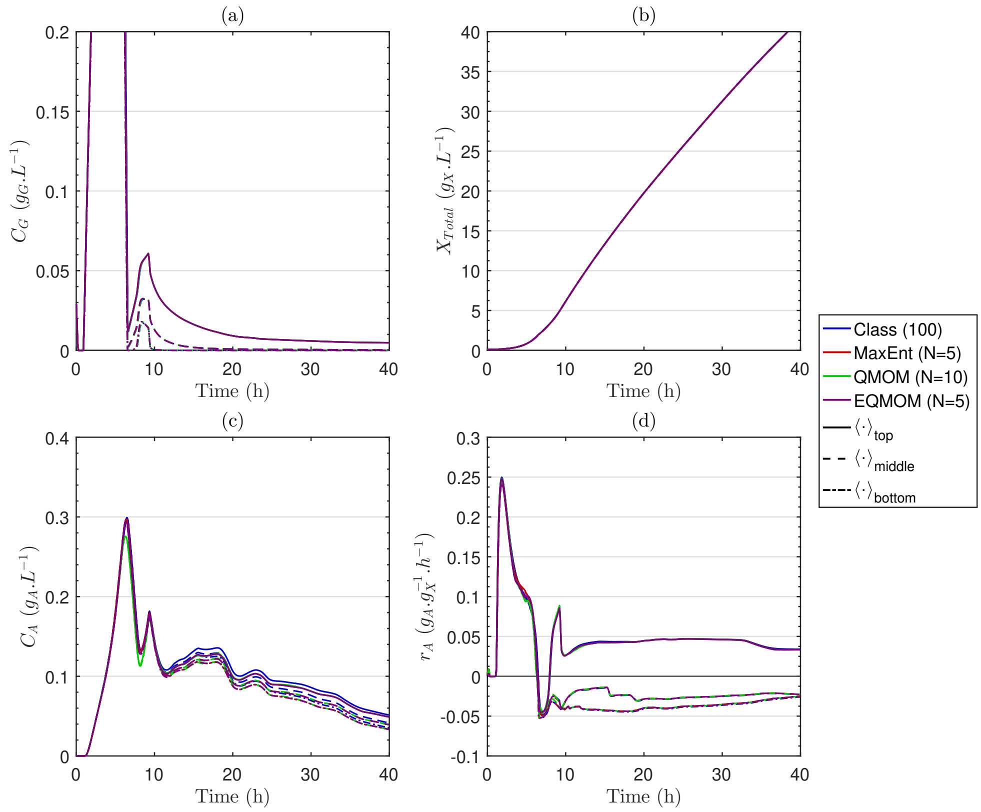


Figure III.7: Simulation results for the different population balance methods in the heterogeneous fedbatch culture. (a) Glucose concentration, (b) Total biomass concentration, (c) Acetate concentration and (d) Acetate specific production rate.

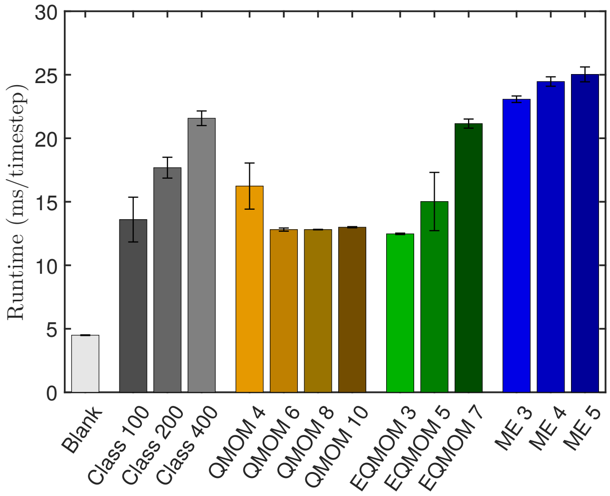


Figure III.8: Mean run-time per time-step for each method (ms/time-step ± standard deviation measured on 5 simulations per method and order).

closely match the results obtained with the class method. This confirms that 5 moments are sufficient to account for biological heterogeneity with the proposed modelling.

In terms of stability and numerical cost, Fig. III.8 is similar to Fig. III.6 by showing run-times per time-step for each method. The analysis of these run-times is similar to the homogeneous chemostat with the only noticeable change being that the Maximum entropy approach could not be tested with more than 5 moments due to stability issues, and its numerical cost is here significantly higher than that of QMOM and analytical EQMOM.

### III.5.3 Conclusion

Overall, all tested quadrature methods gave satisfactory accuracy compared to the class method, but required significantly less resolved variables. In terms of tracked moments, EQMOM and MaxEnt make a better use of the information embedded within the first few integer moments than QMOM which leads to a good accuracy with as few as 3 to 5 moments. Meanwhile, QMOM required twice as much moments as the other methods to reach similar accuracy. Little time after these developments were published [Pigou et al., 2017], an other group performed a similar analysis by comparing QMOM, MaxEnt and a polynomial closure method [Müller et al., 2017]. They observed too the lack of accuracy of QMOM at equivalent number of resolved moments compared to other methods.

A key observation was the lack of stability related to the Maximum Entropy approach. This had already been observed by Vié et al. [2013] and has been confirmed to be a major drawback of this method for simulations of large-scale bioreactors. Therefore, MaxEnt has been discarded from future use in this context.

QMOM and EQMOM shared similar computational costs, but EQMOM required less resolved moments. This implies lower cost for the resolution of coupled mass-balance equations, especially if the PBM is used within a CFD framework, but this also improves the simulation stability. Indeed, the resolution of high order moment sets may lead more easily to the appearance of unrealisable moment sets which would constitute a strong source of instability. Therefore, EQMOM seems to be the most relevant choice for the resolution of Population Balance Equations in the context of large-scale bioreactors.

## III.6 Improving the Extended Quadrature Method of Moments

### III.6.1 Developments background

While performing simulations presented previously when comparing methods to solve PBEs, all variants of EQMOM were tested even though only results based on Gauss-EQMOM have been presented. This is due to instabilities that have been observed for the Log-normal, Gamma and Beta variations. These instabilities led to failures in the quadrature computation when these kernels were used for the EQMOM moment-inversion procedure, only the Gauss kernel did not fail.

Using a mixture of Gaussian kernels –defined on $\mathbb{R}$– to approximate a NDF defined on $\Omega_\mu = [0, \mu_{\text{lim}}]$ or $\Omega_\mu = [0, +\infty[$ obviously presents an intrinsic weakness, and actually led to the appearance of some quadrature nodes with negative growth-rate in results presented previously (see III.5.2). This was corrected by truncating and renormalizing the approximated NDF over $\Omega_\mu$ for the integration of bioreaction rates over the population, this approach was however unsatisfactory.

Results from Pigou et al. [2017] show that 5 moments are sufficient for accurate simulations of large-scale bioreactors when EQMOM is used to provide closure. Consequently, an attempted approach to prevent instabilities was to develop analytical solutions for 2-nodes EQMOM reconstructions for other kernels. Such an analytical solution was only described for the Gauss kernel by Chalons et al. [2010] and supplementary analytical solutions have been identified for Gamma, Laplace and Log-normal kernels. The identification of Laplace and Weibull distributions as kernels compatible with the EQMOM method happened simultaneously with these analytical solution developments.

Unfortunately, using analytical solutions for the Log-normal and Gamma kernels did not provide an efficient response to the previously observed stability issues. The explanation for this came from Nguyen et al. [2016]: the EQMOM reconstruction may happen not to be defined, it may be possible that the last moment may not be conserved by an EQMOM approximation. Nguyen et al. [2016] then proposed a new EQMOM moment-inversion procedure which

- is more stable than previous ones by adding a realisability check at each iteration of the procedure;
- makes use of a faster root-finding algorithm compared to previous implementations, namely the Ridder's method;
- switches from a root-finding algorithm toward a minimization algorithm if no EQMOM reconstruction exists.

If we identified these recent improvements of EQMOM sooner, this would have had an impact on the article presenting the method comparison [Pigou et al., 2017] by

1. allowing the use of LogN-EQMOM or Gamma-EQMOM in place of Gauss-EQMOM;
2. improving the stability even for higher number of moments;

3. reducing the numerical cost observed for EQMOM.

However, the strong focus made on EQMOM to tackle these instabilities, coupled with multiple readings associated to the mathematical fundamentals of Gauss-Quadrature, orthogonal polynomial and moment theories [Chihara, 1978; Dette and Studden, 1997; Gautschi, 2004], led to the identification of a whole new EQMOM moment-inversion procedure of significantly lower numerical cost. Main aspects of this new procedure are laid out hereafter, followed by quantification of gains provided by these new developments.

## III.6.2 Description of EQMOM moment-inversion procedures

The new EQMOM moment-inversion procedure has been fully detailed in Pigou et al. [2018a] and Pigou et al. [2018b]. Main ideas are summarized hereafter but please refer to these articles for complete explanations.

#### Gauss quadratures - QMOM

First, EQMOM is based on the Quadrature Method of Moments or more generally on Gaussian quadratures whose basics are at the core of the new procedure and need to be reminded. Let $n(\mu)$ be a NDF defined on a support $\Omega_\mu$ whose first $N+1$ integer moments are $\boldsymbol{m}_N = [m_0, \dots, m_N]^T$ with

$$m_k = \int_{\Omega_\mu} \mu^k n(\mu) \mathrm{d}\mu \tag{III.45}$$

A P-nodes Gauss-quadrature associated with the NDF $n(\mu)$ is defined by its weights $\boldsymbol{w}_P$ and nodes $\boldsymbol{\xi}_P$ such that

$$\int_{\Omega_\mu} \mu^k n(\mu) \mathrm{d}\mu = m_k = \sum_{i=1}^{P} w_i \xi_i^k \qquad \forall k \in \{0, \dots, 2P-1\} \tag{III.46}$$

The computation of $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ constitutes the core of QMOM and is a two steps process. It is well-known from the theory of orthogonal polynomials [Gautschi, 2004] that any NDF is associated to a sequence of monic orthogonal polynomials denoted $\pi_k$ –with $k$ the order of the polynomial– such that

$$\int_{\Omega_\mu} \pi_i(\mu) \pi_j(\mu) n(\mu) \mathrm{d}\mu = 0 \qquad \text{for } i \neq j. \tag{III.47}$$

These polynomials satisfy a three-terms recurrence relation

$$\pi_{k+1}(\mu) = (\mu - a_k)\pi_k(\mu) - b_k \pi_{k-1}(\mu) \tag{III.48}$$

with $a_k$ and $b_k$ the recurrence coefficients specific to the NDF $n(\mu)$, $\pi_{-1}(\mu) = 0$ and $\pi_0(\mu) = 1$. Let $\boldsymbol{J}_n$ be the $n \times n$ Jacobi matrix associated to $n(\mu)$. This is a tridiagonal matrix defined as

$$\boldsymbol{J}_n = \begin{pmatrix} a_0 & \sqrt{b_1} & & 0 \\ \sqrt{b_1} & a_1 & \ddots & \\ & \ddots & \ddots & \sqrt{b_{n-1}} \\ 0 & & \sqrt{b_{n-1}} & a_{n-1} \end{pmatrix} \tag{III.49}$$

Weights and nodes of the quadrature rule from Eq. (III.46) are obtained from spectral properties of $\boldsymbol{J}_P$. Nodes $\boldsymbol{\xi}_P$ are the eigenvalues of $\boldsymbol{J}_P$ and weights are given by

$$w_i = m_0 v_{1,i}^2 \tag{III.50}$$

where $v_{1,i}$ is the first component of the normalised eigenvector belonging to the eigenvalue $\xi_i$. QMOM and Gauss-quadratures then rely on

1. the computation of recurrence coefficients $\boldsymbol{a}_{P-1} = [a_0, \ldots, a_{P-1}]^T$ and $\boldsymbol{b}_{P-1} = [b_1, \ldots, b_{P-1}]^T$ from $\boldsymbol{m}_{2P-1}$;
2. the computation of eigenvalues and eigenvectors of $\boldsymbol{J}_P$.

The first step is easily performed using the Chebyshev algorithm [Wheeler, 1974]. The second step can be tackled using Francis or Jacobi algorithms [Ford, 2015], but these have a high numerical cost. The following improvements of EQMOM are actually mainly related to reducing the number of call to these algorithms to one single call as further explained hereafter.

### Moment-inversion procedure

EQMOM is a method that approximates a NDF $n(\mu)$ from its first $2P+1$ integer moments $\boldsymbol{m}_{2P}$ as

$$n(\mu) \approx \tilde{n}(\mu) = \sum_{i=1}^{P} w_i \kappa(\mu, \xi_i, \sigma) \tag{III.51}$$

The goal of the EQMOM moment-inversion procedure is to identify the weights $\boldsymbol{w}_P = [w_1, \ldots, w_P]^T$, the nodes $\boldsymbol{\xi}_P = [\xi_1, \ldots, \xi_P]^T$ and the shape parameter $\sigma$ from the first $2P+1$ integer moments $\boldsymbol{m}_{2P}$ of $n(\mu)$, under the constraint $\tilde{\boldsymbol{m}}_{2P} = \boldsymbol{m}_{2P}$ with $\tilde{\boldsymbol{m}}_{2P}$ the moments of $\tilde{n}(\mu)$.

Yuan et al. [2012] noticed a linear relationship $\tilde{\boldsymbol{m}}_n = \boldsymbol{A}_n(\sigma) \cdot \boldsymbol{m}_n^*$ with

$$n^*(\mu) = \sum_{i=1}^{P} w_i \delta(\mu, \xi_i), \quad \boldsymbol{m}_n^* = \begin{bmatrix} m_0^* \\ m_1^* \\ \vdots \\ m_n^* \end{bmatrix}, \quad m_k^* = \int_{\Omega_\mu} \mu^k n^*(\mu) \mathrm{d}\mu \tag{III.52}$$

where $\delta(\mu, \xi)$ is the Dirac distribution defined in Eq. (III.7). $\boldsymbol{A}_n(\sigma)$ is a lower-triangular $(n+1) \times (n+1)$ matrix whose elements depend on the chosen EQMOM kernel. By definition of Gauss quadrature, if one applies QMOM to the vector $\boldsymbol{m}_{2P-1}^*$, the returned weights and nodes will be $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$. Therefore, if $\sigma$ is known, computing $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ is only a matter of (i) computing $\boldsymbol{m}^*_{2P-1} = \boldsymbol{A}^{-1}_{2P-1} \times \boldsymbol{m}_{2P-1}$ and (ii) extract $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ by computing a Gauss quadrature based on $\boldsymbol{m}^*_{2P-1}$. The challenge is then to identify the value of the shape parameter $\sigma$.

The initial approach developed by Yuan et al. [2012] and latter improved by Nguyen et al. [2016] formulates a scalar non-linear function $D_{2P}(s)$, whose root is the shape parameter, $D_{2P}(\sigma) = 0$. The computation of that function is as follow:

1. Compute $\boldsymbol{m}^*_{2P-1}(s) = \boldsymbol{A}^{-1}_{2P-1}(s) \times \boldsymbol{m}_{2P-1}$;
2. Apply the QMOM method to $\boldsymbol{m}^*_{2P-1}(s)$:
    a) Apply the Chebyshev algorithm to $\boldsymbol{m}^*_{2P-1}$ to access recurrence coefficients $\boldsymbol{a}_{P-1}(s)$ and $\boldsymbol{b}_{P-1}(s)$;
    b) Compute eigen-properties of $\boldsymbol{J}_P(s)$ to access $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$.
3. Compute $\tilde{m}_{2P}(s) = \sum_{i=1}^{P} w_i \int_{\Omega_\mu} \mu^{2P} \delta_\sigma(\mu, \xi_i) \mathrm{d}\mu$;
4. Compute $D_{2P}(s) = m_{2P} - \tilde{m}_{2P}(s)$.

By construction, and for any $s$ value, the resulting vectors $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ will verify $\tilde{\boldsymbol{m}}_{2P-1} = \boldsymbol{m}_{2P-1}$, but only the specific value $s = \sigma$ will yield $\tilde{\boldsymbol{m}}_{2P} = \boldsymbol{m}_{2P}$, hence $D_{2P}(\sigma) = 0$. Finding $\sigma$ can then be done by using a root-finding algorithm such as dichotomic or secant-bounded methods, or the more efficient Ridder's method as suggested by Nguyen et al. [2016].

One can notice in this initial approach that it is necessary to compute eigenvalues and eigenvectors of $\boldsymbol{J}_P(s)$ for each tested $s$ value, which is an expensive operation. We then propose a new method which allows identifying $\sigma$ at a lower numerical cost.

The new moment-inversion algorithm is based on the fact that, by construction, $\boldsymbol{m}^*_{2P}(\sigma)$ is the vector of the first $2P+1$ moments of a convex mixture of $P$ Dirac distributions. It is well known that such a moment set has a very specific property: it is strictly located on the boundary of the realisable moment space $\mathcal{M}_{2P}$. This space is the set of all moment sets $m_{2P}$ induced by all possible NDFs defined on $\Omega_\mu$. Therefore, identifying $\sigma$ is equivalent to finding $s$ such that $\boldsymbol{m}^*_{2P}(s)$ is located on the boundary of $\mathcal{M}_{2P}$. By construction, if $s < \sigma$, $\boldsymbol{m}^*_{2P}(s)$ will be within this space, and otherwise, if $s > \sigma$, $\boldsymbol{m}^*_{2P}(s)$ will be outside of the realisable moment space. The question is then to determine whether of moment set is realizable, which actually comes at really low cost.

If the support of the NDF is $\Omega_\mu = \mathbb{R}$, it is known that the realisability of a moment set $\boldsymbol{m}_{2P}$ if equivalent to the positivity of the recurrence coefficients $\boldsymbol{b}_P$ defined previously [Chihara, 1978; Favard, 1935]. If the support is the positive half-line $\Omega_\mu = ]0, +\infty[$, the realisability can be checked by the positivity of elements of $\boldsymbol{\zeta}_{2P} = [\zeta_1, \dots, \zeta_{2P}]^T$ defined by

$$\zeta_{2k} = \frac{b_k}{\zeta_{2k-1}}, \qquad \zeta_{2k+1} = a_k - \zeta_{2k} \tag{III.53}$$

with $\zeta_1 = a_0 = m_1/m_0$ [Shohat and Tamarkin, 1943]. Finally, if the support is the bounded segment $\Omega_\mu = ]0, 1[$, a moment set is realisable if each of its canonical moments $\boldsymbol{p}_{2P} = [p_1, \dots, p_{2P}]^T$ is located in $]0, 1[$ with

$$p_k = \frac{\zeta_k}{1 - p_{k-1}}, \quad p_1 = m_1/m_0 \tag{III.54}$$

Table III.1: Vector of realisability criterion $\boldsymbol{c}_{2P}$ and associated realisability domain $\mathcal{C}_{2P}$ for a vector of $2P+1$ moments on classical supports.

| $\Omega_\xi$ | $\boldsymbol{c}_{2P}$ | $\mathcal{C}_{2P}$ |
|---|---|---|
| $]-\infty,+\infty[$ | $\boldsymbol{b}_P$ | $]0,+\infty[^P$ |
| $]0,+\infty[$ | $\boldsymbol{\xi}_{2P}$ | $]0,+\infty[^{2P}$ |
| $]0,1[$ | $\boldsymbol{p}_{2P}$ | $]0,1[^{2P}$ |

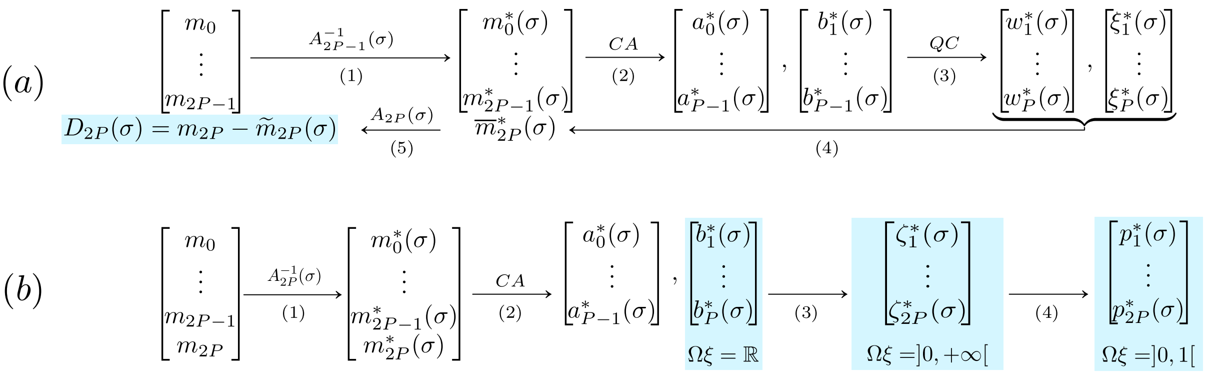


Figure III.9: Steps for the computation of convergence criterion in (a) the original EQMOM moment-inversion procedure [Nguyen et al., 2016] and (b) the new proposed approach. Inspired by Figure 1 from Nguyen et al. [2016]. CA: Chebyshev Algorithm. QC: Quadrature Computation.

It is then possible to define a vector of realisability criterion $\boldsymbol{c}_{2P}$ and its associated realisability domain $\mathcal{C}_{2P}$ for each support as summarized in Table III.1. For each $s$ value, one can compute the vector $\boldsymbol{c}_{2P}^*(s)$ associated to $\boldsymbol{m}_{2P}^*(s)$. If $\boldsymbol{c}_{2P}^*(s) \in \mathcal{C}_{2P}$, then $s < \sigma$. If $\boldsymbol{c}_{2P}^*(s) \notin \mathcal{C}_{2P}$, then $s \geq \sigma$. The situation $s = \sigma$ is characterised by $\boldsymbol{c}_{2P-1}^*(s) \in \mathcal{C}_{2P-1}$ and $c_{2P}^*(s) = 0$.

Complete algorithms based on the Ridder's method to iteratively select new $s$ values to be tested are detailed in Pigou et al. [2018a], as well as multiple subtleties, in particular to handle situations where the EQMOM reconstruction does not exist. Computation steps of both the original objective function $D_{2P}(s)$ and the new vector objective function are illustrated in Fig. III.9. Note that the computation of $\boldsymbol{\zeta}_{2P-1}(s)$ or $\boldsymbol{p}_{2P-1}(s)$ was also necessary in the approach developed by Nguyen et al. [2016] before the quadrature computation step to ensure stability.

### III.6.3 Measuring the performance gains

The new EQMOM moment-inversion procedure only requires the computation of realisability criterion associated to the vector of degenerated moments $\boldsymbol{m}_{2P}^*(s)$ in order to identify $\sigma$. These computations were already performed in the original approach to ensure the realisability of $\boldsymbol{m}_{2P}^*(s)$ prior to the quadrature computation and ulterior steps. Therefore, it is obvious that the new approach will have a lower numerical cost. A strong focus was made in Pigou et al. [2018a] in quantifying these performance gains.

Performance was measured on randomly generated moment sets though three metrics:

- Number of Floating-Point Operations (FLOP) required for the whole moment-inversion procedure;
- Number of tested $s$ values before convergence is reached and $\sigma$ is identified;
- CPU runtime for each call to the procedure.

Both the original and new approaches have been implemented in MATLAB [2016]. All source codes used for these comparisons are available as Supplementary Data of Pigou et al. [2018a] along with the complete methodology.

In terms of computational cost, Table III.2 details the reduction of required FLOP per call to the moment-inversion procedure for five kernels, out of the six available. It is shown that the higher the order of the reconstruction, the more significant the gain provided by the new procedure, ranging from about 60% for $P = 2$ up to 92% for $P = 5$. This significant improvement of EQMOM can be attributed to two phenomena.

Table III.2: Gain in FLOP when using the new EQMOM moment-inversion procedure.

| | $P = 2$ | $P = 3$ | $P = 4$ | $P = 5$ |
|---|---|---|---|---|
| Gauss | $59.1\%_{\pm 12.3\%}$ | $84.2\%_{\pm 3.5\%}$ | $87.9\%_{\pm 2.5\%}$ | $88.0\%_{\pm 13.1\%}$ |
| Laplace | $64.2\%_{\pm 10.7\%}$ | $87.5\%_{\pm 2.7\%}$ | $91.0\%_{\pm 1.7\%}$ | $91.6\%_{\pm 8.9\%}$ |
| Log-normal | $58.9\%_{\pm 20.0\%}$ | $85.6\%_{\pm 5.8\%}$ | $89.1\%_{\pm 4.3\%}$ | $93.3\%_{\pm 3.5\%}$ |
| Gamma | $58.2\%_{\pm 18.8\%}$ | $82.1\%_{\pm 7.3\%}$ | $85.7\%_{\pm 6.8\%}$ | $91.6\%_{\pm 5.4\%}$ |
| Weibull | $67.7\%_{\pm 15.1\%}$ | $87.4\%_{\pm 5.1\%}$ | $90.0\%_{\pm 4.2\%}$ | $94.2\%_{\pm 3.6\%}$ |

The first one is detailed in Table III.3 which presents the reduction in the number of iterations to reach the convergence on $\sigma$. This reduction is mainly due to the fact that in the moment-inversion procedure proposed by Nguyen et al. [2016], if a moment set $\boldsymbol{m}^*_{2P-1}(s)$ happens not to be realisable, the objective function $D_{2P}(s)$ is set to an arbitrarily high negative value such as $-10^{100}$. Using arbitrary value in root-search algorithm slows down the convergence as it leads to a non optimal choice for the next tested value. On the opposite, the way the new procedure makes use of all elements of $\boldsymbol{c}_{2P}(s)$, one after the other, prevents the need of such arbitrary values and therefore allows reaching a converged state in a few iterations.

Table III.3: Reduction in number of tested $s$ values when using the new EQMOM moment-inversion procedure.

| | $P = 2$ | $P = 3$ | $P = 4$ | $P = 5$ |
|---|---|---|---|---|
| Gauss | $8.6\%_{\pm 27.7\%}$ | $40.9\%_{\pm 17.7\%}$ | $54.2\%_{\pm 12.8\%}$ | $53.0\%_{\pm 55.2\%}$ |
| Laplace | $8.6\%_{\pm 28.3\%}$ | $41.1\%_{\pm 17.6\%}$ | $54.3\%_{\pm 12.7\%}$ | $53.2\%_{\pm 54.4\%}$ |
| Log-normal | $8.8\%_{\pm 45.9\%}$ | $21.9\%_{\pm 32.8\%}$ | $30.5\%_{\pm 27.0\%}$ | $49.5\%_{\pm 23.8\%}$ |
| Gamma | $15.5\%_{\pm 38.2\%}$ | $24.7\%_{\pm 31.1\%}$ | $34.9\%_{\pm 30.1\%}$ | $57.2\%_{\pm 24.9\%}$ |
| Weibull | $26.3\%_{\pm 35.0\%}$ | $27.2\%_{\pm 30.2\%}$ | $32.4\%_{\pm 28.1\%}$ | $54.0\%_{\pm 25.8\%}$ |

The second reason for the net decrease of computational cost was expected and is related to the fact that in the former approach, the quadrature computation is performed in almost each iteration loop. More specifically, this step is called if $\boldsymbol{m}^*_{2P-1}(s)$ is happens to be realisable. In the new approach, the quadrature computation is not required any more to identify $\sigma$. It is only performed once when $\sigma$ is identified to obtain $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$.

We could not find a moment set for which the former approach was faster than the new one. Note that only randomly generated sets were used for the comparison. Therefore, the benefits of this new EQMOM version are beyond the specific application case of bioreactors modelling and should benefit all communities using this method.

Moreover, we observed an unexpected yet appreciable feature of the new approach: its stable –bounded– numerical cost. Figs. III.10 and III.11 show the raw and normalized distributions of floating point cost for the Gauss and Log-normal kernels with $P = 4$. In both example, the raw distribution shows a lower numerical cost for the new approach as stated previously. Normalized distribution shows that the standard-deviation of the distribution of numerical costs is lower for the new approach. This is shown on two examples but this has been observed for all kernels and orders. This lower variance in numerical cost is especially marked for EQMOM kernels defined on $\Omega_\mu = ]0, +\infty[$. The fact that the numerical cost is more consistent from one moment set to another will be particularly beneficial to facilitate load-balancing in highly parallelized simulations, for instance for PBE-CFD coupling.

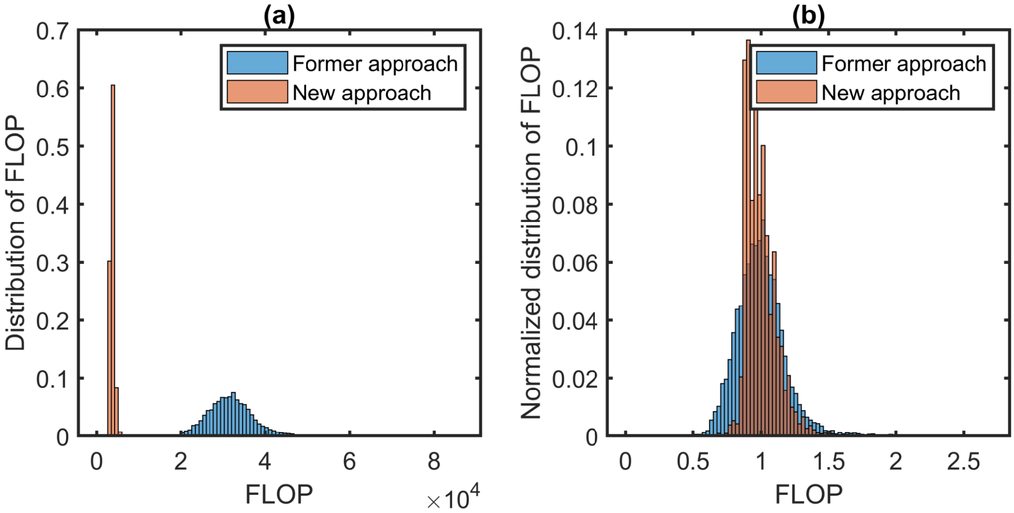


Figure III.10: Distribution of numerical cost (FLOP) over $10^4$ calls to moment-inversion procedures with Gauss-EQMOM and $P = 4$.

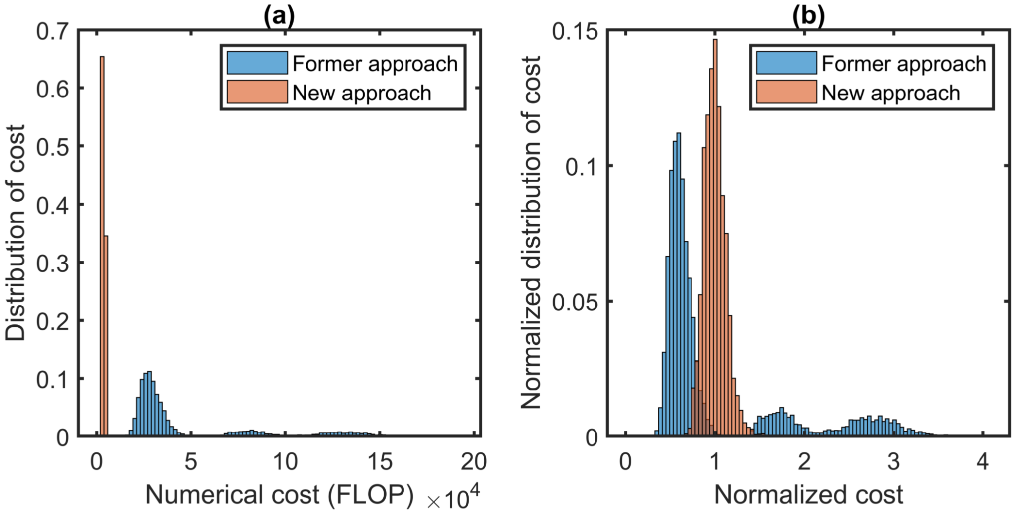


Figure III.11: Distribution of numerical cost (FLOP) over $10^4$ calls to moment-inversion procedures with LogN-EQMOM and $P = 4$.

## III.7 Reducing the 5-D Population Model to a 1-D treatment

This chapter mainly focused on numerical methods that suit mono-dimensional Population Balance Models. However, Chapter V will describe the simulation of industrial cultures of *S. cerevisiae*, whose metabolism is described by accounting for five dynamic biological properties (see Eq. (III.12)). Accounting for multi-variate PBEs for biological systems is actually one of the goal of a project carried out by Quedeville et al. [2018] in parallel to this PhD thesis. For now, we will need to simplify the treatment of the 5D PBE in order to be able to apply 1D methods.

We propose to simplify the PBE treatment by analysing the involved time-scales. The PBM allows tracking the two-way coupling between physical (concentration) heterogeneity with biological heterogeneity. Therefore, we propose to first analyse time-scales associated to concentration fluctuations as perceived by micro-organism against time-scales associated to the evolution of their biological properties.

We first remind that, in industrial bioreactors, order of magnitudes for meso and macro-mixing time scales are respectively about 0.1 to 1s and 10 to 100s [Delafosse, 2008]. The first four biological properties described in the *S. cerevisiae* metabolic model ($\mu_G^{(b)}$, $\mu_E^{(b)}$, $\mu_A^{(b)}$ and $\varphi_O^{(b)}$) evolve over time with a characteristic time-scale $T \approx 1\text{h}$ (see Table II.5). Therefore, we can infer from these time-scales that the slow biological regulations that occur to adapt these biological properties will act as a low-pass filter. The distributions of growth and oxidative capacities will be mainly homogeneous in a large fermenter.

Moreover, Chapter V will focus on the simulation of substrate-limited fedbatch cultures. In short, this implies that while glucose concentration will locally be high at the feeding point, most of the fermenter volume will be depleted in terms of glucose. Therefore, most of the time cells will be in a poor environment and will go through high-glucose concentration areas every few minutes (circulation time) for a few seconds (mean residence time in the feeding area). In these conditions, we can expect the distribution of growth capability on glucose $\mu_G^{(b)}$ to be nearly monodispersed at low values.

The conclusion of this first analysis is that, whilst growth and oxidative capabilities are dynamic properties, it is most likely not required to track their distribution in time and space. The tracking of mean population values might be sufficient.

On the other side, the glycolytic stress $\sigma$ is a peculiar property. In Chapter II, we modelled its evolution such that

- in stressing (high glucose concentration) environment, the stress increases with a time-scale $T_\sigma^{(u)} \approx 1\text{s}$;
- in non-stressing environment, the cell recovers from stress with a time-scale $T_\sigma^{(d)} \approx 1\text{min}$.

Therefore, while other properties are almost unresponsive to the occasional exposition to high glucose concentrations in the feeding area, the stress $\sigma$ will be strongly impacted. A cell travelling in a substrate-limited fedbatch culture will be exposed to the stressing environment of the feeding area for a few second which will be sufficient for its stress level to increase significantly. Multiple passes through this area are separated, in mean, by the fermenter circulation time ($\approx$ 1min) which allows for a recovery from previous stressing events. Depending on trajectories of each cell, areas where recently stressed cells will cross the path of unstressed micro-organisms will exist. Therefore we can *a priori* expect a strong effect of the stress heterogeneity over the biological population.

To account for the dynamic evolution of mean growth and oxidative capabilities, along with the stress heterogeneity, we may simply track uncoupled moments of the distribution $n(\widehat{p})$. Let theses moments be defined as

$$m_{k_1,k_2,k_3,k_4,k_5} = \int_{\Omega\mu_G}\int_{\Omega\mu_E}\int_{\Omega\mu_A}\int_{\Omega\varphi_O}\int_{\Omega\sigma} \mu_G^{k_1}\mu_E^{k_2}\mu_A^{k_3}\varphi_O^{k_4}\sigma^{k_5}\mathrm{d}\sigma\mathrm{d}\varphi_O\mathrm{d}\mu_A\mathrm{d}\mu_E\mathrm{d}\mu_G. \tag{III.55}$$

We will track following moments of the distribution

$$\boldsymbol{M} = \begin{bmatrix} m_{0,0,0,0,0} \\ m_{1,0,0,0,0} \\ m_{0,1,0,0,0} \\ m_{0,0,1,0,0} \\ m_{0,0,0,1,0} \\ m_{0,0,0,0,1} \\ \vdots \\ m_{0,0,0,0,N_{m\sigma}} \end{bmatrix} = \begin{bmatrix} X_L \\ \tilde{\mu}_G^{(b)}X_L \\ \tilde{\mu}_E^{(b)}X_L \\ \tilde{\mu}_A^{(b)}X_L \\ \tilde{\varphi}_O^{(b)}X_L \\ \tilde{\sigma}X_L \\ \vdots \\ m_{0,0,0,0,N_{m\sigma}} \end{bmatrix} \tag{III.56}$$

with $N_{m\sigma}$ the number of moments tracked along the stress direction to account for the heterogeneity of $\sigma$.

From these moments, we will produce a quadrature rule formed by weights $\boldsymbol{w}_Q = [w_1, \ldots, w_Q]$ and vectors of properties $\widehat{\boldsymbol{p}}_Q = [\widehat{p}_1, \ldots, \widehat{p}_Q]$ such that

$$\int_{\Omega\widehat{p}} f(\widehat{p})n(\widehat{p})\mathrm{d}\widehat{p} \approx \sum_{i=1}^{Q} w_i f(\widehat{p}_i). \tag{III.57}$$

Vectors of properties $\widehat{p}_i$ are defined as

$$\widehat{p}_i = \begin{bmatrix} \tilde{\mu}_G^{(b)} \\ \tilde{\mu}_E^{(b)} \\ \tilde{\mu}_A^{(b)} \\ \tilde{\varphi}_O^{(b)} \\ \sigma_i \end{bmatrix} = \begin{bmatrix} m_{1,0,0,0,0}/m_{0,0,0,0,0} \\ m_{0,1,0,0,0}/m_{0,0,0,0,0} \\ m_{0,0,1,0,0}/m_{0,0,0,0,0} \\ m_{0,0,0,1,0}/m_{0,0,0,0,0} \\ \sigma_i \end{bmatrix} \tag{III.58}$$

and $\boldsymbol{w}_Q$ and $\boldsymbol{\sigma}_Q = [\sigma_1, \ldots, \sigma_Q]$ are the weights and nodes of a quadrature formed from the vector of moments $[m_{0,0,0,0,0}, m_{0,0,0,0,1}, \ldots, m_{0,0,0,0,N_{m\sigma}}]$ using either QMOM or EQMOM method.

## III.8 Conclusion

In order to account for the heterogeneity of cell-attached properties in the modelling of bioreactors, the Population Balance Model approach has been selected to facilitate the coupling with the classical mass-balance based modelling framework of reactors in chemical engineering. PBMs are compatible both with Computational Fluid Dynamics, CMA and 0D hydrodynamic model which makes this approach compatible with the modelling of all bioreactor scales.

Population Balance Equations can be solved using either sectional –or class– methods or moment methods. The former are more intuitive and constitute a direct resolution of the equation, at the cost of high number of resolved variables. Therefore, methods of moments have been assessed in order to significantly reduce the size of simulations. Multiple methods to provide closure in moment formulations have been implemented and tested against the class method, these were QMOM, EQMOM and the Maximum Entropy approach. MaxEnt was found to be the most accurate of all methods, but its lack of stability makes this numerical method unsuited for the simulation of large-scale biological cultures where any numerical failure compromises the integrity of the whole simulation. QMOM is the fastest of all three methods, but requires twice has much resolved moments as EQMOM to reach a similar accuracy and cannot provide point-wise evaluation of the NDF. Therefore, EQMOM has been favourably selected as the best suited method. Its proved to be accurate and stable with a reasonable numerical cost.

EQMOM has been developed quite recently [Yuan et al., 2012] and has been improved over time [Madadi-Kandjani and Passalacqua, 2015; Nguyen et al., 2016; Passalacqua et al., 2017]. Using this method and researching the mathematical fundamentals of methods of moments led to the discovery of a new procedure to solve the EQMOM moment-inversion problem. This implies that a new version of EQMOM is now available to whom may be interested in faster computations based on this method [Pigou et al., 2018a].

While Chapter II focused on modelling bioreaction rates for *E. coli* and *S. cerevisiae*, the current Chapter detailed Population Balance Equations for both micro-organisms. Thanks to these, and to the selection of suited numerical methods, previously developed metabolic models can now be applied in dynamic heterogeneous systems at reasonable computational costs.

# IV

# MODÉLISATION DE L'HYDRODYNAMIQUE AUX GRANDES ÉCHELLES

## Résumé

Les Chapitres II et III se sont concentrés sur la modélisation de la phase biologique. Désormais, des modèles sont disponibles pour décrire le comportement métabolique de micro-organismes soumis à des perturbations, ainsi que des méthodes numériques pour intégrer ces comportements à l'échelle d'une population. Afin d'appliquer ces modèles à la simulation de fermenteurs industriels, un aspect manquant est celui de l'hydrodynamique de ces systèmes. Savoir comment les phases gaz et liquide circulent dans un bioréacteur est en effet un pré-requis pour prédire le transport des espèces dissoutes (substrats, sous-produits, ...), le taux de transfert des espèces gazeuses (oxygène, dioxyde de carbone, ...) et le transport de la phase biologique.

Appliquer directement des méthodes expérimentales sur les bioréacteurs industriels pour obtenir les champs de vitesses et la distribution spatiale des deux phases semble difficile, voire impossible, et serait particulièrement onéreux. L'outil alors le plus adapté est la Mécanique des Fluides Numérique (CFD) qui permet de prédire l'hydrodynamique dans les systèmes de grande échelle, seulement à partir de leur géométrie, des conditions opératoires, et du choix de modèles adaptés à la description de cuve agitée gaz-liquide. Au cours de ce travail, deux logiciels de CFD ont été mis en œuvre. La suite applicative ANSYS Workbench a été utilisée pour définir la géométrie d'un fermenteur industriel, pour générer des maillages, et pour effectuer des simulations sous ANSYS Fluent basées sur une approche pseudo-stationnaire. NEPTUNE_CFD, un logiciel développé par les principaux acteurs de l'industrie nucléaire civile française (CEA, EDF, FRAMATOME, IRSN), a été utilisé pour effectuer des simulations plus avancées. Celles-ci sont basées sur une approche instationnaire à maillage tournant en intégrant de nouvelles modélisations pour décrire les interactions gaz/liquide, en particulier à fort taux de gaz. Le coût numérique important de ces simulations instationnaires a nécessité la mise en oeuvre de calculs haute performance sur supercalculateurs. Cela peut, d'une certaine manière, sembler en contradiction avec l'objectif principal de cette thèse de parvenir à simuler des cultures biologiques industrielles de manière rapide et peu coûteuse.

Après avoir détaillé le fermenteur industriel à simuler et les simulations CFD associées, ce chapitre développe une analyse de temps caractéristiques. Il est ainsi montré que, bien qu'une résolution spatiale fine soit requise pour les simulations CFD, le couplage de l'hydrodynamique avec la modélisation de la phase biologique ne requiert pas une telle résolution. Au contraire, une description spatiale fine engendrerait des coûts computationels prohibitifs et contraindrait fortement les pas de temps utilisés pour les simulations. Ainsi, nous proposons d'effectuer le couplage des modèles biologiques (métabolique et bi-

lan de population) avec l'hydrodynamique dans une approche par compartiments. La dernière section de ce Chapitre se concentre donc sur la méthodologie employée pour déduire un modèle de compartiments à partir des résultats de simulations CFD. Cette méthodologie, basée sur l'exploitation de données au format Ensight Gold exportable par ANSYS Fluent et NEPTUNE_CFD, est développée pour intégrer les champs de vitesse aux frontières entre les compartiments, et les champs scalaires (fractions volumiques, variables caractéristiques de la turbulence) sur le volume de chaque compartiment. Lors de l'implémentation de cette méthodologie, nous avons observé une erreur résiduelle non négligeable qui, non résolue, induit un schéma de transport non conservatif. Dès lors, nous proposons une procédure basée sur une optimisation multi-variables et multi-objectifs pour corriger cette erreur et obtenir une modélisation conservative de l'hydrodynamique des bioréacteurs industriels.

# IV

# MODELLING LARGE-SCALE HYDRODYNAMICS

## Summary

Chapters II and III focused on the description of the biological phase. Now models are available to describe the metabolic behaviour of micro-organisms subjected to perturbations along with numerical method to integrate these behaviours at the population-scale. In order to integrate these models in simulations of industrial fermenters, the remaining aspect is that of hydrodynamics. Knowing how gas and liquid phases flow in the system is indeed required to predict the transport of dissolved species (substrates, by-products, . . . ), the local transfer rate of gases (oxygen, carbon dioxide, . . . ) and the transport of the biological phase.

Using directly experimental methods in industrial fermenters to obtain the knowledge of fluid velocities and spatial distribution of both phases is difficult, if not impossible, and would be prohibitively expensive. The most suited tool is the Computational Fluid Dynamics which allows predicting large-scale hydrodynamics solely from the system geometry, operating conditions, and the choice of suitable models to describe gas-liquid turbulent flows in a stirred tank reactor. Over the course of this work, two CFD software have been used. The ANSYS Workbench tool-suite has been employed to define the system geometry, generate meshes, and to perform pseudo-stationary simulations using ANSYS Fluent. NEPTUNE_CFD, a software developed by the main actors of French civil nuclear industry (CEA, EDF, FRAMATOME, IRSN), has been used to perform more detailed simulations. They used an unsteady moving-mesh approach and new models for the description of gas-liquid interaction at high gas flow rates. The numerical cost of these unsteady simulations required the use of supercomputing resources, which somehow contradicts the goal of performing fast –low-cost– simulations for industrial fermentations.

After having detailed the considered industrial set-up and associated CFD simulations, the current Chapter develops a time-scale analysis. While fine CFD simulations are required to access hydrodynamics in industrial bioreactors, it is shown that the coupling with biological models does not require the spatial resolution of CFD simulations. Actually, this spatial resolution would induce a prohibitively high computational cost and would enforce a limiting time-step for simulations. Therefore, we propose to perform the coupling of biological (metabolic and population balance) models with a Compartment Model.

The last part of this Chapter then focuses on describing how to obtain a compartment model starting from CFD results. A methodology, based on the Ensight Gold file format which can be exported from both ANSYS Fluent and NEPTUNE_CFD, is developed to integrate velocity field over contours of compartments, and phase fraction fields over compartment volumes. When implementing this methodology, we noticed non negligible residual error on obtained compartment models. If not resolved, this would lead to non-conservative transport schemes. Therefore, a procedure based on multi-objectives multi-parameter optimisation is proposed to fix this residual error and enforce a conservative hydrodynamic model for large-scale bioreactors.

# IV

# MODELLING LARGE-SCALE HYDRODYNAMICS

## IV.1 Introduction

The modelling of bioreactors has often been limited to either homogeneous descriptions of the system where no gradient are considered, or to heterogeneous descriptions where biological complexity is left apart (see I.8). Only few attempts were made to handle both biological heterogeneity and metabolic complexity within a unified modelling framework that could be applied to all bioreactor scales.

In order to achieve this goal, the Euler-Euler-PBE gas-liquid modelling framework has been selected. It allows integrating biological heterogeneity described by Population Balance Equations (PBEs) (see Chapter III) and metabolic models (see Chapter II) as source/sink terms in mass balance equations. Depending on the scale of the reactor, and on the comparison between mixing and reaction/transfer/adaptation characteristic times, the system will either be homogeneous or heterogeneous with large-scale or local-scale gradients. Depending on the situation, the conservation equations will require to be written over control volumes defined with different length scales. Let $\boldsymbol{C}_k$ be a vector of quantities expressed per unit of volume of the phase $k$, mass balance equations on this quantity can be:

- For homogeneous systems of volume $V$

$$\frac{\partial \langle \alpha_k \rangle V \boldsymbol{C}_k}{\partial t} = \boldsymbol{S}_k \tag{IV.1}$$

- For heterogeneous systems with large-scale gradients, modelled by Compartment Model Approach (CMA)

$$\frac{\partial \langle \alpha_{k,i} \rangle V_i \boldsymbol{C}_{k,i}}{\partial t} = \sum_{j=1}^{N_C} Q_{j,i}^{(k)} \boldsymbol{C}_{k,j} - \boldsymbol{C}_{k,i} \sum_{j=1}^{N_C} Q_{i,j}^{(k)} + \boldsymbol{S}_{k,i} \tag{IV.2}$$

- For heterogeneous systems with short-range gradients, approached by Computational Fluid Dynamics (CFD)

$$\frac{\partial \alpha_k \boldsymbol{C}_k}{\partial t} + \vec{\nabla} \cdot \left( \alpha_k \boldsymbol{C}_k \vec{u}_k - \vec{\nabla} \left( \alpha_k \boldsymbol{\mathcal{D}}_k \circ \boldsymbol{C}_k(\vec{x}, t) \right) \right) = \boldsymbol{S}_k \tag{IV.3}$$

$\alpha_k$ is the local phase fraction and $\langle \alpha_k \rangle$ is its mean value over a control volume. Subscripts $i$ and $j$ denote variables attached to the $i$-th or $j$-th compartment in a compartmental approach with $N_C$ compartments.

Elements of $\boldsymbol{C}_k$ may correspond to liquid-phase concentrations of oxygen, glucose, acetate, … ($\mathrm{kg/m^3_L}$), to gas-phase concentrations of oxygen, carbon dioxide, or other gaseous compounds ($\mathrm{kg/m^3_G}$) but also, for the treatment of PBEs to class-attached biomass concentrations ($\mathrm{kg_X/m^3_L}$) or to moments of the distribution of biological properties (unit depending on moment order and property unit). The compatibility of models and numerical methods developed in Chapters II and III, with these transport equations, is then straightforward. $\boldsymbol{S}_k$ is the vector of source/sink terms associated to each quantity tracked in $\boldsymbol{C}_k$ and can account for mass transfer, reaction, or adaptation/redistribution of biological properties as defined in previous chapters.

Note that the homogeneous modelling is a simplification of a compartment model with a single compartment (or actually, that CMA is an extension of homogeneous models). Therefore, in this chapter we only consider CMA and CFD approaches. Therefore, two questions arise when a specific bioreactor has to be simulated:

- Which approach should be selected between CMA and CFD for the coupling of transport with bioreaction?
- How to access phase presence and velocity fields ($\alpha_k(x)$ and $\vec{u}_k(x)$ for CFD, $\varepsilon_{k,i}$ and $Q^{(k)}_{i,j}$ for CMA)?

As part of this project and in relation with industrial interests of Sanofi, a specific application case was selected for the whole modelling framework and simulation workflow. The presentation of this application will drive the current chapter by illustrating the specific questions that occurred and how they have been tackled to obtain a large-scale hydrodynamics modelling.

First, the industrial bioreactor is briefly described in section IV.2 along with operating conditions. Considering the size of the system, and the culture mode, it will be made clear that the bioreactor will never be strictly homogeneous and that heterogeneity must be accounted for. Therefore, CMA or CFD description of transport will be required, which in turn requires the description of velocity fields. As no sufficient experimental data is available on that industrial geometry, numerical computations of velocity fields are a requirement. CFD simulations are thus presented in section IV.3. Section IV.4 focuses on determining whether the high spatial accuracy achieved in CFD simulations is required for the coupling with biological models. A comparison of time-scales associated to the biological phase and to the CFD transport scheme will show that the CMA approach is more efficient to perform the coupling between transport and biological reactions/adaptation. Therefore a methodology for converting CFD results into Compartment Models is presented in section IV.5.

## IV.2 Production-scale fermenter

### IV.2.1 System geometry

The study of bioreactor modelling and simulation has been applied to an industrial-scale fermenter of *undisclosed volume*. This vessel is equipped with three turbines, whose specifications may not be disclosed for confidentiality reasons. The bottom impeller is a 6-blade radial turbine, responsible for gas dispersion. The other two impellers are 3-blade turbines leading to an hybrid radial and pumping-up axial flow, whose goal is to favour

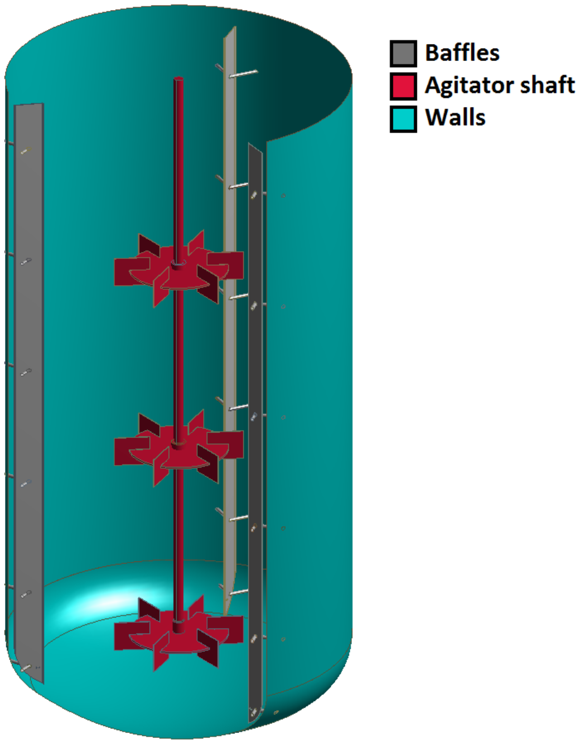


Figure IV.1: Simplified geometry of the considered industrial large-scale bioreactor. 6-blade Rushton turbines are displayed in place of actual turbines.

large-scale mixing and gas dispersion. Central vortex is prevented by three baffles. An incomplete geometry, where actual impellers are replaced by Rushton turbines, is shown in Fig. IV.1. The fermenter *dimensions are undisclosed.* Gas is injected *at undisclosed location.*

### IV.2.2 Operating conditions

The goal of the study and of simulations presented in Chapter V is to assess the effect of stirring speed and gas flow rate on observed metabolic dysfunctions of the baker yeast (see Chapter II). Therefore, CFD computations will be performed for two sets of operating conditions summed up in Table IV.1.

Table IV.1: Operating conditions to be simulated. *Normalized by Operating Condition 1 for confidentiality reasons.*

| | Operating condition 1 | Operating condition 2 |
|---|---|---|
| Impeller rotation speed | 1 | 1.4 |
| Gas flow rate | 1 | 1.85 |
| Liquid volume | 1 | 1 |

This fermenter is used for fedbatch cultures. This implies that the liquid volume increases over-time during the course of the fermentation process which spans several days. Instead of solving hydrodynamics over such a long period, it has been decided to perform pseudo-stationary simulations of the overall system at an operating point corresponding to *an undisclosed* liquid volume. This point corresponds to a specific time in the fedbatch process, associated to a specific biomass concentration and substrate feeding rates which will be used for simulations in Chapter V.

## IV.3 CFD simulations

### IV.3.1 Goal and constraint on bioreactor hydrodynamic computations

Leaving aside all biological considerations and on a strictly hydrodynamics point of view, the numerical simulation of industrial fermenters poses multiple challenges.

The first one comes from the geometric complexity of these systems. In order to monitor the process, to control temperature and pH or to allow injection of gas and liquid sources of feeding material, bioreactors are filled with numerous apparatus which can locally affect the circulation of fluids.

The second aspect is the presence of macro-instabilities. Large-scale bioreactors are operated in turbulent flow regime which implies small scales for velocity fluctuations, while the overall system is large. The continuous stirring also induces periodic fluctuations of velocity fields associated to the cyclic movement of turbines. Solving simultaneously small and large scales poses a numerical challenge.

A third source of complexity originates from physical properties of culture broth. The presence of high cell concentrations increases the fluid viscosity. This viscosity is even more significantly increased locally around the glucose feeding point if concentrated syrup is directly injected in the system. Moreover, the numerous chemical and biological species in the culture broth influence the gas-liquid surface tension.

Finally, a fourth challenge comes from the gas-liquid aspect of these systems. Indeed, high gas flow rates applied to the considered fermenter prevent the use of spargers to disperse gas in the system. Instead, an exhaust-pipe injects the gas and the bottom turbine is responsible for breaking gas pockets into dispersed bubbles. This implies that a correct description of the gas-liquid hydrodynamics requires handling both a continuous gas phase under the bottom turbine, a dispersed gas phase carried by the liquid, as well as breakage and coalescence phenomenon which will affect the distribution of bubble sizes and shapes. Note that breakage and coalescence will be impacted by the local evolutions of viscosity and surface tensions as pointed out in the third consideration.

Overall, bioreactor hydrodynamics is a complex topic indeed. Nevertheless, a perfect description of these complexities will not be required for the upcoming coupling of transport by fluids with biological reactions. The hydrodynamic description will then be simplified by estimating whether each aspect is expected to have a significant effect on the overall coupling. Said otherwise, we will only describe phenomenon whose consequences are perceived by the biological modelling, both in terms of metabolism and of biological adaptation.

### IV.3.2 Simplifying the hydrodynamics modelling

This thesis is not focused on performing complex –state-of-the-art– simulations of industrial fermenters hydrodynamics. All previously listed aspects do have an impact on overall bioreactor performances, but quantifying these impacts requires describing all phenomenon, including biological-phase related aspects, with the same high accuracy and at the expense of inaccessible numerical cost. Therefore, we focused on obtaining rough hydrodynamics, both in terms of velocity field and in terms of spatial distribution of the gas phase.

Having the correct order of magnitude for these variables (velocities and phase fractions) will yield correct meso and macro-mixing times, as well as local oxygen transfer rate. These in turn will impact the biological model, and allow a reverse coupling between large-scale gradient/local substrate availabilities, biological state and metabolic behaviours.

To facilitate CFD simulations while preserving meso/macro-mixing time-scales, the fermenter geometry has been simplified to its bare minimum. All small elements that only slightly disturb the flow have been removed from the geometry as they only induce local perturbations that do not significantly influence macromixing and gradients. Fig. IV.1 illustrates which elements were kept: vessel walls, the three baffles, and the agitator shaft. Baffles and turbine blades were even more simplified by describing them using a thin-wall approach, therefore their thickness was not described, which allows for a slightly coarser meshing and improves the mesh quality in terms of aspect ratios.

In terms of turbulent velocity fluctuations, solving them would require performing Direct Numerical Simulations (DNSs) where all scales of turbulence are resolved, down to the Kolmogorov scale, or at least Large-Eddies Simulations (LESs) where only the smallest –sub-grid– scales are modelled. However, the size of the industrial fermenter prevents such simulations due to a prohibitive numerical cost. This will not be a major issue, considering that an accurate description of turbulent velocity fluctuations is not required in the first place. Indeed, the biological model only responds to potential lack of substrate due to micro-mixing limitations, and not to actual velocity fluctuations. Evaluating the flux of substrate available through micro-mixing is feasible using integral properties of turbulence, such as the kinetic turbulent energy $k$ and the dissipation rate of that energy $\epsilon$, in the $k-\epsilon$ Reynolds-Averaged Navier-Stokes (RANS) approach.

It is known that in stirred vessels, the instantaneous velocity field $U(\vec{x},t)$ can be decomposed into a mean value $\bar{u}(\vec{x})$, a periodic perturbation associated to agitation $\tilde{u}(\vec{x},t)$ and chaotic fluctuations associated to turbulence $u'(\vec{x},t)$ [Escudié and Liné, 2004].

$$U(\vec{x},t) = \bar{u}(\vec{x}) + \tilde{u}(\vec{x},t) + u'(\vec{x},t) \tag{IV.4}$$

It is clear that the mean velocity will strongly impact meso and macro-mixing and therefore large-scale gradients. As stated previously, smallest scales of turbulent fluctuations are related to micro-mixing and will have an impact on substrate consumption by micro-organisms and this aspect will be modelled using variables of the $k-\epsilon$ approach in Chapter V. However, the need for describing periodic fluctuations $\tilde{u}(\vec{x},t)$ is more difficult to assess beforehand. For sake of simplicity, we first used in ANSYS Fluent [2015] the Multiple Reference Frames (MRF) approach which consists in discarding these periodic fluctuations, in line with the approach used by Haringa et al. [2016] among other examples. Other simulations using a rotating mesh approach have been performed under an other software, NEPTUNE_CFD [2016], in order to account for these periodic fluctuations of velocity fields.

*Part of this section is edited out for confidentiality reasons.*

Due to the breakage of large gas caps, and to classical breakage and coalescence of dispersed bubbles, strongly poly-dispersed Bubble Size Distributions (BSDs) are expected in industrial fermenters. Despite the fact that models are available to track this BSD

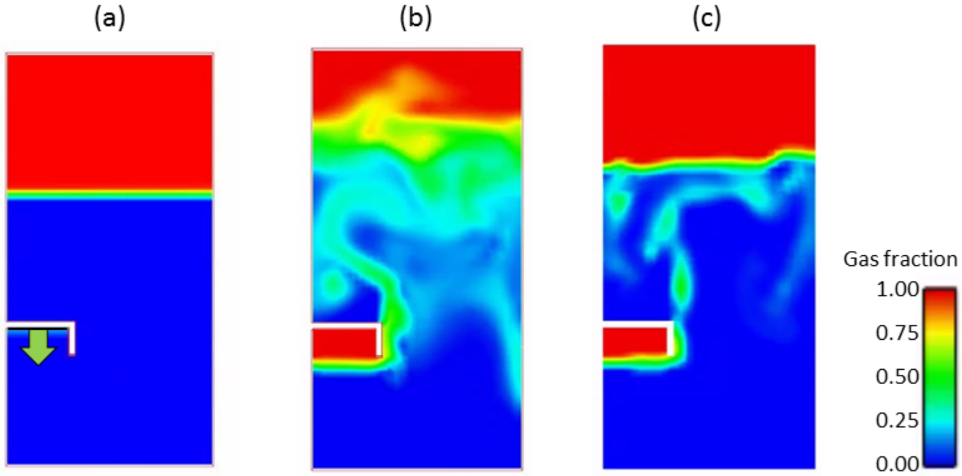


Figure IV.2: Illustration of differences between dispersed gas model and Generalized Large-Interface Model (GLIM) on simple 2D CFD simulations performed under Neptune CFD. (a) Initial conditions. The green arrow indicates the source and direction of gas injection. (b) Dispersed model. (c) GLIM model.

through a Population Balance Model approach [Laakkonen et al., 2007; Li et al., 2018], or to resolve the transport of the interfacial area, we did not integrate these in our simulations. Instead, we performed simulations by enforcing a constant bubble diameter.

### IV.3.3 Momentum exchange between phases

For interaction modelling between gas and liquid phases, we followed general guidelines of ANSYS Fluent [2015] user guide for bubbly flows. Drag was described using the universal drag laws for bubble-liquid flows [Kolev, 2007]. This law is applicable to most gas-flow regimes, and even to non-spherical bubbles, under the condition that the hydraulic diameter of the flow domain is significantly larger than the characteristic size of the dispersed phase, which is here the case. The lift force is modelled using the variant of Tomiyama [1998] law proposed by Frank et al. [2004]. For stability reasons, virtual mass force modelling was not enabled in performed simulations.

For simulations that ran under NEPTUNE_CFD [2016], we made use of the Generalized Large-Interface Model [Merigoux et al., 2016] which is designed to describe flows where a phase exists both as continuous pockets and dispersed bubbles [Mer et al., 2018]. This model detects large gas-liquid interfaces by comparing gas fraction gradients to a threshold value and by adding a special treatment at the interface location. This ensures preserving the consistency of these large interfaces (large pockets or free surface) which is otherwise lost when using purely dispersed models. This is illustrated in Fig. IV.2 where a simple 2D set-up is simulated using either dispersed or GLIM modelling. The free surface consistency is quickly lost using the dispersed modelling (Fig. IV.2b). The gas pockets predicted by GLIM can be dispersed by turbulent flow which is consistent with experimental observations of aerated bioreactors [Moilanen et al., 2008].

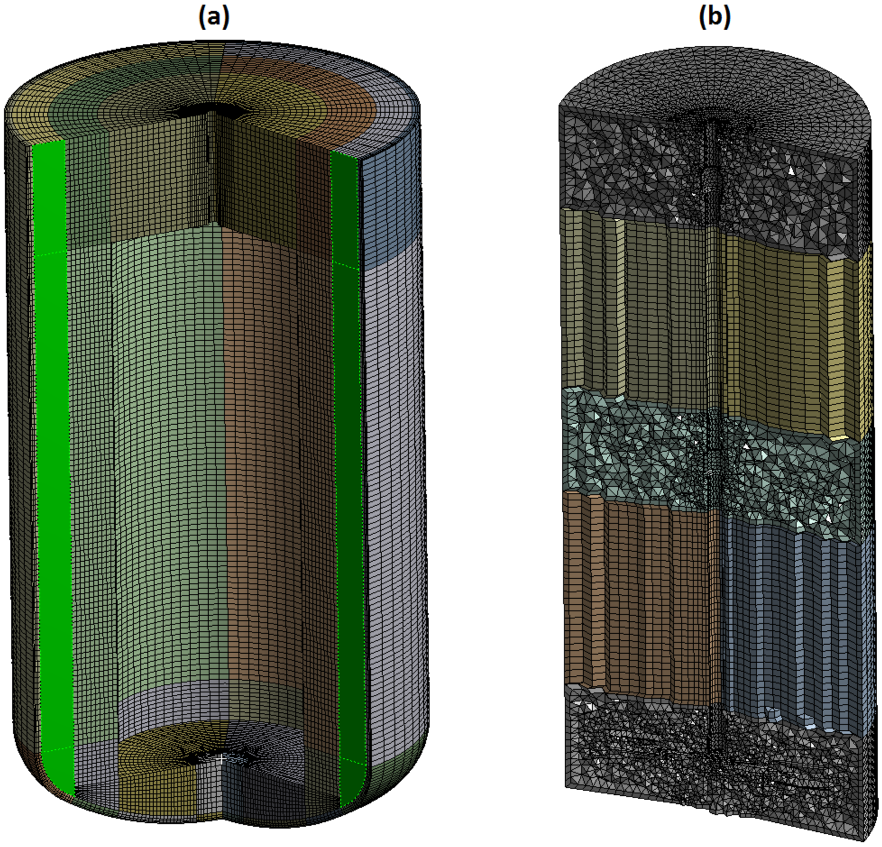


Figure IV.3: (a) Mesh of stator zone, discretized using ≈ 300000 hexahedral elements. Green surfaces represent two out of three thin-wall baffles. (b) Sectional view of rotor zone mesh, discretized using ≈ 1000000 tetrahedral and prism elements.

### IV.3.4 Geometry meshing

The same mesh has been used on both ANSYS Fluent [2015] and NEPTUNE_CFD [2016]. In order for it to be compatible with MRF as well as rotating-mesh approaches, two zones were defined: a rotor encompassing the agitator shaft, and a stator. These meshes are illustrated on Fig. IV.3. The stator zone could be entirely meshed by hexahedral elements. The complex geometry of impellers prevented such a regular meshing and therefore required an unstructured tetrehadron-based meshing. Zones in-between impellers were meshed using prism elements. A similar quadrangle-based mesh was enforced on the contact surface between rotor and stator meshes.

## IV.4 Time-scale analysis

As stated previously, the instantaneous velocity field $U(\vec{x}, t)$ can be decomposed into a mean value $\bar{u}(\vec{x})$, a periodic fluctuation $\tilde{u}(\vec{x}, t)$ and a chaotic fluctuations $u'(\vec{x}, t)$.

Turbulent fluctuations are filtered by the RANS approach which is closed by applying the $k-\epsilon$ model for the transport and dissipation of turbulent kinetic energy. Moreover, the periodic perturbation is filtered out too by the MRF approach which models the system as pseudo-stationary. Therefore, smallest time-scales of the hydrodynamics have not been resolved by previously described CFD simulations. Chapter V describes how micromixing and turbulent fluctuations can be accounted for in the coupling of transport with biological reaction.

If only mean velocity fields have been solved, and therefore only meso and macromixing are well described by these fields, one can wonder whether the resolution of the "fine" mesh used for CFD simulations is still required for the coupling of transport with metabolic and population balance models described in Chapters II and III. The smallest time-scale associated to these models is the one associated to the increase of glycolytic stress, $\sigma$, if cells are exposed to environment rich in terms of glucose concentration. This time-scale was identified to be around 3 seconds. Therefore, in order to correctly resolve the biological model, the simulation time-step does not need to be significantly lower than that value.

If the coupling of transport and biological model is directly performed with transport described by a finite-volume scheme using the CFD mesh, the time integration will have to respect the CFL condition [Courant et al., 1928]. Let $c_i$ be the Courant number associated to the $i$-th volume element of the mesh, $i \in \{1, \ldots, N_{\text{mesh}}\}$:

$$c_i = \frac{|\vec{u}(\vec{x}_i)|\delta t}{\delta x_i} \tag{IV.5}$$

with $\vec{u}(\vec{x}_i)$ the velocity of a phase at the center of the $i$-th volume element, $\delta t$ the simulation time-step, and $\delta x_i$ the characteristic length of the volume element. It is well known that the stability of the transport scheme requires that

$$\max_i c_i \leq 1$$

Having obtained a pseudo-stationary velocity field for each phase through the RANS-MRF approach, it is possible to deduce from this stability condition a maximum time-step $\delta t_{\max}$ that should be used to perform coupled simulations within the CFD framework.

This maximum time-step has been measured using ANSYS Fluent built-in post-processing tools. The characteristic length $\delta x_i$ was approximated by the cubic root of the $i$-th cell's volume. Obtained maximum time-steps are 2.0ms and 1.5ms for the two operating conditions presented in Section IV.2.

The time-scale associated to numerical stability of the transport terms is significantly lower than the smallest time-scale associated to other aspects of bioreactor modelling. Therefore, keeping the spatial accuracy of CFD simulations is not a requirement to perform coupled simulations of transport with transfer, reaction, and biological adaptation. It will be more efficient to compute a Compartment Model from CFD results as it will allow to significantly reduce the numerical cost on two aspects.

Typical number of compartments in Compartment Model Approach range from dozens [Vrábel et al., 1999] to thousands [Delafosse et al., 2014]. This is significantly lower than the number of volume-elements used for CFD simulations, which was here of 1.2 million cells. Therefore, a compartment model yields a much lower number of resolved variables, hence limited memory requirements, and decreases the numerical cost by reducing the number of calls to the metabolic model (Chapter II) and to numerical methods as Quadrature Method of Moments (QMOM) or Extended Quadrature Method of Moments (EQMOM) for the resolution of PBEs (Chapter III).

Moreover, the modelling will be used to perform simulations of long-term biological cultures. Whether these simulations span a whole culture (hours/days of simulated time), or attempt to identify a steady or pseudo-stationary state as shown in Chapter V, it will be necessary to simulate multiple hours of the system. The simulation of 10h of culture with a time-step limited by the Courant condition of $\delta t_{\max} \approx 2.0$ms would require $\approx 1.8\ 10^9$ time-steps. Though being clearly infeasible, we still attempted to perform these simulations in order to estimate the numerical cost that would be required. These simulations ran on a computing cluster using 240 allocated CPU cores using the software NEPTUNE_CFD [2016] without accounting for biological heterogeneity (unsegregated approach, see I.6.3), and by disabling the resolution of RANS and $k-\epsilon$ equations, therefore using a frozen flow approach. Only mean values for biological properties were tracked, and the metabolic model was called once per mesh volume-element. This over-simplified modelling, and the use of significant computing resources, still yield a mean computation time of 1.5s per time-step. A 10h long simulation would then require about a year of computations. It is therefore obvious that handling the transport of species carried by each phase within the framework of CFD simulations is not reasonable, if particular if one expect quick answer from simulations for an engineering application, and even more if the complete modelling, integrating biological heterogeneity, is considered.

For all previously detailed reasons, the Compartment Model Approach will be used to perform coupled simulations in Chapter V. The lower number of compartments compared to the number of mesh elements will decrease both memory and numerical costs. Moreover, the higher volume of compartments will allow to relax the Courant condition to higher stable time-steps, thus significantly reducing the total number of time-steps required. Following sections therefore focus on the computation of a suitable Compartment Model from CFD results.

# IV.5 From CFD to Compartment Models

## IV.5.1 State of the art, constraints and selected approach

The idea of degrading results of CFD simulations to obtain compartment models and associated flow maps is certainly not new [Bezzo et al., 2003]. It has been successfully applied to waste-water treatment [Le Moullec et al., 2010] and to bioreactor modelling in previous works [Delafosse et al., 2010, 2014; Nauha et al., 2018].

Two main approaches exist to perform the compartmentalization of CFD results. The first one consists in identifying sub-volumes in which some parameters are near constant through either stochastic or more usually deterministic methods. This approach leads to compartments of homogeneous turbulence or gas fraction for instance. However, the shape of these compartments will tend to be irregular and will strongly differ between operating conditions.

The second approach to obtain a compartment model from CFD is the manual zoning where geometric bounds of each compartment are defined by-hand. For the simulation of bioreactors, having regular compartments seems to be a more interesting option for two reasons. The first one is the easier comparison of simulation results when operating conditions are modified. For instance, assessing the effect of stirring speed or aeration rate on the overall culture is easier if one is sure that the compartmentalization was performed in exactly the same way between simulations (see Chapter V). The second reason why manual zoning might be more interesting is related to an unexplored idea. If one wants to integrate varying operating conditions or fluid properties over time such as a liquid viscosity depending on the biomass concentration, or on a product concentration (*i.e.* xanthan production), the approach would consist in

- performing CFD simulations for multiple conditions;
- computing compartment models associated to each CFD simulation;
- dynamically compute a compartment model as a convex mixture of previously defined models.

The last step is by far easier and less error prone if compartment models are defined on a regular –manual– zoning.

Note that Delafosse et al. [2010] obtained a better agreement in terms of mixing time between CFD simulations and manual zoning than when automatic zoning was used. This manual zoning approach has also been favourably selected by Delafosse et al. [2014] and Nauha et al. [2018]. For all these reasons, this will be the retained approach for following developments. The question is then to determine the technical solution for computing the compartment model from CFD results.

By CFD results, we imply results coming from simulation performed either on ANSYS Fluent or on NEPTUNE_CFD. These could be single-phase simulations or, more generally for aerated bioreactors, gas-liquid simulations. While we initially aimed at generating transient compartment model based on time-resolved hydrodynamics, only pseudo-stationary hydrodynamics are handled by following developments.

To handle results coming from both software, ANSYS Fluent and NEPTUNE_CFD, the selected approach has been to create a code able to read and post-process files in the Ensight Gold format. Both software can export simulation results in this format, and a

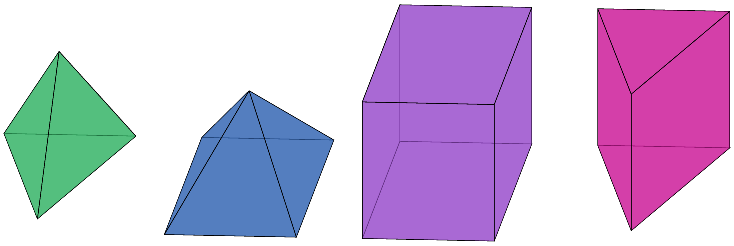


Figure IV.4: Volume mesh element types allowed by Ensight Gold format. From left to right: 4-nodes tetrahedron, 5-nodes pyramid, 8-nodes hexahedron and 6-nodes pentahedron.

full documentation of this standard is available [Ensight, 2016]. Following developments would not have been possible without the help of Dr. Bastien POLIZZI who wrote C++ libraries to handle reading operations on datafile in Ensight Gold format.

### IV.5.2 Numerical integrations

Simulation data exported by ANSYS Fluent and Neptune CFD are cell-centered values of scalar and vector fields. Exported fields are

- Velocity fields $\vec{u}_G(\vec{x})$ and $\vec{u}_L(\vec{x})$;
- Phase fraction fields $\alpha_G(\vec{x})$ and $\alpha_L(\vec{x})$;
- Liquid-phase turbulence characteristic variables $k(\vec{x})$ and $\epsilon(\vec{x})$

The mesh generated in section IV.3 is unstructured and contains following volume element types illustrated in Fig. IV.4: tetrahedron, pyramids, pentahedron and hexahedron.

The mesh being unstructured and mostly irregular, the challenge is to identify a method to automatically integrate scalar fields on zone volumes, and vector fields on zone boundaries to obtain the compartment model, as illustrated in Fig. IV.5.

Similarly to Delafosse et al. [2014], we chose to define compartments on a regular cylindrical grid. Let $r_{\min} = 0$, $r_{\max}$, $a_{\min} = -\pi$, $a_{\max} = +\pi$, $z_{\min}$ and $z_{\max}$ be respectively the radius, angle and height extents of the system. This cylindrical domain will be split into $N_r$, $N_a$ and $N_z$ regular divisions along each component. Therefore, the total number of compartments will be $N_C = N_r * N_a * N_z$. This compartmentalization is illustrated in Fig. IV.6.

The integration procedure has hardly been documented in previous compartmentalization works. Therefore, we detail hereafter the chosen approach which is numerically efficient and is compatible with complex geometries to compute mean values of scalar fields over each compartment volume, and flux of velocity fields over inter-compartment limits.

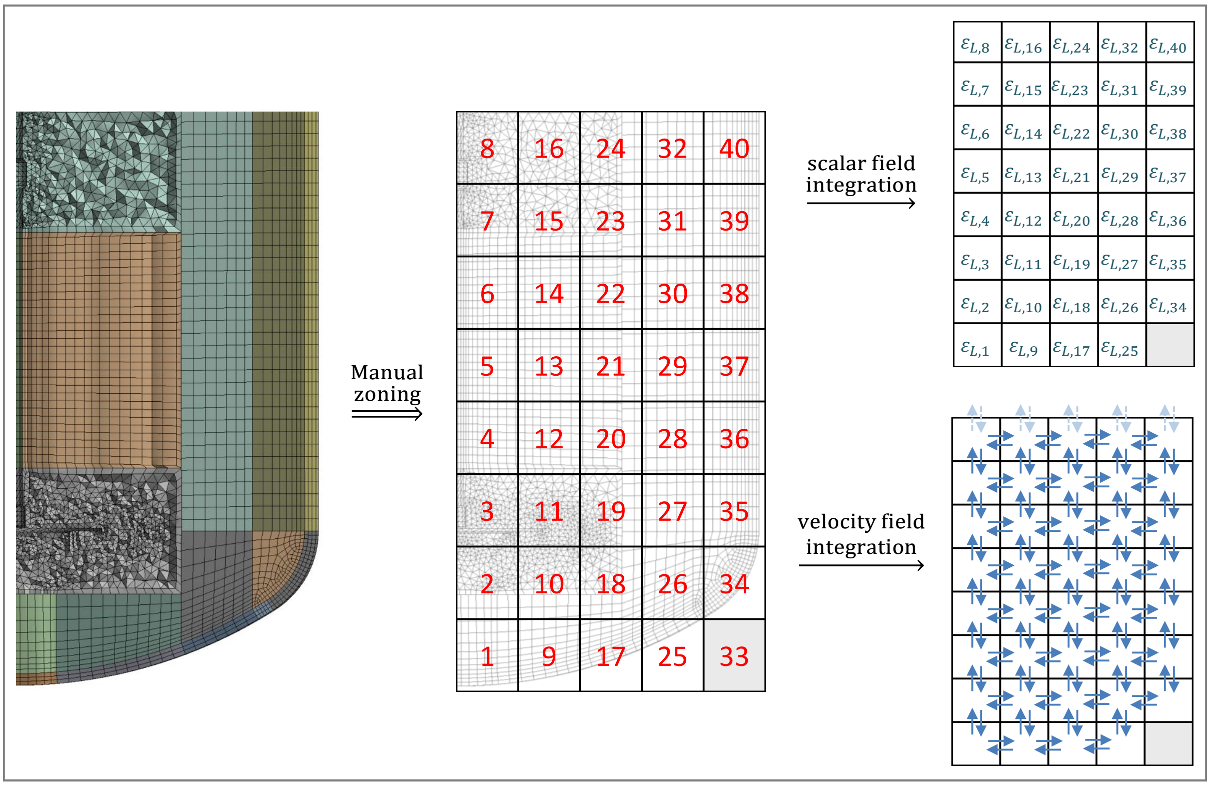


Figure IV.5: Steps of the compartmentalization procedure of CFD results.

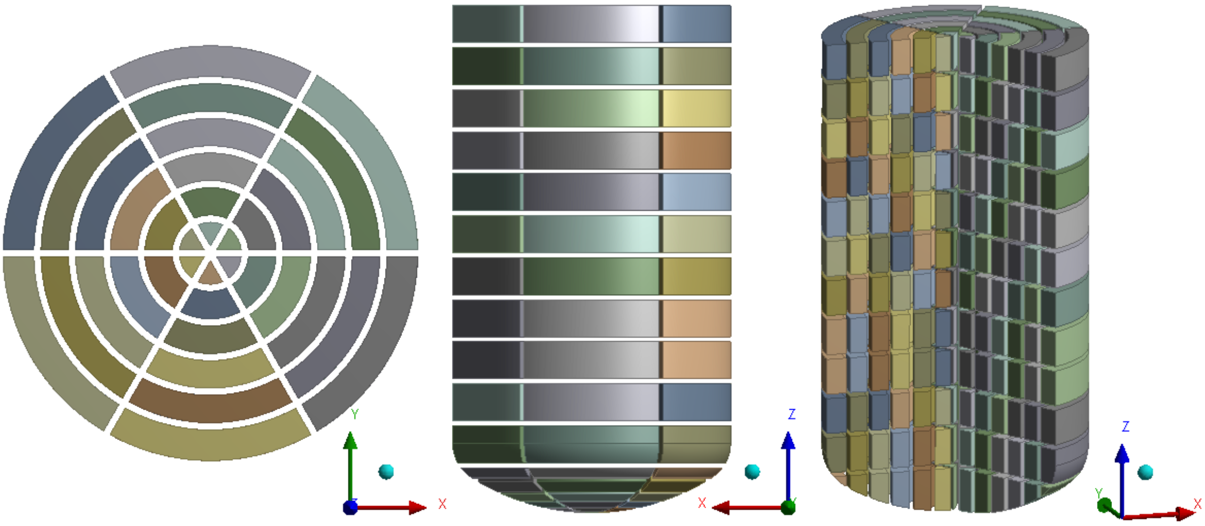


Figure IV.6: Compartmentalization of the industrial geometry. (left) top view, (middle) side view, (right) interior view.

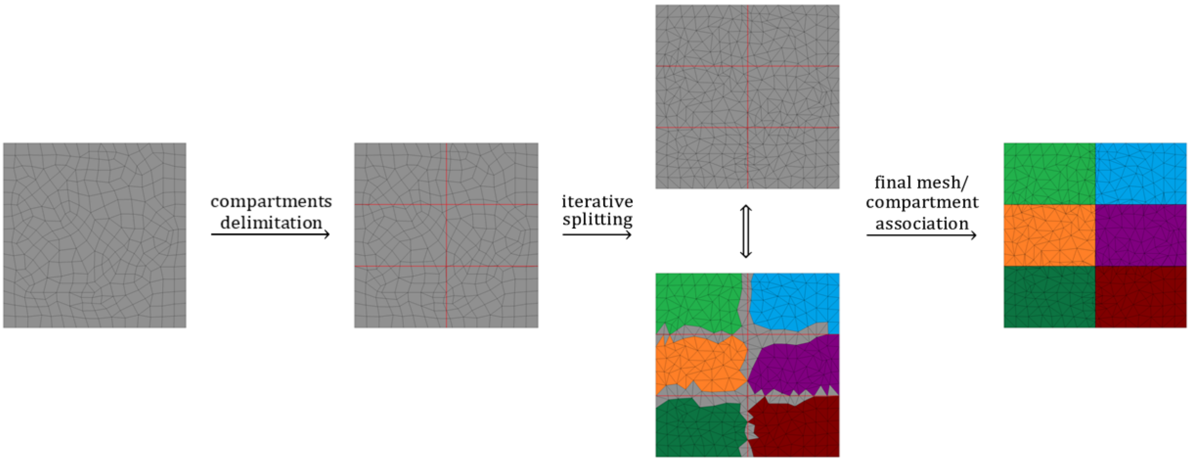


Figure IV.7: Illustration of splitting procedure to 2D meshes. Note that the initial mesh (left-hand side) contains both triangles and quadrangles, some of them located in multiple compartments. The final mesh (right-hand side) contains only triangles strictly located within a single compartment.

### Mesh splitting

Some authors used natural limits of mesh elements to perform the integration of fields. This was done by identifying neighbour cells whose center lie in different compartments, and then integrating vector fields over the facet joining these cells. While being seemingly straightforward, this approach has a main numerical drawbacks: it requires the knowledge of the mesh associativity, *i.e.* which mesh elements are connected by shared vertices and facets. This information is not directly available in Ensight Gold geometry files. Associativity can be reconstructed at moderate cost if the whole geometry is meshed in a unique block (or “part’ is the Ensight Gold format terminology), but the cost significantly increases when multiples parts constitute the complete mesh. As explained in Section IV.3.4, the simulation of stirred reactors requires to mesh independently rotor and stator zones, which translates as two different parts when data is exported to Ensight Gold format. Moreover, the accuracy of this approach can be questioned as soon as the mesh is strongly irregular.

In order to obtain low-cost and accurate compartmentalization even for complex geometry, we chose to perform a mesh splitting. The basic idea is to split each mesh-element located on a compartment limit into multiple sub-elements whose vertices are all located into a single compartment or strictly on a compartment limit. This is illustrated in Fig. IV.7 for a 2D-mesh where at-limit elements are split into sub-elements strictly located in each compartment. This idea is easily generalizable to 3D-meshes too by splitting volume mesh elements into sub-elements.

To facilitate the splitting operation of mesh-elements by surfaces that limit compartments, we first reduce all types of elements to tetrahedron. This is illustrated by Fig. IV.8 for element types handled by the Ensight Gold format shown in Fig. IV.4. Indeed, splitting in two a hexahedron, pentahedron or pyramid can yield complex shapes, however a tetrahedron can only be split by a plane into sub-elements that are tetrahedrons, pyramids or pentahedrons. These elements can in turn be re-divided into tetrahedrons allowing for an easy iterative procedure.

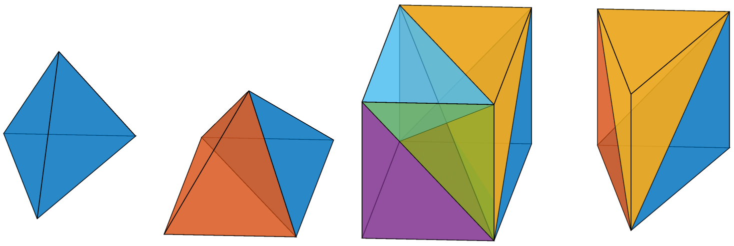


Figure IV.8: Splitting of volume elements from Fig. IV.4 into tetrahedrons.

Therefore, one only needs to

1. define how to split pyramids, pentahedrons and hexahedrons into tetrahedron;
2. identify tetrahedrons which are located in multiple compartments;
3. compute coordinates of sub-parts of the tetrahedron which are located in each compartment.

The first step is straightforward and is illustrated by Fig. IV.8. The second step is easily done by checking whether all vertices of a tetrahedron are located in the same compartment. If at least one vertex is in a different compartment than others, then the tetrahedron needs to be split into two sub-volumes. Finally, the third step is only a matter of detecting which edges cross the compartment limits, computing the intersection coordinates and forming sub-elements.

Note that in this approach, one only needs to know coordinates of vertices for each initial mesh element. This information is directly available in geometry files in Ensight Gold format. Iterating over mesh elements, and splitting them into sub-volumes is only a matter of simple linear algebra and computations of edge/surface intersections.

These steps have been implemented into a non-optimized and unparallelized C++ code. When applied to the industrial bioreactor mesh, with a 432 compartments ($N_r = 6$, $N_a = 6$, $N_z = 12$) model as shown in Fig. IV.6, the mesh splitting procedure only required about 3 seconds on a mid-end desktop while an associativity based approach required multiple minutes, which illustrates the efficiency of this approach.

After the mesh-splitting step, all facets located at the limit between compartments are triangles and should have been identified. Similarly, all volumes elements should be tetrahedrons fully located in each compartment. Therefore, integrating vector and scalar fields will easily be done as explained hereafter.

As illustrated in Fig. IV.5, applying the manual zoning on a complex geometry can yield compartments that do not contain any volume element (compartment n°33) or partially filled compartments (n°1, 9, 17, 25 and 34). Therefore, the actual volume of compartments will not be defined by their geometric boundaries shown in Fig. IV.6, but will rather be the sum of elementary volumes of each tetrahedron contained in each compartment. By automatically discarding empty compartments, the developed compartmentalization approach can be applied to any complex geometry.

### Integrating scalar fields

Let $N_{T,i}$ be the number of tetrahedrons in the $i$-th compartment ($i \in \{1, \ldots, N_C\}$), $V_{T,i,j}$ the volume of the $j$-th tetrahedron associated to that compartment ($j \in \{1, \ldots, N_{T,i}\}$), and $S_{T,i,j}$ the value of a scalar field at the center of this tetrahedron.

The total volume of the $i$-th compartment, $V_i$, is simply defined as the sum of elementary tetrahedron volumes:

$$V_i = \sum_{j=1}^{N_{T,i}} V_{T,i,j} \tag{IV.6}$$

Let $\langle S_i \rangle$ be the volumetric mean value of the scalar field in the $i$-th compartment, it is defined as

$$\langle S_i \rangle = \frac{1}{V_i} \sum_{j=1}^{N_{T,i}} S_{T,i,j} V_{T,i,j} \tag{IV.7}$$

One main drawback of manual zoning approach is that created compartments might be strongly spatially heterogeneous in terms of some scalar fields. If one cannot afford to lose this heterogeneity information because it has a significant impact on other aspects of the modelling, approaches developed in Chapter III can be reapplied here. Indeed, it has been observed that methods of moments are efficient when one wants to describe heterogeneous systems by storing only few information. Let $\boldsymbol{M}_{S,N,i} = [M_{S,0,i}, \ldots, M_{S,N,i}]$ be moments of the volumetric distribution $v_i(S)$ where $\int_{S_1}^{S_2} v_i(S)\mathrm{d}S$ is the sub-volume of compartment $i$ associated to scalar values in-between $S_1$ and $S_2$. These moments can be computed by:

$$M_{S,k,i} = \int_{S_\mathrm{min}}^{S_\mathrm{max}} S^k v(S)\mathrm{d}S \approx \sum_{j=1}^{N_{T,i}} S_{T,i,j}^k V_{T,i,j} \tag{IV.8}$$

where $S_{i,\mathrm{min}} = \min_j S_{T,i,j}$ and $S_{i,\mathrm{max}} = \max_j S_{T,i,j}$ are limits of $S$ values observed in the $i$-th compartment. Note that $M_{S,0,i} = V_i$ and $M_{S,1,i} = V_i \langle S_i \rangle$.

Using first few moments of $v(S)$, one can precompute a quadrature rule such that each nodes of this rule is a sub-volume of the $i$-th compartment associated to a discrete value of the scalar field $S$. This gives access to sub-grid –or sub-compartment– heterogeneity and helps circumventing the drawback of manual zoning.

The moment approach is suggested as it allows gathering most of heterogeneity information within a few numerical values, but it is not strictly speaking required as the volumetric distribution $v_i(S)$ is known thank to the small-scale CFD meshing. Therefore, if multiple uncorrelated heterogeneous fields, say $S^{(1)}$ and $S^{(2)}$, coexist in a compartment, is is possible to directly post-process the multivariate distribution $v_i(S^{(1)}, S^{(2)})$ to obtain a sub-compartment quadrature.

These developments about sub-grid heterogeneity have not been applied, due to time limitations in the overall project. However, it is worth noting that when done right, the compartmentalization process can retain most of useful information obtained through the expensive CFD computations without even using a high number of compartments.

As stated previously, we exported from CFD simulations multiple scalar fields: local phase fractions $\alpha_G$ and $\alpha_L$ and liquid-phase turbulent variables $k$ and $\epsilon$. Therefore, we directly integrated fields of local phase fractions to obtain phase volumes $V_{G,i}$ and $V_{L,i}$, and we computed mean turbulence variables $\langle k \rangle_i$ and $\langle \epsilon \rangle_i$:

$$V_{G,i} = \sum_{j=1}^{N_{T,i}} \alpha_{GT,i,j} V_{T_i,j} \tag{IV.9}$$

$$V_{L,i} = \sum_{j=1}^{N_{T,i}} \alpha_{LT,i,j} V_{T_i,j} \tag{IV.10}$$

$$\langle k \rangle_i = \frac{1}{V_{L,i}} \sum_{j=1}^{N_{T,i}} \alpha_{LT,i,j} k_{T,i,j} V_{T_i,j} \tag{IV.11}$$

$$\langle \epsilon \rangle_i = \frac{1}{V_{L,i}} \sum_{j=1}^{N_{T,i}} \alpha_{LT,i,j} \epsilon_{T,i,j} V_{T_i,j} \tag{IV.12}$$

### Integrating vector fields

For the simulation of bioreactors, we are interested in computing the flow of each phase through all faces delimiting compartments, to obtain the volumetric flow rates, $Q_{i,j}^{(k)}$, required in Eq. (IV.2).

Let $S$ be the surface separating the $i$-th and $j$-th compartments and $\vec{n}$ the vector normal to that surface, oriented toward the $j$-th compartment. In order to account for the fact that phases can flow in both directions across the surface, which introduce back-mixing between compartments, we will distinguish two volumetric flow-rates $Q_{i,j}^{(k)}$ and $Q_{j,i}^{(k)}$ defined as

$$\begin{aligned} Q_{i,j}^{(k)} &= \iint_S \alpha_k \max\left(0, \vec{u}_k \cdot \vec{n}\right) \mathrm{d}S \\ Q_{j,i}^{(k)} &= \iint_S \alpha_k \max\left(0, -\vec{u}_k \cdot \vec{n}\right) \mathrm{d}S \end{aligned}$$

with $\alpha_k$ the local phase fraction of phase $k \in \{G, L\}$, and $\vec{u}_k$ its velocity.

Using the manual zoning described previously, surfaces delimiting compartments are defined at constant radius, angle or height. Therefore, the normal vector of these surface are defined as one of the basis vector of a cylindrical coordinate system. Let $c \in \{r, a, z\}$ be the constant cylindrical coordinate of the surface $S$, previous definitions of $Q_{i,j}^{(k)}$ and $Q_{j,i}^{(k)}$ can be simplified into

$$\begin{aligned} Q_{i,j}^{(k)} &= \iint_S \alpha_k \max\left(0, u_{k,c}\right) \mathrm{d}S \\ Q_{j,i}^{(k)} &= \iint_S \alpha_k \max\left(0, -u_{k,c}\right) \mathrm{d}S \end{aligned}$$

where $u_{k,c} \in \{u_{k,r}, u_{k,a}, u_{k,z}\}$ is the component of the phase velocity orthogonal to the surface $S$. Vector fields are defined in terms of Cartesian coordinates in Ensight Gold format, $\vec{u}_k = [u_{k,x}, u_{k,y}, u_{k,z}]^T$, and from these values, one can easily deduce components in terms of a cylindrical coordinate basis $u_{k,r}$ and $u_{k,a}$; $u_{k,z}$ is left unchanged between Cartesian and Cylindrical coordinate basis.

$$r = \sqrt{x^2 + y^2} \qquad\qquad u_{k,r} = \cos(a) u_{k,x} + \sin(a) u_{k,y}$$

$$a = \tan_2^{-1}(y, x) \qquad\qquad u_{k,a} = -\sin(a)u_{k,x} + \cos(a)u_{k,y}$$

where $\tan_2^{-1}$ is the generalization of the $\tan^{-1}$ function which yields angle in-between $-\pi$ and $+\pi$ depending on the quadrant in which the point of coordinates $(x, y)$ is located.

Thank to this simplification allowed by a cylindrical manual zoning, the approximation of flow-rates is straightforward. Let $N_t$ be the number of triangles associated to the surface $S$ following the mesh-splitting operation, $\alpha_{t_n,k}$ and $u_{t_n,k,c}$ variables attached to the $n$-th triangle ($n \in \{1, \ldots, N_t\}$) and $S_{t_n}$ its surface. Flow rates are simply approximated as:

$$Q_{i,j}^{(k)} \approx \sum_{n=1}^{N_t} \alpha_{t_n,k} \max\left(0, \quad u_{t_n,k,c}\right) S_{t_n}$$

$$Q_{i,j}^{(k)} \approx \sum_{n=1}^{N_t} \alpha_{t_n,k} \max\left(0, -u_{t_n,k,c}\right) S_{t_n}$$

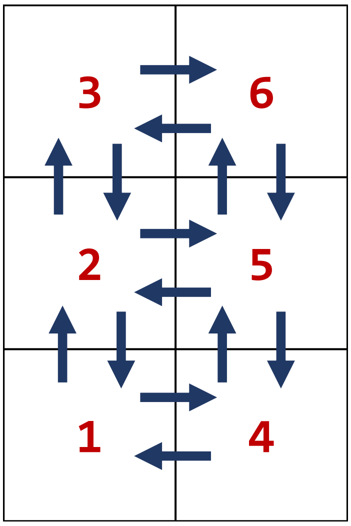
The remark formulated on the integration of scalar fields is also applicable here for velocity fields: one could track the distribution of velocity over the surface, for instance by tracking the moments of that distribution. However, this seems less relevant for multiple reasons. The first one is that, as detailed hereafter (see IV.5.3), when performing the integration of velocity fields, we noticed non negligible error on computed flows. If the first-order integration is vitiated with manifest error, one could question the accuracy of higher order information about the velocity distribution on the surface. Moreover, the integration of weighted velocity fields $\alpha_k \vec{u}_k$ only accounts for mean velocity fields as $\vec{u}_k$ is here defined as the Reynolds averaged velocity, and a pseudo-stationary –the MRF– approach was used in ANSYS Fluent. As previously suggested, integrating periodic and turbulent fluctuations of velocity fields in the compartment model might be of prior interest before considering the tracking of high order information on the mean velocity distribution on inter-compartment surfaces.

### IV.5.3 Cleaning error from circulation map

Flows between compartments are more easily handled if stored into a circulation matrix. Consider the simplified compartment model illustrated in Fig. IV.9. The matrix $\boldsymbol{M}$ is the circulation matrix associated to this simple compartment model and contains flows between compartments. The flow on the $i$-th row and $j$-th column is the volumetric flow rate going from the $i$-th compartment to the $j$-th. Note that this value contains only positive values, or null values for unconnected compartments. This matrix $\boldsymbol{M}$ is defined for both phases but following observations and developments can be applied similarly to each phase. Therefore, the phase dependency does not appear in following notations.

The integration methodology developed in Section IV.5.2 was applied to steady-state velocity and scalar fields. Therefore, phases should not accumulate into compartments and inflows associated to each compartment should balance outflows. Let $F_i^{(in)}$ and $F_i^{(out)}$ be the total inflow and outflow associated to the $i$-th compartment, these can be computed by summing all elements on the $i$-th column or row of $M$:

$$F_i^{(in)} = \sum_{j=1}^{N_C} M_{j,i}, \qquad F_i^{(out)} = \sum_{j=1}^{N_C} M_{i,j} \tag{IV.13}$$



$$M = \begin{array}{|c|c|c|c|c|c|}
\hline
 & Q_{1,2} & & Q_{1,4} & & \\
\hline
Q_{2,1} & & Q_{2,3} & & Q_{2,5} & \\
\hline
 & Q_{3,2} & & & & Q_{3,6} \\
\hline
Q_{4,1} & & & & Q_{4,5} & \\
\hline
 & Q_{5,2} & & Q_{5,4} & & Q_{5,6} \\
\hline
 & & Q_{6,3} & & Q_{6,5} & \\
\hline
\end{array}$$

Figure IV.9: Illustration of a 6 compartments model and associated circulation matrix.

One would expect that $F_i^{(in)} = F_i^{(out)}$, $\forall i \in \{1, \dots, N_C\}$. To check the accuracy of the computation of flow rates as detailed previously, we measured the relative error $\bar{e}_i$ on the flow balance on each compartment, defined as

$$\bar{e}_i = \frac{2|F_i^{(in)} - F_i^{(out)}|}{F_i^{(in)} + F_i^{(out)}} \tag{IV.14}$$

For both phases, and operating conditions described in Section IV.2, $\bar{e}_i$ was usually low, around 1% of numerical error, but could go as high as 25% in some compartments.

The reason why this error is of major importance and must be tackled is that it leads to a non-conservative transport scheme. If only transport were considered, the total mass of transported species would not change over-time, but the system would never reach homogeneity, which is unrealistic. Moreover, the modelling of bioreactors is not limited to transport but must also account for mass transfer and bioreaction. Therefore, using a non-conservative transport scheme would lead to an error that would cumulate over time. To prevent any bias in simulation results, it is of prior importance to "clean" the circulation map from these imbalances.

The source of the error has not been perfectly identified. It did not seem to come from a lack of convergence in CFD as pursuing these simulations did not resolve the issue. The C++ implementation of the integration method could be to blame as it has not been directly checked against an other code implementing the same approach. However the software Ensight has been used to manually compute the flow rate imbalance over a few compartments, and a similar –thought slightly lower– numerical error was again observed.

Consequently, it has been decided to keep the flawed circulation maps $\boldsymbol{M}$ as "raw material" to produce cleaned circulation maps $\boldsymbol{M'}$ by applying a transformation dictated by the two following constraints:

1. $\boldsymbol{M'}$ must be a balanced circulation maps, which translates into errors on each compartments, $\bar{e}'_i$, being close to the machine error under the considered precision.

2. $\boldsymbol{M'}$ should be as close as possible to $\boldsymbol{M}$. A metric could be to minimize $\|\boldsymbol{M} - \boldsymbol{M'}\|$ while respecting the first constraint.

The first constraint makes $\boldsymbol{M'}$ suitable for a coupling with transfer and bioreactions without cumulating numerical error over time. The second constraint ensures that the circulation of matter as predicted by CFD computations is conserved, as much as possible, even if one has to arbitrarily modify the circulation map $\boldsymbol{M}$.

Let $Q'_{ij}$ be the cleaned flow rates forming the matrix $\boldsymbol{M'}$ and $w_{ij}$ the correction weights such that $Q'_{ij} = w_{ij}Q_{ij}$. This simply translates into $\boldsymbol{M'} = \boldsymbol{W} \circ \boldsymbol{M}$. The problem of identifying the matrix of correction weights $\boldsymbol{W} \in \mathbb{R}_+^{N_C^2}$ under the previously defined two constraints can be translated into the following multi-objective and multi-parameter optimization problem:

$$\begin{aligned} \underset{\boldsymbol{W}\in\mathbb{R}_+^{N_C^2}}{\text{minimize}} \quad & (1-w_{ij})^2,\ (i,j) \in \{1,\dots,N_C\}^2 \\ \text{subject to} \quad & \sum_{j=1}^{N_C} (w_{ij}Q_{ij} - w_{ji}Q_{ji}) = 0,\ i \in \{1,\dots,N_C\} \end{aligned}$$

The optimization goal "minimize $(1-w_{ij})^2$" translates the second constraint of changing $\boldsymbol{M}$ as less as possible by trying to keep correction weights as close as possible from 1. This is done under the constraint that corrected flows must satisfy the balance of flows on all compartments.

This optimization system can easily be solved using the "*fmincon*" function available in MATLAB [2016] optimization toolbox. This function uses a deterministic linear approach to find a minimum to the problem, under specified constraints. Note that Jacobian matrices associated to the vector of minimization objectives, and to the vector of equality constraints, are easily expressed in terms of $w_{ij}$ values which facilitates the identification of a minimum.

The computation time associated to the minimization process was about a few minutes. Correction weights identified through this optimization process led to circulation matrices $\boldsymbol{M'}$ for each phase and operating condition that were balanced with an absolute error on flow balances always lower than $1.8\ 10^{-15}$, which is only 8 time the machine error under double precision. Note that this value is the maximum observed error, but for most compartments, this error was even lower. Therefore, the balance constraint is indeed abided to, thank to the resolution of the constrained minimization problem.

Even though circulation matrix $\boldsymbol{M'}$ complies with the balance constraint, the high dimension of the problem prevents ensuring that the identified minimum is global and not only local. Therefore, there may exist a better solution to the problem than the one identified. As illustrated if Fig. IV.10, for all cases the correction weights were kept close to 1, with a standard deviation of only 4.5%. Highest corrections were around 15% which we considered to be sufficiently low to maintain macro and meso-mixing times, and main circulation loops, during the correction process. Therefore, we did not attempt to improve the correction by verifying whether the identified minimum is global, or only local. If needed, it will be more efficient to improve the numerical integration of flows in order for $\boldsymbol{M}$ to be closer from a balanced circulation matrix, rather than trying to identify the extrema of a high dimension optimization problem.

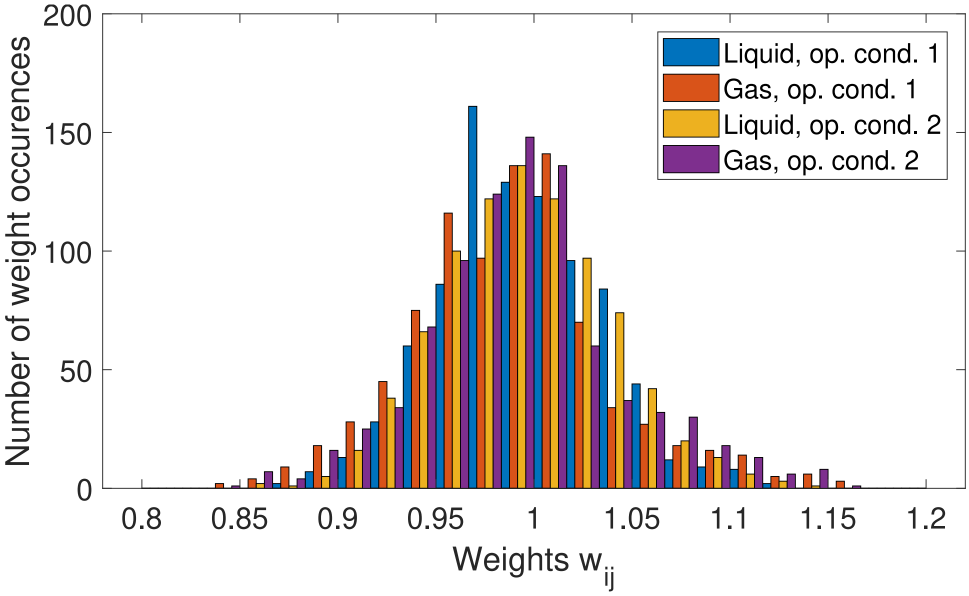


Figure IV.10: Distribution of correction weights $w_{ij}$ on compartment models associated to each phase and operating conditions described in section IV.2.

## IV.5.4 Conclusion and outlooks on the compartmentalization process

Using previous developments and pseudo-stationnary simulation results from ANSYS Fluent [2015], the production of a conservative compartment model is a simple two-step process:

1. Use the developed C++ tool to integrate scalar and vector fields over compartments.
2. "Clean" the obtained circulation maps by solving a multi-dimensional optimisation problem.

For now, the second step was performed using MATLAB [2016] and functions available in its optimisation toolbox, but certainly libraries are available to integrate this step directly in the C++ tool. Obtaining a compartment model now simply consists in defining the number of compartments along each direction.

For arbitrary reasons, we decided to fix the number of orthoradial divisions to $N_a = 6$, by simply considering the number of baffles (three) and blades for bottom and top turbines (respectively six and three blades). From there, we defined the number of radial divisions $N_r = 6$ and axial divisions $N_z = 12$ to preserve similar compartment lengths along each direction. After the compartmentalization process, we measured the mean residential time associated to each compartment. Obtained min and max values are 0.281s and 4.31s for a operation condition 1, and 0.196s and 3.08s for operating conditions 2 (see Table IV.1). These values being of the same order of magnitude than the smallest time-scale associated to the biological model ($\tau_\sigma^{(u)} = 3\text{s}$), we considered this compartmentalization to be sufficient.

Developments related to the compartmentalization process came late over the course of this PhD thesis. After obtaining a conservative compartment model, focus was quickly shifted toward coupled simulations presented in Chapter V. Therefore, some aspects would still require supplementary work. The first aspect, which would not need much investment, would be to assess the sensitivity of coupled simulations to the number of divisions by increasing and decreasing the total number of compartments. The second aspect lies in the exploitation of CFD results from NEPTUNE_CFD [2016].

NEPTUNE_CFD [2016] was used to perform unsteady simulations of the fermenter hydrodynamics. The simulated time of transient hydrodynamic cannot be directly used to perform simulation spanning a whole fermentation culture (several days). Fields of velocity, phase fractions and turbulence variables evolve through time and space and can be decomposed into mean fields, periodic fields, and noise fields. Exploiting them to obtain a unstationary compartment model would require to filter out noise fields and to ensure the periodicity of field by enforcing a cyclic pattern over a short period of time that would be repeated to span days of fermentation culture. This work might require coupling Proper Orthogonal Decomposition (POD) of these fields [Bergmann, 2004; Liné et al., 2013] with Fourier transform. These are only raw ideas and would require maybe a year of dedicated work. In particular, we already observed an unbalance of phase mass balances with stationary CFD results. This problem will necessarily occur again with unsteady hydrodynamics, and will be more difficult to tackle considering that the previous constraint on circulation maps,

$$\sum_{j=1}^{N_C} Q_{ij} - Q_{ji} = 0, i \in \{1, \ldots, N_C\},$$

would become

$$\sum_{j=1}^{N_C} Q_{ij}(t) - Q_{ji}(t) = V_i \frac{\partial \langle \alpha_i \rangle (t)}{\partial t}, i \in \{1, \ldots, N_C\}$$

which is not a simple scalar problem any-more. A first step would then be to attempt these developments on liquid-only CFD simulation results. Therefore, the exploitation of the simulation results database from NEPTUNE_CFD [2016] to obtain compartment models has not been tackled. This database is kept for ulterior exploitation.

## IV.6 Conclusion

Large-scale bioreactor always tend to be heterogeneous. This fact is particularly true for the substrate-limited fedbatch culture which constitutes the application example of the overall modelling framework developed over this thesis. To account for spatial heterogeneity and gradients, one needs to model the transport of species carried by both phases. This, in turns, requires the knowledge of fluid velocities and spatial distribution of both phases.

In order to access hydrodynamics information, Computational Fluid Dynamics is the most suited tool as it gives a good idea of organised flow structures, associated mixing times and gas dispersion, without other knowledge than the system geometry and operating conditions. Therefore, the considered bioreactor has been simulated using two software and different approaches to account for the rotating agitator shaft, and to describe the gas phase.

ANSYS Fluent has been used to perform pseudo-stationary simulations using the MRF approach. By virtually adding source terms to momentum conservation equation associated to centrifugal and Coriolis pseudo-forces in an rotating area, this modelling approach is used to describe the effect of agitators without actually obtaining a transient hydrodynamics. In these simulations, the gas injection had to be simplified by not describing a continuous gas phase at its injection point, to ensure computation stability.

NEPTUNE_CFD was used to perform more detailed simulations, by accounting both for the transient –periodic– effect of agitators on velocity fields through a moving-mesh approach, and for the presence of both continuous and dispersed gas-phase using the Generalized Large-Interface Model [Coste, 2013; Merigoux et al., 2016]. Though being available for future developments related to this work, these simulation results have not been further exploited due to difficulties associated to the formulation of dynamic compartment models.

CFD is a powerful tool to obtain large-scale hydrodynamics, and both used software allow the resolution of mass balance equation of species carried by phases. These equations could account for source/sink terms in order to model bioreactions as well as the evolution of biological properties. Therefore, it could have been possible to integrate models from Chapters II and III into CFD simulations. However, it has been observed that performing this coupling within the CFD simulation framework is actually infeasible due to prohibitively expensive computations. This numerical cost comes from two origins: (i) the high number of mesh cells in which all variables must be resolved, and for which the metabolic model must be computed multiple times due to biological heterogeneity; (ii) the Courant condition [Courant et al., 1928] which enforces a small time-step to ensure the stability of the transport scheme, while simulations must span dozens of hours of biological culture. Therefore, the use of Compartment Model Approach seems to be a more interesting choice to perform the coupling of species transport with other aspects of the modelling.

Multiple works already focused on using CFD simulation results to form compartment models. Two main methodologies exist: (i) the automatic zoning, which aims toward homogeneous compartments in terms of some variables (*i.e.* turbulent kinetic energy), and (ii) the manual zoning where compartments are defined by setting arbitrary geometric boundaries. The second option has been favoured as it will simplify the comparison between different operating conditions, and was found not to produce significantly different results compared to the automatic zoning [Delafosse et al., 2010]. Moreover, manual zoning is a requirement if one aims at producing dynamic compartment models.

As CFD simulations ran using two different software and with a complex geometry and associated meshing, specific developments were required to perform the integration of scalar and vector fields over the manually defined compartmentalization. Both ANSYS Fluent and NEPTUNE_CFD can export results in the same –documented– Ensight Gold format which enabled the development of a common tool to perform these integrations. An approach based on mesh refinement –or splitting– has been proposed and successfully applied to pseudo-stationary CFD results originating from ANSYS Fluent. Numerical error led to a non-conservative compartment model but this could be easily solved by a simple optimization-based correction of the circulation map. The formulation of this optimization problem was facilitated by the pseudo-steady state associated to the MRF approach. While the use of the same integration tool could be directly applied to unsteady CFD results from Neptune CFD, the fact that local phase fractions evolve over time complicates the post-integration correction of computed flow to obtain a conservative compartment model. This task has not been tackled over the course of this work and is left for future developments.

Overall, through the use of numerical tools, it has been possible to obtain a conservative gas-liquid compartment model for multiple operation conditions on an industrial fermenter. Through Chapter II to IV, all scales of interest for the description of bioreactors have been modelled, using matched accuracy. Simulation coupling transport, biological heterogeneity and dynamic metabolism are therefore accessible as it will be demonstrated in the following chapter.

# V

# SIMULATION D'UN BIORÉACTEUR INDUSTRIEL

## Résumé

En couplant les développements issus des Chapitres précédents, en particulier le modèle métabolique pour *S. cerevisiae* (Chapitre II), la résolution d'une équation de bilan de population par la méthode EQMOM (Chapitre III) et les modèles de compartiments d'un bioréacteur industriel (Chapitre IV), il est désormais possible de réaliser la simulation de cultures industrielles et d'illustrer les capacités de la structure proposée pour la modélisation des bioréacteurs. Ces simulations sont le sujet principal de ce Chapitre.

Dans un premier temps, les modèles et méthodes numériques précédemment développés sont résumés afin de souligner en quoi leur usage couplé induit une modélisation fermée des bioréacteurs intégrant les aspects listés au Chapitre I : fonctionnement cellulaire interne, gradients aux grandes échelles et leur couplage bidirectionnel au travers de l'hétérogenéité biologique. De plus, la manière dont est modélisée l'articulation entre limitations biologiques et environnementales est détaillée ici, et s'appuie sur des considérations abordées dans les Chapitres précédents.

Le système simulé est un fermenteur industriel opérant en culture fedbatch. Ces cultures sont parmis les plus difficiles à simuler car leurs comportements sont sensibles aussi bien aux conditions initiales du système qu'aux conditions opératoires tout au long de la culture. Afin de se défaire de ces sensibilités, nous simulerons un état pseudo-stationnaire de la culture. Ces simulations visent à quantifier l'effet des conditions opératoires de mélange et d'injection de gaz, ainsi que l'importance de la stratégie d'injection du substrat, sur l'apparition de comportements métaboliques indésirables, *i.e.* les productions d'éthanol et d'acétate.

Il est ainsi montré que dans des conditions opératoires de référence, éthanol et acétate sont produits en conséquence (i) d'un transfert gaz-liquide d'oxygène insuffisant et (ii) d'un dysfonctionnement métabolique attribué à l'effet Crabtree. En multipliant les points d'injection du substrat principal, nous montrons que la production d'éthanol peut être réduite de 48% sans coût énergétique supplémentaire. En augmentant la dépense énergétique par un mélange plus rapide et un débit de gaz injecté plus élevé, la production d'éthanol peut être entièrement inhibée. Ces simulations, servant de preuves de concept à la mise en œuvre de la modélisation proposée des bioréacteurs, montrent l'intérêt de disposer de tels outils facilitant les prises de décision sur les problématiques d'ingénierie.

# V

# SIMULATION OF AN INDUSTRIAL FERMENTER

## Summary

By coupling developments from previous Chapters, namely the *S. cerevisiae* metabolic model from Chapter II, the resolution of the population balance using EQMOM as detailed in Chapter III and the compartment models from Chapter IV, it is now possible to perform simulations of an industrial fermenter to show the capabilities of the proposed bioreactor modelling framework. These simulations are the core topic of this Chapter.

Previously detailed models and numerical methods are first summarized to underline how their conjoint use form a closed bioreactor model that accounts for all aspects listed in Chapter I: inner cell functioning, large-scale gradients, and their two-way coupling through biological heterogeneity. Moreover, the articulation between biological and environmental metabolic limitations is discussed here as this aspect requires considerations from all three previous Chapters.

The simulated system is an industrial fermenter used in fedbatch culture mode. These cultures are the most difficult to simulate as their courses depend both on initial conditions and operating conditions over time. In order to discard any effect of initial conditions, we choose to simulate a pseudo-stationary state of the system. These simulations aim at quantifying the effect of mixing and gas-flow rate operating conditions, as well as the importance of the substrate injection strategy, on the occurrence of undesired metabolic behaviours, *i.e.* the productions of ethanol or acetate.

It is shown that under reference operating conditions, ethanol and acetate are produced as a consequence of both (i) poor gas-liquid mass transfer of oxygen and (ii) metabolic dysfunction due to the Crabtree effect. By changing the substrate injection design, we show that the ethanol production can be reduced by 48% at no extra energetic cost. By increasing the energetic expense through faster mixing and higher gas-flow rates, this ethanol production can be entirely prevented. These proof-of-concept simulations of a large-scale bioreactor show the interest of such a decision-assisting tool for engineering purposes.

# V

# SIMULATION OF AN INDUSTRIAL FERMENTER

## V.1 Introduction

The simulation of all phenomenon in bioreactors is a challenging topic, and will require a computational power that will probably not be accessible within the next few years or even decades. However, as it has been thoroughly discussed throughout this thesis, describing all details of these phenomenon is not actually required, which allows for a simplified yet complete enough modelling of industrial fermenters. This Chapter will focus on demonstrating that, indeed, accurate enough simulations of large-scale fermenters can be performed, at reasonable numerical cost, to obtain an engineering assisting tool. Doing so will require to integrate developments of previous three Chapters.

Chapters II, III and IV focused on the development or applications of models and numerical methods that, when coupled, yield a bioreactor model complete enough to account simultaneously for large-scale gradients, biological heterogeneity, and metabolism complexity. More specifically, Chapter IV presented the methodology used to obtain a gas-liquid hydrodynamic model of an actual industrial fermenter. Chapter III focused on numerical methods to efficiently account for biological heterogeneity in terms of some dynamic parameters whose effect on metabolism is non negligible. Finally, Chapter II detailed a structure for metabolic models that can account for most cell behaviour observed experimentally, but at a very low numerical cost. Section V.2 sums-up these models, methods, and the way they can easily be used together.

Developed models will be applied to the simulation of a the culture of *S. cerevisiae*, in the bioreactor presented in Chapter IV. Section V.3 details these simulations in terms of culture mode, operating conditions, and simulation tool. Section V.4 presents results of these simulations and underlines how these could be used for engineering purposes.

## V.2 Summary of developed model

Chapter I details most approaches that can be used to model bioreactors. We selected those that appeared to be the most numerically efficient, while being accurate enough, to obtain a model that do not neglect spatial nor biological heterogeneities, and accounts for metabolic consequences of these heterogeneities. Hereafter, these approaches are briefly reminded to clarify which set of equations is solved, and which closures are used. Some modelling aspects, related to limitations of mass transfer between phases have not been particularly considered in previous Chapters and will be detailed in this section.

### V.2.1 Mass balance equations

Simulations detailed in Section V.4 are about fedbatch cultures of *S. cerevisiae* in an industrial-scale gas-liquid fermenter. The model developed for this yeast in Chapter II accounts for four species dissolved in the liquid phase that can be ingested or produced by cells: glucose, ethanol, acetate and oxygen. Out of these, only oxygen also exists in the gas phase. It has been explained in Chapter IV that spatial heterogeneity will be accounted for by a Compartment Model Approach, and in chapter III that biological heterogeneity will be tracked using moment methods applied to a 5 variables Population Balance Equation (PBE).

Therefore, simulations will track three vectors: liquid-phase concentrations $\boldsymbol{C}_L = [G_L,\ E_L,\ A_L,\ O_L]^T$, gas-phase concentrations $\boldsymbol{C}_G = [O_G]$ and moments of biological-phase properties distribution $\boldsymbol{M}$, in each compartment of the hydrodynamic model:

$$\frac{\partial \varepsilon_{L,i} V_i \boldsymbol{C}_{L,i}(t)}{\partial t} = \sum_{j=1}^{N_C} Q_{j,i}^{(L)} \boldsymbol{C}_{L,j}(t) - \boldsymbol{C}_{L,i}(t) \sum_{j=1}^{N_C} Q_{i,j}^{(L)} + \boldsymbol{\Phi}_{BL,i} + \boldsymbol{\Phi}_{GL,i} \tag{V.1}$$

$$\frac{\partial \varepsilon_{G,i} V_i \boldsymbol{C}_{G,i}(t)}{\partial t} = \sum_{j=1}^{N_C} Q_{j,i}^{(G)} \boldsymbol{C}_{G,j}(t) - \boldsymbol{C}_{G,i}(t) \sum_{j=1}^{N_C} Q_{i,j}^{(G)} + \boldsymbol{\Phi}_{LG,i} \tag{V.2}$$

$$\frac{\partial \varepsilon_{L,i} V_i \boldsymbol{M}_i(t)}{\partial t} = \sum_{j=1}^{N_C} Q_{j,i}^{(L)} \boldsymbol{M}_j(t) - \boldsymbol{M}_i(t) \sum_{j=1}^{N_C} Q_{i,j}^{(L)} + \boldsymbol{Z}_i \tag{V.3}$$

All terms appearing in these balance equations are detailed hereafter.

### V.2.2 Compartment models for industrial fermenters

The transport of dissolved or biological species by gas and liquid phases is described by a Compartment Approach. This model describes the industrial fermenter as $N_C = 432$ compartments as described in Chapter IV and illustrated in Fig. IV.6 through a manual zoning approach.

This models defines $V_i$ as the $i$-th compartment volume (m$^3$); $\varepsilon_{L,i}$ and $\varepsilon_{G,i}$ respectively the liquid and gas phase fraction ($\mathrm{m_L^3/m^3}$ and $\mathrm{m_G^3/m^3}$); $Q_{i,j}^{(L)}$ and $Q_{i,j}^{(G)}$ the volumetric flow rate from the $i$-th to the $j$-th compartment for the liquid and gas phases ($\mathrm{m_L^3/s}$ and $\mathrm{m_G^3/s}$).

Chapter IV detailed how these model parameters could be obtained from a pseudo-stationary Computational Fluid Dynamics (CFD) simulation of the industrial fermenter. Therefore, for upcoming simulations, these parameters will be kept constant over time. Further work will be required to obtain transient compartment models. Note that in the compartmentalization process, fields of turbulent kinetic energy $k$ ($\mathrm{J/kg_L}$) and its dissipation rate $\epsilon$ ($\mathrm{W/kg_L}$) were post-processed so that following models may access $k_i$ and $\epsilon_i$ in each compartment if required. This will be of particular interest in models for inter-phase mass transfer limitation.

### V.2.3 Population Balance Equations tracking biological heterogeneity

The metabolic model described in Chapter II and briefly recalled hereafter accounts for five cell-attached dynamic variables:

- $\mu_G^{(b)}$: glucose-based growth capability ($\mathrm{g_X/g_X \cdot h}$);
- $\mu_E^{(b)}$: ethanol-based growth capability ($\mathrm{g_X/g_X \cdot h}$);
- $\mu_A^{(b)}$: acetate-based growth capability ($\mathrm{g_X/g_X \cdot h}$);
- $\varphi_O^{(b)}$: oxidative capability ($\mathrm{mol_O/g_X \cdot h}$);
- $\sigma$: glycolytic stress.

Let $\widehat{p} = [\mu_G^{(b)},\ \mu_E^{(b)},\ \mu_A^{(b)},\ \varphi_O^{(b)},\ \sigma]^T$ be the vector of these properties, and $n_i(\widehat{p})$ the statistical distribution of these properties over cells located in the $i$-th compartment. This distribution is defined such that

$$\int_{\Omega\widehat{p}} n_i(\widehat{p})\mathrm{d}\widehat{p} = X_{L,i} \tag{V.4}$$

where $X_{L,i}$ is the biomass concentration $(kg_X/m_L^3)$ in the $i$-th compartment.

A PBE tracking $n_i(\widehat{p})$ is

$$\frac{\partial \varepsilon_{L,i} V_i n_i(\widehat{p},t)}{\partial t} + \varepsilon_{L,i} V_i \widehat{\nabla} \cdot \left(\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_{L,i}) n(\widehat{p})\right) = \sum_{j=1}^{N_C} \left(Q_{j,i}^{(L)} n_j(\widehat{p},t) - n_i(\widehat{p},t) Q_{i,j}^{(L)}\right) + \int_{\Omega\widehat{p}} r_X(\widehat{p}', \boldsymbol{C}_{L,i}) \beta(\widehat{p}, \widehat{p}') n(\widehat{p}') \mathrm{d}\widehat{p}' \tag{V.5}$$

where $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_{L,i})$ is the vector of adaptation rates of biological properties, $r_X(\widehat{p}, \boldsymbol{C}_L)$ is the actual growth rate of cells of properties $\widehat{p}$ in a liquid environment of composition $\boldsymbol{C}_L$, and $\beta(\widehat{p}, \widehat{p}')$ the probability that a cell of properties $\widehat{p}'$ yields a daughter cell of properties $\widehat{p}$. The modelling of $\widehat{\zeta}(\widehat{p}, \boldsymbol{C}_{L,i})$ and $r_X(\widehat{p}, \boldsymbol{C}_L)$ is handled hereafter by the metabolic model.

Due to a lack of experimental data, $\beta(\widehat{p}, \widehat{p}')$ will simply be defined as a 5-dimensional Dirac distribution

$$\int_{\Omega\widehat{p}} f(\widehat{p}')\beta(\widehat{p}, \widehat{p}')\mathrm{d}\widehat{p}' = f(\widehat{p}) \tag{V.6}$$

which implies that daughter cells inherit properties of their mother.

The use of moment methods has been selected in Chapter III for their efficiency. Let moments of $n(\widehat{p})$ be defined as

$$m_{k_1,k_2,k_3,k_4,k_5} = \int_{\Omega\mu_G} \int_{\Omega\mu_E} \int_{\Omega\mu_A} \int_{\Omega\Phi_O} \int_{\Omega\sigma} \mu_G^{k_1} \mu_E^{k_2} \mu_A^{k_3} \varphi_O^{k_4} \sigma^{k_5} n(\widehat{p}) \mathrm{d}\sigma \mathrm{d}\varphi_O \mathrm{d}\mu_A \mathrm{d}\mu_E \mathrm{d}\mu_G \tag{V.7}$$

Due to the fact that only monovariate numerical methods were considered in Chapter III, only following uncoupled moments will be resolved:

$$\boldsymbol{M} = \begin{bmatrix} m_{0,0,0,0,0} \\ m_{1,0,0,0,0} \\ m_{0,1,0,0,0} \\ m_{0,0,1,0,0} \\ m_{0,0,0,1,0} \\ m_{0,0,0,0,1} \\ \vdots \\ m_{0,0,0,0,N_{m\sigma}} \end{bmatrix} = \begin{bmatrix} X_L \\ \tilde{\mu}_G X_L \\ \tilde{\mu}_E X_L \\ \tilde{\mu}_A X_L \\ \tilde{\varphi}_O X_L \\ \tilde{\sigma} X_L \\ \vdots \\ m_{0,0,0,0,N_{m\sigma}} \end{bmatrix} \tag{V.8}$$

with $N_{m\sigma}$ the number of moments resolved along the stress dimension of the distribution. $\tilde{x}$ denotes a population mean variable. The rational behind the choice of these moments has been detailed in Section III.7.

The Extended Quadrature Method of Moments (EQMOM) will be applied on moments $m_{0,0,0,0,k_5}$, $k_5 \in \{0, \dots, N_{m\sigma}\}$ to obtain a quadrature rule $\boldsymbol{w}_Q = [w_1, \dots, w_Q]^T$, $\boldsymbol{\sigma}_Q = [\sigma_1, \dots, \sigma_Q]^T$, such that

$$\int_{\Omega_{\widehat{p}}} f(\widehat{p}) n(\widehat{p}) \mathrm{d}\widehat{p} \approx \sum_{j=1}^{Q} w_j f(\widehat{p}_j) \tag{V.9}$$

with $\widehat{p}_j = [\tilde{\mu}_G \ \tilde{\mu}_E \ \tilde{\mu}_A \ \tilde{\Phi}_O \ \sigma_j]^T$.

### V.2.4 Metabolic model for *S. cerevisiae*

The metabolic model of *S. cerevisiae* allows the computation of $\boldsymbol{r}_{\boldsymbol{C}_L}$, $r_X$ and $\widehat{\zeta}$, considering biological properties $\widehat{p}$ and the environment composition $\boldsymbol{C}_L$, with

$$\boldsymbol{r}_{\boldsymbol{C}_L}(\widehat{p}, \boldsymbol{C}_L) = \begin{bmatrix} r_G(\widehat{p}, \boldsymbol{C}_L) \\ r_E(\widehat{p}, \boldsymbol{C}_L) \\ r_A(\widehat{p}, \boldsymbol{C}_L) \\ r_O(\widehat{p}, \boldsymbol{C}_L) \end{bmatrix}, \qquad \widehat{\zeta}(\widehat{p}, \boldsymbol{C_L}) = \begin{bmatrix} \zeta_{\mu_G}(\widehat{p}, \boldsymbol{C_L}) \\ \zeta_{\mu_E}(\widehat{p}, \boldsymbol{C_L}) \\ \zeta_{\mu_A}(\widehat{p}, \boldsymbol{C_L}) \\ \zeta_{\Phi_O}(\widehat{p}, \boldsymbol{C_L}) \\ \zeta_{\sigma}(\widehat{p}, \boldsymbol{C_L}) \end{bmatrix} \tag{V.10}$$

$r_i$ is the rate of consumption or production of $i \in \{G, E, A, O, X\}$ $(\mathrm{kg_i}/\mathrm{kg_X} \cdot \mathrm{h})$, and $\zeta_i$ the rate of adaptation of the biological property $i \in \{\mu_G, \mu_e, \mu_A, \Phi_O, \sigma\}$ ([unit of i]/h). The model structure proposed in Chapter II performs these computations at very low numerical cost, between 135 and 150 floating-point operations by set of biological and environmental properties $(\widehat{p}, \boldsymbol{C}_L)$.

$\widehat{\zeta}(\widehat{p}, \boldsymbol{C_L})$ and $r_X(\widehat{p}, \boldsymbol{C_L})$ are directly used in the moment-tracking equation. Coupling the quadrature rule in Eq. (V.9) with $\boldsymbol{r}_{\boldsymbol{C}_L}(\widehat{p}, \boldsymbol{C}_L)$ yields the mass transfer rate from liquid phase to biological phase, $\boldsymbol{\Phi}_{LB,i}$ in Eq. (V.1):

$$\boldsymbol{\Phi}_{LB,i} = \int_{\Omega_{\widehat{p}}} \boldsymbol{r}_{\boldsymbol{C}_L}(\widehat{p}, \boldsymbol{C}_{L,i}) n_i(\widehat{p}) \mathrm{d}\widehat{p} \approx \sum_{j=1}^{Q} w_{i,j} \boldsymbol{r}_{\boldsymbol{C}_L}(\widehat{p}_{i,j}, \boldsymbol{C}_{L,i}) \tag{V.11}$$

with $w_{i,j}$ and $\widehat{p}_{i,j}$ the quadrature weights and nodes computed from moments $\boldsymbol{M}_i$, $j \in \{1, \ldots, Q\}$.

### V.2.5 Mass transfer between gas and liquid phases

As briefly described in Section I.6.1, the model for gas-liquid mass-transfer will only consider the resistance to transfer in the liquid phase. The transfer rate of oxygen will be described by

$$\Phi_{GL,O} = k_L a^* V (H\ O_G - O_L) \tag{V.12}$$

with

- $k_L$ the liquid-film resistance to transfer (m/s);
- $a^*$ the specific inter-facial area ($\mathrm{m}^2/\mathrm{m}^3$);
- $V$ the considered volume of control ($\mathrm{m}^3$);
- $H$ the Henry constant for oxygen at the culture temperature ($\mathrm{kg_{O_L}}/\mathrm{kg_{O_G}}$).

The specific transfer area will be deduced from an estimated mean bubble diameter of $d_b \approx 1\mathrm{cm}$ and the local gas fraction $\varepsilon_G$:

$$a^* = \frac{6\varepsilon_G}{d_b} \tag{V.13}$$

The resistance to transfer, $k_L$, is modelled using the following correlation from Kawase et al. [1992]; Laakkonen et al. [2006] suited for mass transfer in bioreactors:

$$k_L = 0.3\sqrt{\mathcal{D}_{O,L}} \left( \frac{\epsilon}{\varepsilon_G \nu_G + \varepsilon_L \nu_L} \right)^{1/4} \tag{V.14}$$

with $\mathcal{D}_{O,L} = 1.98\ 10^{-9}\mathrm{m}^2/\mathrm{s}$ the oxygen diffusivity in water, and $\nu_k = \mu_k/\rho_k$ the kinematic viscosity of phase $k$ ($\mathrm{m}^2/\mathrm{s}$).

The solubility was set to $H = 0.0272\mathrm{kg_{O_L}}/\mathrm{kg_{O_G}}$ which corresponds to a saturation oxygen concentration of $7.8\mathrm{g_O}/\mathrm{m_L^3}$ for water at equilibrium with surrounding air under atmospheric pressure and a temperature of $20°C$.

Oxygen being the only compound that transfers between gas and liquid phases, vectors of gas-liquid transfer rates in Eqs. (V.1-V.2) are simply defined as

$$\boldsymbol{\Phi}_{GL} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Phi_{GL,O} \end{bmatrix}, \qquad \boldsymbol{\Phi}_{LG} = [-\Phi_{GL,O}] \tag{V.15}$$

## V.2.6 Mass transfer toward the biological phase

### Integrating biological and physical regimes in the modelling

As for now, the coupling of a metabolic model with the description of a variable environment has not been properly addressed while this topic is key to the modelling of industrial bioreactors.

In Chapter II, the metabolic model was formulated by assuming that one knows the flux of matter (glucose, ethanol, acetate, oxygen) that cells are able to assimilate. These values were defined as cell-attached properties, either constant for carbonaceous substrates ($\varphi_G^{(b)}$, $\varphi_E^{(b)}$ and $\varphi_A^{(b)}$) or variable for the oxidative capability ($\varphi_O^{(b)}$). These fluxes are meaningful when the cell functioning, or bioreactions in general, are in the biological regime: when growth is limited by cells capabilities to process substrate available in the environment. Biological regime occurs when the environment is rich in substrates compared to a cell capability to process that substrate.

The inverse situation is the physical regime, which occurs when the environment is poor compared to cell capabilities. The coupled modelling must be able to account for both biological and physical regimes, as well as the transitional regime where both limitation occurs. This will allow the model to deal with poor –limiting– environments.

### Initial proposition based on Monod kinetics

In previous work [Pigou and Morchain, 2015], we considered the modelling of *E. coli* and described the glucose-based growth capability $\mu_G^{(b)}$ as a dynamic variable attached to cells, and compared it to the Monod equilibrium growth rate $\mu_G^*$, to assess whether the cell was in the biological or the physical regime. Under this approach, the actual growth rate $\mu_G^{(a)}$ was defined as

$$\mu_G^{(a)} = \min(\mu_G^{(b)}, \mu_G^*) \tag{V.16}$$

which exhibited expected asymptotic behaviours:

$$\begin{cases} \mu_G^{(a)} = \mu_G^{(b)} & G \gg K_G \\ \mu_G^{(a)} = 0 & G \mapsto 0 \end{cases} \tag{V.17}$$

with $K_G$ the affinity of cells toward glucose ($\mathrm{kg_G/m_L^3}$).

Despite its correct asymptotic behaviour, the approach based on Monod kinetic law is unsatisfactory on multiple regards.

1. This approach associates instantaneous metabolic behaviours to variables defined after a long-term adaptation of a population to its environment, *i.e.* the equilibrium growth-rate $\mu^*$.
2. As $0 \leq \mu_G^* \leq \mu_G^{(\max)}$ by definition of Monod kinetics, this formulation forbids cells to have an actual growth rate higher than $\mu_G^{(\max)}$, despite experimental proofs that exceeding this Monod maximum growth-rate is actually feasible at the cell-scale.
3. This formulation does not explicit why cells actually slow their growth rate down in poor environments.

For these reasons, a more satisfactory and predictive description of the switch between physical and biological regimes has been looked for.

## A micro-mixing issue

It is easily understood that transport of substrate in the culture medium up to a cell membrane precedes its uptake and use in anabolic or catabolic reactions. Bioreactions can then be limited by either (i) external transport, (ii) mass transfer through the biological membrane or (iii) internal transport and/or availability of reaction catalyzers (*i.e.* enzymes).

In models developed in Chapter II for *E. coli* and *S. cerevisiae*, limitations of transport through the biological membrane and of reactive capabilities are encompassed within (i) maximum consumption flow rates $\varphi^{(b)}_{\{G,E,A,O\}}$ and (ii) growth capabilities $\mu^{(b)}_{\{G,E,A\}}$. A lack of these capabilities will induce a biological regime for growth or reaction.

Therefore, including physical regime to the model requires to account for limitations of external transport too. On many regards, biological systems can be analysed under the point of view of heterogeneous catalysis. Micro-organisms behave as reactive particles in a turbulent flow that catalyse the transformation of substrates into products and by-products, and also happen to generate new reactive particles through cell division or budding. To analyse heterogeneous catalysis systems, a classical approach in chemical engineering is to compare the reaction rate to the transport rate of reactants. This is done through the so-called Damköhler number Da:

$$\mathrm{Da} = \frac{\text{(bio)reaction rate}}{\text{transport rate}} = \frac{\text{transport characteristic time-scale}}{\text{(bio)reaction characteristic time-scale}} \tag{V.18}$$

where Da $\ll 1$ denotes a reactive –or biological– regime, and Da $\gg 1$ denotes a physical regime. Detecting the regime then requires the identification of the limiting transport phenomenon. Considering the cell size ($\approx 5-10\mu$m), the liquid medium surrounding cells will be renewed through micromixing that encompasses both engulfment micromixing at the Kolmogorov scale, and diffusive mixing at the Batchelor scale [Delafosse, 2008]. Therefore, we previously proposed [Morchain et al., 2016] to base the physical limitation of bioreactions on the micro-mixing time-scale associated to the lifespan of engulfing eddies [Baldyga and Bourne, 1999]:

$$\tau_E \approx 17\sqrt{\frac{\mu_L}{\rho_L \epsilon}} \tag{V.19}$$

with $\mu_L$ and $\rho_L$ the dynamic viscosity (Pa $\cdot$ s) and density ($\mathrm{kg_L/m^3_L}$) of the liquid phase, and $\epsilon$ the local turbulent-energy dissipation rate ($\mathrm{W/kg_L}$).

Overall, transport by micro-mixing and assimilation by cells are two serial phenomenon whose limiting rate must be simultaneously considered to compute a maximum achievable consumption rate for substrates. Let $\varphi^{(b)}_S$ be the maximum specific consumption rate of substrate $S \in \{G, E, A, O\}$ considering biological limitations as already defined in Chapter II, and $\varphi^{(e)}_S$ be the maximum specific consumption rate considering environmental limitations ($\mathrm{kg_S/kg_X \cdot h}$). We propose the following model for $\varphi^{(e)}_S$:

$$\varphi^{(e)}_S \approx \frac{S_L}{X_L \tau_E} \tag{V.20}$$

with $S_L$ the substrate concentration ($\mathrm{kg_S/m_L^3}$), $X_L$ the biomass concentration ($\mathrm{kg_X/m_L^3}$) and $\tau_E$ the micro-mixing time-scale as defined previously. This formulation was chosen as it is the most simple that exhibits the three following behaviours:

1. If the environment is depleted in terms of substrate $S$, environmental limitation occurs: $\lim\limits_{S\mapsto 0}\varphi_S^{(e)} = 0$.
2. At low turbulence, micro-mixing is inefficient and a limitation occurs: $\lim\limits_{\epsilon\mapsto 0}\varphi_S^{(e)} = 0$.
3. At constant substrate concentration, if the amount of biomass increases, less substrate will be available per cell as we assume that substrate is distributed evenly between micro-organisms: $\lim\limits_{X\mapsto +\infty}\varphi_S^{(e)} = 0$

Once maximum consumption rates associated to environment limitations $\varphi_S^{(e)}$ and to biological limitations $\varphi_S^{(b)}$ are known, one has to compute a global maximum consumption rate $\varphi_S^{(g)}$ which cannot be higher than either $\varphi_S^{(e)}$ or $\varphi_S^{(b)}$. We chose to apply the following formulation, based on serial resistance law:

$$\frac{1}{\varphi_S^{(g)}} = \frac{1}{\varphi_S^{(e)}} + \frac{1}{\varphi_S^{(b)}} \Leftrightarrow \varphi_S^{(g)} = \frac{\varphi_S^{(e)}\varphi_S^{(b)}}{\varphi_S^{(e)} + \varphi_S^{(b)}} \tag{V.21}$$

This formulation has the same asymptotic behaviours than a simple $\varphi_S^{(g)} = \min\left(\varphi_S^{(e)}, \varphi_S^{(b)}\right)$ but offers a smoother transition between biological and physical regimes.

Note that the oxygen consumption rate has been described in a slightly different manner. Indeed, when biological oxygen consumption exceeds gas-liquid mass transfer, the liquid phase gets depleted and oxygen consumption enters a physical regime. Supposedly, an equilibrium should be reached between gas-liquid mass transfer and biological consumption of oxygen. However, the high sensitivity of the metabolic behaviour over oxygen consumption leads to a stiff set of Ordinary Differential Equations (ODEs) as far as oxygen mass balance equations are concerned.

In order to relax the stiffness of this problem, $\varphi_O^{(e)}$ is defined as

$$\varphi_O^{(e)} \approx \frac{O_L}{X_L\tau_E} + \frac{\Phi_{GL,O}}{X_L} \tag{V.22}$$

with $\Phi_{GL,O}$ defined previously in Eq. (V.12). This formulation implies that oxygen can be transported to cells either by micro-mixing when dissolved oxygen is available, or directly from the gas-phase to cells. This could be justified by considering that, when the liquid phase is mainly depleted in terms of oxygen, micro-organisms close to bubbles will directly assimilate freshly dissolved oxygen.

When the liquid phase is close to oxygen saturation, the gas-liquid transfer rate is negligible and only micro-mixing brings oxygen to cells. On the other side, when low oxygen concentrations are observed, cells are still able to grow using aerobic pathways, with oxygen consumption being equal to the gas-liquid transfer rate.

This formulation makes $\varphi_O^{(g)}$ less sensitive to oxygen concentration variations in micro-aerobic environments, and strongly stabilises simulations, without altering the physical meaning of the overall model.

### V.2.7 Short summary

Overall, the bioreactor modelling consists in three sets of Ordinary Differential Equations per compartment recalled hereafter

$$\frac{\partial \boldsymbol{C}_{L,i}(t)}{\partial t} = \frac{1}{\varepsilon_{L,i} V_i}\left[\sum_{j=1}^{N_C} Q_{j,i}^{(L)} \boldsymbol{C}_{L,j}(t) - \boldsymbol{C}_{L,i}(t)\sum_{j=1}^{N_C} Q_{i,j}^{(L)} + \boldsymbol{\Phi}_{BL,i} + \boldsymbol{\Phi}_{GL,i}\right] \tag{V.23}$$

$$\frac{\partial \boldsymbol{C}_{G,i}(t)}{\partial t} = \frac{1}{\varepsilon_{G,i} V_i}\left[\sum_{j=1}^{N_C} Q_{j,i}^{(G)} \boldsymbol{C}_{G,j}(t) - \boldsymbol{C}_{G,i}(t)\sum_{j=1}^{N_C} Q_{i,j}^{(G)} + \boldsymbol{\Phi}_{LG,i}\right] \tag{V.24}$$

$$\frac{\partial \boldsymbol{M}_{i}(t)}{\partial t} = \frac{1}{\varepsilon_{L,i} V_i}\left[\sum_{j=1}^{N_C} Q_{j,i}^{(L)} \boldsymbol{M}_{j}(t) - \boldsymbol{M}_{i}(t)\sum_{j=1}^{N_C} Q_{i,j}^{(L)} + \boldsymbol{Z}_{i}\right] \tag{V.25}$$

where $i$ denotes a variable attached to the $i$-th compartment, and with

- $\varepsilon_L$ the liquid phase fraction ($\mathrm{m_L^3/m^3}$);
- $\varepsilon_G$ the gas phase fraction ($\mathrm{m_G^3/m^3}$);
- $V$ the compartment volume ($\mathrm{m^3}$);
- $Q_{i,j}^{(k)}$ the volumetric flow rate of the phase $k \in \{G, L\}$ from the $i$-th to the $j$-th compartment ($\mathrm{m_k^3/h}$);
- $\boldsymbol{C}_L = [G_L,\ E_L,\ A_L,\ O_L]^T$ the vector of liquid-phase attached concentrations ($\mathrm{kg/m_L^3}$);
- $\boldsymbol{C}_G = [O_G]$ the vector of gas-phase attached concentrations ($\mathrm{kg/m_G^3}$);
- $\boldsymbol{\Phi}_{BL} = [\Phi_{BL,G},\ \Phi_{BL,E},\ \Phi_{BL,A},\ \Phi_{BL,O}]^T$ the vector of mass transfer rates from the biological phase to the liquid phase (kg/h);
- $\boldsymbol{\Phi}_{GL} = [0,\ 0,\ 0,\ \Phi_{GL,O}]^T$ the vector of mass transfer rates from the gas phase to the liquid phase (kg/h);
- $\boldsymbol{\Phi}_{LG} = [-\Phi_{GL,O}]$ the vector of mass transfer rates from the liquid phase to the gas phase (kg/h);
- $\boldsymbol{M}$ the vector of moments of the distribution of biological properties;
- $\boldsymbol{Z}$ the vector of rates of change of elements of $\boldsymbol{M}$.

$\varepsilon_L$, $\varepsilon_G$, $V$ are defined for each compartment, along with circulation maps $Q_{i,j}^{(k)}$, by the compartment model computed in Chapter IV. The computation of transport terms, *i.e.* summations over $N_C$ compartments in Eqs. (V.23-V.25), is straightforward once the circulation map is known.

In each compartments, following computations are performed at each time-step:

1. Compute $\boldsymbol{\Phi}_{GL}$ and $\boldsymbol{\Phi}_{LG}$ by simply computing $\Phi_{GL,O}$ as detailed in Eq. (V.12).
    - Note that this computation is based on $k_L a^* V$ values which depend on local phase fractions and turbulent dissipation rates.

2. Compute the vector of local maximum specific consumption rates considering environmental limitations $\boldsymbol{\varphi}^{(e)}$ based on the local micro-mixing time-scale defined in Eq. (V.19):

$$\boldsymbol{\varphi}^{(e)} = \begin{bmatrix} \varphi_G^{(e)} \\ \varphi_E^{(e)} \\ \varphi_A^{(e)} \\ \varphi_O^{(e)} \end{bmatrix} = \begin{bmatrix} G_L/(X_L\tau_E) \\ E_L/(X_L\tau_E) \\ A_L/(X_L\tau_E) \\ (O_L/\tau_E + \Phi_{GL,O})/X_L \end{bmatrix} \tag{V.26}$$

3. Compute a quadrature rule $\boldsymbol{w}_Q$ and $\widehat{\boldsymbol{p}}_Q$ based on moments $\boldsymbol{M}$ as detailed in Section V.2.3, using the improved EQMOM method detailed in Chapter III.

4. Integrate the metabolic behaviour over quadrature nodes to obtain $\boldsymbol{\Phi}_{BL}$:

$$\boldsymbol{\Phi}_{BL} = \sum_{i=1}^{Q} w_i \boldsymbol{\varphi}^{(a)}(\boldsymbol{C}_L, \boldsymbol{\varphi}^{(e)}, \widehat{p}_i) \tag{V.27}$$

   were $\boldsymbol{\varphi}^{(a)}(\boldsymbol{C}_L, \boldsymbol{\varphi}^{(e)}, \widehat{p}_i)$ represents the actual consumption or production rates given as output of the metabolic model detailed in Chapter II.

## V.3 Simulations description

Now that the complete modelling framework has been extensively detailed in previous Chapters, and summarized in Section V.2, we shall present the illustrative simulations in which the framework has been applied. This section details a fedbatch culture and elements of comparison between multiple operating conditions.

### V.3.1 A pseudo-stationary fedbatch culture

**Culture mode**

In the context of baker's yeast production, many studies have shown that the excess of free glucose in the environment tends to induce aerobic ethanol production [Yamanè and Shimizu, 1984], which was referred to in Chapter II as the Crabtree effect. This ethanol production is identified as the main cause for lower conversion yields of glucose into cells.

In order to reduce this effect, a usual strategy is to perform fedbatch cultures of *S. cerevisiae*. In this cultivation mode, a pre-culture –usually in batch– allows to reach a high cell-density. After this first step, the system is continuously fed with a limiting quantity of substrate. This is expected to limit the amount of free glucose, and thus the occurrence of aerobic ethanol formation. However, this is not always the case and the Crabtree effect may still be triggered even in glucose-limited high cell-density fedbatch cultures. This is usually attributed to poor macro-mixing of the system through the short-term Crabtree effect.

For the industrial case of interest, the culture goal is actually not cell production, and for undisclosed reasons, the culture medium needs to be fed both with glucose and ethanol. Note that ethanol should only be a substrate, and for efficiency reason, should not be a by-product of glucose consumption. Therefore, both for baker's yeast production and for

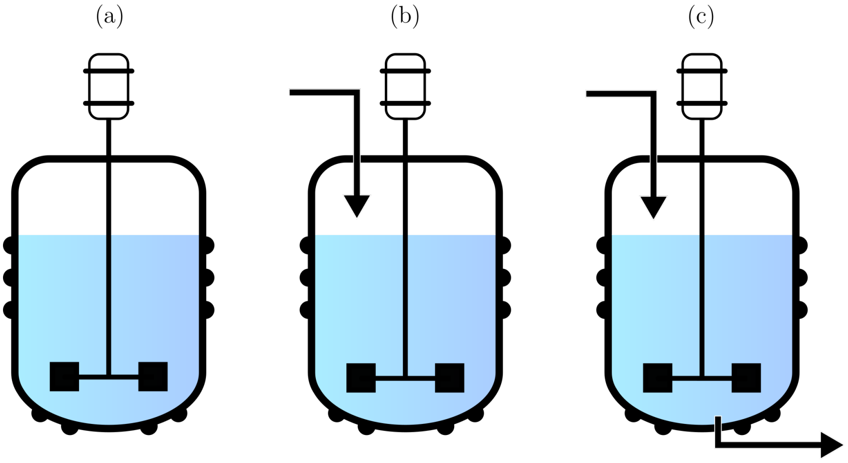


Figure V.1: Illustration of bioreactor culture modes. (a) Batch. (b) Fedbatch. (c) Chemostat.

the industrial production, it will be of interest to be able to predict the appearance of the Crabtree effect in large-scale fedbatch cultures of *S. cerevisiae*.

### Difficulties associated with the simulation of fedbatch processes

There exists three classical culture modes for bioreactors, illustrated in Fig. V.1:

- Chemostat: continuous cultures with constant feeding and withdrawal of culture medium;
- Batch: discontinuous cultures where all nutrients are initially in the culture medium;
- Fedbatch: semi-continuous cultures where the system is closed but continuously fed.

In terms of consequences of the culture mode on numerical simulation, one can easily understand that

- In chemostat cultures, initial conditions have no effect on steady-state conditions. This steady-state is controlled by the feeding flow rate and its composition.
- In batch cultures, only initial conditions will influence the course of the culture.
- In fedbatch cultures, both initial conditions and operating conditions will impact the course of the culture.

For these reasons, simulating a whole fedbatch culture is particularly difficult. Any error on initial concentration, any uncertainty on the initial state of cells over the biological population, or errors on feeding flow rate or composition, will accumulate over time and can yield unrealistic results, mainly after considering strong interactions between gradients and the heterogeneous metabolic behaviour.

Moreover, due to the continuous feeding, the liquid volume keeps increasing over time. Describing this effect would require a Compartment Model with dynamic compartment volumes and most certainly dynamic number of compartments. While being easily feasible, this would still require to perform CFD simulations with multiple increasing liquid volumes, and would need some supplementary developments that were not performed over the course of this thesis.

**Proposed approach: simulating a pseudo-stationnary state in the fedbatch process**

The goal of simulations will not be to predict the evolution of concentrations over time for the whole fedbatch process. We will assess the effect of operating conditions –*i.e.* stirring speed, gas-flow rate and injection mode of glucose– on the occurrence of ethanol formation due to the Crabtree effect.

Thank to the glucose-limited culture mode, the biomass concentration is expected not to increase significantly over time. Moreover, high glucose and ethanol concentrations will be expected near their respective feeding points, but the overall system will be poor in terms of these substrates. Therefore, large-scale gradients should remain relatively constant over time. This will allow overcoming previously listed problems encountered with the simulation of a fedbatch process.

The chosen approach will be to describe glucose and ethanol feeding as volumetric mass source terms $F_G$ and $F_E$ ($\mathrm{kg}_{\{\mathrm{G,E}\}}/\mathrm{m}^3_\mathrm{L}$) in compartments corresponding to their respective feeding points, to set a biomass concentration $X_L$ corresponding to these feeding rates of substrate, and to disable the increase of biomass concentration over time.

Using mass source terms simply removes the need for describing the volume increase. Moreover, disabling accumulation of biomass will let the simulation converge toward concentration fields close to those that would be experimentally encountered when the fedbatch process reaches the biomass concentration $X_L$. This approach can only be used thanks to the fact that concentration fields are expected to evolve over-time in a slower manner than local distributions of biological properties.

The same set of values $X_L$, $F_G$ et $F_E$ will be used with all operating conditions detailed hereafter. Note that actual values cannot be disclosed for confidentiality reasons.

### V.3.2 Assessed variables

The simulation tool will be used to assess the effect of following operating conditions on the occurrence of the Crabtree effect:

- Stirring speed: 35 or 49 rotations per minute;
- Gas flow rate: 2700 or 5000$\mathrm{Nm}^3/\mathrm{h}$;
- Glucose feeding strategy: unique injection point (surface feeding) or multiple injection points.

These simulated conditions are summarized in Table V.1.

Stirring speed and gas flow rates are variables that can be easily changed, in a certain range of values, on an existing industrial installation. Therefore, quantifying their impact on the Crabtree effect means being able to improve the efficiency of an existing process. This is done by identifying the good trade-off between the process performance loss due to this undesired metabolic behaviour, and the energetic expense associated to mixing and gas injection. Investigating the effect of the glucose injection strategy should generally be reserved for process design purposes but we expect to show how critical the glucose injection is.

Table V.1: Operating conditions (OC) to be simulated. *Values are normalized by Operating Conditions 1 for confidentiality reasons.*

| | Stirring speed (rpm) | Gas flow rate ($Nm^3/h$) | Glucose feeding points |
|---|---|---|---|
| Operating conditions 1 | 1 | 1 | 1 |
| Operating conditions 2 | 1.4 | 1.85 | 1 |
| Operating conditions 3 | 1 | 1 | 3 |

### V.3.3 Numerical tool

All following simulations were performed using a short (about 600 lines) Matlab code. This code is organised as following:

1. Load pre-defined circulation maps.
2. Precompute constants defined globally (*e.g.* associated to metabolic mode) or locally (*e.g.* local $k_L a$ values).
3. Define the matrix of initial conditions.
4. Integrate over time using a first-order explicit Euler scheme.

A convergence criteria was introduced to automatically detect the steadiness of the solution and stop the simulation. This criteria is defined by

$$\left\| \frac{\partial \boldsymbol{V}_i}{\partial t} \oslash \boldsymbol{V}_i \right\| < 10^{-4}\mathrm{h}^{-1}, \quad \forall \boldsymbol{V} \in \{\boldsymbol{C}_L, \boldsymbol{C}_G, \boldsymbol{M}\}, \forall i \in \{1, \ldots, N_C\} \tag{V.28}$$

which implies that the simulation stops as soon as the relative variation of all resolved variable is lower than 0.01% over the course of an hour.

## V.4 Simulation results and analysis

### V.4.1 Concentration fields presentation

Due to the high number of compartments, only some concentration fields are shown hereafter. Full concentration fields are given as six sub-figures, each representing a slice of compartments at constant angle. An example is given in Fig. V.2 which illustrate glucose concentration fields under Operating Conditions 1. Note that in these conditions, glucose is injected on top of the reactor, in the compartments of coordinates ($r_5$, $a_2$, $z_{10}$). Due to the counter-clockwise rotation of impellers, the liquid phase mainly moves in the direction of decreasing angle values. This explains why concentrations of glucose decrease along with the angle indexes ($\langle G \rangle_{a_2} > \langle G \rangle_{a_1} > \langle G \rangle_{a_6} > \langle G \rangle_{a_5} > \langle G \rangle_{a_4} > \langle G \rangle_{a_3}$).

### V.4.2 Quantifying metabolic dysfunction

As stated previously, the biological system is cultivated under a glucose-limited fedbatch mode. This limitation is indeed visible in Fig. V.2 where glucose is only available in

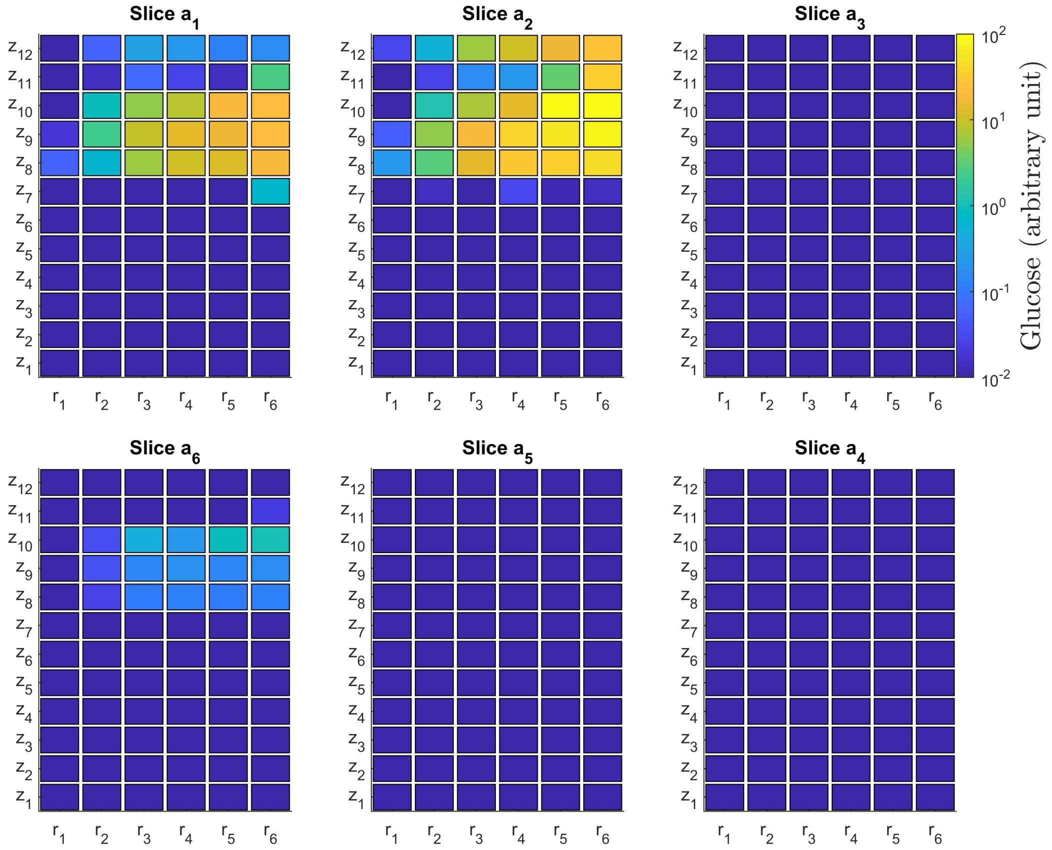


Figure V.2: Glucose concentration fields in steady fedbatch culture under Operating Conditions 1.

the downstream of the feeding point and is depleted in most of the reactor volume. In these conditions, the population mean growth-rate is spatially homogeneous: $\langle \mu_G^{(b)} \rangle = 0.0118\ \text{h}^{-1}$ and $\max\left(\mu_G^{(b)}\right) - \min\left(\mu_G^{(b)}\right) < 4.10^{-6}\ \text{h}^{-1}$ which is consistent with predictions from Chapter III where we decided not to track the diversity of growth-rates as this would not be significant.

The low growth-rate, coupled with the periodic exposure to high glucose concentrations is expected to induce the short-term Crabtree effect through the increase of the glycolytic stress $\sigma$. We recall that this stress is indeed induced by a disequilibrium between the uptake rate (high in glucose-rich regions) and the glucose utilization rate through anabolism (spatially homogeneous). The mean value of this stress is shown in Fig. V.3 and ranges between $\min(\sigma) = 0.09$ and $\max(\sigma) = 0.66$. This corresponds to a loss of 12% to 50% of efficiency for oxidative metabolic modes according to Eq. (II.18).

The metabolic dysfunctional can be illustrated by concentration fields of ethanol and acetate, shown in Figs. V.4 and V.5. The presence of ethanol can be partially attributed to its direct injection as a co-substrate in the compartment of coordinates ($r_5$, $a_4$, $z_9$). However, the fact that highest concentrations span the whole top area of the reactor implies that ethanol is also produced as a by-product of glucose fermentation. This fermentation is due to the short-term Crabtree effect, which limits the use of oxidative metabolisms, and is reinforced by an insufficient gas-liquid mass transfer rate of oxygen which causes oxygen depletion in the upper part of the reactor, as shown by low residual oxygen concentrations in Fig. V.6.

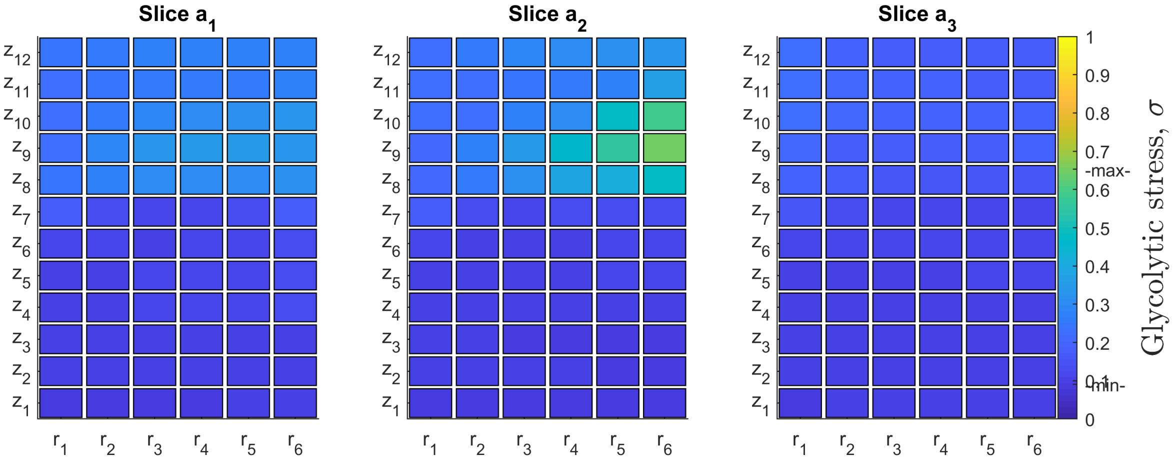


Figure V.3: Field of mean glycolytic stress in steady fedbatch culture under Operating Conditions 1.

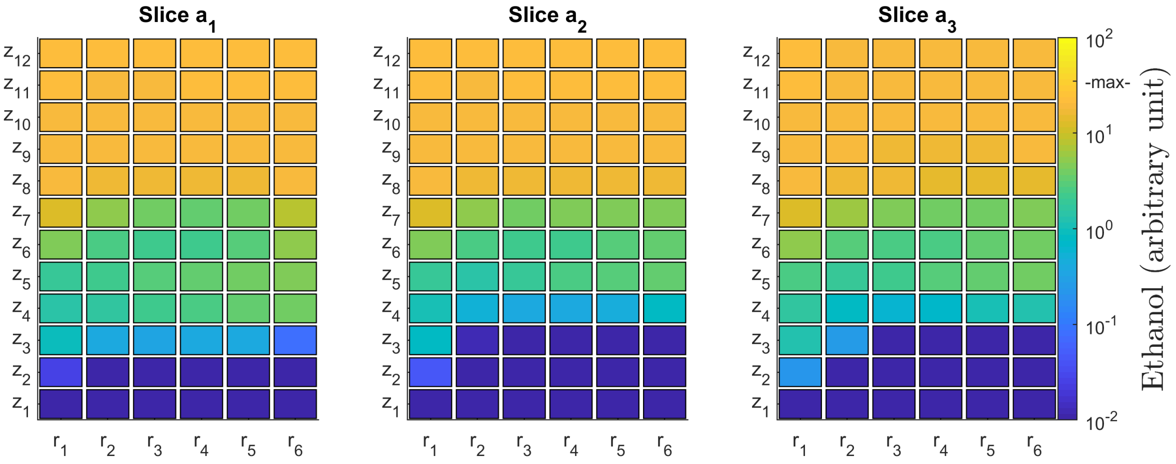


Figure V.4: Field of ethanol concentration in steady fedbatch culture under Operating Conditions 1.

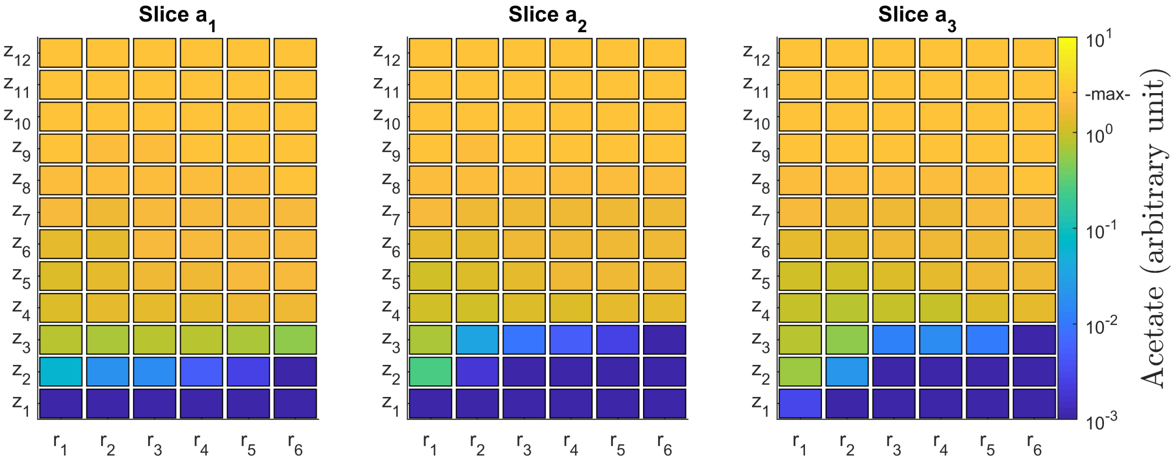


Figure V.5: Field of acetate concentration in steady fedbatch culture under Operating Conditions 1.

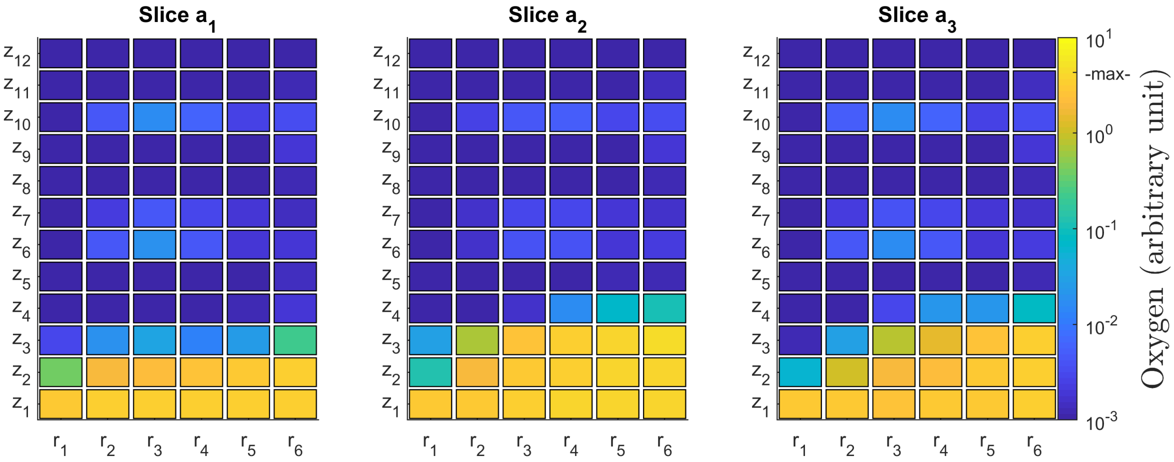


Figure V.6: Field of oxygen concentration in steady fedbatch culture under Operating Conditions 1.

In the bottom part of the reactor, oxygen is more available since glucose is depleted whilst oxygen mass-transfer is high. Therefore ethanol is oxidized into acetate which, in turn, is consumed through acetate-based growth of cells. Overall, all carbonaceous substrates are fully used for cell production, as their concentrations are negligible at the bottom of the reactor. However, the production of ethanol and acetate and their subsequent re-consumption induces yield losses and should be avoided. More than concentration fields, it appears that the most interesting information that can be retrieved from these simulations are the local production and consumption rates [Pigou and Morchain, 2015].

### V.4.3 Reducing metabolic dysfunction: energy or design?

The metabolic dysfunction of *S. cerevisiae*, mainly associated to ethanol production, can be attributed to both (i) the Crabtree effect and (ii) a limiting gas-liquid mass transfer of oxygen. Operating Conditions 2 and 3 represent two proposals to improve the functioning of the bioreactor.

Operating Conditions 2 correspond to the use of supplementary energy to increase the stirring speed and the injected gas flow rate. This should improve both mixing, thus reducing the magnitude of the short-term Crabtree effect, and gas-liquid mass-transfer by increasing local gas fractions. Note that these conditions required specific CFD simulations as detailed in Chapter IV to account for the different hydrodynamic. On the other-hand, Operating Conditions 3 propose a change in the bioreactor design, by injecting the glucose through three feeding pipes. In doing so, glucose should be more evenly distributed which should decrease the overshoot of substrate perceived by cells when travelling near feed points, and thus the short-term Crabtree effect.

Figures V.7, V.8, V.9, V.10 and V.11 show the comparison of glucose, ethanol, aceate and oxygen concentration fields, and the field of population mean stress, between the three operating conditions. Similarly, some noteworthy macroscopic metrics are summed-up in Table V.2. $m_E$ and $m_A$ are total masses of ethanol and acetate in the bioreactor.

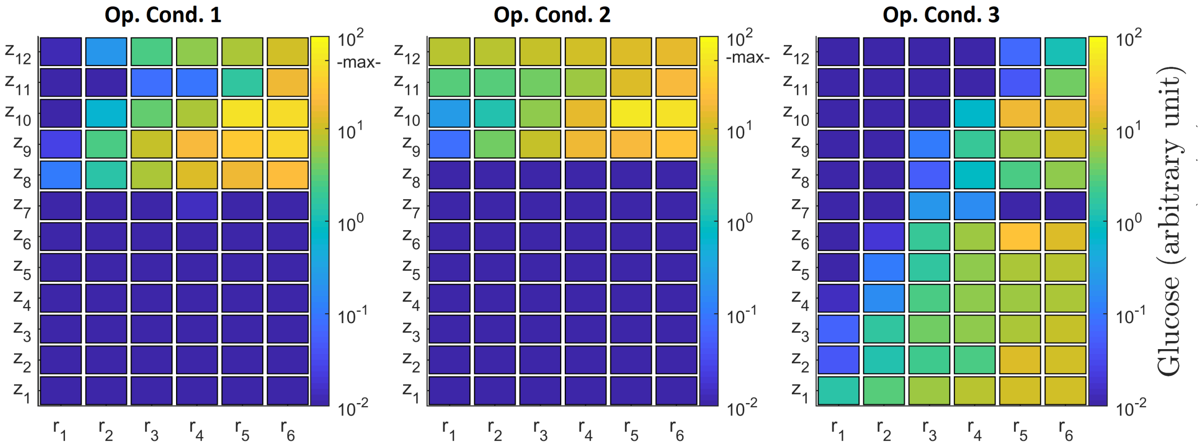


Figure V.7: Comparison of glucose concentration fields between simulated operating conditions on compartment slice $a_2$.

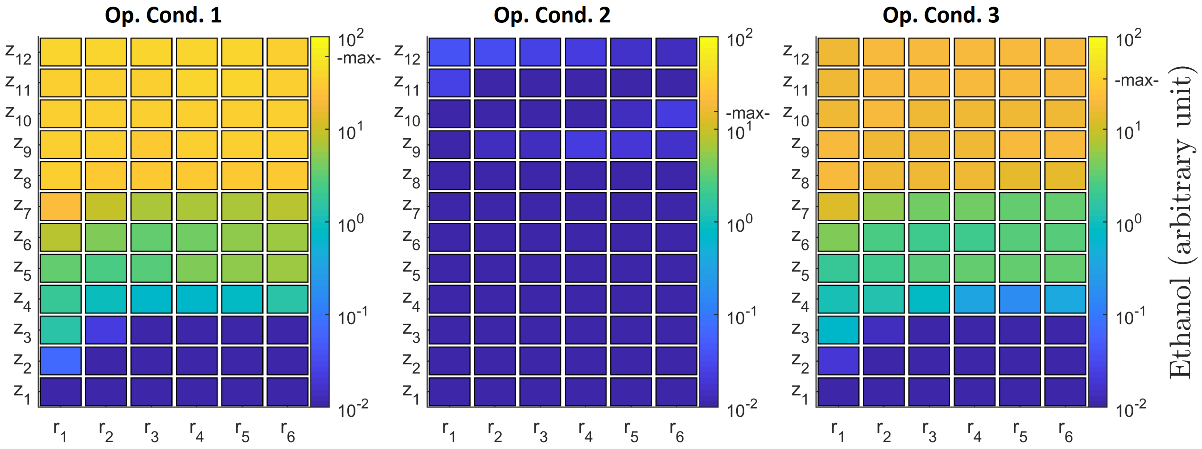


Figure V.8: Comparison of ethanol concentration fields between simulated operating conditions on compartment slice $a_2$.

$m_\sigma$ is expressed in terms of mass of biomass but no physical meaning should be looked for this variable. It is defined as

$$m_\sigma = \iiint_V \int_{\Omega\sigma} \sigma n(\sigma) \mathrm{d}\sigma \mathrm{d}V, \text{ with } n(\sigma) = \int_{\Omega\mu_G} \int_{\Omega\mu_E} \int_{\Omega\mu_A} \int_{\Omega\Phi_O} n(\widehat{p}) \mathrm{d}\Phi_O \mathrm{d}\mu_A \mathrm{d}\mu_E \mathrm{d}\mu_G \tag{V.29}$$

which is the volumic integral of $m_{0,0,0,0,1}$. $\sigma$ was defined as an adimensional quantity measuring the stress level of cell between 0 and 1. By its definition, $m_\sigma$ is expressed as a mass of biomass, but should be interpreted as a "quantity of stress" in the reactor.

By simply looking at ethanol and acetate concentration fields, Operating Conditions 2 which correspond to higher stirring and gas-flow rates are the more favourable in terms of performances and reduction of metabolic dysfunctions. The explanation for this mainly lies in the elevated oxygen concentrations. These are the consequence of an enhanced gas-liquid mass transfer of oxygen thanks to (i) higher local gas fractions and inter-facial area, and (ii) increased turbulence which favourably affects the mass transfer coefficient $k_L$.

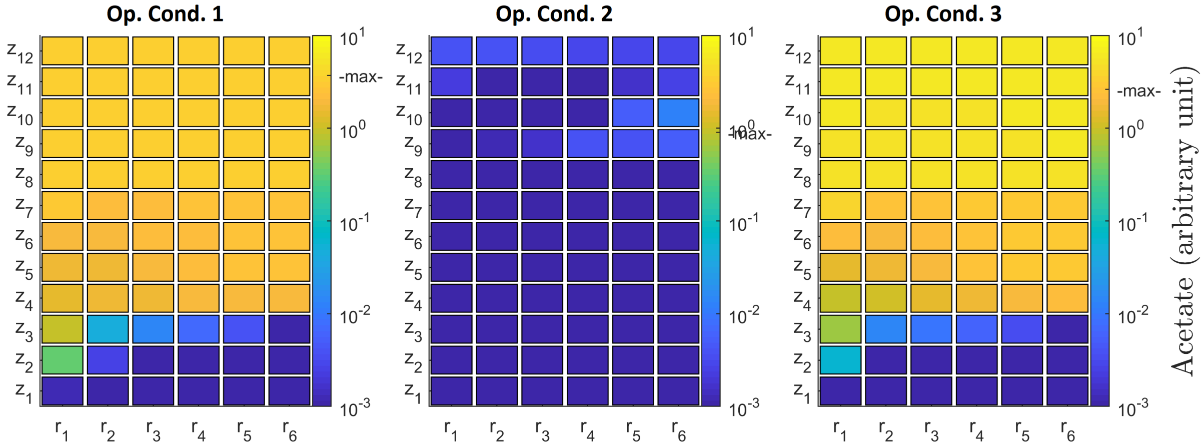


Figure V.9: Comparison of acetate concentration fields between simulated operating conditions on compartment slice $a_2$.

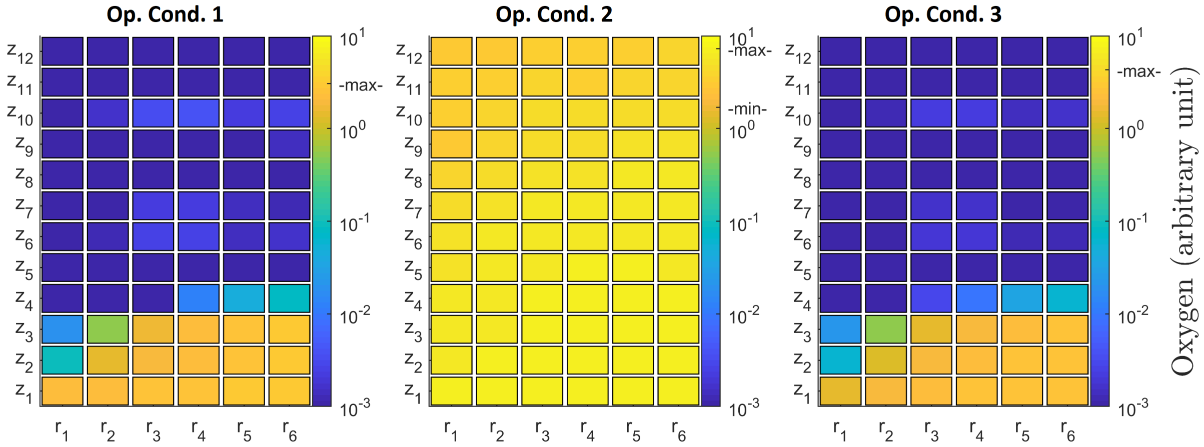


Figure V.10: Comparison of oxygen concentration fields between simulated operating conditions on compartment slice $a_2$.

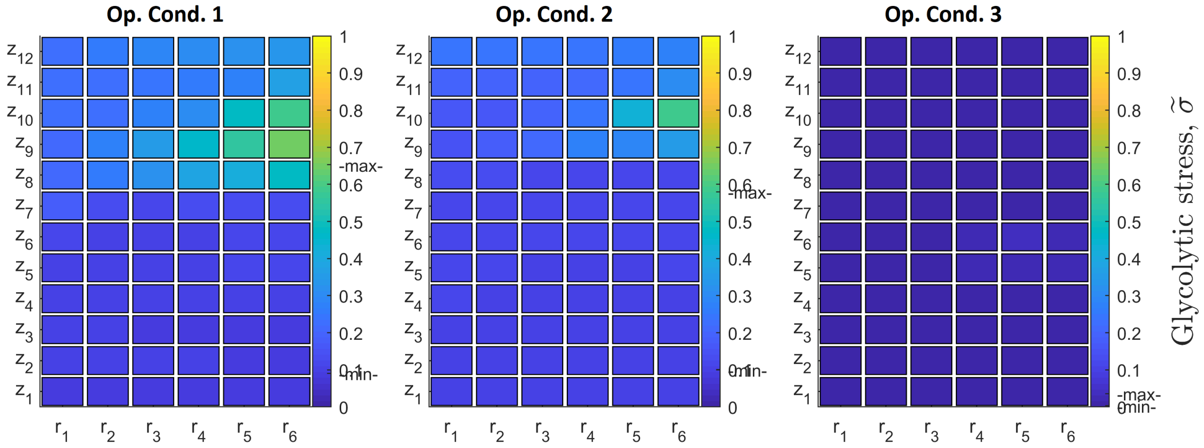


Figure V.11: Comparison of fields of population mean glycolytic stress between simulated operating conditions on compartment slice $a_2$.

Table V.2: Macroscopic metrics comparing performances of the three simulated operating conditions. *Values normalized by Operating Conditions 1 results (except for* $\min(\sigma)$ *and* $\max(\sigma)$*). Units are those of raw (not normalized) values.*

| | Op. Cond. 1 | Op. Cond. 2 | Op. Cond. 3 |
|---|---|---|---|
| $\max(G_L)$ ($\mathrm{g_G/L}$) | 1.00 | 1.13 | 0.47 |
| $m_E$ ($\mathrm{kg_E}$) | 1.00 | 0.02 | 0.52 |
| $m_A$ ($\mathrm{kg_A}$) | 1.00 | 0.03 | 1.44 |
| $\langle O_L \rangle$ ($\mathrm{mg_O/L}$) | 1.00 | 16.7 | 1.03 |
| $\min(\sigma)$ | 0.09 | 0.10 | 0.00 |
| $\max(\sigma)$ | 0.65 | 0.57 | 0.04 |
| $m_\sigma$ ($\mathrm{kg_X}$) | 1.00 | 0.76 | 0.04 |

Under Operating Conditions 2, the stress level is slightly lower than in the reference case (Op. Cond. 1). Indeed the higher stirring rate induces a shorter residence time in the feeding zones where the maximum glucose concentration is also smaller. Both effects decrease cell exposure to stressing conditions. However, this decrease in stress levels is not especially significant compared to the reference case. Coupling this observation with the low amount of ethanol observed under Op. Cond. 2 allows inferring that the ethanol produced in reference conditions is mainly attributed to the limiting oxygenation of the liquid phase. This conclusion is reinforced by results of simulations under Op. Cond. 3.

When changing the reference case by injecting glucose through three different feeding points (located in compartments of heights $z_2$, $z_6$ and $z_{10}$, radius $r_5$ and angle $a_2$), it becomes visible in Fig. V.11 that the stress level has fallen down to zero everywhere in the reactor. This can be explained by the fact that a significant part of the bioreactor volume now contains glucose (Fig. V.7) in noticeable amounts, and that over-concentrations at feeding points are less stressing (see $\max(G)$ in Tab. V.2). Therefore, maximum stress levels under Op. Cond. 3 are lower than minimum stress levels in the two other conditions. Resolving the stress issue and the occurrence of the Crabtree effect partially resolves the ethanol production issue by decreasing its quantity by 48% (see $m_E$ in Tab. V.2).

A surprising result from these simulations is the slight increase of acetate quantities due to the multiple point injection of glucose. This can be explained by subtleties of the metabolic models from Chapter II: the presence of a significant amount of glucose in a larger fraction of the bioreactor volume, under Op. Cond. 3, strongly inhibits growth on ethanol. However, the currently proposed metabolic model assumes a constant ethanol uptake rate unless the environment is limiting. By inhibiting growth on ethanol while not varying the uptake rate of this substrate, the metabolic model predicts that this substrate will be diverted toward ethanol oxidation into acetate (metabolic mode $\mathcal{E}_2$). Whether this behaviour is consistent with experimental data is yet to be checked. If it is not, this shows that the metabolic model could still be improved, in particular by considering varying biological limits on uptake rates as currently considered by Quedeville et al. [2018].

Overall, these simulations show the interest of a modelling framework for bioreactors which allows for fast simulations. Here, we could identify the cause for ethanol production in an industrial fermenter, and therefore a way to decrease these negative effects, *i.e.* the increase of oxygen mass-transfer rate. We could also quantify the effect of a better reactor design by showing a reduction of ethanol concentrations by 50% simply by distributing the glucose feeding over three points instead of a single one. Operating conditions at higher stirring and gas-flow rates have shown to significantly decrease ethanol production, but the study could be pushed further by identifying the optimal trade-off between this undesired ethanol production, and the energetic cost of the process. In particular, energetic expense is known to increase as the cube of the stirring speed, therefore one could use the proposed simulation framework to identify the lowest stirring speed which still prevents ethanol production. It is easy to see how this tool can find its place in technical studies offices as an engineering or decision-assisting tool.

### V.4.4 Numerical cost

Presented simulations ran in almost real-time using a simple Matlab code. Matlab is a programming language known for its ease-of-use and the accessibility of advanced routines for scientific computing, but this is an interpreted language, which consequently does not make an efficient use of computing resources. This is in particular true when one uses multiple routines, each defining its own variables, as this has been the case with the metabolic model and the EQMOM moment-inversion procedure in previously described simulations. For this reason, it might be beneficial to estimate the gain that would be provided by switching simulations toward a compiled language.

First, we estimate the numerical cost (in floating-point operations, flop) associated to computations performed at each time-step and in each compartment. This corresponds to following operations:

- 1 call to a 2-nodes EQMOM moment-inversion procedure $\approx 800$ flop [Pigou et al., 2018a],
- 5 calls to the metabolic model for *S. cerevisiae* $\approx 5 \times 140 = 700$ flop,
- time-advancement of variables, computation of stopping criteria, computation of gas-liquid mass-transfer rate, $\ldots \approx 50$ flop

. To simulate one hour of culture, using timesteps $\delta t = 1.10^{-5}$ h, the number of operation required is then $N_{\text{flop}} \approx 10^5 \times 432 \times (800+700+50) \approx 10^{10}$ flop. This number of operations is not high even for mid-range CPUs present on current desktop computers. The actual computing throughput of a CPU is always hard to predict, as different codes will not have the same vectorizability, especially when iterative algorithms such as the EQMOM moment-inversion procedure are involved. However, any current CPU should be able to process $10^{10}$ flop within a few seconds with a tight memory management. Over the course of this PhD thesis, an engineer wrote a C++ code that reproduced simulations from Pigou and Morchain [2015]. Our Matlab code needed one day to perform associated simulations while the C++ code obtained the same results within only ten seconds.

Therefore, the coupling of compartment, population balance and mode-based metabolic models does allow for fast simulations on industrial bioreactors. A “slow” Matlab code was able to perform these simulations in real-time, but translating this interpreted code into a compiled language could allow for much faster than real-time simulations. This opens the door for many engineering applications, from parametric optimization of the process to real-time simulation-based control of an industrial installation.

## V.5 Conclusion

The modelling of industrial bioreactors is a challenging topic as one must account for large-scale hydrodynamics as well as the peculiarities of the biological phase which is made of dynamic living systems and therefore tends to become heterogeneous. To perform predictive simulations while using limited computational resources, we choose to couple a metabolic designed in Chapter II, a population balance model tackled by a numerical method improved in Chapter III and compartment models deduced from CFD simulations in Chapter IV. Numerical methods and procedures associated to these models were designed, improved or selected for their efficiency in order to reach fast simulation times.

One aspect that was missing from previous Chapters developments is the limitation of substrate uptake by cells due to limiting environments. This aspects has been introduced in the modelling by following preliminary developments detailed in Morchain et al. [2016] based on micro-mixing considerations and classical chemical engineering approaches. Closure laws have been suggested to simultaneously account for environmental and biological limits on uptake rates.

By applying these developments to an industrial example, we could show the capabilities of this modelling framework. Using a simple MATLAB [2016] implementation, we could simulate in real-time an industrial bioreactor to quantitatively assess the effect of some operating conditions, namely the stirring speed and the gas-flow rate, and to illustrate the use of this tool for bioreactor design purposes. This shows the feasibility of simulating large-scale bioreactors while accounting for both physical and biological heterogeneities, on standard computing resources, and opens the door to the development of engineering software.

# VI

# OUTLOOKS AND CONCLUSIONS

## VI.1 Introduction

The simulation of industrial biological systems form a vast topic encompassing a variety of strikingly different aspects, from the comprehension of the living organisms chemistry and of advanced numerical methods, to high-performance computing and chemical engineering. Many of these aspects were, at least partially, considered over the course of this PhD thesis as demonstrated in previous Chapters.

To cover both metabolic modelling, advanced numerical methods for population balance, the mastering of Computational Fluid Dynamics (CFD) software and the development of numerous processing and simulation tools, over the course of three years, choice had to be made about what could, and could not, be reasonably done. Therefore, for each aspect developed in previous Chapters, there is still room left for numerous improvements and supplementary studies and developments. Though not being exhaustive, section VI.2 details some of these outlooks.

Research is an “activity undertaken in order to increase the stock on knowledge [...] and the use of this knowledge to devise new applications” [OECD, 2008]. On this regard, section VI.3 serves as a general conclusion to this report with an overview of major achievements of this work.

## VI.2 Outlooks

The major goal of this thesis was to develop a model framework or structure that allows fast, yet accurate, simulations of bioreactors. On that regard, we mainly aimed at performing proof-of-concept simulations coupling developed models structures, numerical methods and tools, rather than focusing on obtaining an over-parametrized model and fit it to few experimental data. Now that these proof-of-concept simulations have been obtained as presented in Chapter V, it will be interesting to improve the accuracy of proposed models to enhance their predictive capabilities, or to slightly modify the modelling framework to expand its capabilities. These aspects are discussed hereafter, by keeping a structure close to this thesis: prospect considerations for Chapter II to V are discussed one-by-one in dedicated sections.

### VI.2.1 Metabolic modelling

The metabolic model structure developed in Chapter II is satisfactory on multiple regards. In particular, it allows describing dynamic cultures, including switches between multiple elementary modes, while keeping a strikingly low numerical cost. This is mainly done by outsourcing dynamic aspects to the biological heterogeneity part of the overall modelling, which track the evolution and diversity of biological properties. If the metabolic model structure was to be improved, the most crucial improvement might be to reduce the constraint on the sorting of metabolic modes. A comparison with cybernetic models with a view of tracking the origin of metabolic mode shifts would be profitable. If done, the proposed solution should not significantly increase the model numerical cost, for instance by not relying on iterative optimization methods.

Apart from the metabolic model structure definition, two main aspects should be consider to build upon developments presented in Chapter II. The first one is in the direct continuation of the modelling of *Saccharomyces cerevisiae*. Chapter II describes a model for this yeasts which has been fitted against multiple experimental data sets. While being satisfactory, this fitting was performed by-hand through a manual iterative procedure. The goal was to demonstrate capabilities of the model structure under dynamic perturbations but we did not search for the highest possible accuracy. On top of accuracy, an other aspect is that of parameter identifiability. While designing the model, we attempted to keep the number of parameters as low as possible to limit the risk of over-parametrization and prevent in-identifiability of parameters. We did manage to limit the number of parameters to 35 where other models required more parameters for a final model with less capabilities to predict the course of dynamic large-scale cultures (*e.g.* 87 parameters for the model from Rizzi et al. [1997]). Nevertheless, tools and methods exist to automatically improve a model fitting, and to assess the identifiability of parameters [Steuer, 2007; Balsa-Canto et al., 2010; Kravaris et al., 2013; Sánchez et al., 2014]. Applying these approaches, using existing tools, to the currently proposed *S. cerevisiae* model would be easily feasible and could significantly improve its predictive capabilities. Once these developments are integrated to reinforce the model definition method, one could easily enrich the wild strain *S. cerevisiae* model which new elementary modes corresponding to specific synthetic pathways corresponding to industrial interests.

A large part of this thesis work was dedicated to automating complex tasks. When improving the Extended Quadrature Method of Moments in Chapter III, we automated the benchmark of these method implementations. As detailed in Chapter IV, we used CFD software which, though requiring a certain level of expertise, were designed to be as user-friendly as possible. To make use of CFD results, we developed a C++ tool which automatically integrates scalar and vector fields into a compartment model, and a Matlab tool to clean these models from numerical error. Even though all these automated tools would need to be interconnected under a single platform –which has been started under the ADENON project briefly detailed hereafter– one aspect has not been automated at all: the design of elementary-mode based metabolic models. For now, designing metabolic models for *E. coli* and *S. cerevisiae* required analysing experimental data from lab-scale cultures [van Dijken and Scheffers, 1986; Flikweert et al., 1999; Sweere et al., 1988], reading analysis of their metabolism dysfunctional [Pronk et al., 1996; Xu et al., 1999] and browsing databases of metabolic reactions [Heavner et al., 2013; Orth et al., 2011] to define a minimal set of required elementary modes. This work is time-consuming and widely depends on the availability of detailed resources about a specific micro-organism.

To improve on this aspect, and possibly reach an almost fully automated modelling and simulation platform for bioreactors, tools will be required to simplify the design of metabolic models. One approach would be to rely on detailed metabolic network to determine simplified networks [Gerdtzen et al., 2004] from which elementary modes can be extracted [Trinh et al., 2008].

## VI.2.2 Biological heterogeneity

The idea of tracking dynamics properties of micro-organisms is not new [Sweere et al., 1988] but coupling this description to the transport of a biological phase in an non-homogeneous system is a more demanding problem. Some attempts used Lagrangian approaches [Lapin et al., 2006] but did not provide a reverse coupling between these dynamic cell properties and local consumptions rates which influence gradients. Other attempts were limited to a rough description of spatial heterogeneity by considering simple two-compartments set-up [Heins et al., 2015]. Our current Population Balance-based modelling results from work by Morchain [2000]; Morchain et al. [2013] who applied Population Balance Model (PBM) approaches to biological systems, thus enabling a full coupling between a fine treatment of the biological phase and the description of time-evolving gradients [Pigou and Morchain, 2015].

In this context, we did not diverge from these ideas and, as shown in Chapter III, we kept a Population Balance treatment of the biological phase. We shifted from class-approach toward less expensive moment methods but we restricted ourselves to monovariate numerical methods. This is partially due to results from Pigou and Morchain [2015] where a single biological property was deemed sufficient on some regards, and to discussion with Lebaz et al. [2016] who, in a different context, also dealt with dynamic systems using monovariate Population Balance Equations (PBEs). This has shown some limits as the metabolic model designed for *S. cerevisiae* happened to require at least five dynamic biological properties. Consequently, a significant improvement of the modelling framework for industrial fermenters would be the ability to handle multi-variate description of biological heterogeneity. Two approaches could be suggested to do so while keeping the numerical cost reasonable.

### Multi-variate population balance models

Numerical methods to treat multivariate population balance models do exist. For obvious reasons of numerical costs, discretization methods (finite-volume, finite-element) will not be considered. However, multi-variate moment methods are available such as the multidimensional moment-constrained maximum-entropy method [Abramov, 2007], the Conditional Quadrature Method of Moments [Yuan and Fox, 2011] or a multivariate version of the Gauss-EQMOM approach [Pollack et al., 2016; Chalons et al., 2017]. We could attempt to implement and apply them to the modelling of biological systems, but for multiples reasons, this choice might not be the most advisable.

The choice of moment-methods in place of class method for mono-variate PBMs was motivated in Chapter III by their extensive comparison. In particular, we could use the class-method as a reference to ensure the accuracy for mono-variate moment methods. However, such a reference is not available for multi-variate PBMs unless one implements finite-volume scheme to obtain a reference resolution of the PBE under controlled accuracy.

If a metabolic model happens to require multiple dynamic variables evolving with short time-scales and whose evolutions are tightly coupled, accuracy in the resolution of coupled moments, and on the computation of quadrature rules, will be critical. However, it is known that increasing the number of moments does not necessarily improves significantly the accuracy of the method [Tagliani, 2001; Gzyl and Tagliani, 2010]. This could already be observed in comparisons performed in Chapter III (see also Pigou et al. [2017], Appendix B) between numerical methods, where results from MaxEnt and EQ-MOM methods were only slightly sensitive to the number of resolved moments. The capability of multi-variate moment methods to handle any set of biological properties whose evolution may or may not be tightly coupled can therefore be questioned. At least, this aspect can hardly be assessed in a rational manner.

A final aspect is that of stability. Numerical methods tested in Chapter III are applicable only if resolved moments remain within the realisable moment space, *i.e.* a complex-geometry convex space in which moment sets correspond to actual –realisable– probability distributions. In Chapter V, we already faced instability issues when resolving more than 5 moments over the distribution of glycolytic stress $\sigma$. These were due to a time-integration scheme which did not ensure moment realisability. Some possible development could then be to derive such a realisability-preserving scheme to perform time-integration of moment sets in a Compartment Model framework, or to identify one if this already exists. Note that while the geometry and mathematical properties of the realisable mono-variate moment space are well understood [Dette and Studden, 1997; Gautschi, 2004], as much cannot be said for the multi-variate realisable moment space. Therefore, deriving a realisability-preserving scheme for multi-variate problems is expected to be a more challenging task.

### Pseudo-Lagrangian/Monte-Carlo approaches

Though suited for mono-variate problems, the application of moment methods is more challenging for multi-variate models. Class methods have also been considered for multi-variate situations, but they basically consist in a $N+3$-dimensions finite volume scheme to solve the Population Balance Equation with $N$ the number of biological properties and 3 the number of geometrical dimensions. This would turn to be expensive, both in terms of meshing due to the extra dimensions, and in terms of global numerical cost as this has been discussed in previous chapters.

A solution to tackle multi-variate problems for the simulation of large-scale bioreactors might lies in an approach that was discarded in Chapter I to consider biological heterogeneity: the Lagrangian approach. We oriented developments toward Population Balances and moment-methods following a lineage other works [Morchain and Fonade, 2009; Morchain et al., 2013; Pigou and Morchain, 2015] and due to a synergy between multiple projects with this thesis [Lebaz, 2015; Quedeville et al., 2018]. The Lagrangian framework solves issues associated to class and moment methods: its cost does not increase with the number of biological properties, its accuracy is easily controlled by the number of considered particles and is assessed through convergence studies. Moreover, only relevant parts of the phase-space are explored by particles which removes the drawbacks of class-methods where numerous classes happen to be empty most of the time. Finally, computations over a large-number of particles constitutes an embarrassingly parallel problem which can strongly benefit from current advances in parallel computations.

The reason why we did not consider Lagrangian approaches sooner is because, for now, Euler-Lagrange simulations of bioreactors were limited to using the Lagrange phase as a non-reactive phase which would only record the events occurring along the trajectories of each particles [Lapin et al., 2004, 2006; Haringa et al., 2016]. The main difficulty is to allow a reverse coupling: having access to the biological heterogeneity is useful only if its effect on local uptake rates and on concentration gradients can be accounted for. Doing so within a CFD framework, which is the natural framework for Lagrangian simulations, is a major challenge.

So could we overcome these remaining difficulties? An approach that seems to resolve all issues is to couple the principle of particle-tracking along with Compartment Modelling. Under this idea, the "location" of a particle, or micro-organism, is no longer a vector of spatial coordinates $\vec{x}$, but the index of a compartment. The location of the micro-organism within the compartment is not resolved, hence the pseudo-Lagrangian approach.

The "tracking" of particles in a compartmental framework has been suggested by Delafosse et al. [2015] which compared "trajectories" of particles with stochastic motion in a compartment model against experimental trajectories obtained through optical trajectography. Note that in their work, the compartment model was deduced from CFD simulations. A particle located in the $i$-th compartment would leave this compartment after a time $\delta t$ taken randomly out of the Residential Time Distribution (RTD) of an homogeneous volume:

$$\delta t \sim \lambda e^{-\lambda t}, \text{ with } \lambda = \frac{\sum_{j=1}^{N_C} Q_{ij}}{V_i} \tag{VI.1}$$

Once a particle reaches its local residence time, the probability that it enters the $j$-th compartment is given by

$$p_j = \frac{Q_{ij}}{\sum_{j=1}^{N_C} Q_{ij}} \tag{VI.2}$$

Delafosse et al. [2015] have shown the interest of this method to effectively track the rough locations and trajectories of particles in a stirred-tank reactor under a Compartment Model framework. Note however that they used between 19000 to 35000 compartments in their zonal model which is significantly high compared to the 432 compartments model from Chapters IV and V. We may question the relevancy of previous closures for $\delta t$ and $p_j$ if the number of compartments is significantly lowered.

To improve upon previous developments, we propose to deduce laws for $\delta t$ and $p_j$ using actual Lagrangian tracking in CFD software. It is quite simple to perform particle tracking in ANSYS Fluent [2015] and to export trajectories in an Ensight Gold format datafile. Therefore, an easily feasible development is to improve the tool presented in Chapter IV that converts CFD results into a compartment model. The goal would be to detect all events of a particle moving from one compartment $i$ to another $j$, and to record both the time of the event $t$, the source and destinations compartments $i$ and $j$. From this information, on numerous events, it would be possible to construct probability distributions from which one can randomly pick the time $\delta t$ during which a particle that entered compartment $j$ while coming from compartment $i$ will remain in $j$ before jumping into the compartment $k$. By extracting these easily accessible distributions from CFD results, it should be possible to accurately mimic trajectories of particles, within a compartment model, without requiring as many compartments as Delafosse et al. [2015]. This also opens the door to Monte-Carlo simulations where one does not actually solve the trajectories of particles.

Once pseudo-Lagrangian simulations are shown to be feasible within compartment model framework, it remains to explain how they actually solve previously met problems. In Chapter V, moments of the distribution of biological properties were used to create quadrature rules that allow integrating the biological behaviour over the population (see Eq. (V.27)):

$$\boldsymbol{\Phi}_{BL} = \sum_{i=1}^{Q} w_i \boldsymbol{\varphi}^{(a)}(\boldsymbol{C}_L, \boldsymbol{\varphi}^{(e)}, \widehat{p}_i) \tag{VI.3}$$

Using pseudo-Lagrangian approaches, this quadrature is simply replaced by:

$$\boldsymbol{\Phi}_{BL} = \frac{X_L}{N_p} \sum_{i=1}^{N_p} \boldsymbol{\varphi}^{(a)}(\boldsymbol{C}_L, \boldsymbol{\varphi}^{(e)}, \widehat{p}_i) \tag{VI.4}$$

with $N_p$ the number of particles in the considered compartment, $X_L$ the local biomass concentration ($\mathrm{g_X/L}$) and $\widehat{p}_i$ the vector of biological properties of the $i$-th particle. These properties simply evolve over time as described by the following Ordinary Differential Equation:

$$\frac{\partial \widehat{p}_i}{\partial t} = \widehat{\zeta}(\widehat{p}_i, \boldsymbol{C}_L) \tag{VI.5}$$

Overall, this Lagrangian approach easily replaces the quadrature methods of moment to account for biological heterogeneity. However, they have the advantage of being directly compatible with multi-variate descriptions of the biological phase, and they only span parts of the phase-space that are relevant for the described system, thus tackling drawbacks of both the class and moment approaches. Moreover, as stated previously, the accuracy of the method can easily be assessed through convergence studies by varying the total number of particles.

The main drawback of this approach will be its numerical cost that increases linearly with the number of particles. In particular, the metabolic model (computation of $\boldsymbol{\varphi}(a)$ and $\widehat{\zeta}$) must be called for each particle individually. On that regard, the low-cost of the model structure proposed in Chapter II is beneficial, but does not prevent the fact that, to reach convergence, the number of particle per compartment $N_p$ will be higher than the number of quadrature nodes in moment method approaches. This is the cost for enabling multi-variate biological models. Note however that this can be balanced by the use of parallelisation programming techniques which are easily applicable here as the metabolic model can be computed in parallel between multiple particles.

### VI.2.3 Large-scale hydrodynamics

This thesis was not about improving the current state-of-the-art in terms of hydrodynamics modelling for bioreactors. There is still many aspects to improve upon in terms of modelling of rheology, of gas-liquid interactions, of turbulence, ... Improving this modelling can be done in parallel of its coupling with biological heterogeneity and behaviour as this will only impact the obtained compartment model. Nevertheless, the work presented in Chapter IV can be improved on two aspects. The first one is specific to the application case of the industrial fermenter, and the second one is a long-term improvement of the proposed modelling framework.

For the CFD simulations of the industrial fermenter presented in Chapter IV, the choice of turbulent model and of closures for momentum exchange between gas and liquid phases has been somehow overlooked: we did not try to choose the most recent models on that regard. Moreover, we imposed a mono-dispersed bubble size distribution onto the gas phase, which is clearly a disputable choice considering the high gas flow rate and the large role of coalescence and breakage in large-scale aerated reactors [Laakkonen et al., 2006, 2007]. Therefore, to improve coupled simulations from Chapter V, a first step could be to improve the hydrodynamic modelling that yields the used compartment model in Chapter IV. On that regard, a critical aspect would be to enable models for interfacial area transport which would significantly improve gas-liquid interaction description, as well as the prediction of gas-liquid oxygen mass transfer in coupled simulations [Hibiki and Ishii, 2000].

In coupled simulations of an industrial fermenter shown in Chapter V, most of the numerical cost is associated to the treatment of the biological phase, to account for its heterogeneity and to compute its metabolic behaviour. It would make sense to also dedicate part of the computational power toward an improvement of the hydrodynamic description. In particular, for now, we presented simulations using a steady compartment model, with circulation maps and local phase-fractions that were constant over time. As already briefly suggested in Chapter IV, it should be possible to obtain dynamic compartment models from time-resolved CFD simulations. Doing so could improve the description of interactions between the biological phase and its environment, in particular in stressing zones, by improving the prediction of exposure time of cells to these stressing situations. Over the course of this thesis, a significant time has been spent on obtaining time-resolved CFD simulations using NEPTUNE_CFD [2016], the development of dynamic compartment models could rely upon these available results.

### VI.2.4 Simulation of industrial fermenters

The simulation of industrial fermenters has been the core topic of this thesis. Beyond modelling considerations that have been thoroughly explored previously, the future of bioreactor simulation can be summed-up to two questions: why perform such simulations? and how to perform these simulations? We already started answering the first question, by illustrating the use of the proposed modelling framework for design and engineering purposes, the second question is far from being tackled.

Engineer toolboxes lack an integrated solution to easily perform simulations of large-scale bioreactors and most developments from this thesis could be integrated into such a software tool-suite. Since the prototype work that led to this PhD thesis [Pigou and Morchain, 2015], we have been developing the ADENON software which is a MATLAB [2016] based solution for the dynamic simulation of large-scale bioreactors. This tool provides a user-friendly graphical interface to set-up simulations coupling compartment models, metabolic modelling and population balance models solved through class or moment methods. It was designed with modularity in mind allowing for new models (compartment, metabolism, . . . ) to be easily added to the database of already available models. This tools has been used to perform simulations presented in multiple published articles [Pigou and Morchain, 2015; Morchain et al., 2016; Pigou et al., 2017].

We could build upon this first software by (i) converting it into a compiled language thus improving significantly its performances and (ii) by adding developments from Chapter IV to help integrating compartment models based on CFD results. Moreover, as suggested previously, this tool-suite could integrate a tool formulating elementary-mode based metabolic model using detailed metabolic networks.

Such tools already exist in the Chemical Engineering community to assist the design and the operation of chemical processes (Aspen HYSYS, ProSim, . . . ). It is easy to see how the biochemical engineering community could also benefit from similar software.

## VI.3 General conclusion

The modelling and simulation of industrial bioreactors is a challenging topic. This last sentence has been a leitmotiv scattered throughout this thesis. To explain these challenges, our main postulate is that micro-organisms may face both internal (*i.e.* biological) and external (*i.e.* environmental) limitations and that due to the scale of industrial fermenters, these limitations will be at disequilibrium, inducing the well-known metabolic dysfunctions in large-scale cultures. To address these challenges, we propose to model bioreactors using three distinct yet complementary modelling-blocks. These are illustrated in the graphical abstract of an introductory article reproduced in Fig. VI.1 [Pigou and Morchain, 2015]. We proposed a modelling for *E. coli* cultures based on (i) a simplified metabolic model, (ii) a mono-variate population balance model and (iii) an hydrodynamic compartmental model. By using Population Balance modelling to track the biological state, we could apply the same metabolic model to the simulation of lab-scale and industrial-scale cultures while correctly predicting the metabolic disequilibrium between cells and their environment. This PhD thesis was a direct continuation of this preliminary work and aimed at improving the proposed framework to expend its capabilities while preserving a numerical cost sufficiently low to allow for fast simulations. To do so, we compared and improved existing numerical methods, used engineering CFD tools, developed methodologies to obtain compartment models from CFD results, defined a low-cost architecture for metabolic model, and illustrated the capabilities of the obtained modelling framework on an industrial application case.

Chapters II to V show multiple developments improving the modelling of bioreactor but also what could be a generic methodology to model these systems:

1. Identify metabolic features of the micro-organism and develop a low-cost dynamic metabolic model.

2. Select a suited approach and numerical method to account for the heterogeneity of dynamic biological properties between micro-organisms.

3. Use CFD software to describe the hydrodynamic and deduce a compartment model from resulting fields.

4. Couple previously obtained models to gain insight in the bioreactor functioning.

Of course, these developments are still on-going works and do no yet fully address the modelling of bioreactors. This thesis still does not change the statement from Takors [2012] according to which we are still not able to “accurately predict – a priori — large-scale process performance”. In particular, future modellings will require experimental advances to fully analyse core mechanisms that need to be accounted for. Overall, this thesis was at the junction of multiple disciplines and we attempted to cover all of these,

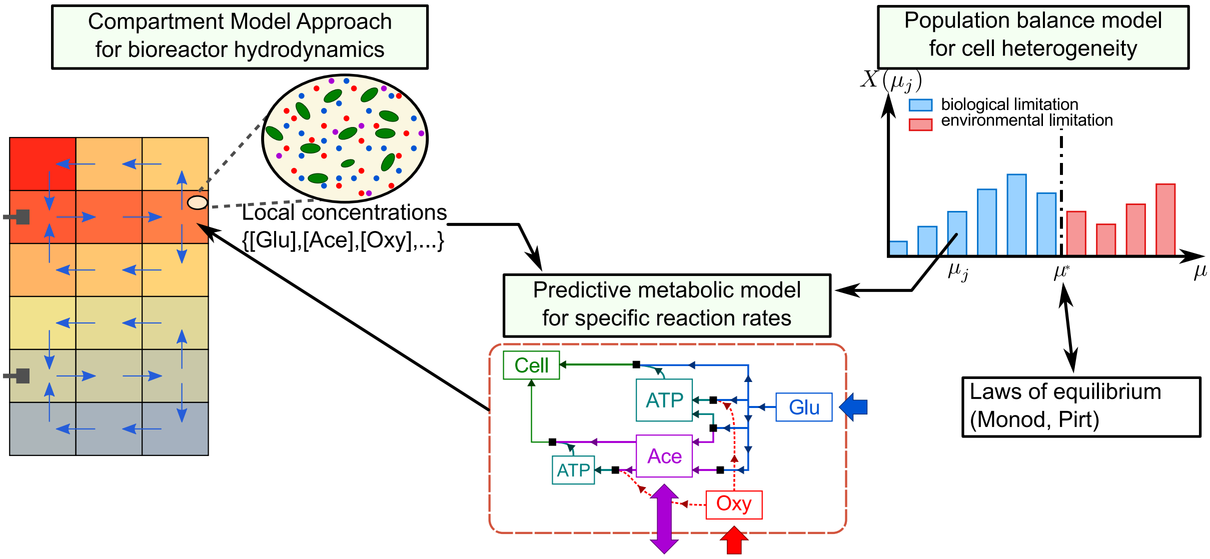


Figure VI.1: Graphical abstract of Pigou and Morchain [2015] illustrating the three complementary modelling blocks forming the proposed modelling framework for bioreactors: metabolic, population balance and hydrodynamic.

not only by using existing tools, but also by improving the connectivity between multiple modelling blocks.

In terms of metabolic modelling, we tried to find the best trade-off between the complexity of the metabolic network, the model numerical cost, and its capability to handle dynamic situations. We applied the proposed metabolic model structure to *E. coli* and *S. cerevisiae*, and were able to challenge the *S. cerevisiae* model against experimental data from numerous different culture modes and yeast strains. On that regard, the developed model structure can be seen as a rationalisation of the one described in Pigou and Morchain [2015] to facilitates its application to new micro-organisms and to improve its modularity.

In terms of biological heterogeneity, we used the Population Balance modelling framework in continuation to previous developments [Morchain et al., 2013; Pigou and Morchain, 2015]. We focused on numerical methods suiting mono-variate problems only to realise latter that this would generally not be sufficient for micro-organisms and that multi-variate population balances, or possibly Lagrangian methods, should be considered as well. Nevertheless, we could challenge multiple moment methods (QMOM, EQMOM, Maximum Entropy) against the class method already used in Pigou and Morchain [2015]. The main conclusion from this comparison being that QMOM is the fastest method but requires numerous moments to preserve a good accuracy while EQMOM offers the best trade-off between accuracy, stability and number of resolved moments. An unexpected achievement has been the significant improvement of the core iterative algorithm of EQMOM inducing a significant performance and stability gain that will benefit all communities using this method.

Finally, in terms of bioreactor hydrodynamics, we mainly used available CFD software to access the gas-liquid hydrodynamics of an industrial bioreactors. Our main contribution has been on coupling the CFD results with the biological modelling. For that purpose, we developed a procedure to convert CFD results into compartment models, in a way that exploits the information contained in the standard Ensight Gold file format, so that a single software implementation of that procedure can be used for the compartmentalization of results from ANSYS Fluent [2015], NEPTUNE_CFD [2016], OpenFOAM, and other CFD software.

The final aspect in this work has been the development of a bioreactor simulation tool-suite, ADENON, whose purpose is to integrate and couple metabolic models and numerical methods to solve population balances along with compartmental approaches to enable engineering studies using the proposed modelling framework. The relevancy of this tool has been demonstrated on scale-up issues and on actual industrial systems for design and optimisation purposes. For now, a MATLAB version of ADENON allows for real-time simulations but future versions based on compiled languages shall enable significantly faster simulations along with new capabilities such as parametric studies or anticipated process control.

# BIBLIOGRAPHY

Abramov, R.V., 2007. An improved algorithm for the multidimensional moment-constrained maximum entropy problem. Journal of Computational Physics 226, 621–644. doi:`10.1016/j.jcp.2007.04.026`.

Abulesz, E.M., Lyberatos, G., 1989. Periodic operation of a continuous culture of baker's yeast. Biotechnology and Bioengineering 34, 741–749. doi:`10.1002/bit.260340603`.

Adamberg, K., Lahtvee, P.J., Valgepea, K., Abner, K., Vilu, R., 2009. Quasi steady state growth of *Lactococcus lactis* in glucose-limited acceleration stat (a-stat) cultures. Antonie van Leeuwenhoek 95, 219–226. doi:`10.1007/s10482-009-9305-z`.

Alopaeus, V., Laakkonen, M., Aittamaa, J., 2006. Solution of population balances with breakage and agglomeration by high-order moment-conserving method of classes. Chemical Engineering Science 61, 6732–6752. doi:`10.1016/j.ces.2006.07.010`.

Alopaeus, V., Laakkonen, M., Aittamaa, J., 2007. Solution of population balances with growth and nucleation by high order moment-conserving method of classes. Chemical Engineering Science 62, 2277–2289. doi:`10.1016/j.ces.2006.12.070`.

Alopaeus, V., Laakkonen, M., Aittamaa, J., 2008. Solution of population balances by high order moment-conserving method of classes: reconstruction of a non-negative density distribution. Chemical Engineering Science 63, 2741–2751. doi:`10.1016/j.ces.2008.02.027`.

Alves, S., Maia, C., Vasconcelos, J., Serralheiro, A., 2002. Bubble size in aerated stirred tanks. Chemical Engineering Journal 89, 109–117. doi:`10.1016/s1385-8947(02)00008-6`.

Anane, E., López C, D.C., Neubauer, P., Bournazou, M.N.C., 2017. Modelling overflow metabolism in *Escherichia coli* by acetate cycling. Biochemical Engineering Journal 125, 23–30. doi:`10.1016/j.bej.2017.05.013`.

ANSYS Fluent, 2015. User's Guide, release 16.2. ANSYS, Inc., Southpointe, 2600 ANSYS Drive, Canonsburg, PA 15317.

Askari, E., Proulx, P., Passalacqua, A., 2018. Modelling of bubbly flow using CFD-PBM solver in OpenFOAM: Study of local population balance models and extended quadrature method of moments applications. ChemEngineering 2, 8. doi:`10.3390/chemengineering2010008`.

Athanassoulis, G., Gavriliadis, P., 2002. The truncated Hausdorff moment problem solved by using kernel density functions. Probabilistic Engineering Mechanics 17, 273–291. doi:`10.1016/S0266-8920(02)00012-7`.

Bailey, J., 1998. Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. Biotechnology Progress 14, 8–20. doi:`10.1021/bp9701269`.

Baldyga, J., Bourne, J.R., 1999. Turbulent Mixing and Chemical Reactions. John Wiley and Sons.

Balsa-Canto, E., Alonso, A.A., Banga, J.R., 2010. An iterative identification procedure for dynamic modeling of biochemical networks. BMC Systems Biology 4, 11. doi:`10.1186/1752-0509-4-11`.

Barford, J.P., Hall, R.J., 1979. An examination of the crabtree effect in *Saccharomyces cerevisiae*: the role of respiratory adaptation. Journal of General Microbiology 114, 267–275. doi:`10.1099/00221287-114-2-267`.

Barigou, M., Greaves, M., 1992. Bubble-size distributions in a mechanically agitated gas-liquid contactor. Chemical Engineering Science 47, 2009 – 2025. doi:`10.1016/0009-2509(92)80318-7`.

Bellgardt, K.H., 1994. Analysis of synchronous growth of baker's yeast. part II: Comparison of model prediction and experimental data. Journal of Biotechnology 35, 35–49. doi:`10.1016/0168-1656(94)90188-0`.

Bergmann, M., 2004. Optimisation aérodynamique par réduction de modèle POD et contrôle optimal : application au sillage laminaire d'un cylindre circulaire. Ph.D. thesis. Institut National Polytechnique de Lorraine.

Bezzo, F., Macchietto, S., Pantelides, C.C., 2003. General hybrid multizonal/CFD approach for bioreactor modeling. AIChE Journal 49, 2133–2148. doi:`10.1002/aic.690490821`.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1462. doi:`10.1126/science.277.5331.1453`.

Bolotin, A., Quinquis, B., Sorokin, A., dusko Ehrlich, S., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561. doi:`10.1099/mic.0.28048-0`.

Brand, E., Junne, S., Anane, E., Cruz-Bournazou, M.N., Neubauer, P., 2018. Importance of the cultivation history for the response of *Escherichia coli* to oscillations in scale-down experiments. Bioprocess and Biosystems Engineering doi:`10.1007/s00449-018-1958-4`.

Brognaux, A., Han, S., Sørensen, S.J., Lebeau, F., Thonart, P., Delvigne, F., 2013. A low-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. Microbial Cell Factories 12, 100. doi:`10.1186/1475-2859-12-100`.

Buffo, A., Vanni, M., Marchisio, D., 2012. Multidimensional population balance model for the simulation of turbulent gas–liquid systems in stirred tank reactors. Chemical Engineering Science 70, 31–44. doi:`10.1016/j.ces.2011.04.042`.

Camarasa, C., Chiron, H., Daboussi, F., Valle, G.D., Dumas, C., Farines, V., Floury, J., Gagnaire, V., Gorret, N., Leonil, J., Mouret, J.R., O'Donohue, M.J., Sablayrolles, J.M., Salmon, J.M., Saulnier, L., Truan, G., 2018. INRA's research in industrial biotechnology: for food, chemicals, materials and fuels. Innovative Food Science & Emerging Technologies 46, 140–152. doi:`10.1016/j.ifset.2017.11.008`.

Caudal, J., Fiorina, B., Massot, M., Labégorre, B., Darabiha, N., Gicquel, O., 2013. Characteristic chemical time scales identification in reactive flows. Proceedings of the Combustion Institute 34, 1357 – 1364. doi:`doi.org/10.1016/j.proci.2012.06.178`.

Chalons, C., Fox, R., Laurent, F., Massot, M., Vié, A., 2017. Multivariate gaussian extended quadrature method of moments for turbulent disperse multiphase flow. Multiscale Modeling & Simulation 15, 1553–1583. doi:`10.1137/16M109209X`.

Chalons, C., Fox, R.O., Massot, M., 2010. A multi-gaussian quadrature method of moments for gas-particle flows in a LES framework, in: Proceedings of the 2010 Summer Program, Center for turbulence Research. Stanford University. pp. 347–358.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnology and Bioengineering 79, 53–73. doi:`10.1002/bit.10288`.

Chihara, T., 1978. An introduction to orthogonal polynomials. Number vol. 13 in Mathematics and its applications, Gordon and Breach.

Coste, P., 2013. A large interface model for two-phase CFD. Nuclear Engineering and Design 255, 38–50. doi:`10.1016/j.nucengdes.2012.10.008`.

Couderc, J.P., Gourdon, C., Liné, A., 2008. Phénomènes de transfert en génie des procédés. Lavoisier.

Courant, R., Friedrichs, K., Lewy, H., 1928. Über die partiellen differenzengleichungen der mathematischen physik. Mathematische Annalen 100, 32–74. doi:`10.1007/BF01448839`.

Delafosse, A., 2008. Analyse et Étude Numérique des Effets de Mélange dans un Bioréacteur. Ph.D. thesis. INSA de Toulouse.

Delafosse, A., Calvo, S., Collignon, M.L., Delvigne, F., Crine, M., Toye, D., 2015. Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – Comparison to experimental flow characterization and particle tracking. Chemical Engineering Science 134, 457–466. doi:`10.1016/j.ces.2015.05.045`.

Delafosse, A., Collignon, M.L., Calvo, S., Delvigne, F., Crine, M., Thonart, P., Toye, D., 2014. CFD-based compartment model for description of mixing in bioreactors. Chemical Engineering Science 106, 76–85. doi:`10.1016/j.ces.2013.11.033`.

Delafosse, A., Delvigne, F., Collignon, M.L., Crine, M., Thonart, P., Toye, D., 2010. Development of a compartment model based on CFD simulations for description of mixing in bioreactors. Biotechnology, Agronomy, Society and Environment 14, 517–522.

Delvigne, F., Goffin, P., 2014. Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. Biotechnology Journal 9, 61–72. doi:`10.1002/biot.201300119`.

Delvigne, F., Zune, Q., Lara, A.R., Al-Soud, W., Sørensen, S.J., 2014. Metabolic variability in bioprocessing: implications of microbial phenotypic heterogeneity. Trends in Biotechnology 32, 608–616. doi:`10.1016/j.tibtech.2014.10.002`.

Dette, H., Studden, W.J., 1997. The theory of canonical moments with applications in statistics, probability, and analysis. John Wiley & Sons, New York; Chichester.

van Dijken, J., Bauer, J., Brambilla, L., Duboc, P., Francois, J., Gancedo, C., Giuseppin, M., Heijnen, J., Hoare, M., Lange, H., Madden, E., Niederberger, P., Nielsen, J., Parrou, J., Petit, T., Porro, D., Reuss, M., van Riel, N., Rizzi, M., Steensma, H., Verrips, C., Vindeløv, J., Pronk, J., 2000. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. Enzyme and Microbial Technology 26, 706–714. doi:`10.1016/S0141-0229(00)00162-9`.

van Dijken, J.P., Scheffers, W., 1986. Redox balances in the metabolism of sugars by yeasts. FEMS Microbiology Letters 32, 199–224. doi:`10.1111/j.1574-6968.1986.tb01194.x`.

Elqotbi, M., Vlaev, S., Montastruc, L., Nikov, I., 2013. CFD modelling of two-phase stirred bioreaction systems by segregated solution of the euler-euler model. Computers & Chemical Engineering 48, 113–120. doi:`10.1016/j.compchemeng.2012.08.005`.

Enfors, S.O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H.Y., Neubauer, P., van der Lans, R., Luyben, K., Vrábel, P., Å. Manelius, 2001. Physiological responses to mixing in large scale bioreactors. Journal of Biotechnology 85, 175–185. doi:`10.1016/S0168-1656(00)00365-5`. twenty years of the European Federation of Biotechnology.

Ensight, 2016. version 10.2. Computational Engineering International, Inc., 2166 N. Salem Street, Suite 101, Apex, NC 27523 USA. URL: `http://www3.ensight.com/EnSight10_Docs/UserManual.pdf`.

Escudié, R., Liné, A., 2004. Experimental analysis of hydrodynamics in a radially agitated tank. AIChE Journal 49, 585–603. doi:`10.1002/aic.690490306`.

Farzan, P., Lerapetritou, M.G., 2017. Integrated modeling to capture the interaction of physiology and fluid dynamics in biopharmaceutical bioreactors. Computers & Chemical Engineering 97, 271–282. doi:`10.1016/j.compchemeng.2016.11.037`.

Favard, J., 1935. Sur les polynomes de Tchebicheff. C. r. hebd. séances Acad. sci. 200, 2052–2053.

Ferenci, T., 1996. Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. FEMS Microbiology Reviews 18, 301–317. doi:`10.1111/j.1574-6976.1996.tb00246.x`.

Ferenci, T., 1999. Regulation by nutrient limitation. Current Opinion in Microbiology 2, 208–213. doi:`10.1016/S1369-5274(99)80036-8`.

Ferenci, T., 2001. Hungry bacteria - definition and properties of a nutritional state. Environmental Microbiology 3, 605–611. doi:`10.1046/j.1462-2920.2001.00238.x`.

Flikweert, M.T., Kuyper, M., Maris, A.J.A.v., Kötter, P., Dijken, J.P.v., Pronk, J.T., 1999. Steady-state and transient-state analysis of growth and metabolite production in a *Saccharomyces cerevisiae* strain with reduced pyruvate-decarboxylase activity. Biotechnology and Bioengineering 66, 42–50. doi:`10.1002/(SICI)1097-0290(1999)66:1<42::AID-BIT4>3.0.CO;2-L`.

Ford, W., 2015. Chapter 19 - the symmetric eigenvalue problem, in: Numerical Linear Algebra with Applications. Academic Press, Boston, pp. 439–468. doi:`10.1016/B978-0-12-394435-1.00019-3`.

Frank, T., Shi, J.M., Burns, A.D., 2004. Vaalidation of eulerian multiphase flow models for nuclear safety applications., in: 3rd International Symposium on two-phase flow modelling and experimentations.

Fredrickson, A.G., Tsuchiya, H.M., 1963. Continuous propagation of microorganisms. AIChE Journal 9, 459–468. doi:`10.1002/aic.690090410`.

Gautschi, W., 2004. Orthogonal Polynomials: Computation and Approximation. Numerical mathematics and scientific computation, Oxford University Press.

Gerdtzen, Z.P., Daoutidis, P., Hu, W.S., 2004. Non-linear reduction for kinetic models of metabolic reaction networks. Metabolic Engineering 6, 140–154. doi:`10.1016/j.ymben.2003.11.003`.

Goldrick, S., Ştefan, A., Lovett, D., Montague, G., Lennox, B., 2015. The development of an industrial-scale fed-batch fermentation simulation. Journal of Biotechnology 193, 70–82. doi:`10.1016/j.jbiotec.2014.10.029`.

Gosset, G., 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate: sugar phosphotransferase system. Microbial Cell Factories 4, 14. doi:`10.1186/1475-2859-4-14`.

Guillou, V., Plourde-Owobi, L., Parrou, J.L., Goma, G., François, J., 2004. Role of reserve carbohydrates in the growth dynamics of saccharomyces cerevisiae. FEMS Yeast Research 4, 773–787. doi:`10.1016/j.femsyr.2004.05.005`.

Gzyl, H., Tagliani, A., 2010. Hausdorff moment problem and fractional moments. Applied Mathematics and Computation 216, 3319–3328. doi:`10.1016/j.amc.2010.04.059`.

Haringa, C., Tang, W., Deshmukh, A.T., Xia, J., Reuss, M., Heijnen, J.J., Mudde, R.F., Noorman, H.J., 2016. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. Engineering in Life Sciences doi:`10.1002/elsc.201600061`.

Hatzis, C., Srienc, F., Fredrickson, A., 1995. Multistaged corpuscular models of microbial growth: Monte Carlo simulations. Biosystems 36, 19–35. doi:`10.1016/0303-2647(95)01524-o`.

Heavner, B.D., Smallbone, K., Barker, B., Mendes, P., Walker, L.P., 2012. Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. BMC Systems Biology 6. doi:`10.1186/1752-0509-6-55`.

Heavner, B.D., Smallbone, K., Price, N.D., Walker, L.P., 2013. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. Database 2013. doi:`10.1093/database/bat059`.

Heins, A.L., Fernandes, R.L., Gernaey, K.V., Lantz, A.E., 2015. Experimental and in silico investigation of population heterogeneity in continuous *Sachharomyces cerevisiae* scale-down fermentation in a two-compartment setup. Journal of Chemical Technology and Biotechnology 90, 324–340. doi:`10.1002/jctb.4532`.

Hempfling, W.P., Mainzer, S.E., 1975. Effects of varying the carbon source limiting growth on yield and maintenance characteristics of *Escherichia coli* in continuous culture. Journal of bacteriology 123, 1076–1087.

Hibiki, T., Ishii, M., 2000. One-group interfacial area transport of bubbly flows in vertical round tubes. International Journal of Heat and Mass Transfer 43, 2711–2726. doi:`10.1016/s0017-9310(99)00325-7`.

Hjersted, J.L., Henson, M.A., 2006. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. Biotechnology Progress 22, 1239–1248. doi:`10.1021/bp060059v`.

Hristov, H., Mann, R., Lossev, V., Vlaev, S., 2004. A simplified CFD for three-dimensional analysis of fluid mixing, mass transfer and bioreaction in a fermenter equipped with triple novel geometry impellers. Food and Bioproducts Processing 82, 21–34. doi:`10.1205/096030804322985281`.

John, V., Angelov, I., Öncül, A., Thévenin, D., 2007. Techniques for the reconstruction of a distribution from a finite number of its moments. Chemical Engineering Science 62, 2890–2904. doi:`10.1016/j.ces.2007.02.041`.

Jung, W.M., Kang, S.H., Kim, W.S., Choi, C.K., 2000. Particle morphology of calcium carbonate precipitated by gas-liquid reaction in a Couette-Taylor reactor. Chemical Engineering Science 55, 733 – 747. doi:`10.1016/S0009-2509(99)00395-4`.

Kafarski, P., 2012. Rainbow code of biotechnology. CHEMIK 66, 811–816.

Kajihata, S., Matsuda, F., Yoshimi, M., Hayakawa, K., Furusawa, C., Kanda, A., Shimizu, H., 2015. 13c-based metabolic flux analysis of saccharomyces cerevisiae with a reduced crabtree effect. Journal of Bioscience and Bioengineering 120, 140–144. doi:`10.1016/j.jbiosc.2014.12.014`.

Kätterer, L., Allemann, H., Käppeli, O., Fiechter, A., 1986. Transient responses of continuously growing yeast cultures to dilution rate shifts: A sensitive means to analyze biology and the performance of equipment. Biotechnology and Bioengineering 28, 146–150. doi:`10.1002/bit.260280126`.

Kawase, Y., Halard, B., Moo-Young, M., 1992. Liquid-phase mass transfer coefficients in bioreactors. Biotechnology and Bioengineering 39, 1133–1140. doi:`10.1002/bit.260391109`.

Kim, S.R., Park, Y.C., Jin, Y.S., Seo, J.H., 2013. Strain engineering of *Saccharomyces cerevisiae* for enhanced xylose metabolism. Biotechnology Advances 31, 851–861. doi:`10.1016/j.biotechadv.2013.03.004`.

Klumpp, S., Scott, M., Pedersen, S., Hwa, T., 2013. Molecular crowding limits translation and cell growth. Proceedings of the National Academy of Sciences 110, 16754–16759. doi:`10.1073/pnas.1310377110`.

Kolev, N.I., 2007. Multiphase Flow Dynamics 2. Springer Berlin Heidelberg. doi:`10.1007/3-540-69835-3`.

Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. Computers & Chemical Engineering 51, 111–123. doi:`10.1016/j.compchemeng.2012.06.001`.

Kumar, S., Ramkrishna, D., 1996a. On the solution of population balance equations by discretization - I. a fixed pivot technique. Chemical Engineering Science 51, 1311 – 1332. doi:`10.1016/0009-2509(96)88489-2`.

Kumar, S., Ramkrishna, D., 1996b. On the solution of population balance equations by discretization - II. a moving pivot technique. Chemical Engineering Science 51, 1333–1342. doi:`10.1016/0009-2509(95)00355-X`.

Kumar, S., Ramkrishna, D., 1997. On the solution of population balance equations by discretization - III. nucleation, growth and aggregation of particles. Chemical Engineering Science 52, 4659–4679. doi:`10.1016/S0009-2509(97)00307-2`. festshrift for Professor M. M. Sharma.

Kuschel, M., Siebler, F., Takors, R., 2017. Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. Bioengineering 4, 27. doi:`10.3390/bioengineering4020027`.

Laakkonen, M., Moilanen, P., Alopaeus, V., Aittamaa, J., 2006. Dynamic modeling of local reaction conditions in an agitated aerobic fermenter. AIChE Journal 52, 1673–1689. doi:`10.1002/aic.10782`.

Laakkonen, M., Moilanen, P., Alopaeus, V., Aittamaa, J., 2007. Modelling local bubble size distributions in agitated vessels. Chemical Engineering Science 62, 721 – 740. doi:`10.1016/j.ces.2006.10.006`.

Lapin, A., Müller, D., Reuss, M., 2004. Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler–Lagrange methods: Traveling along the lifelines of single cells. Industrial & Engineering Chemistry Research 43, 4647–4656. doi:`10.1021/ie030786k`.

Lapin, A., Schmid, J., Reuss, M., 2006. Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor—a structured–segregated approach. Chemical Engineering Science 61, 4783 – 4797. doi:`10.1016/j.ces.2006.03.003`.

Launder, B., Spalding, D.B., 1972. Lectures in Mathematical Models of Turbulence. Academic Press Inc, London, England.

Laupsien, D., 2017. Hydrodynamics, Mass Transfer and Mixing induced by Bubble Plumes in Viscous Fluids. Ph.D. thesis. INSA de Toulouse.

Le Moullec, Y.and Gentric, C., Potier, O., Leclerc, J.P., 2010. Comparison of systemic, compartmental and CFD modelling approaches: application to the simulation of a biological reactor of wastewater treatment. Chemical Engineering Science 65, 343–350. doi:`10.1016/j.ces.2009.06.035`.

Lebaz, N., 2015. Modélisation de l'hydrolyse enzymatique de substrats lignocellulosiques par bilan de population. Ph.D. thesis. INSA de Toulouse. URL: `http://www.theses.fr/2015ISAT0030`.

Lebaz, N., Cockx, A., Spérandio, M., Morchain, J., 2016. Reconstruction of a distribution from a finite number of its moments: A comparative study in the case of depolymerization process. Computers & Chemical Engineering 84, 326–337. doi:`10.1016/j.compchemeng.2015.09.008`.

Leegwater, M.P.M., Neijssel, O.M., Tempest, D.W., 1982. Aspects of microbial physiology in relation to process control. Journal of Chemical Technology and Biotechnology 32, 92–99. doi:`10.1002/jctb.5030320113`.

Lei, F., Rotbøll, M., Jørgensen, S.B., 2001. A biochemically structured model for *Saccharomyces cerevisiae*. Journal of Biotechnology 88, 205–221. doi:`10.1016/S0168-1656(01)00269-3`.

Lencastre Fernandes, R., Carlquist, M., Lundin, L., Heins, A.L., Dutta, A., Sørensen, S.J., Jensen, A.D., Nopens, I., Lantz, A.E., Gernaey, K.V., 2012. Cell mass and cell cycle dynamics of an asynchronous budding yeast population: Experimental observations, flow cytometry data analysis, and multi-scale modeling. Biotechnology and Bioengineering 110, 812–826. doi:`10.1002/bit.24749`.

Li, L., Chen, N., Xiang, K., Xiang, B., 2018. A comparative CFD study on gas-liquid dispersion in a stirred tank with low and high gas loadings. International Journal of Chemical Reactor Engineering doi:`10.1515/ijcre-2017-0147`.

Liné, A., Gabelle, J.C., Morchain, J., Anne-Archard, D., Augier, F., 2013. On POD analysis of PIV measurements applied to mixing in a stirred vessel with a shear thinning fluid. Chemical Engineering Research and Design 91, 2073 – 2083. doi:`10.1016/j.cherd.2013.05.002`. mixing.

Linkès, M., Fede, P., Morchain, J., Schmitz, P., 2014. Numerical investigation of subgrid mixing effects on the calculation of biological reaction rates. Chemical Engineering Science 116, 473–485. doi:`10.1016/j.ces.2014.05.005`.

Löffler, M., Simen, J.D., Müller, J., Jäger, G., Laghrami, S., Schäferhoff, K., Freund, A., Takors, R., 2017. Switching between nitrogen and glucose limitation: Unraveling transcriptional dynamics in *Escherichia coli*. Journal of Biotechnology 258, 2–12. doi:`10.1016/j.jbiotec.2017.04.011`.

Machon, V., Pacek, A., Nienow, A., 1997. Bubble sizes in electrolyte and alcohol solutions in a turbulent stirred vessel. Chemical Engineering Research and Design 75, 339–348. doi:`10.1205/026387697523651`.

Madadi-Kandjani, E., Passalacqua, A., 2015. An extended quadrature-based moment method with log-normal kernel density functions. Chemical Engineering Science 131, 323–339. doi:`10.1016/j.ces.2015.04.005`.

Mantzaris, N.V., Daoutidis, P., Srienc, F., 2001. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. Computers & Chemical Engineering 25, 1411–1440. doi:`10.1016/S0098-1354(01)00709-8`.

Mantzaris, N.V., Daoutidis, P., Srienc, F., Fredrickson, A.G., 1999. Growth processes in a cascade of bioreactors: Mathematical models. AIChE Journal 45, 164–176. doi:`10.1002/aic.690450114`.

Marchisio, D., Fox, R., 2013. Computational Models for Polydisperse Particulate and Multiphase Systems. Cambridge Series in Chemical Engineering, Cambridge University Press.

Marchisio, D.L., Pikturna, J.T., Fox, R.O., Vigil, R.D., Barresi, A.A., 2003a. Quadrature method of moments for population-balance equations. AIChE Journal 49, 1266–1276. doi:`10.1002/aic.690490517`.

Marchisio, D.L., Vigil, R.D., Fox, R.O., 2003b. Implementation of the quadrature method of moments in CFD codes for aggregation-breakage problems. Chemical Engineering Science 58, 3337–3351. doi:`10.1016/S0009-2509(03)00211-2`.

MATLAB, 2016. version 9.0 (R2016a). The MathWorks, Inc., Natick, Massachusetts, United States.

Matsuoka, Y., Shimizu, K., 2013. Catabolite regulation analysis of *Escherichia coli* for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation. Journal of Biotechnology 168, 155 – 173. doi:`10.1016/j.jbiotec.2013.06.023`. special issue: Biotechnology for a healthy and green world.

Mayr, B., Horvat, P., Nagy, E., Moser, A., 1993. Mixing-models applied to industrial batch bioreactors. Bioprocess Engineering 9, 1–12. doi:`10.1007/BF00389534`.

McGraw, R., 1997. Description of aerosol dynamics by the quadrature method of moments. Aerosol Science and Technology 27, 255–265. doi:`10.1080/02786829708965471`.

Mead, L.R., Papanicolaou, N., 1984. Maximum entropy in the problem of moments. Journal of Mathematical Physics 25, 2404–2417.

Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metabolic Engineering 12, 150–160. doi:`10.1016/j.ymben.2009.07.006`.

Mer, S., Praud, O., Neau, H., Merigoux, N., Magnaudet, J., Roig, V., 2018. The emptying of a bottle as a test case for assessing interfacial momentum exchange models for euler–euler simulations of multi-scale gas-liquid flows. International Journal of Multiphase Flow 106, 109–124. doi:`10.1016/j.ijmultiphaseflow.2018.05.002`.

Merigoux, N., Lavieville, J., Mimouni, S., Guingo, M., Baudry, C., 2016. A generalized large interface to dispersed bubbly flow approach to model two-phase flows in nuclear power plant, in: MA-USA, C. (Ed.), CFD4NRS-6, pp. 1–20. Paper 11-1.

Moilanen, P., Laakkonen, M., Visuri, O., Aittamaa, J., 2007. Modeling local gas-liquid mass transfer in agitated viscous shear-thinning dispersions with CFD. Industrial & Engineering Chemistry Research 46, 7289–7299. doi:`10.1021/ie070566x`.

Moilanen, P., Laakkonen, M., Visuri, O., Alopaeus, V., Aittamaa, J., 2008. Modelling mass transfer in an aerated 0.2 m3 vessel agitated by rushton, phasejet and combijet impellers. Chemical Engineering Journal 142, 95–108. doi:`10.1016/j.cej.2008.01.033`.

Monod, J., 1952. Theory and Application of the Technique of Continuous Culture. Defense Technical Information Center.

Morchain, J., 2000. Étude et modélisation des couplages entre cinétiques physiques et biologiques dans les réacteurs de grand volume. Ph.D. thesis. INSA de Toulouse.

Morchain, J., 2017. Bioreactor Modeling. Elsevier. doi:`10.1016/c2015-0-05985-8`.

Morchain, J., Fonade, C., 2009. A structured model for the simulation of bioreactors under transient conditions. AIChE Journal 55, 2973–2984. doi:`10.1002/aic.11906`.

Morchain, J., Gabelle, J.C., Cockx, A., 2013. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. AIChE Journal 59, 369–379. doi:`10.1002/aic.13820`.

Morchain, J., Gabelle, J.C., Cockx, A., 2014. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AIChE Journal 60, 27–40. doi:`10.1002/aic.14238`.

Morchain, J., Pigou, M., Lebaz, N., 2016. A population balance model for bioreactors combining interdivision time distributions and micromixing concepts. Biochemical Engineering Journal , – doi:`10.1016/j.bej.2016.09.005`.

Müller, L., Klar, A., Schneider, F., 2017. A numerical comparison of the method of moments for the population balance equation. `arXiv:arXiv:1706.05854`.

Natarajan, A., Srienc, F., 2000. Glucose uptake rates of single *E. coli* cells grown in glucose-limited chemostat cultures. Journal of Microbiological Methods 42, 87–96. doi:`10.1016/S0167-7012(00)00180-9`. microbial Analysis at the Single Cell Level.

Nauha, E.K., Kálal, Z., Ali, J.M., Alopaeus, V., 2018. Compartmental modeling of large stirred tank bioreactors with high gas volume fractions. Chemical Engineering Journal 334, 2319 – 2334. doi:`/10.1016/j.cej.2017.11.182`.

NEPTUNE_CFD, 2016. version 3.2.0 user guide. EDF SA, 6 quai Watier, 78401 Chatou, France.

Neubauer, P., Häggström, L., Enfors, S.O., 1995. Influence of substrate oscillations on acetate formation and growth yield in escherichia coli glucose limited fed-batch cultivations. Biotechnology and Bioengineering 47, 139–146. doi:`10.1002/bit.260470204`.

Nguyen, T., Laurent, F., Fox, R., Massot, M., 2016. Solution of population balance equations in applications with fine particles: Mathematical modeling and numerical schemes. Journal of Computational Physics 325, 129–156. doi:`10.1016/j.jcp.2016.08.017`.

Nielsen, J., Villadsen, J., 1992. Modelling of microbial kinetics. Chemical Engineering Science 47, 4225–4270. doi:`10.1016/0009-2509(92)85104-J`.

Nobs, J.B., Maerkl, S.J., 2014. Long-term single cell analysis of *S. pombe* on a microfluidic microchemostat arraynobs. PLoS ONE 9. doi:`10.1371/journal.pone.0093466`.

Nöh, K., Grönke, K., Luo, B., Takors, R., Oldiges, M., Wiechert, W., 2007. Metabolic flux analysis at ultra short time scale: Isotopically non-stationary 13c labeling experiments. Journal of Biotechnology 129, 249–267. doi:`10.1016/j.jbiotec.2006.11.015`.

Noorman, H.J., Heijnen, J.J., 2017. Biochemical engineering´s grand adventure. Chemical Engineering Science 170, 677 – 693. doi:`10.1016/j.ces.2016.12.065`. 13th International Conference on Gas-Liquid and Gas-Liquid-Solid Reactor Engineering.

OECD, 2008. OECD Glossary of Statistical Terms. OECD Publishing.

Ordoñez, M.C., Raftery, J.P., Jaladi, T., Chen, X., Kao, K., Karim, M.N., 2016. Modelling of batch kinetics of aerobic carotenoid production using saccharomyces cerevisiae. Biochemical Engineering Journal 114, 226–236. doi:`10.1016/j.bej.2016.07.004`.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism – 2011. Molecular Systems Biology 7. doi:`10.1186/1752-0509-6-55`.

Paddon, C.J., Keasling, J.D., 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nature Reviews Microbiology 12, 355–366. doi:`10.1038/nrmicro3240`.

Pagliolico, S., Marchisio, D., Battesi, A.A., 1999. Influence of operating conditions on BaSO4 crystal size and morphology in a continuous couette precipitator. Journal of Thermal Analysis and Colorimetry 56, 1423–1433.

Passalacqua, A., Laurent, F., Madadi-Kandjani, E., Heylmun, J., Fox, R., 2017. An open-source quadrature-based population balance solver for OpenFOAM. Chem. Eng. Sci. doi:`10.1016/j.ces.2017.10.043`.

Petrik, M., Käppeli, O., Fiechter, A., 1983. An expanded concept for the glucose effect in the yeast *Saccharomyces uvarum*: Involvement of short- and long-term regulation. Journal of General Microbiology 129, 43–49. doi:`10.1099/00221287-129-1-43`.

Pigou, M., Morchain, J., 2015. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. Chemical Engineering Science 126, 267–282. doi:`10.1016/j.ces.2014.11.035`.

Pigou, M., Morchain, J., Fede, P., Penet, M.I., Laronze, G., 2017. An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors. Chemical Engineering Science 171, 218–232. doi:`10.1016/j.ces.2017.05.026`.

Pigou, M., Morchain, J., Fede, P., Penet, M.I., Laronze, G., 2018a. New developments of the extended quadrature method of moments to solve population balance equations. Journal of Computational Physics 365C, 243–268. doi:`10.1016/j.jcp.2018.03.027`.

Pigou, M., Morchain, J., Fede, P., Penet, M.I., Laronze, G., 2018b. A new moment-inversion procedure for the extended quadrature method of moments, in: 6th Population Balance Modelling Conference.

Pirt, S.J., 1965. The maintenance energy of bacteria in growing cultures. Royal Society of London Proceedings Series B 163, 224–231. doi:`10.1098/rspb.1965.0069`.

Pollack, M., Salenbauch, S., Marchisio, D., Hasse, C., 2016. Bivariate extensions of the extended quadrature method of moments (EQMOM) to describe coupled droplet evaporation and heat-up. Journal of Aerosol Science 92, 53–69. doi:`10.1016/j.jaerosci.2015.10.008`.

Postma, E., Verduyn, C., Scheffers, W.A., Dijken, J.P.V., 1989. Enzymatic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 55, 468–477. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC184133/`.

Pronk, J.T., Yde Steensma, H., Van Dijken, J.P., 1996. Pyruvate Metabolism in *Saccharomyces cerevisiae*. Yeast 12, 1607–1633. doi:`10.1002/(SICI)1097-0061(199612)12:16<1607::AID-YEA70>3.0.CO;2-4`.

Quedeville, V., Ouazaite, H., Polizzi, B., Fox, R., Villedieu, P., Fede, P., Létisse, F., Morchain, J., 2018. A two-dimensional population balance model for cell growth including multiple uptake systems. Chemical Engineering Research and Design 132, 966 – 981. doi:`10.1016/j.cherd.2018.02.025`.

Reuss, M., 1991. Structured modeling of bioreactors. Annals of the New York Academy of Sciences 646, 284–299. doi:`10.1111/j.1749-6632.1991.tb18591.x`.

Ribeiro, C.P., Lage, P.L., 2004. Population balance modeling of bubble size distributions in a direct-contact evaporator using a sparger model. Chemical Engineering Science 59, 2363–2377. doi:`10.1016/j.ces.2004.01.060`.

Rieger, M., Käppeli, O., Fiechter, A., 1983. The role of limited respiration in the incomplete oxidation of glucose by *Saccharomyces Cerevisiae*. Microbiology 129, 653–661. doi:`10.1099/00221287-129-3-653`.

Rizzi, M., Baltes, M., Theobald, U., Reuss, M., 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model. Biotechnology and Bioengineering 55, 592–608. doi:`10.1002/(SICI)1097-0290(19970820)55:4<592::AID-BIT2>3.0.CO;2-C`.

Robert, L., Hoffmann, M., Krell, N., Aymerich, S., Robert, J., Doumic, M., 2014. Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism. BMC Biology 12, 17. doi:`10.1186/1741-7007-12-17`.

Rodríguez, J., Perner, I., Schmidt, K., Posten, C., 2004. Simple metabolic model of *Saccharomyces cerevisiae* in fed-batch culture to study the cellular nitrogen uptake, in: Pons, M.N., Van Impe, J. (Eds.), Computer Applications in Biotechnology 2004, International Federation of Automatic Control. Elsevier Science Ltd.

Rosas-Lemus, M., Uribe-Alvarez, C., Chiquete-Félix, N., Uribe-Carvajal, S., 2014. In *Saccharomyces cerevisiae* fructose-1,6-bisphosphate contributes to the crabtree effect through closure of the mitochondrial unspecific channel. Archives of Biochemistry and Biophysics 555-556, 66–70. doi:`10.1016/j.abb.2014.05.027`.

Sánchez, B.J., Pérez-Correa, J.R., Agosin, E., 2014. Construction of robust dynamic genome-scale metabolic model structures of *Saccharomyces cerevisiae* through iterative re-parameterization. Metabolic Engineering 25, 159–173. doi:`10.1016/j.ymben.2014.07.004`.

Sander, R., 2015. Compilation of Henry's law constants (version 4.0) for water as solvent. Atmospheric Chemistry and Physics 15, 4399–4981. doi:`10.5194/acp-15-4399-2015`.

Schlegel, A., Böhm, A., Lee, S., Peist, R., Decker, K., Boos, W., 2002. Network regulation of the *Escherichia coli* maltose system. Journal of Molecular Microbiology and Biotechnology 4, 301–307.

Schmideder, A., Severin, T.S., Cremer, J.H., Weuster-Botz, D., 2015. A novel milliliter-scale chemostat system for parallel cultivation of microorganisms in stirred-tank bioreactors. Journal of Biotechnology 210, 19 – 24. doi:`10.1016/j.jbiotec.2015.06.402`.

Schütze, J., Hengstler, J., 2006. Assessing aerated bioreactor performance using CFD, in: 12th European Conference on Mixing, Bologna, Italy.

Shohat, J., Tamarkin, J., 1943. The Problem of Moments. Mathematical Surveys and Monographs. 4 ed., American Mathematical Society.

Silveston, P., Budman, H., Jervis, E., 2008. Forced modulation of biological processes: A review. Chemical Engineering Science 63, 5089–5105. doi:`10.1016/j.ces.2008.06.017`. 5TH Unsteady-State Processes in Catalysis: a Special Issue of Chemical Engineering Science.

Simen, J.D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., Müller, J., Takors, R., 2017. Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. Microbial Biotechnology 10, 858–872. doi:`10.1111/1751-7915.12713`.

Steuer, R., 2007. Computational approaches to the topology, stability and dynamics of metabolic networks. Phytochemistry 68, 2139–2151. doi:`10.1016/j.phytochem.2007.04.041`.

Sunya, S., Delvigne, F., Uribelarrea, J.L., Molina-Jouve, C., Gorret, N., 2012. Comparison of the transient responses of *Escherichia coli* to a glucose pulse of various intensities. Applied Microbiology and Biotechnology 95, 1021–1034. doi:`10.1007/s00253-012-3938-y`.

Sweere, A.P.J., Giesselbach, J., Barendse, R., de Krieger, R., Honderd, G., Luyben, K.C.A.M., 1988. Modelling the dynamic behaviour of *Saccharomyces cerevisiae* and its application in control experiments. Applied Microbiology and Biotechnology 28, 116–127. doi:`10.1007/BF00694298`.

Tagliani, A., 1999. Hausdorff moment problem and maximum entropy: A unified approach. Applied Mathematics and Computation 105, 291–305. doi:`10.1016/S0096-3003(98)10084-X`.

Tagliani, A., 2001. Numerical aspects of finite Hausdorff moment problem by maximum entropy approach. Applied Mathematics and Computation 118, 133 –149. doi:`10.1016/S0096-3003(99)00210-6`.

Takors, R., 2012. Scale-up of microbial processes: Impacts, tools and open questions. Journal of Biotechnology 160, 3–9. doi:`10.1016/j.jbiotec.2011.12.010`.

Tartakovsky, B., Sheintuch, M., Hilmer, J.M., Scheper, T., 1997. Modelling of *E.coli* fermentations: comparison of multicompartment and variable structure models. Bioprocess Engineering 16, 323–329. doi:`10.1007/pl00008947`.

Theisen, M., Liao, J.C., 2016. Industrial Biotechnology: *Escherichia coli* as a Host. Wiley-Blackwell. chapter 5. pp. 149–181. doi:`10.1002/9783527807796.ch5`.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., Reuss, M., 1997. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. experimental observations. Biotechnology and Bioengineering 55, 305–316. doi:`10.1002/(SICI)1097-0290(19970720)55:2<305::AID-BIT8>3.0.CO;2-M`.

Tomiyama, A., 1998. Struggle with computational bubble dynamics. Proc. Int. Conf. Multiphase Flow URL: `https://ci.nii.ac.jp/naid/10010263641/en/`.

Trinh, C.T., Wlaschin, A., Srienc, F., 2008. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Applied Microbiology and Biotechnology 81, 813. doi:`10.1007/s00253-008-1770-1`.

United Nations, 1992. Convention on biological diversity. URL: `https://www.cbd.int/doc/legal/cbd-en.pdf`.

Varma, A., Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* w3110. Applied and environmental microbiology 60, 3724–3731.

Vié, A., Laurent, F., Massot, M., 2013. Size-velocity correlations in hybrid high order moment/multi-fluid methods for polydisperse evaporating sprays: Modeling and numerical issues. Journal of Computational Physics 237, 177–210. doi:`10.1016/j.jcp.2012.11.043`.

Vikas, V., Wang, Z., Passalacqua, A., Fox, R., 2011. Realizable high-order finite-volume schemes for quadrature-based moment methods. Journal of Computational Physics 230, 5328–5352. doi:`10.1016/j.jcp.2011.03.038`.

Vrábel, P., van der Lans, R.G., van der Schot, F.N., Luyben, K.C., Xu, B., Enfors, S.O., 2001. CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. Chemical Engineering Journal 84, 463–474. doi:`10.1016/S1385-8947(00)00271-0`.

Vrábel, P., der Lans, R.V., Cui, Y., Luyben, K., 1999. Compartment model approach: Mixing in large scale aerated reactors with multiple impellers. Chemical Engineering Research and Design 77, 291–302. doi:`10.1205/026387699526223`.

Wheeler, J.C., 1974. Modified moments and gaussian quadratures. Rocky Mountain J. Math. 4, 287–296. doi:`10.1216/RMJ-1974-4-2-287`.

Winter, G., Krömer, J.O., 2013. Fluxomics - connecting `omics analysis and phenotypes. Environmental Microbiology 15, 1901–1916. doi:`10.1111/1462-2920.12064`.

Xu, B., Jahic, M., Enfors, S.O., 1999. Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. Biotechnology Progress 15, 81–90. doi:`10.1021/bp9801087`.

Yamanè, T., Shimizu, S., 1984. Fed-batch techniques in microbial processes, in: Bioprocess Parameter Control, Springer Berlin Heidelberg. pp. 147–194.

Yang, N., Xiao, Q., 2017. A mesoscale approach for population balance modeling of bubble size distribution in bubble column reactors. Chemical Engineering Science 170, 241–250. doi:`10.1016/j.ces.2017.01.026`.

Yasuda, K., 2011. Algebraic and geometric understanding of cells: Epigenetic inheritance of phenotypes between generations, in: Müller, S., Bley, T. (Eds.), High Resolution Microbial Single Cell Analytics. Springer Berlin Heidelberg. volume 124 of *Advances in Biochemical Engineering / Biotechnology*, pp. 55–81. doi:`10.1007/10_2010_97`.

Young, T.B., Bruley, D.F., Bungay, H.R., 1970. A dynamic mathematical model of the chemostat. Biotechnology and Bioengineering 12, 747–769. doi:`10.1002/bit.260120506`.

Yuan, C., Fox, R., 2011. Conditional quadrature method of moments for kinetic equations. Journal of Computational Physics 230, 8216–8246. doi:`10.1016/j.jcp.2011.07.020`.

Yuan, C., Laurent, F., Fox, R., 2012. An extended quadrature method of moments for population balance equations. Journal of Aerosol Science 51, 1–23. doi:`10.1016/j.jaerosci.2012.04.003`.

Zahradník, J., Mann, R., Fialová, M., Vlaev, D., Vlaev, S., Lossev, V., Seichter, P., 2001. A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. Chemical Engineering Science 56, 485–492. doi:`10.1016/s0009-2509(00)00252-9`.

Zhang, H., Zhang, K., Fan, S., 2009. CFD simulation coupled with population balance equations for aerated stirred bioreactors. Engineering in Life Sciences 9, 421–430. doi:`10.1002/elsc.200800074`.

Zhao, W., Buffo, A., Alopaeus, V., Han, B., Louhi-Kultanen, M., 2016. Application of the compartmental model to the gas-liquid precipitation of CO2-ca(OH)2aqueous system in a stirred tank. AIChE Journal 63, 378–386. doi:`10.1002/aic.15567`.

Zieringer, J., Takors, R., 2018. In silico prediction of large-scale microbial production performance: Constraints for getting proper data-driven models. Computational and Structural Biotechnology Journal 16, 246–256. doi:`10.1016/j.csbj.2018.06.002`.

# A

# INVESTIGATING THE INTERACTION BETWEEN PHYSICAL AND BIOLOGICAL HETEROGENEITIES IN BIOREACTORS USING COMPARTMENT, POPULATION BALANCE AND METABOLIC MODELS

Chemical Engineering Science 126 (2015) 267–282

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

# Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models

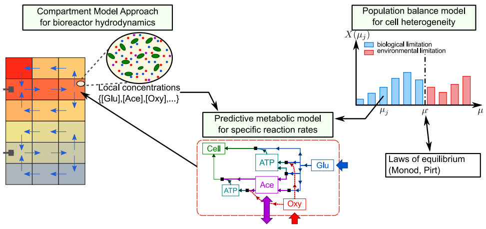
Maxime Pigou, Jérôme Morchain*

*Université de Toulouse, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INSA/CNRS/INRA, 135 Avenue de Rangueil, 31077 Toulouse Cedex, France*

## HIGHLIGHTS

- Population balance, metabolic and compartment models are coupled.
- The disequilibrium between substrate uptake and utilization rate is introduced.
- Simulations at lab and large scales compare favorably to *E. coli* culture data.
- Mixing limitation induces substrate gradients affecting metabolism.
- Through scale-up, yield loss and by-product formation are predicted.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

*Article history:*
Received 21 July 2014
Received in revised form
3 November 2014
Accepted 14 November 2014
Available online 26 November 2014

*Keywords:*
Two-way coupling
Population balance
Metabolic model
Compartment model
Bioreactor scale-up

## ABSTRACT

A generic model for the description of biological population dynamics in industrial bioreactors is detailed. Hydrodynamics, mass transfer between the cell and the surrounding fluid, population heterogeneity, metabolism and biological adaptation have to be considered with equal interest and, if possible, simultaneously. This model couples a hydrodynamic model, a population balance model for the growth rate adaptation and a metabolic model predicting the reaction rates depending on the state of the individuals. This approach dissociates the growth rate from local concentrations leading to a good understanding of the effects of a changing environment on a microbial population. Our model is applied to *Escherichia coli* for which experimental data exist in the literature for batch and fed-batch cultures. The considered strain is known for producing acetate when exposed to heterogeneities. When simulating a large bioreactor using a compartment model approach for hydrodynamics, our coupled model could predict that, under certain conditions, acetate is simultaneously produced and consumed in different areas of the reactor.

## 1. Introduction

The modeling of intensified bioreactors is a current challenge for both the academic and industrial communities. This is principally due to the complexity of such processes that combine mixing, transfer and reaction over a very large range of time and length scales. Compared to other chemical processes, the case of bioprocesses is even more complex because of the dynamic response of biosystems (Enfors et al., 2001). The chemical composition of the liquid phase constraints the cell potential, the biological uptakes modify the concentration fields and microorganisms adapt to the concentration changes experienced along their trajectories. Aiming at higher productivities always pushes the bioreactor towards a more severe competition between mixing and substrate uptake (Linkès et al., 2014). The main consequence is the formation of large scale concentration gradients which in turn expose cells to fluctuating concentration signals further triggering cell

---

[*] Corresponding author. Tel.: +33 561559774; fax: +33 561559400.
*E-mail address:* jerome.morchain@insa-toulouse.fr (J. Morchain).

http://dx.doi.org/10.1016/j.ces.2014.11.035
0009-2509/

adaptation and metabolic dysfunctioning. It has been shown that periodic exposure to excessive substrate concentration induces some modifications of the cell metabolism and leads to: (i) the over assimilation of substrates, (ii) the formation of by-products, and (iii) a decrease in the overall reactor performance. As a result, the modeling and simulation of industrial bioreactors leads to a strong two-way coupling issue illustrated in Fig. 1. Indeed, in biological reactors, the reaction rates co-evolve with concentration fields due to a permanent adaptation of the physiological state of cells.

Comparing the characteristic times of the various processes involved is of great help to identify the potential issues. In a previous work, mixing, substrate uptake and biological adaptation times were considered (Morchain et al., 2014). An important result concerns the concept of local equilibrium between the biophase and the liquid phase. A cell may adapt to any change in its environment (Ferenci, 1996, 1999) through a large variety of response systems having different characteristic times (Wick et al., 2001, 2002; Franchini and Egli, 2006; Ryall et al., 2012). As an illustration, growth rate adaptation is a slow process whereas the change of biovolume due to an osmotic shock is very fast. In between, the characteristic time of substrate uptake regulation is in the range of a few seconds. In most, if not all, studies coupling hydrodynamics and bioreactions, it is assumed that the microorganisms are at an equilibrium with their environment: the reaction rates are calculated from the local concentrations in the liquid phase via a kinetic (Altintas et al., 2006; Peskov et al., 2012) or a metabolic model (Xu et al., 1999; Meadows et al., 2010; Matsuoka and Shimizu, 2013). The consequence is that the actual history of cells is not considered: all cells are supposed to behave as if they were adapted to that local environment. In other words, instantaneous adaptation of the cell functioning is postulated. In order to account for the cell diversity and the ability to be out of equilibrium, the outgoing approach is to consider population balance modeling. Cells are segregated according to one or more internal properties. They are presumably different from one another. The vector of internal properties can be used to define the behavior of any cell in terms of reaction rates. This approach is very powerful to address the issues related to bioreactor dynamics. Unfortunately, cell ensembles (Henson et al., 2002; Mantzaris, 2005, 2006, 2007) and class methods (Lencastre Fernandes et al., 2012) are not easy to implement in the framework of Computational Fluid Dynamics softwares and would lead to prohibitively large computational times (Lapin et al., 2006).

Nevertheless, hydrodynamics, mass transfer between the cell and the surrounding fluid, population heterogeneity, metabolism and biological adaptation have to be considered with equal interest and, if possible, simultaneously. One possibility to combine all these aspects into a tractable model is to reduce the size of the problem through the use of a simplified hydrodynamic model named compartment approach (Vrábel et al., 1999, 2000; Lencastre Fernandes et al., 2013). Bezzo et al. (2003), Moullec et al. (2010) and Delafosse et al. (2014) have proposed different techniques to transpose the 3D-CFD results into a reduced compartment model.

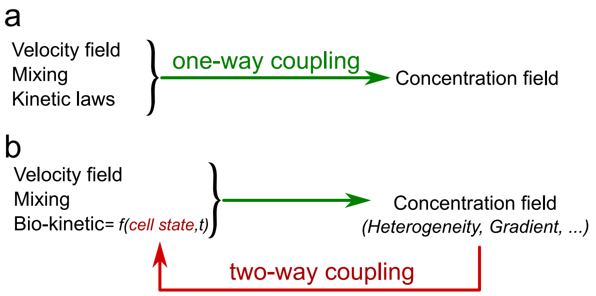


**Fig. 1.** Illustration of (a) the one-way coupling (as in chemical reactors) and (b) the two-way coupling in biochemical reactors: concentration gradients impact the cell state and induce biological heterogeneity.

In this paper, we propose to combine a compartment model approach for the hydrodynamics, a population balance model for the population dynamics and a metabolic model for the description of bioreactions. The problem formulation is closed by setting the mass transfer law between cells and the liquid phase. The first part of the paper deals with the model presentation with a minimum details since most aspects have already been published elsewhere (Morchain and Fonade, 2009; Morchain et al., 2013, 2014). Then the metabolic part of the model (adapted from Xu et al., 1999) is validated against experimental results obtained in a small scale (15 L) batch cultivation of *Escherichia coli*. In the third part the whole model combining the aforementioned aspects is compared to some experimental results from a 20 $m^3$ fed-batch cultivation of the same strain (Xu et al., 1999a; Vrábel et al., 2001). In that case, we will rely upon the reactor compartimentation proposed by the authors. A significant improvement of the predictive capacities is obtained with our two-way coupled approach, without parameter adjustment between the two scales. Namely, the occurrence of a disequilibrium between the cell and the environment allow the formation of large amount of by-products when assimilation rates exceed the internal utilization rates for growth and energy production. On the opposite, by-products are uptaken and used as secondary substrate in zones where the main substrate is depleted. Owing to the population balance approach the threshold between substrate excess and substrate limiting conditions is relative to the physiological state of each microorganism. The actual behavior of each subgroup of cells results from the disequilibrium between its own potential and that offered by the local environment.

## 2. Materials and methods

### 2.1. Hydrodynamic model

The principle of the Compartment Model Approach (CMA) is briefly recalled hereafter. More details can be found in the original papers of Hristov et al. (2004) and Zahradník et al. (2001). In such a model, the reactor's volume is split into $N$ sub-volumes, referred as “compartments” and considered as perfectly homogeneous. We then need to define a “circulation map” representing the flow pattern circulating between those compartments. This circulation map may be deduced from CFD simulations (Delafosse et al., 2010, 2014) or experimental observations (Vrábel et al., 1999).

To mathematically implement such a model, we define a matrix of volume flow rates, $M^f$, which is a $N$-by-$N$ matrix such as the $M^f_{m,n}$ entry of this matrix is the value of the flow, expressed in $m^3.h^{-1}$, going from the $m$-th compartment to the $n$-th one with $m$ and $n \in \{1, \dots, N\}$.

Let $C_i^n$ be the mass concentration of any species, referred as $i$, in the compartment $n \in \{1, \dots, N\}$. Then, the mass conservation equation for this species in the $n$-th compartment is given in the following equation:

$$\frac{\partial C_i^n}{\partial t} = R_i + T_i + \sum_{m=1}^{N} (M^f_{m,n} C_i^m) - C_i^n \sum_{m=1}^{N} (M^f_{n,m}) \qquad (1)$$

The term $T_i$ is a gas–liquid transfer rate and $R_i$ an overall reaction rate, both expressed in $g_i\,L^{-1}\,h^{-1}$. The calculation procedure for the bioreaction rates will be detailed in following parts through a population balance model and a metabolic model.

The mass balances, for each species and over each compartment, lead to a set of Ordinary Differential Equations. The ODEs sets were solved using a program written with MATLAB® 7.9 (R2009b) using the ODE solver *ode23* with relative and absolute error tolerances set respectively to $10^{-3}$ and $10^{-6}$.

### 2.2. Population balance

The population balance model, detailed in previous papers (Morchain et al., 2013, 2014), discriminates individuals depending on their growth capabilities represented by a biological growth rate $\mu$. The application of the general formulation of a Population Balance Equation (PBE) proposed by Ramkrishna (2000), adapted to our case in a homogeneous volume, is given by the following equation:

$$\frac{\partial X(\mu,t)}{\partial t} = -\frac{\partial}{\partial \mu}(X(\mu,t)\zeta) + h(\mu,t) \tag{2}$$

In this formulation, $\zeta$ represents the rate of variation of the internal variable $\mu$ and $h$ is the rate of production of cell with a given value of $\mu$.

We consider that the cell division will lead to the formation of daughter cells which inherit of their mother's growth rate. Under that assumption, the PBE given in Eq. (2) becomes

$$\frac{\partial X(\mu,t)}{\partial t} = -\frac{\partial}{\partial \mu}(X(\mu,t)\zeta(\mu)) + \mu X(\mu,t) \tag{3}$$

The first term on the right hand side of Eq. (3) is the net flux of biomass moving in $\mu$-space. $\zeta$ represents the velocity in that space.

The model actually distinguishes an upward velocity, $\zeta^u$ computed using a time constant $T^u$ characterizing the upward adaptation, and a downward velocity $\zeta^d$ computed with a similar time constant $T^d$.

$$\zeta^u(\mu) \approx \left(T^{u-1} + \mu\right)(\mu^* - \mu) \tag{4}$$

$$\zeta^d(\mu) \approx \left(T^{d-1} + \mu\right)(\mu^* - \mu) \tag{5}$$

This formulation implies that the population will tend to reach a balanced growth rate $\mu^*$ representing the optimal growth rate considering local concentrations. We can assume that the relationship between $\mu^*$ and those local concentrations is known.

In order to solve this PBE, we use a discretization method. Previous equations have then been discretized and their exact formulations are given in previous work (Morchain et al., 2013). The biological population is segregated within $J$ classes referred by $j \in \{1,\ldots,J\}$. Each class is characterized by its biological growth rate $\mu_j$:

$$\mu_j = \mu_{\max}\frac{j-1}{J-1} \tag{6}$$

The biomass concentration within the class $j$ is noted $X_j$, the total biomass concentration $X_T$ is then given by

$$X_T = \sum_{j=1}^{J} X_j \tag{7}$$

Each concentration $X_j$ is transported by the liquid phase as dissolved species, which allow applying Eq. (1) to those $J$ concentrations.

The migration of biomass between classes is presented in Fig. 2. A change in the environment results in a modification of the optimal growth rate, $\mu^*$, represented by the double-arrow in Fig. 2. The population does not adapt immediately to the new environment, but migrates toward the two classes surrounding $\mu^*$ at a rate controlled by the time constants $T^u$ and $T^d$.

The second term of Eq. (3) represents the formation of new cells with a growth rate equal to $\mu$. After discretizing this term, we introduce the actual growth rate $\mu_j^a$. Indeed, if the environmental conditions are favorable, cells will be able to achieve the growth rate $\mu_j$ corresponding to their class. If the environment is limiting, cells will not be able to grow at their potential growth rate $\mu_j$ and will be limited to the maximum growth rate possible in that environment, $\mu^*$. Thus, the actual growth rate in each class is calculated as the minimum of the two: $\mu_j^a = \min\left(\mu_j, \mu^*\right)$. In the discretized version of Eq. (3), the rate of cell production is therefore defined as $\mu_j^a X_j$. See also Appendix A for further details.

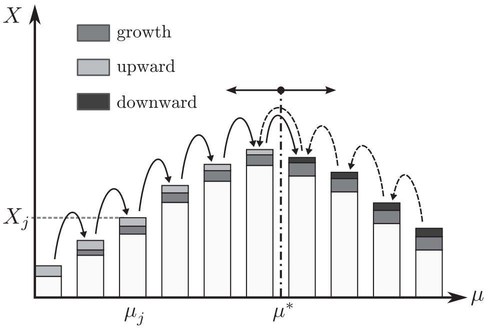


**Fig. 2.** Schematic representation of mechanisms affecting the specific growth rate distribution.

It is noteworthy that cells with a specific growth rate smaller than $\mu^*$ are limited by their own biological capabilities whereas those having a specific growth rate higher than $\mu^*$ will be limited by the environment. Although the environment is the same for all cells, their metabolic behaviors are therefore expected to be different. The population balance model thus introduces inertia, or time delay, in the dynamic response of the cell population to a changing environment. It decouples the actual growth rate of the population from local concentrations.

### 2.3. Metabolic model

#### 2.3.1. General description

The metabolic model is invoked for each class of individuals described in the population balance model. It allows the calculation of the specific reaction rates for each group of individuals. This step only requires the knowledge of the concentrations in the liquid phase and the specific growth rate of individuals (related to their class index).

Our metabolic model for *E. coli* is based on the one described by Xu et al. (1999): their model details the growth of *E. coli* in aerobic conditions in the presence of glucose or acetate. They considered a metabolism, called “overflow metabolism”, leading to acetate formation when the oxidative pathway for energy production was saturated.

Based on their work, we added a fermentative pathway for energy production triggered in anaerobic conditions, or when the oxidative pathway is saturated. We also dissociated the overflow metabolism from energy or biomass production, which was not the case in Xu et al.'s model. Now, this metabolism is triggered when there is a disequilibrium between glucose assimilation and needs. Finally, we used the Pirt's formulation (Pirt, 1965) to estimate the conversion yield of glucose in biomass as explained in Appendix A.

Then, our metabolic model is based on four categories of biological reactions:

- Anabolism (*ana*): biomass production through glucose (or acetate) and energy consumption.
- Oxidative catabolism (*oxy*): energy production through oxidative pathway.
- Fermentative catabolism (*ferm*): energy production through fermentative pathway (mixed-acid fermentation).
- Metabolism (*over*): production of acetate through glucose overconsumption.

The oxidative and fermentative catabolism pathways allow to produce the energy needed for biomass growth. In presence of oxygen, the oxidative pathway will be preferably used for its better energetic yield. The fermentative pathway is triggered only when the energetic need for growth is not fulfilled by the oxidative pathway, it leads to acetate formation.

It is known that *E. coli* is able to grow using glucose ($G$), or acetate ($A$) if glucose is depleted, as carbon source. If a cell consumes more glucose than what it is able to use, the amount in excess will be converted in acetate through the overflow metabolism (Matsuoka and Shimizu, 2013). In our work, we assimilated the energy to molecules of Adenosine TriPhosphate (ATP, noted $E$) which is the main energy source used for cellular functions.

Those considerations lead to the following set of reactions, represented in Fig. 3:

$$G + Y_{EG}E \xrightarrow{q^G_{ana}} Y_{XG}X \tag{R1}$$

$$A + Y_{EA}E \xrightarrow{q^A_{ana}} Y_{XA}X \tag{R1$'$}$$

$$G + Y_{OG}O \xrightarrow{q^G_{oxy}} Y^o_{EG}E \tag{R2}$$

$$A + Y_{OA}O \xrightarrow{q^A_{oxy}} Y^o_{EA}E \tag{R2$'$}$$

$$G \xrightarrow{q^G_{ferm}} Y^f_{EG}E + Y_{AG}A \tag{R3}$$

$$G \xrightarrow{q^G_{over}} Y_{AG}A \tag{R4}$$

With $Y_{BA}$ being stoichiometric molar coefficients in $\text{mol}_B\ \text{mol}_A^{-1}$ while $q^G_\alpha$ and $q^A_\alpha$ are the specific reaction rates on $G$ and $A$ respectively expressed in $\text{mol}_G\ g_X^{-1}\ \text{h}^{-1}$ and $mol_A\ g_X^{-1}\ \text{h}^{-1}$.

### 2.3.2. Solution procedure

The improvements made between Xu et al.'s model and ours do not rely only on the new fermentative reaction, but mainly on its strong coupling with the population balance model. Our goal is here to estimate the reaction rates for individuals in each class. When solving the ODEs set, the following procedure will be called for each class of the population balance model in each compartment of the hydrodynamic model. Although modern metabolic flux calculations now refer to time-consuming optimization methods, our procedure is a hierarchic method, consisting in a set of tests, similar to the method used by Xu et al. (1999). The advantage of the simple method adopted here is to allow a direct and rapid calculation of the reaction rates.

First, let us dissociate the overflow reaction ($R_4$) from "useful reactions". By useful reaction, we mean the reactions whose goal is the production of new cells, through direct production (anabolism: $R_1$ and $R_1'$) or energy production (catabolism: $R_2$, $R_2'$ and $R_3$).

Our main hypothesis is that there is an energetic balance within a cell: the rate of ATP production through catabolism reactions must be equal to the rate of ATP consumption by anabolism. Under that assumption, we will be able to define a method to estimate the useful reaction rates of a cell, just by knowing its specific growth rate $\mu_j$, and liquid-phase concentrations ($G$, $A$, $O$). This method is represented by the hereunder function $f$, and is detailed in Appendix A.

$$f : [0,\ \mu_{\max}] \times \mathbb{R}^3_+ \longrightarrow \mathbb{R}^5 (\mu_j, G, A, O) \mapsto (q^G_{ana}, q^A_{ana}, q^G_{oxy}, q^A_{oxy}, q^G_{ferm}) \tag{8}$$

With this method defined, the overflow reaction rate is the only one missing. In our model, this metabolism is triggered when a cell consumes (or uptakes) more glucose than what is needed for its growth.

Let us note that $\Phi^u_G$ the useful glucose uptake rate, i.e. the amount of glucose which is used by anabolism and catabolism reactions. This rate is then defined as

$$\Phi^u_{G,j} = q^G_{ana,j} + q^G_{oxy,j} + q^G_{ferm,j} \tag{9}$$

The reaction rate of the overflow metabolism, will simply be the difference between the total glucose uptake rate $\Phi_{G,j}$ and the useful part of this rate $\Phi^u_{G,j}$:

$$q^G_{over,j} = \Phi_{G,j} - \Phi^u_{G,j} \tag{10}$$

Closing this problem finally requires an estimation of the total glucose uptake rate $\Phi_{G,j}$. It has been observed that the glucose uptake rate, in continuous cultures of *E. coli*, is not correlated to the growth rate (Leegwater et al., 1982; Natarajan and Srienc, 2000). In other words, the glucose uptake rate will be the same for all cells, no matter their class-index.

As cells dynamically adapt their uptake capacity in response to substrate fluctuations (Ferenci, 1996), our proposal is to consider that the regulation of the uptake system is fast compared to the characteristic time of concentration fluctuations. This adaptation of the uptake system will be made in order to uptake the amount of glucose that would be needed for internal reactions if the cell were at equilibrium with the environment.

By definition, cells at equilibrium do not produce overflow metabolites and therefore $\Phi_{G,j}$ can be expressed as

$$\forall j \in \{1, \ldots, J\}, \Phi_{G,j} = \Phi^*_G = \Phi^u_G(\mu^*) \tag{11}$$

In other words, the glucose uptake rate equals the sum of glucose internal utilization rates of a cell whose biological growth rate is $\mu_j = \mu^*$.

The last requirement in this procedure is an expression for $\mu^*$. This equilibrium growth rate represents the mean growth rate of a biological population adapted to its environment. Usually, this

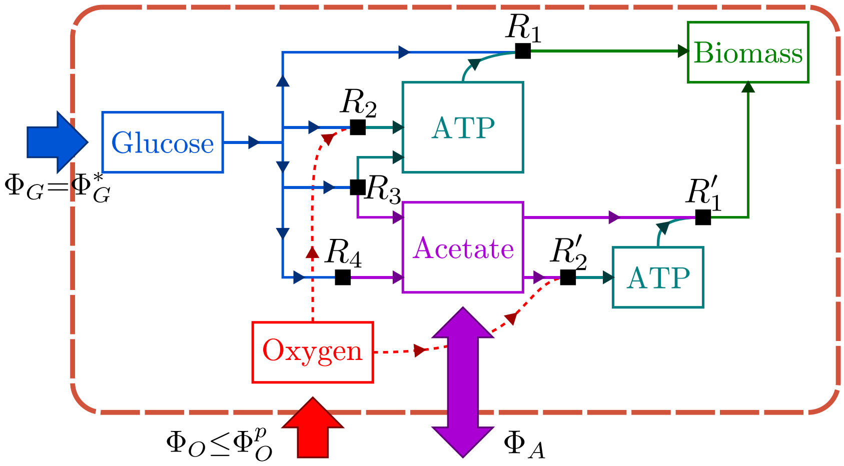


**Fig. 3.** Representation of the internal reactions considered in the metabolic model.

growth rate is modelized using an empirical law such as a Monod formulation which we used here.

We consider growth on two different substrates: glucose and acetate. The presence of glucose strongly inhibits the use of acetate (Xu et al., 1999). Moreover, acetate is known for its inhibition properties. This leads to the following formulation for $\mu^*$ :

$$\mu^{G*} = \mu^G_{max} \frac{G}{G+K_G} \frac{O}{O+K_O} \frac{K_{i,A}}{A+K_{i,A}} \tag{12}$$

$$\mu^{A*} = \mu^A_{max} \frac{A}{A+K_A} \frac{O}{O+K_O} \frac{K_{i,G}}{G+K_{i,G}}. \tag{13}$$

$$\mu^* = \mu^{G*} + \mu^{A*} \tag{14}$$

Summing up, the calculation process requires the following steps:

(i) Estimate the equilibrium growth rate from local concentrations:

$$G, A, O \mapsto \mu^*$$

(ii) Estimate the glucose consumption rate of a virtual cell at equilibrium with its environment:

$$(q^{G*}_\alpha, q^{A*}_\alpha) = f(\mu^*, G, A, O),\ \alpha \text{ stands for any reaction (anabolism, catabolism, etc.)}$$

(iii) Deduce the population glucose uptake rate from this balanced state:

$$\Phi_G = \Phi^u_G(\mu^*) = q^{G*}_{ana} + q^{G*}_{oxy} + q^{G*}_{ferm}$$

(iv) In each class $j \in \{1, \ldots, J\}$, estimate the internal reaction rates for anabolism and catabolism:

$$(q^G_{\alpha,j}, q^A_{\alpha,j}) = f(\mu_j, G, A, O)$$

(v) In each class, deduce from precedent results the overflow reaction rate:

$$q^G_{over,j} = \Phi_G - q^G_{ana,j} - q^G_{oxy,j} - q^G_{ferm,j}$$

It must be understood that all these reaction rates will strongly depend on the class-index of the considered cell. In the same environment, two cells of different classes will exhibit very different behaviors. An illustration of the variety of behaviors encountered in a same population, for a particular environment, is given in Appendix A after the presentation of the method $f$.

### 2.3.3. Overall reaction rates

Knowing the reaction rates for each reaction, the specific consumption or production rates $r_{i,j}$ in $g_i\, g_{X_j}^{-1}\, h^{-1}$ are calculated:

$$r_{X,j} = \left(q^G_{ana,j} Y_{XG} + q^A_{ana,j} Y_{XA}\right) M_X = \mu^a_j \tag{15}$$

$$r_{G,j} = -\left(q^G_{ana,j} + q^G_{oxy,j} + q^G_{ferm,j} + q^G_{over,j}\right) M_G \tag{16}$$

$$r_{A,j} = \left(q^G_{ferm,j} Y_{AG} + q^G_{over,j} Y_{AG} - q^A_{oxy,j} - q^A_{ana,j}\right) M_A \tag{17}$$

$$r_{O,j} = -\left(q^G_{oxy,j} Y_{OG} + q^A_{oxy,j} Y_{OA}\right) M_O \tag{18}$$

The volumetric reaction rates needed in the mass conservation equation (Eq. (1)), $R_i$ in $g_i\, \mathrm{L}^{-1}\, h^{-1}$ with $i \in \{X, G, A, O\}$, are easily obtained from a summation of the specific rates over the entire population:

$$R_i = \sum_{j=1}^{J} r_{i,j} X_j \tag{19}$$

For all specific variables, i.e. variables expressed per gram of biomass (subscript $j$), we define a mean variable over the population using the notation $\tilde{}$. Thus, the population growth rate, noted $\tilde{\mu^a}$, is given by

$$\tilde{\mu^a} = \frac{1}{X_T} \sum_{j=1}^{J} (\mu^a_j X_j) \tag{20}$$

Note that $\tilde{\mu^a}$ is the growth rate experimentally observed.

Model parameters used in our simulations of *E. coli* cultivations are given in Tables 1 and 2.

## 3. Results

### 3.1. Batch culture

First, our biological model has been challenged by a set of experimental data coming from a batch culture conducted in a 15 liters stirred tank reactor. This culture and related measurement methods have been described by Xu et al. (1999). From these data, it can be assumed that mixing and oxygen mass transfer were not limiting, so that the reactor is treated as perfectly mixed. Thus, the associated Compartment Model consisted in a single compartment ($M^f = 0$). In the experiment, the oxygen concentration was regulated around a value corresponding to 30% of saturation. Accordingly it was decided to discard the conservation equation for oxygen in the liquid phase and to impose a constant value in the simulations.

During a 12 h culture, measurements of glucose, biomass and acetate concentrations were carried out each hour. Plus, a respirometric monitoring allowed to determine the Oxygen Consumption Rate (*OCR* in $\mathrm{mmol}_O\ \mathrm{L}^{-1}\ \mathrm{h}^{-1}$). In that paper the authors proposed a metabolic model for aerobic cultivation of *E. coli* in which overflow metabolism resulted from the saturation of the oxidative capacity of cells. Our formulation is slightly different: overflow results from an extra assimilation of substrate compared to the cell needs. The latter may include energy production by fermentation if the oxidative capacity is saturated. Therefore it was essential to check

**Table 1**
Model constants for *E. coli*.

| Symbol | Value | Unit | Source |
|---|---|---|---|
| Molar masses | | | |
| $M_X$ | 113.1 | $g_X\ \mathrm{mol}_X^{-1}$ | [a] |
| $M_G$ | 180.2 | $g_G\ \mathrm{mol}_G^{-1}$ | |
| $M_O$ | 32.0 | $g_O\ \mathrm{mol}_O^{-1}$ | |
| $M_A$ | 59.0 | $g_A\ \mathrm{mol}_A^{-1}$ | |
| Affinity and inhibition constants | | | |
| $K_G$ | 0.05 | $g_G\ \mathrm{L}^{-1}$ | $0.05^b$<br>$0.095^c$ |
| $K_A$ | 0.05 | $g_A\ \mathrm{L}^{-1}$ | $0.05^b$ |
| $K_O$ | 0.1 | $mg_O\ \mathrm{L}^{-1}$ | $0.0768^c$<br>$0.1^d$ |
| $K_{i,G}$ | 0.2 | $g_G\ \mathrm{L}^{-1}$ | |
| $K_{i,A}$ | $3.0^e$ | $g_A\ \mathrm{L}^{-1}$ | $4.0–5.0^b$ |
| $K^o_{i,A}$ | 4.0 | $g_A\ \mathrm{L}^{-1}$ | $4.0^b$ |

[a] The biomass is represented by the typical chemical formula $C_5H_7NO_2$.
[b] Xu et al. (1999).
[c] Meadows et al. (2010).
[d] Morchain et al. (2013).
[e] Curve fitting.

**Table 2**
Model constants for *E. coli*

| Symbol | Value | Unit | Source |
|---|---|---|---|
| Molar yields | | | |
| $Y_{EG}$ | 12.05 | $\text{mol}_E\ \text{mol}_G^{-1}$ | |
| $Y_{XG}^{\max}$ | 1.32 | $\text{mol}_X\ \text{mol}_G^{-1}$ | $1.25_{\pm 0.05}$[a] |
| $m$ | 250 | $\mu\text{mol}_G\ g_X^{-1}.h^{-1}$ | $310^{b}$<br>$220^{c}$<br>$350_{\pm 40^{d}}$ |
| $Y_{OG}$ | 6.0 | $\text{mol}_O\ \text{mol}_G^{-1}$ | 6.0[c,d] |
| $Y_{EG}^{o}$ | 20.0 | $\text{mol}_E\ \text{mol}_G^{-1}$ | 18.7[d] |
| $Y_{EG}^{f}$ | 3.0 | $\text{mol}_E\ \text{mol}_G^{-1}$ | 3.0[e] |
| $Y_{AG}$ | 2.0 | $\text{mol}_A\ \text{mol}_G^{-1}$ | 2.0[c] |
| $Y_{EA}$ | 4.0 | $\text{mol}_E\ \text{mol}_A^{-1}$ | |
| $Y_{XA}$ | 0.40[f] | $\text{mol}_X\ \text{mol}_A^{-1}$ | 0.21[c] |
| $Y_{OA}$ | 2.0 | $\text{mol}_O\ \text{mol}_A^{-1}$ | 2.0[c,d] |
| $Y_{EA}^{o}$ | 4.67 | $\text{mol}_E\ \text{mol}_A^{-1}$ | 4.67[d] |
| Biological limitations | | | |
| $\mu_{\max}^{G}$ | 0.663 | $g_X.g_X^{-1}\ \text{h}^{-1}$ | $0.676_{\pm 0.038}$[g] |
| $\mu_{\max}^{A}$ | 0.032 | $g_X\ g_X^{-1}\ \text{h}^{-1}$ | $0.052_{\pm 0.028}$[g] |
| $\Phi_O^{\max}$ | 15.6 | $\text{mmol}_O\ g_X^{-1}\ \text{h}^{-1}$ | $15.3_{\pm 19^{c}}$<br>$20^{d,h}$ |
| Growth rate adaptation | | | |
| $T^{u}$ | 1.9 | $h$ | 1.9[i] |
| $T^{d}$ | 6.7[f] | $h$ | 1.9[i] |

[a] Values extracted from Xu's overall yield $Y_{X/S,of}$ and $Y_{X/S,ov}$ (Xu et al., 1999) coupled with our $Y_{EG}^{\varnothing/f/o}$ values.
[b] Russell and Cook (1995).
[c] Xu et al. (1999).
[d] Varma et al. (1993).
[e] Wang et al. (2010).
[f] Curve fitting.
[g] Extracted from Xu's $qA_{c,\max}$, $qS_{\max}$, $Y_{X/A}$ and $Y_{X/S,ox}$ (Xu et al., 1999).
[h] Meadows et al. (2010).
[i] Morchain and Fonade (2009).

that our modification constitutes a real improvement while preserving the original model capacities.

Results for batch culture simulations with both models, as well as experimental data, are shown in Fig. 4a,b,e,f. The initial concentrations used for simulations are detailed in Table 3. The initial biomass concentration, $X_0$, was not reported by Xu et al. (1999). After implementing the original model described by these authors, the initial biomass concentration could be identified through a curve fitting of the data. The value obtained was also used as the initial condition for our simulations.

We decided to initialize the population in the class corresponding to $\mu_j = 0.4\ \text{h}^{-1}$. This value is chosen as it gives the best fitting of experimental data. The reason for this choice is further detailed at the beginning of the discussion section.

The results presented in Fig. 4 were obtained using 100 classes in our population balance model. Unpublished data show that the simulation results are independent from the number of classes as long as this number is higher than 60.

Fig. 4a and b represents the concentrations of glucose and acetate over time. The substrate consumption is perfectly represented by both models, however those models differ in the way this substrate is used, leading to a difference in acetate production. When glucose is depleted, the acetate is re-consumed at a rate consistent with experimental observation. Our model provides a much better fit of the experimental data which is not the result of a parameter adjustment as will be discussed later on. Fig. 4e and f shows a good prediction of oxygen consumption as long as glucose is present. This prediction becomes less accurate as soon as this substrate is entirely consumed. In such a situation, some phenomena like endogenous metabolisms are known to participate in the oxygen consumption. Yet, such details are not implemented in our model, leading to those prediction errors.

The experimental concentrations have been post-processed to determine the experimental apparent growth rate $\tilde{\mu^a}$ and acetate variation rate $\tilde{r_A}$ as detailed in the following equations:

$$\tilde{\mu^a}\left(\frac{t_i+t_{i+1}}{2}\right) = \frac{X_{i+1}-X_i}{t_{i+1}-t_i}\left(\frac{X_i+X_{i+1}}{2}\right)^{-1} \tag{21}$$

$$\tilde{r_A}\left(\frac{t_i+t_{i+1}}{2}\right) = \frac{A_{i+1}-A_i}{t_{i+1}-t_i}\left(\frac{X_i+X_{i+1}}{2}\right)^{-1} \tag{22}$$

With $A$ and $X$ being the experimental concentrations of Acetate (in $g_A\ \text{L}^{-1}$) and Biomass (in $g_X\ \text{L}^{-1}$). The subscripts $i$ and $i+1$ refer to two consecutive samples and analysis.

As shown in Fig. 4b, our model predicts acetate formation almost perfectly compared to Xu et al's model. This is mainly explained by the way our model differentiates the two acetate origins: the fermentation, and the overflow metabolism.

At initial conditions, glucose and oxygen concentrations are high compared to their affinity constants, meaning that culture conditions are optimal (here, $\mu_0^* \approx 0.64\ h^{-1}$). Meanwhile, it is observed experimentally that the specific rate of acetate production is high too. However, Xu et al.'s model is not able to predict that high initial acetate production (Fig. 4d): only a constant production rate is predicted, and is correlated to an oxygen deficiency caused by a low value of the maximum oxygen uptake rate. On the opposite, our own model differentiate two phases in the acetate production. In our model the glucose uptake rate is correlated to $\mu^*$, which happens to be high initially. However, we initialized the biological population in a unique class such as $\tilde{\mu} = 0.4\ \text{h}^{-1}$. Then, for the first 2 h of the culture, a disequilibrium exists between the population and its environment ($\tilde{\mu} < \mu^*$) leading to acetate formation through the overflow metabolism (see $\tilde{r}_{Aover}$ in Fig. 5b). This production decreases along with the progressive adaptation of the population to the environment.

Around 2 h, a switch between the two acetate origins occurs: the high value of $\tilde{\mu^a}$ leads to high energetic needs, saturating the oxidative catabolism pathway. Then, a fraction of the energy is produced through fermentation.

It may be observed that the curve of acetate produced by the fermentation pathway ($A_{ferm}$ in Fig. 5a) looks alike the curve of acetate predicted by Xu et al.'s model (Fig. 4b). This tends to show that the so-called overflow metabolism used in their model consists in a fermentative metabolism. Indeed, their model links acetate production with energy and cell production as does our fermentation reaction ($R_3$).

Here, we chose to initialize the population with a Dirac distribution by concentrating all biomass in a unique class. However, many other distributions could have led to similar results such as, maybe, a Gaussian distribution centered around $\mu = 0.45\ \text{h}^{-1}$ or even a multimodal distribution. Here, we want to point out that there is a lack of experimental data about the state of the population in the inoculum used in biological cultures. As measurements are usually carried out at the population scale, hardly no data exists about the distribution of biological parameters even though such data appear more and more mandatory for the good comprehension of biological behaviors observed in bio-reactors (Dhar and McKinney, 2007).

The inhibition of respiration by acetate is underlined in Fig. 4e as the specific oxygen consumption rate $\tilde{\Phi}_O$ decreases significantly between 2 and 8 h. This inhibition strengthens the acetate production through fermentation by limiting the use of the oxidative pathway.

Those results allow to validate our biological model in a reactor in which hydrodynamics has no visible effect on the biological

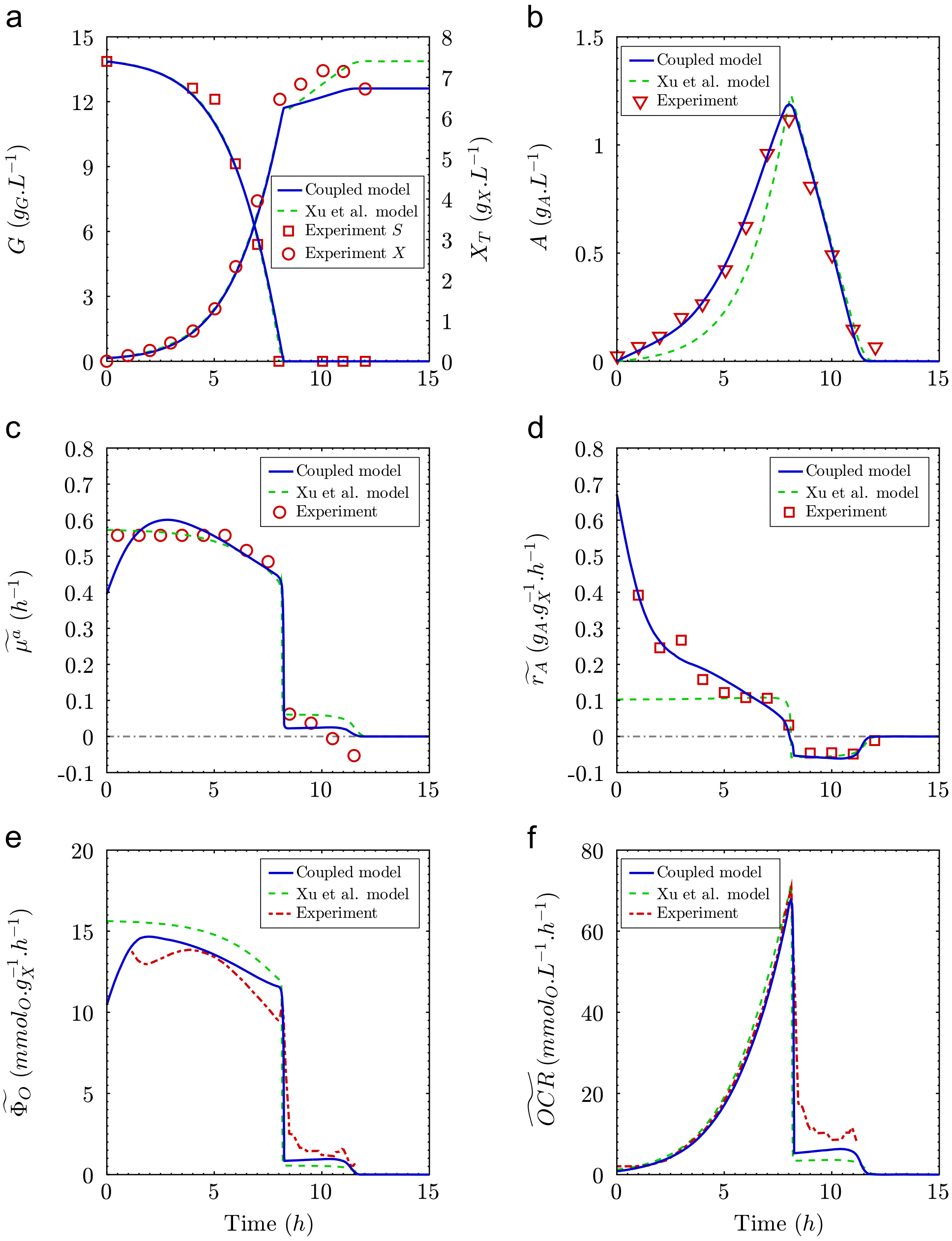


**Fig. 4.** Comparison of experimental data with simulation results in a batch reactor.

**Table 3**
Initial concentrations for batch culture.

| Symbol | Value | Unit |
|---|---|---|
| $X_0$ | 0.077 | $g_X\ \mathrm{L}^{-1}$ |
| $A_0$ | 0.00 | $g_A\ \mathrm{L}^{-1}$ |
| $G_0$ | 13.86 | $g_G\ \mathrm{L}^{-1}$ |
| $O_0$ | 2.70 | $mg_O\ \mathrm{L}^{-1}$ |

population. However, in large bioreactors, heterogeneities may appear under certain conditions depending on the quality of mixing and the intensity of biological reactions. Thus, the ability of our model to predict the complex transition between homogeneous and heterogeneous concentration fields has to be challenged against experimental data.

### 3.2. Fedbatch culture

The validation of our model in large bioreactors was performed using data provided by Xu et al. (1999a) on a fed-batch culture conducted in a 30 $\mathrm{m}^3$ reactor stirred with four impellers (Rushton turbines). This reactor has been described previously (Xu et al., 1999a; Larsson et al., 1996) and its matrix of volume flow rates, $M^f$, was computed using the compartment model proposed by Vrábel et al. (1999, 2000, 2001) with 70 compartments. The Appendix B details the calculation of $M^f$ for this reactor.

The validation of our implementation of this compartment model was performed through the simulation of a tracer pulse-response. Fig. 6 details the predicted relative concentration of the tracer on top, middle and bottom positions of the reactor by injecting the tracer in the top-stirrer compartment. These curves are very similar to those presented by Vrábel et al. (1999) which validates our implementation of the flow map at the reactor scale.

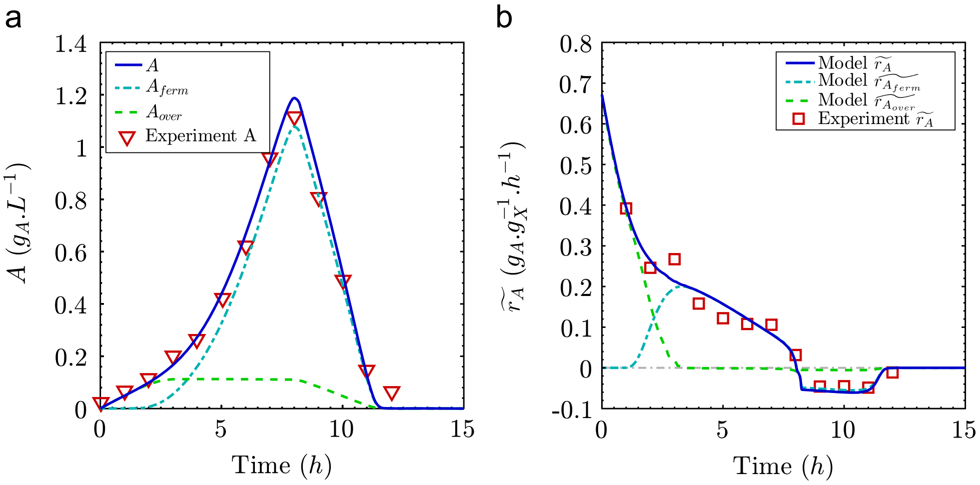


**Fig. 5.** Differentiation of acetate production pathway throughout the culture. $A_{\text{ferm}}$: acetate produced through fermentative catabolism, $A_{\text{over}}$: acetate produced by overflow metabolism.

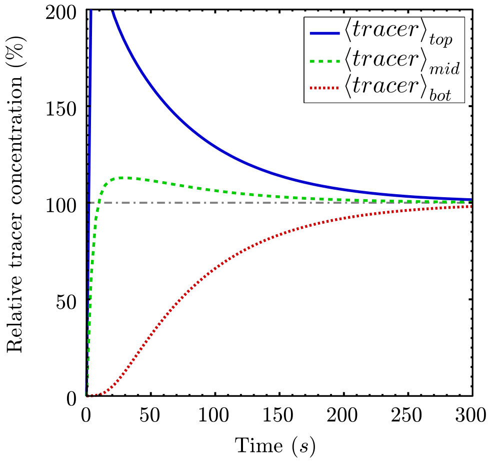


**Fig. 6.** Pulse-response monitoring after tracer injection in the feeding compartment.

**Table 4**
Fed-batch culture—initial conditions.

| | Value | Unit |
|---|---|---|
| Volume $V_0$ | 22 | m$^3$ |
| Biomass $X_0$ | 0.10 | $g_X$ L$^{-1}$ |
| Glucose $G_0$ | 0.029 | $g_G$ L$^{-1}$ |
| Acetate $A_0$ | 0.05 | $g_A$ L$^{-1}$ |

We also deduce from this tracer experiment that the overall mixing time in this reactor is about $t_{mix} \approx 250$ s.

In the following, the notation $\langle\rangle$ describes the volumetric average of variables. By default, the volumetric average is taken over the entire reactor volume, but a subscript may be used to specify a sub-part of the reactor (top, middle or bottom). The Fig. 7 compares the results from our simulation, the experimental data (Xu et al., 1999a) and the curves coming from Xu et al.'s simulations using Xu et al.'s model (Vrábel et al., 2001). Initial conditions are reported in Table 4. The biomass was initially put in the class corresponding to $\mu_j = 0.63$ h$^{-1}$ as this choice leads to the best curve fitting. As for the batch culture, consequences of this choice are detailed at the beginning of the discussion.

It is important to note that the simulation of the fedbatch culture was carried out without further adjustments of our model parameters compared to the batch culture simulation.

Fig. 7a details the evolution of glucose concentrations on top, middle and bottom positions. It should be noticed that the way Vrábel et al. (2001) modeled the feed leads to an overestimation of glucose concentration during the first hours of culture (Fig. 7a) which is not consistent with experimental measurements. Nevertheless, the same sudden shift of the feed flow rate at $t = 0.8$ h was imposed in our simulation in order to allow the assessment of the results. Both models give very similar results regarding the glucose concentration. It is likely that a modification of the affinity constant for glucose could produce better agreement between the experimental data and the simulation results. However this does not constitute an objective for this work which is more focused on the minimal structure of a two-way coupling approach for the simulation of heterogeneous bioreactors. It will appear clearly in the next paragraphs that the benefits of the coupled approach is more related to the structure of the model than to the accuracy of the parameter identification procedure.

The first observation of interest is on the total biomass concentration $\langle X_T \rangle$ given in Fig. 7b. The mixing-time, $t_{mix} \approx 250$ s, is negligible compared to the characteristic time of biological growth (Morchain et al., 2014), then the total biomass concentration is independent of the position within the reactor. Vrábel et al.'s simulation over-predicted the biomass production and an unjustified adjustment of the conversion yield $Y_{XG}$ by a factor as large as 0.76 was necessary to fit experimental data. Meanwhile, our model correctly predicts this production by taking into account the known phenomena of yield decrease at low growth rate using the Pirt's formulation (Pirt, 1965) as explained in Appendix A.

The next remarkable point is related to the acetate production. The acetate concentration is presented in Fig. 7c and its specific variation rate is given in Fig. 7d. In that case a noticeable difference between the model predictions appears. Whereas the model of Xu et al. predicted no residual acetate, our coupled model qualitatively predicts the transient accumulation before 10 h and the formation of a spatial gradient as the biomass concentration increases.

As shown in Fig. 7d, the acetate is produced everywhere in the reactor before 7 h. After that, it appears that acetate is produced in the upper part of the reactor and is consumed in the lower part of the reactor where the glucose concentration is very low. Moreover the data in Fig. 7d also allow the identification of the mechanisms involved. The thick dotted-dashed line corresponds to the

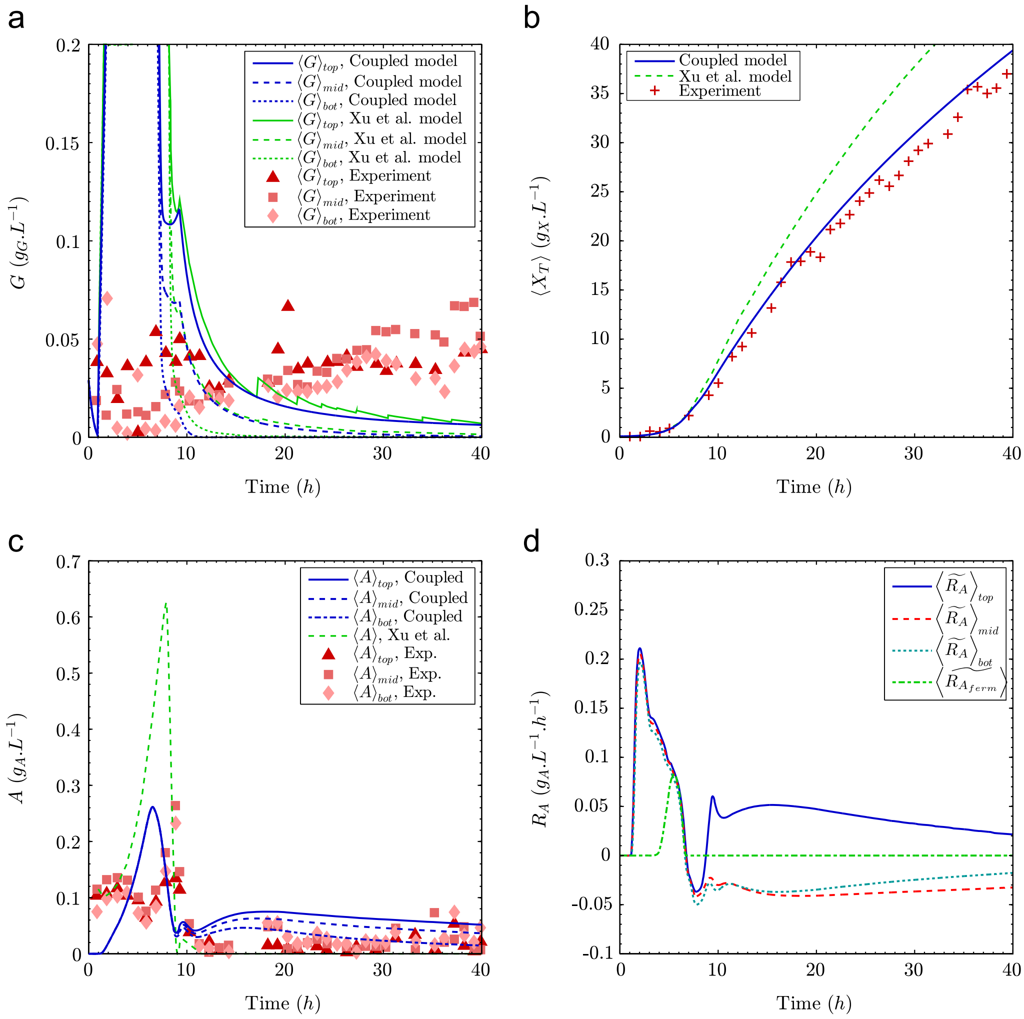


**Fig. 7.** Experimental data and simulation results in the 30 m$^3$ fedbatch reactor.

production through fermentation. The fermentation metabolism is only triggered between 4 and 7 h leading to the acetate peak observable at the same time in Fig. 7c. This demonstrates that, apart from that period of time, acetate is produced through overflow metabolism.

## 4. Discussion

Our modeling approach, based on population balance model, is clearly dependent on the initial conditions, obviously $X$ and $G$ concentrations, but also the initial growth rate distribution whose choice may have a real impact on the simulation results. This is a common feature of biological systems which has been discarded in most previous modeling approaches. From that initial distribution will depend the global kinetic of the culture, but also the initial disequilibrium between the population and its environment which leads to different metabolic behaviors.

In batch and fedbatch cultures, we decided to concentrate all the biological population in one class. The choice of that class was made in order to fit the results with experimental data.

For the batch culture, the initial population growth rate imposed in the simulation ($\mu_j = 0.4\ \mathrm{h}^{-1}$) is different from the experimental one reported by Xu et al. (1999). Based on their experimental measurements, Xu et al. (1999) estimated the initial growth rate around $\mu = 0.56\ \mathrm{h}^{-1}$. However, since initial biomass concentrations are negligible, this estimation of the initial growth rate may be flawed and we estimate, by an error analysis, that the actual growth rate is lies between 0.3 h$^{-1}$ and 0.8 h$^{-1}$.

With our model, the choice of the initial distribution has two main consequences: the first one is about the initial disequilibrium between the biological population and the environment. Actually, the value of $\mu$ returned by Xu et al.'s model is very similar to what we call the equilibrium growth rate, $\mu^*$. Then, by initializing the entire population at $\mu_j = 0.4\ \mathrm{h}^{-1}$, we artificially create the initial disequilibrium that explains, with our model, the high value of $\tilde{r_A}$. Many different initial distributions could produce similar results in terms of $\tilde{r_A}$. The fact is that the experimental information on the actual distribution of growth rate at $t=0$ is not available.

The second consequence is about the global kinetic of the culture. If we initialized the biomass at $\mu_j = 0.2\ \mathrm{h}^{-1}$, we would have increased the disequilibrium, leading to an higher production of acetate, but we also would have add a lag in the biomass growth (glucose consumption would have been slower). We choose the value of $\mu_j = 0.4\ \mathrm{h}^{-1}$ as it was the best compromise between the global kinetic and the acetate production.

The same comment applies to the initial distribution of fedbatch culture. Moreover, in continuous culture (e.g. fedbatch or chemostat), the medium is continuously renewed and the duration of the culture exceeds the time scale of biological growth-rate adaptation. Then, after a few hours, the population tends to “forget” its initial state. For example, in our fedbatch simulation, the choice of the initial distribution only has an effect for the first hours of the simulation: it is found that the initial conditions are forgotten when the culture enters the “mixing/reaction competition zone” after 7 h. So it has no fundamental impact on the observations made when the reactor becomes spatially heterogeneous.

On the results of the models, the main difference between ours and Xu et al.'s one is shown on fedbatch acetate curves. Indeed, their model was not able to predict the acetate residual concentration that is experimentally observed after 10 h of fedbatch culture while ours predicts the apparition of an acetate gradient over the reactor.

To understand this capability of our model, one must distinguish two different phases in the culture. For the 7 first hours of culture, the reactor appears homogeneous while, after 10 h, vertical gradients appear for glucose and acetate. It is worth investigating more specifically what happens during the transitional phase between 7 h and 10 h. This can be done through the time scale analysis of the various processes involved.

In a previous paper (Morchain et al., 2014), we introduced the time scale of the substrate assimilation, $t_S$:

$$t_G = \frac{\langle G \rangle}{\langle R_G \rangle} \approx \frac{M_X Y_{XG} \langle G \rangle}{M_G \mu_{max} \langle X_T \rangle} \tag{23}$$

Fig. 8 shows the evolution of this assimilation time-scale compared to the mixing time. It can be seen that a switch occurs around 7 h with the assimilation time becoming smaller than the mixing time. This means that at the beginning of the culture, the glucose poured into the reactor is quickly mixed compared to the reaction rate, leading to an homogeneous reactor. Then a regime switch occurs, the substrate consumption gets faster (due to an increasing biomass concentration) and substrate is now consumed before reaching the bottom of the reactor, leading to the observed gradients.

A gradient on acetate concentration appears around 2 h after the glucose gradient. This delay of 2 h results from the dynamic adaptation of the population (Eq. (4)). This clearly illustrates the of dynamic responses at the reactor scale: the acetate gradient is a consequence of the biomass adaptation to the heterogeneous glucose concentration field induced by a mixing limitation.

Those different situations may be visualized and understood in Figs. 9 and 10. It is shown that around 8 h the reactor changes from a homogeneous to a heterogeneous state. After the transitional phase, glucose is mainly present around the feeding point while a major fraction of the reactor has little or no substrate as shown in Fig. 9 top. This leads to the strong spatial variation of the equilibrium growth rate (Fig. 9 bottom). In our model, the uptake rate is closely related to local concentrations whereas the utilization rate for growth is determined by the volume average concentration. This is consistent with the fact that the adaptation of uptake systems is rather fast compared to that of the growth rate. This difference in the response dynamics is thought to be responsible for the cell disequilibrium.

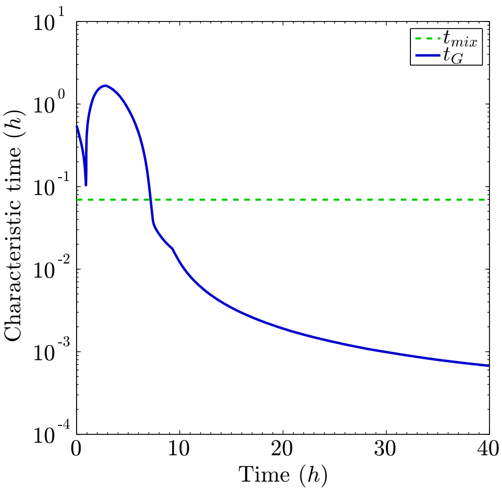


**Fig. 8.** Comparison of mixing and substrate assimilation time scales during the fedbatch culture.

As explained previously, the overflow metabolism is triggered when the biomass is in a rich-in-substrate environment while being adapted to a limiting environment. This is what happens during the heterogeneous phase: the biomass is accustomed to a mean concentration but happens to be in a high concentration zone around the feeding point, and in low concentration zones elsewhere, leading to a permanent disequilibrium (Fig. 10 top). This unbalanced population reacts by producing acetate through overflow in the feeding area. This acetate is afterward transported in the areas with low glucose concentrations and is consumed there by the starving biomass, leading to the acetate gradient as underlined in Fig. 10 bottom.

Furthermore, it is interesting to observe that, the time constant of growth rate adaptation being larger than the circulation time, the mean biological growth rate $\tilde{\mu}$ is related to the volume average concentration of glucose $\langle G \rangle$. Thus, as already demonstrated in a previous work, the population distribution in the $\mu$-space is independent of the location in the reactor, but the difference between the population state and the local environment is spatially dependent (Morchain et al., 2014). Although the time constant related to the specific growth rate adaptation is much larger than the mixing time, the repeated exposure to high and low concentrations over hours has a significant effect on the width of specific growth rate distribution of the population. Indeed, Fig. 11 shows the evolution of the population distribution over classes compared with the equilibrium growth rate in different areas of the reactor. The width of the distribution is related to the width of the colored zone around the volumetric mean. In this figure $\beta$ is defined by

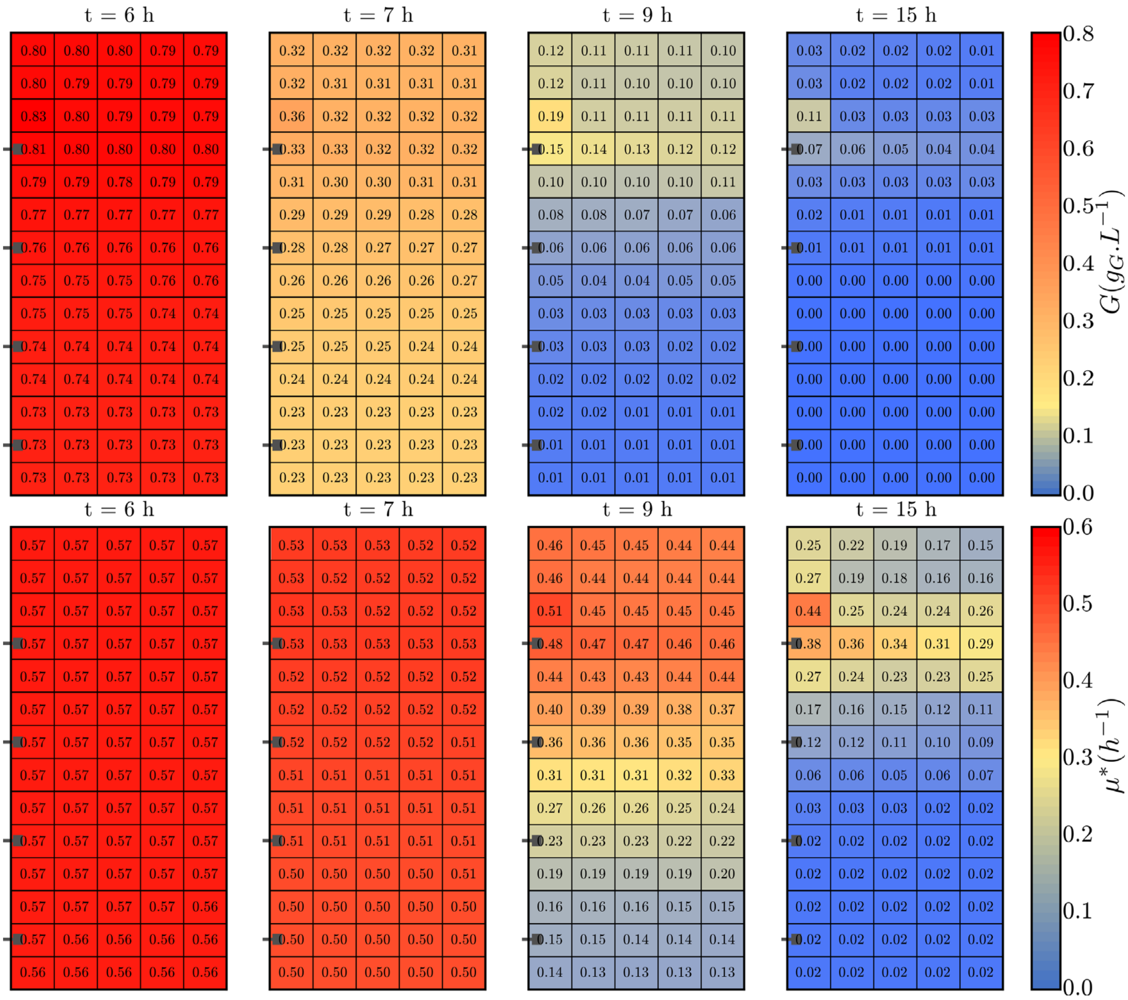
$$\beta(t, \mu_j) = \frac{X_j(t)}{\max_{j=1}^{J}(X_j(t))} \tag{24}$$

It is indicative of the population growth rate distribution width (note that only values higher than 5% are shown). Thus it is visible that the population is always distributed (and almost centered) around the mean volumetric growth rate.

It may be seen that after 10 h, when a major volumetric fraction of this reactor has low glucose concentrations, the population gets used to these low concentrations by moving to lowest classes. Then, this population is no longer capable to use the glucose efficiently when being in the feeding area, leading to these high local acetate productions.

Such phenomena may only be predicted by considering the disequilibrium between the substrate consumption and the cell needs. The quantification of this disequilibrium relies upon the use of a population balance model to describe the dynamic adaptation of the growth rate.

In the present work, we considered in our modeling that biological heterogeneity in terms of growth rate distribution is induced by the environment (exogenous heterogeneity). In a chemostat, this characteristic of the model ensures that the population growth rate will tend, in the long term, to the dilution rate. In a constant environment, all cells will be distributed in the two classes surrounding $\mu^*$ as shown in Morchain et al. (2013). However, recent experimental data indicate that an interdivision time distribution exists in a chemostat (Nobs and Maerkl, 2014). It implies that endogenous source of heterogeneity is also present.



t = 6 h

| | | | | |
|---|---|---|---|---|
| 0.80 | 0.80 | 0.80 | 0.79 | 0.79 |
| 0.80 | 0.79 | 0.79 | 0.79 | 0.79 |
| 0.83 | 0.80 | 0.79 | 0.79 | 0.79 |
| 0.81 | 0.80 | 0.80 | 0.80 | 0.80 |
| 0.79 | 0.79 | 0.78 | 0.79 | 0.79 |
| 0.77 | 0.77 | 0.77 | 0.77 | 0.77 |
| 0.76 | 0.76 | 0.76 | 0.76 | 0.76 |
| 0.75 | 0.75 | 0.75 | 0.76 | 0.76 |
| 0.75 | 0.75 | 0.75 | 0.74 | 0.74 |
| 0.74 | 0.74 | 0.74 | 0.74 | 0.74 |
| 0.74 | 0.74 | 0.74 | 0.74 | 0.74 |
| 0.73 | 0.73 | 0.73 | 0.73 | 0.73 |
| 0.73 | 0.73 | 0.73 | 0.73 | 0.73 |
| 0.73 | 0.73 | 0.73 | 0.73 | 0.73 |

t = 7 h

| | | | | |
|---|---|---|---|---|
| 0.32 | 0.32 | 0.32 | 0.32 | 0.31 |
| 0.32 | 0.31 | 0.31 | 0.31 | 0.31 |
| 0.36 | 0.32 | 0.32 | 0.32 | 0.32 |
| 0.33 | 0.33 | 0.32 | 0.32 | 0.32 |
| 0.31 | 0.30 | 0.30 | 0.31 | 0.31 |
| 0.29 | 0.29 | 0.29 | 0.28 | 0.28 |
| 0.28 | 0.28 | 0.27 | 0.27 | 0.27 |
| 0.26 | 0.26 | 0.26 | 0.26 | 0.27 |
| 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| 0.25 | 0.25 | 0.25 | 0.24 | 0.24 |
| 0.24 | 0.24 | 0.24 | 0.24 | 0.24 |
| 0.23 | 0.23 | 0.23 | 0.23 | 0.23 |
| 0.23 | 0.23 | 0.23 | 0.23 | 0.23 |
| 0.23 | 0.23 | 0.23 | 0.23 | 0.23 |

t = 9 h

| | | | | |
|---|---|---|---|---|
| 0.12 | 0.11 | 0.11 | 0.11 | 0.10 |
| 0.12 | 0.11 | 0.10 | 0.10 | 0.10 |
| 0.19 | 0.11 | 0.11 | 0.11 | 0.11 |
| 0.15 | 0.14 | 0.13 | 0.12 | 0.12 |
| 0.10 | 0.10 | 0.10 | 0.10 | 0.11 |
| 0.08 | 0.08 | 0.07 | 0.07 | 0.06 |
| 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
| 0.05 | 0.04 | 0.04 | 0.05 | 0.05 |
| 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| 0.03 | 0.03 | 0.03 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.01 | 0.01 | 0.01 |
| 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |

t = 15 h

| | | | | |
|---|---|---|---|---|
| 0.03 | 0.02 | 0.02 | 0.02 | 0.01 |
| 0.03 | 0.02 | 0.02 | 0.02 | 0.01 |
| 0.11 | 0.03 | 0.03 | 0.03 | 0.03 |
| 0.07 | 0.06 | 0.05 | 0.04 | 0.04 |
| 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| 0.02 | 0.01 | 0.01 | 0.01 | 0.01 |
| 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

t = 6 h

| | | | | |
|---|---|---|---|---|
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.57 |
| 0.57 | 0.57 | 0.57 | 0.57 | 0.56 |
| 0.57 | 0.56 | 0.56 | 0.56 | 0.56 |
| 0.56 | 0.56 | 0.56 | 0.56 | 0.56 |

t = 7 h

| | | | | |
|---|---|---|---|---|
| 0.53 | 0.53 | 0.53 | 0.52 | 0.52 |
| 0.53 | 0.52 | 0.52 | 0.52 | 0.52 |
| 0.53 | 0.53 | 0.52 | 0.52 | 0.52 |
| 0.53 | 0.53 | 0.53 | 0.53 | 0.53 |
| 0.52 | 0.52 | 0.52 | 0.52 | 0.52 |
| 0.52 | 0.52 | 0.52 | 0.52 | 0.52 |
| 0.52 | 0.52 | 0.52 | 0.52 | 0.51 |
| 0.51 | 0.51 | 0.51 | 0.51 | 0.51 |
| 0.51 | 0.51 | 0.51 | 0.51 | 0.51 |
| 0.51 | 0.51 | 0.51 | 0.51 | 0.51 |
| 0.50 | 0.50 | 0.50 | 0.50 | 0.51 |
| 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |

t = 9 h

| | | | | |
|---|---|---|---|---|
| 0.46 | 0.45 | 0.45 | 0.44 | 0.44 |
| 0.46 | 0.44 | 0.44 | 0.44 | 0.44 |
| 0.51 | 0.45 | 0.45 | 0.45 | 0.45 |
| 0.48 | 0.47 | 0.47 | 0.46 | 0.46 |
| 0.44 | 0.43 | 0.43 | 0.44 | 0.44 |
| 0.40 | 0.39 | 0.39 | 0.38 | 0.37 |
| 0.36 | 0.36 | 0.36 | 0.35 | 0.35 |
| 0.31 | 0.31 | 0.31 | 0.32 | 0.33 |
| 0.27 | 0.26 | 0.26 | 0.25 | 0.24 |
| 0.23 | 0.23 | 0.23 | 0.22 | 0.22 |
| 0.19 | 0.19 | 0.19 | 0.19 | 0.20 |
| 0.16 | 0.16 | 0.16 | 0.15 | 0.15 |
| 0.15 | 0.15 | 0.14 | 0.14 | 0.14 |
| 0.14 | 0.13 | 0.13 | 0.13 | 0.13 |

t = 15 h

| | | | | |
|---|---|---|---|---|
| 0.25 | 0.22 | 0.19 | 0.17 | 0.15 |
| 0.27 | 0.19 | 0.18 | 0.16 | 0.16 |
| 0.44 | 0.25 | 0.24 | 0.24 | 0.26 |
| 0.38 | 0.36 | 0.34 | 0.31 | 0.29 |
| 0.27 | 0.24 | 0.23 | 0.23 | 0.25 |
| 0.17 | 0.16 | 0.15 | 0.12 | 0.11 |
| 0.12 | 0.12 | 0.11 | 0.10 | 0.09 |
| 0.06 | 0.06 | 0.05 | 0.06 | 0.07 |
| 0.03 | 0.03 | 0.03 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |

**Fig. 9.** Glucose concentration field (top) and related equilibrium growth rates (bottom) in the 30 $m^3$ reactor during the cultivation. The grid corresponds to the 70 compartments of the hydrodynamic model. The numerical data are local instantaneous values.

Taking this fact into account in the modeling constitutes a potential way for future improvements. Still, the current formulation proposed in this work remains meaningful in a large scale bioreactor in which concentration gradients are present and play a dominant role in the production of heterogeneity within the population.

## 5. Conclusions

In this work, we focused on the dynamic simulation of bioreactors. Due to the strong two-way coupling between the biological behavior and the spatial and temporal variations of concentration fields, we introduced a modeling using a metabolic and a population balance models. It leads to the description of a possible disequilibrium between the biomass and its environment, interpreted as an imbalance between a cell needs and its actual glucose consumption. This modelization has been applied to the well known bacteria *Escherichia coli* in order to challenge our model with data extracted from the literature.

We integrated this biological model with a compartment model describing the hydrodynamics of two reactors of different scales: (i) a 15 L homogeneous reactor and (ii) a 30 $m^3$ stirred reactor. Our model was consistent enough to allow good predictions of experimental measurements at two reactor scales without modifying its parameters. Thus, we shown that the coupling of a metabolic and a population balance models strongly improves the prediction of acetate production, as well as the overall biological behavior, compared to the standalone metabolic model. In particular, we observed that the formation of a glucose gradient necessarily induces the production and the consumption of acetate in different zones of the reactor. Such phenomena could not be predicted without introducing an unbalanced biomass in our modeling. So it is believed that these concepts leading to a two-way coupled approach is a promising way to address the issue of simulating industrial bioreactors.

## Notation

Roman

| | |
|---|---|
| $A$ | acetate |
| $E$ | energy, molecule of ATP |
| $G$ | glucose |
| $J$ | number of classes |
| $K$ | affinity or inhibition biological constant ($g\,L^{-1}$) |
| $M$ | molar mass ($g\,mol^{-1}$) |
| $m$ | maintenance rate ($mol_G\,g_X^{-1}.h^{-1}$) |
| $M^f$ | matrix of volume flow rates |

| | |
|---|---|
| $N$ | number of compartments |
| $O$ | oxygen |
| $q$ | specific reaction rate (mol $g_X^{-1}$ $h^{-1}$) |
| $R$ | volumetric reaction rate (g $L^{-1}$ $h^{-1}$) |
| $r$ | specific reaction rate (g $g_X^{-1}$ $h^{-1}$) |
| $T$ | volumetric transfer rate (g $L^{-1}$ $h^{-1}$) |
| $T^d$ | time constant for downward transfer between classes (h) |
| $T^u$ | time constant for upward transfer between classes (h) |
| $t$ | time (h) |
| $X$ | cells, biomass |
| $Y$ | stoichiometric molar coefficient (mol $mol^{-1}$) |

*Subscript and superscript*

| | |
|---|---|
| $*$ | variable describing a cell at equilibrium with its environment |
| 0 | initial conditions |
| $a$ | actual or achieved |
| $f$ | fermentative reaction |
| $i$ | generic notation for any species ($A$, $E$, $G$, $O$ or $X$) |
| $j$ | class index |
| $m$ | compartment index |
| max | maximum value of a biological constant |
| $n$ | compartment index |
| $o$ | oxidative reaction |
| $T$ | total over the biological population |

*Greek*

| | |
|---|---|
| $\alpha$ | generic notation for any biological reaction |
| $\Phi$ | biological uptake rate (mol $g_X^{-1}$ $h^{-1}$) |
| $\mu$ | specific growth rate of a cell ($h^{-1}$) |
| $\zeta$ | rate of change of specific growth rate ($h^{-2}$) |

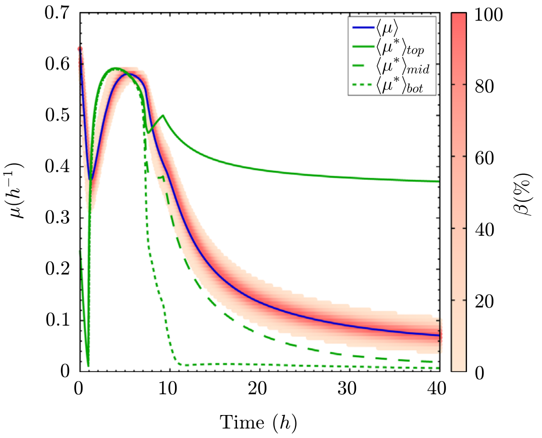
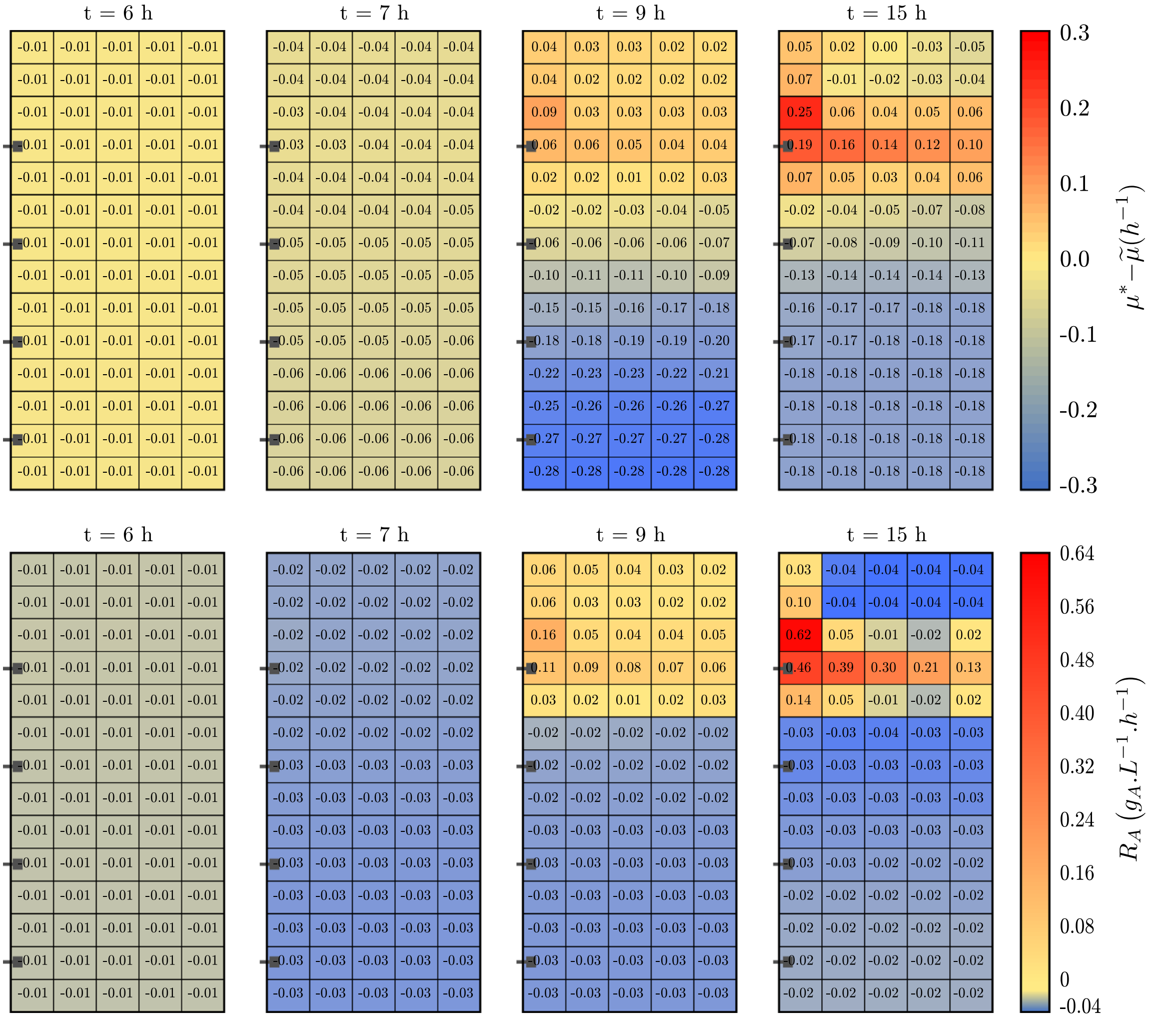


**Fig. 11.** Evolution of biological population width, and equilibrium growth rates at different locations in the reactor.



t = 6 h

| | | | | |
|---|---|---|---|---|
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |

t = 7 h

| | | | | |
|---|---|---|---|---|
| -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| -0.03 | -0.04 | -0.04 | -0.04 | -0.04 |
| -0.03 | -0.03 | -0.04 | -0.04 | -0.04 |
| -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| -0.04 | -0.04 | -0.04 | -0.04 | -0.05 |
| -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| -0.05 | -0.05 | -0.05 | -0.05 | -0.06 |
| -0.06 | -0.06 | -0.06 | -0.06 | -0.06 |
| -0.06 | -0.06 | -0.06 | -0.06 | -0.06 |
| -0.06 | -0.06 | -0.06 | -0.06 | -0.06 |
| -0.06 | -0.06 | -0.06 | -0.06 | -0.06 |

t = 9 h

| | | | | |
|---|---|---|---|---|
| 0.04 | 0.03 | 0.03 | 0.02 | 0.02 |
| 0.04 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.09 | 0.03 | 0.03 | 0.03 | 0.03 |
| 0.06 | 0.06 | 0.05 | 0.04 | 0.04 |
| 0.02 | 0.02 | 0.01 | 0.02 | 0.03 |
| -0.02 | -0.02 | -0.03 | -0.04 | -0.05 |
| -0.06 | -0.06 | -0.06 | -0.06 | -0.07 |
| -0.10 | -0.11 | -0.11 | -0.10 | -0.09 |
| -0.15 | -0.15 | -0.16 | -0.17 | -0.18 |
| -0.18 | -0.18 | -0.19 | -0.19 | -0.20 |
| -0.22 | -0.23 | -0.23 | -0.22 | -0.21 |
| -0.25 | -0.26 | -0.26 | -0.26 | -0.27 |
| -0.27 | -0.27 | -0.27 | -0.27 | -0.28 |
| -0.28 | -0.28 | -0.28 | -0.28 | -0.28 |

t = 15 h

| | | | | |
|---|---|---|---|---|
| 0.05 | 0.02 | 0.00 | -0.03 | -0.05 |
| 0.07 | -0.01 | -0.02 | -0.03 | -0.04 |
| 0.25 | 0.06 | 0.04 | 0.05 | 0.06 |
| 0.19 | 0.16 | 0.14 | 0.12 | 0.10 |
| 0.07 | 0.05 | 0.03 | 0.04 | 0.06 |
| -0.02 | -0.04 | -0.05 | -0.07 | -0.08 |
| -0.07 | -0.08 | -0.09 | -0.10 | -0.11 |
| -0.13 | -0.14 | -0.14 | -0.14 | -0.13 |
| -0.16 | -0.17 | -0.17 | -0.18 | -0.18 |
| -0.17 | -0.17 | -0.18 | -0.18 | -0.18 |
| -0.18 | -0.18 | -0.18 | -0.18 | -0.18 |
| -0.18 | -0.18 | -0.18 | -0.18 | -0.18 |
| -0.18 | -0.18 | -0.18 | -0.18 | -0.18 |
| -0.18 | -0.18 | -0.18 | -0.18 | -0.18 |

t = 6 h

| | | | | |
|---|---|---|---|---|
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |
| -0.01 | -0.01 | -0.01 | -0.01 | -0.01 |

t = 7 h

| | | | | |
|---|---|---|---|---|
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |

t = 9 h

| | | | | |
|---|---|---|---|---|
| 0.06 | 0.05 | 0.04 | 0.03 | 0.02 |
| 0.06 | 0.03 | 0.03 | 0.02 | 0.02 |
| 0.16 | 0.05 | 0.04 | 0.04 | 0.05 |
| 0.11 | 0.09 | 0.08 | 0.07 | 0.06 |
| 0.03 | 0.02 | 0.01 | 0.02 | 0.03 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |

t = 15 h

| | | | | |
|---|---|---|---|---|
| 0.03 | -0.04 | -0.04 | -0.04 | -0.04 |
| 0.10 | -0.04 | -0.04 | -0.04 | -0.04 |
| 0.62 | 0.05 | -0.01 | -0.02 | 0.02 |
| 0.46 | 0.39 | 0.30 | 0.21 | 0.13 |
| 0.14 | 0.05 | -0.01 | -0.02 | 0.02 |
| -0.03 | -0.03 | -0.04 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| -0.03 | -0.03 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |
| -0.02 | -0.02 | -0.02 | -0.02 | -0.02 |

**Fig. 10.** Overall population disequilibrium expressed as $\mu^* - \tilde{\mu}$ (top) and acetate production/consumption rates (bottom). The grid corresponds to the 70 compartments of the hydrodynamic model. The numerical data are local instantaneous values.

## Acknowledgments

The authors would like to thank Toulouse White Biotechnology UMS INRA 1337, UMS CNRS 3582 for its financial support. The authors declare no conflict of interest.

## Appendix A. Metabolic reaction rates calculation

This appendix details explicitly the calculation procedure of biological reaction rates. It presents a method to calculate these reactions rates, starting from the cell biological growth rate $\mu_j$ and the liquid-phase concentrations (in glucose ($G$), acetate ($A$) and oxygen ($O$)). In its principle, this method is similar to that proposed by Xu et al. (1999): it consists in a set of tests (hierarchic method). This procedure defines the rates of the following reactions:

$$G + Y_{EG}E \xrightarrow{q_{ana}^{G}} Y_{XG}X \tag{R1}$$

$$A + Y_{EA}E \xrightarrow{q_{ana}^{A}} Y_{XA}X \tag{R1$'$}$$

$$G + Y_{OG}O \xrightarrow{q_{oxy}^{G}} Y_{EG}^{o}E \tag{R2}$$

$$A + Y_{OA}O \xrightarrow{q_{oxy}^{A}} Y_{EA}^{o}E \tag{R2$'$}$$

$$G \xrightarrow{q_{ferm}^{G}} Y_{EG}^{f}E + Y_{AG}A \tag{R3}$$

This estimation is here possible by assuming an energetic balance within a cell: the rate of ATP production, through catabolism pathways, is strictly equal to the rate of ATP consumption through anabolism:

$$Y_{EG}^{o}q_{oxy}^{G} + Y_{EG}^{f}q_{ferm}^{G} = Y_{EG}q_{ana}^{G} \tag{A.1}$$

$$Y_{EA}^{o}q_{oxy}^{A} = Y_{EA}q_{ana}^{A} \tag{A.2}$$

First, we must evaluate the actual growth rate of the considered cell. We consider that growth may be limited either by the biological growth rate (internal limitation), or by the environment concentrations (external limitation). Those concentrations allow to compute an environment permitted (or equilibrium) growth-rate $\mu^*$ :

$$\mu^{G*} = \mu_{max}^{G}\frac{G}{G+K_G}\frac{O}{O+K_O}\frac{K_{i,A}}{A+K_{i,A}} \tag{A.3}$$

$$\mu^{A*} = \mu_{max}^{A}\frac{A}{A+K_A}\frac{O}{O+K_O}\frac{K_{i,G}}{G+K_{i,G}} \tag{A.4}$$

$$\mu^{*} = \mu^{G*} + \mu^{A*} \tag{A.5}$$

Then, the actual growth rate is defined as the minimum between the biological, and the equilibrium, growth rates. We define an actual growth rate for glucose-based growth ($\mu_j^{a,G}$), and an other one for acetate-based growth ($\mu_j^{a,A}$). The sum of them leads to the actual growth rate of the cell:

$$\mu_j^{a,G} = \min\left(\mu_j, \mu^{G*}\right) \tag{A.6}$$

$$\mu_j^{a,A} = \min\left(\mu_j - \mu_j^{a,G}, \mu^{A*}\right) \tag{A.7}$$

$$\mu_j^{a} = \mu_j^{a,G} + \mu_j^{a,A} = \min\left(\mu_j, \mu^{*}\right) \tag{A.8}$$

*Anabolism*

The anabolism reaction rates are directly deduced from the actual growth rates, on $G$ (Eq. (A.10)) and $A$ (Eq. (A.11)), of the $j$-th class. The conversion yield of glucose in biomass, $Y_{XG}$, is computed using the Pirt's formulation (Pirt, 1965). This formulation, given in Eq. (A.9), reflects the fact that maintenance operations take a major role in substrate consumption at low growth rate while having a negligible effect at high growth rate (Russell and Cook, 1995):

$$\frac{1}{Y_{XG}} = \frac{m \times M_X}{\mu_j^{a}} + \frac{1}{Y_{XG}^{max}} \tag{A.9}$$

$m$ (in $mol_G\, g_X^{-1}\, h^{-1}$) is the maintenance rate, representing the amount of substrate needed for maintenance. In the present work, $m$ was kept constant at 250 $\mu$mol$_G$ g$_X^{-1}$ $h^{-1}$, but for a more detailed model it has been suggested that this rate is partially linked to the growth rate (Holms, 1996; Meadows et al., 2010).

$$q_{ana,j}^{G} = \frac{\mu_j^{a,G}}{Y_{XG}M_X} \tag{A.10}$$

$$q_{ana,j}^{A} = \frac{\mu_j^{a,A}}{Y_{XA}M_X} \tag{A.11}$$

Those rates represent a cell-production rate that will effectively be achieved. But to allow this cell production, energy must be produced as well. We then must estimate the rate of energy production through the oxidative pathway, and if needed, through the fermentative pathway.

*Catabolism*

As the oxidative pathway provides a better energetic yield, we make the assumption that the energy needed for anabolism will be only produced through that pathway if possible. The fermentative pathway will only be triggered if there is an oxygen deficiency.

First, lets estimate the amount of oxygen that would be needed (upper-script n) to produce all the energy needed. In that case, the oxidative catabolism reaction rates would be

$$q_{oxy,j}^{G,n} = \frac{Y_{EG}q_{ana,j}^{G}}{Y_{EG}^{o}} \tag{A.12}$$

$$q_{oxy,j}^{A,n} = \frac{Y_{EA}q_{ana,j}^{A}}{Y_{EA}^{o}} \tag{A.13}$$

This leads to a needed oxygen consumption rate, expressed in mol$_O$ g$_X^{-1}$ h$^{-1}$, in Eq. (A.16).

$$\Phi_{O,j}^{G,n} = Y_{OG}q_{oxy,j}^{G,n} \tag{A.14}$$

$$\Phi_{O,j}^{A,n} = Y_{OA}q_{oxy,j}^{A,n} \tag{A.15}$$

$$\Phi_{O,j}^{n} = \Phi_{O,j}^{G,n} + \Phi_{O,j}^{A,n} \tag{A.16}$$

Beside of the oxygen need, we estimate the oxygen availability which does not depend on the considered class and relies on liquid phase concentrations. This maximum possible oxygen consumption rate (upper-script $p$) is assumed to follow a Monod model. This model implements an inhibition of oxygen uptake rate by acetate as observed by Xu et al. (1999):

$$\Phi_{O}^{p} = \Phi_{O}^{\max}\frac{O}{O+K_O}\frac{K_{i,A}^{O}}{A+K_{i,A}^{O}} \tag{A.17}$$

If the oxygen availability excess the cell needs ($\Phi_O^p \geq \Phi_{O,j}^n$), then all energy is produced by oxidative pathway:

$$q_{oxy,j}^{G} = q_{oxy,j}^{G,n} \tag{A.18}$$

$$q_{oxy,j}^{A} = q_{oxy,j}^{A,n} \tag{A.19}$$

$$q_{ferm,j}^{G} = 0 \tag{A.20}$$

Otherwise, if oxygen is a limiting factor ($\Phi_O^p < \Phi_{O,j}^n$), the fermentative catabolism is triggered to fulfill energetic needs. Then, the fermentation reaction rate is deduced from the energy balance (Eq. (A.1)):

$$q_{oxy,j}^G = q_{oxy,j}^{G,n} \frac{\Phi_O^p}{\Phi_{O,j}^n} \tag{A.18′}$$

$$q_{oxy,j}^A = q_{oxy,j}^{A,n} \frac{\Phi_O^p}{\Phi_{O,j}^n} \tag{A.19′}$$

$$q_{ferm,j}^G = \frac{Y_{EG} q_{ana,j}^G - Y_{EG}^o q_{oxy,j}^G}{Y_{EG}^f} \tag{A.20′}$$

Using Eqs. (A.1)–(A.20)′, we define the function $f$ introduced in Section 2.3.2:

$$\begin{aligned} f &: [0, \mu_{\max}] \times \mathbb{R}_+^3 \longrightarrow \mathbb{R}^5 \\ &(\mu_j, G, A, O) \mapsto \left( q_{ana}^G, q_{ana}^A, q_{oxy}^G, q_{oxy}^A, q_{ferm}^G \right) \end{aligned} \tag{A.21}$$

By using both this function and the calculation procedure of the overflow reaction rate, we may plot all the reaction rates throughout a population in a given environment as done in Fig. A1. This figure has been computed with the following concentrations: $G = 0.2\ g_G\ \mathrm{L}^{-1}$, $O = 3.6\ \mathrm{mg}_O\ \mathrm{L}^{-1}$ and $A = 0.06\ g_A\ \mathrm{L}^{-1}$. Some constants were also modified to make the figure more readable: $\mu_{\max}^G = 0.6\ h^{-1}$, $\mu_{\max}^A = 0.3\ \mathrm{h}^{-1}$ and $\Phi_O^{\max} = 8\ \mathrm{mmol}_O\ g_X^{-1}\ \mathrm{h}^{-1}$.

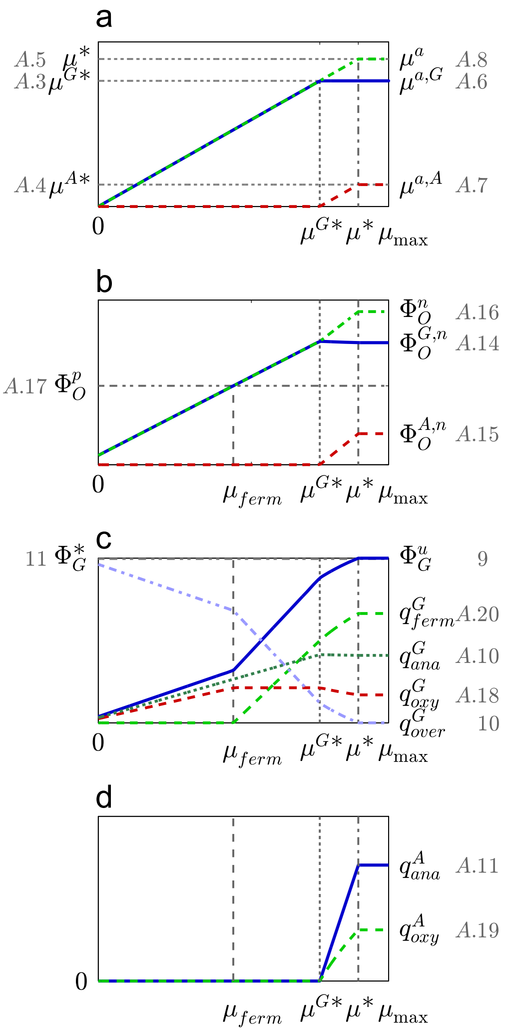


**Fig. A1.** Evolution of (a) growth rates, (b) oxygen uptake rates, (c) glucose uptake and reactions rates and (d) acetate reaction rates, throughout the population. Left hand side: legends of variables depending only on the environment (independent from individuals). Right hand side: legends for corresponding curves. Legends also detail the equations in which are given expressions of corresponding variables.

In this figure, each abscissa corresponds to the behavior of a particular class.

Fig. A1a shows the evolution of the actual growth rates both on glucose and acetate following Eqs. (A.6)–(A.8). Thus, three different regimes may be observed:

- $\mu_j \in [0; \mu^{G*}]$: only glucose in consumed due to strong internal limitations.
- $\mu_j \in [\mu^{G*}, \mu^*]$: both glucose and acetate are consumed, but growth is still limited by internal capabilities.
- $\mu_j \in [\mu^*, \mu_{\max}]$: availabilities of glucose and acetate limit growth.

Fig. A1b shows the evolution of the needed oxygen consumption rate $\Phi_O^n$, expressed in Eq. (A.16). When the oxygen needs ($\Phi_O^n$) exceed the oxygen availability ($\Phi_O^p$, Eq. (A.17)), the oxygen consumption is saturated and a part of the energy is produced through fermentation. It leads to the following regimes:

- $\mu_j \in [0; \mu_{ferm}]$: Energy produced only through the oxidative pathway.
- $\mu_j \in [\mu_{ferm}, \mu_{\max}]$: Energy produced with both oxidative and fermentative pathways.

These two first sub-figures allow to identify, for each value of $\mu_j$, what is (or are) the factor(s) limiting growth. From that, we can deduce the actual reaction rates, presented in Fig. A1c and d. In particular, these inter-correlations may be observed:

1. The anabolism rates, $q_{ana}^G$ and $q_{ana}^A$, are directly proportional to $\mu^{a,G}$ (Eq. (A.10)) and $\mu^{a,A}$ (Eq. (A.11)).
2. The energy used for growth on acetate is always produced through acetate oxidation, then, $q_{oxy}^A$ is directly proportional to $q_{ana}^A$.
3. The rate of energy production for growth on glucose are calculated such as the maximum part of this energy in produced through the oxidative pathway, in the limit of oxygen availabilities.
4. The useful part of glucose uptake, $\Phi_G^u$ is the sum of $q_{ana}^G$, $q_{oxy}^G$ and $q_{ferm}^G$.
5. The overflow reaction rate, $q_{over}^G$, is defined as the difference between $\Phi_G^u(\mu^*)$ and $\Phi_G^u(\mu)$

Again, each abscissa represents the behavior of a particular class. It is important to note that depending on the state of the population, not all these behaviors will be exhibited. Our biological model couples the population balance with this metabolic model, leading to a wide variety of different possible behaviors at the scale of the population. In usual models, an instantaneous adaptation of bacteria to their environments is postulated. When using our metabolic model, such an hypothesis would constrain the observed behaviors to the one described in Fig. A1 at the $\mu^*$ abscissa.

## Appendix B. Computation of the matrix of volume flow rates for the fedbatch reactor

Vrábel et al. (1999, 2000, 2001) designed, challenged and used a compartment model to run simulations compared with the behaviors observed within their large bioreactor (30 m$^3$). We used the description of their compartment model to compute the matrix of volume flow rates, $M^f$, needed to run our own

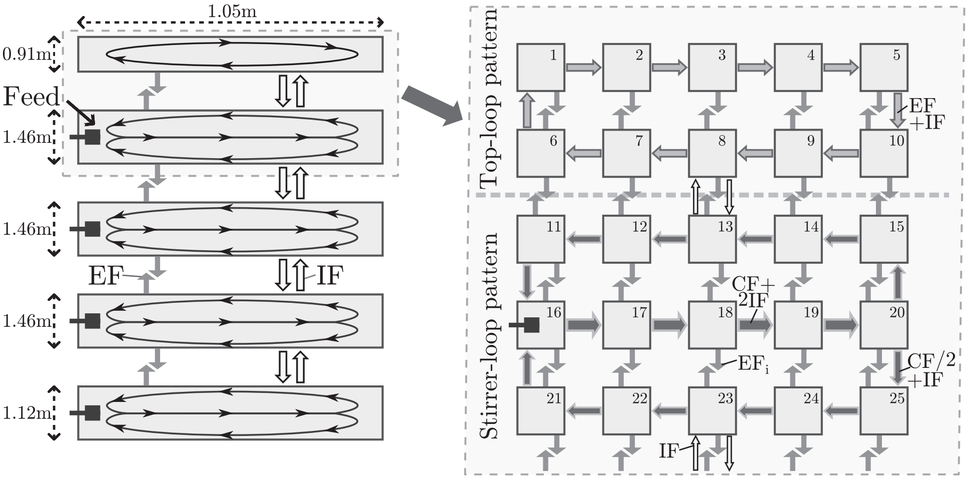


**Fig. B1.** Representation of the macroscopic flow patterns in the fedbatch reactor (left) and detail of its compartimentation and specific flows on the top of the reactor (right).

simulations. This model is based on the definition of three flows induced either by the use of Rushton turbines, or by the aeration.

Vrábel et al. observed that each of the four stirrers induces two circulation loops plus a supplementary loop on the top of the reactor as shown in Fig. B1 left. The flow in those loops is named Circulation Flow (CF). The presence of bubble for the aeration creates a flow that connects the circulation loops on their middles. This aeration Induced Flow (IF) also increases the velocity of the liquid within the circulation loops. Finally, the mixing and the aeration leads to the apparition of vertical turbulence flow (EF).

With these observations, Vrábel et al. modelized this hydrodynamics by splitting the reactor's volume in 14 rows (3 per stirrer plus 2 for the top-loop), and each row in 5 columns, as shown in the right of Fig. B1. The last figure shows a numeration of compartments as well as the flows going through each of them.

By definition, the matrix of volume flow rates is here a 70-by-70 matrix. Its value on the $m$-th row and $n$-th column represents the flow going from the compartment $m$ to the compartment $n$. For example $M^f_{18,19} = CF + 2IF$ as shown in Fig. B1.

Once $M^f$ has been designed based on values of CF, IF and EF, an estimation of these flows is needed. The calculation method used in estimating those flows has already been detailed by Vrábel et al. (1999, 2000). However, their procedure needs the knowledge of the stirrer speed and the gas-holdup throughout the culture. We then considered a constant stirrer speed of 1 $s^{-1}$, a gas flow of $1/60\ m^3\ s^{-1}$ and an overall gas-holdup of 8%. Those values are consistent with the range of values used by Vrábel et al. (1999, 2000, 2001).

These values lead to the following flows:

- $CF = 0.32\ m^3\ s^{-1}$
- $IF = 0.04\ m^3\ s^{-1}$
- $EF = 0.28\ m^3\ s^{-1}$

## References

Altintas, M.M., Eddy, C.K., Zhang, M., McMillan, J.D., Kompala, D.S., 2006. Kinetic modeling to optimize pentose fermentation in *Zymomonas mobilis*. Biotechnol. Bioeng. 94 (2), 273–295.
Bezzo, F., Macchietto, S., Pantelides, C.C., 2003. General hybrid multizonal CFD approach for bioreactor modeling. AIChE J. 49 (8), 2133–2148.
Delafosse, A., Collignon, M.-L., Calvo, S., Delvigne, F., Crine, M., Thonart, P., Toye, D., 2014. CFD-based compartment model for description of mixing in bioreactors. Chem. Eng. Sci. 106 (0), 76–85.
Delafosse, A., Delvigne, F., Collignon, M.-L., Crine, M., Thonart, P., Toye, D., et al., 2010. Development of a compartment model based on CFD simulations for description of mixing in bioreactors. Biotechnol. Agron. Soc. Environ. 14 (2), 517–522.
Dhar, N., McKinney, J.D., 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. Curr. Opin. Microbiol. 10 (1), 30–38.
Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H.-Y., Neubauer, P., van der Lans, R., Luyben, K., Vrábel, P., Manelius, Å., 2001. Physiological responses to mixing in large scale bioreactors. J. Biotechnol. 85 (2), 175–185 (Twenty years of the European Federation of Biotechnology).
Ferenci, T., 1996. Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. FEMS Microbiol. Rev. 18 (4), 301–317.
Ferenci, T., 1999. Regulation by nutrient limitation. Curr. Opin. Microbiol. 2 (2), 208–213.
Franchini, A.G., Egli, T., 2006. Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology 152 (7), 2111–2127.
Henson, M.A., Müller, D., Reuss, M., 2002. Cell population modelling of yeast glycolytic oscillations. Biochem. J. 368 (2), 433–446.
Holms, H., 1996. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. FEMS Microbiol. Rev. 19 (2), 85–116.
Hristov, H., Mann, R., Lossev, V., Vlaev, S., 2004. A simplified CFD for three-dimensional analysis of fluid mixing, mass transfer and bioreaction in a fermenter equipped with triple novel geometry impellers. Food Bioprod. Process. 82 (1), 21–34 (Mixing, Heat and Mass Transfer).
Lapin, A., Schmid, J., Reuss, M., 2006. Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor—a structured—segregated approach. Chem. Eng. Sci. 61 (14), 4783–4797.
Larsson, G., Törnkvist, M., Wernersson, E., Trägårdh, C., Noorman, H., Enfors, S.-O., 1996. Substrate gradients in bioreactors: origin and consequences. Bioprocess. Biosyst. Eng. 14 (6), 281–289.
Leegwater, M.P.M., Neijssel, O.M., Tempest, D.W., 1982. Aspects of microbial physiology in relation to process control. J. Chem. Technol. Biotechnol. 32 (1), 92–99.
Lencastre Fernandes, R., Jensen, A.D., Nopens, I., Gernaey, K. V., 2013. The effect of bioreactor compartmentalization on yeast population dynamics during continuous cultivation. In: Proceedings of 5th International Conference on Population Balance Modelling, p. 5.
Lencastre Fernandes, R., Nierychlo, M., Lundin, L., Pedersen, A.E., Puentes Tellez, P.E., Dutta, A., Carlquist, M., Bolic, A., Schäpper, D., Brunetti, A.C., Helmark, S., Heins, A.L., Jensen, A.D., Nopens, I., Rottwitt, K., Szita, N., van Elsas, J.D., Nielsen, P.H., Martinussen, J., Sorensen, S.J., Lantz, A.E., Gernaey, K.V., 2012. Experimental methods and modeling techniques for description of cell population heterogeneity. Biotech. Adv. 29 (6), 575–599.
Linkès, M., Fede, P., Morchain, J., Schmitz, P., 2014. Numerical investigation of subgrid mixing effects on the calculation of biological reaction rates. Chem. Eng. Sci. 116, 473–485.
Mantzaris, N.V., 2005. A cell population balance model describing positive feedback loop expression dynamics. Comput. Chem. Eng. 29 (4), 897–909 (Control of Multiscale and Distributed Process Systems).
Mantzaris, N.V., 2006. Stochastic and deterministic simulations of heterogeneous cell population dynamics. J. Theor. Biol. 241 (3), 690–706.
Mantzaris, N.V., 2007. From single-cell genetic architecture to cell population dynamics: quantitatively decomposing the effects of different population heterogeneity sources for a genetic network with positive feedback architecture. Biophys. J. 92 (12), 4271–4288.

Matsuoka, Y., Shimizu, K., 2013. Catabolite regulation analysis of *Escherichia coli* for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation. J. Biotechnol. 168 (2), 155–173 (Special Issue: Biotechnology for a healthy and green world).

Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab. Eng. 12 (2), 150–160 (Metabolic Flux Analysis for Pharmaceutical Production Special Issue).

Morchain, J., Fonade, C., 2009. A structured model for the simulation of bioreactors under transient conditions. AIChE J. 55 (11), 2973–2984.

Morchain, J., Gabelle, J.-C., Cockx, A., 2013. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. AIChE J. 59 (2), 369–379.

Morchain, J., Gabelle, J.-C., Cockx, A., 2014. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AIChE J. 60 (1), 27–40.

Moullec, Y.L., Gentric, C., Potier, O., Leclerc, J., 2010. Comparison of systemic, compartmental and CFD modelling approaches: application to the simulation of a biological reactor of wastewater treatment. Chem. Eng. Sci. 65 (1), 343–350 (20th International Symposium in Chemical Reaction Engineering–Green Chemical Reaction Engineering for a Sustainable Future).

Natarajan, A., Srienc, F., 2000. Glucose uptake rates of single *E. coli* cells grown in glucose-limited chemostat cultures. J. Microbiol. Methods 42 (1), 87–96.

Nobs, J.-B., Maerkl, S.J., 2014. Long-term single cell analysis of *S. pombe* on a microfluidic microchemostat array. PLoS ONE 9 (April (4)), e93466.

Peskov, K., Mogilevskaya, E., Demin, O., 2012. Kinetic modelling of central carbon metabolism in *Escherichia coli*. FEBS J. 279 (18), 3374–3385.

Pirt, S.J., 1965. The maintenance energy of bacteria in growing cultures. Proc. R. Soc. B 163 (October), 224–231.

Ramkrishna, D., 2000. Population Balances: Theory and Applications to Particulate Systems in Engineering. Academic Press.

Russell, J.B., Cook, G.M., 1995. Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol. Rev. 59 (1), 48–62.

Ryall, B., Eydallin, G., Ferenci, T., 2012. Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. Microbiol. Mol. Biol. Rev. 76 (3), 597–625.

Varma, A., Boesch, B.W., Palsson, B.O., 1993. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. Appl. Environ. Microbiol. 59 (8), 2465–2473.

Vrábel, P., van der Lans, R., Cui, Y., Luyben, K., 1999. Compartment model approach: mixing in large scale aerated reactors with multiple impellers. Chem. Eng. Res. Des. 77 (4), 291–302.

Vrábel, P., vanderLans, R.G., Luyben, K.C., Boon, L., Nienow, A.W., 2000. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements. Chem. Eng. Sci. 55 (23), 5881–5896.

Vrábel, P., vanderLans, R.G., vanderSchot, F.N., Luyben, K.C., Xu, B., Enfors, S.-O., 2001. CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. Chem. Eng. J. 84 (3), 463–474.

Wang, Q., Ou, M.S., Kim, Y., Ingram, L., Shanmugam, K., 2010. Metabolic flux control at the pyruvate node in an anaerobic *Escherichia coli* strain with an active pyruvate dehydrogenase. Appl. Environ. Microbiol. 76 (7), 2107–2114.

Wick, L.M., Quadroni, M., Egli, T., 2001. Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. Environ. Microbiol. 3 (9), 588–599.

Wick, L.M., Weilenmann, H., Egli, T., 2002. The apparent clock-like evolution of *Escherichia coli* in glucose-limited chemostats is reproducible at large but not at small population sizes and can be explained with monod kinetics. Microbiology 148 (9), 2889–2902.

Xu, B., Jahic, M., Blomsten, G., Enfors, S.-O., 1999a. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. Appl. Microbiol. Biotechnol. 51 (5), 564–571.

Xu, B., Jahic, M., Enfors, S.-O., 1999. Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. Biotechnol. Progr. 15 (1), 81–90.

Zahradník, J., Mann, R., Fialová, M., Vlaev, D., Vlaev, S., Lossev, V., Seichter, P., 2001. A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. Chem. Eng. Sci. 56 (2), 485–492 (16th International Conference on Chemical Reactor Engineering).

# B

# AN ASSESSMENT OF METHODS OF MOMENTS FOR THE SIMULATION OF POPULATION DYNAMICS IN LARGE-SCALE BIOREACTORS

Chemical Engineering Science 171 (2017) 218–232

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

# An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors

Maxime Pigou [a,b,*], Jérôme Morchain [a], Pascal Fede [b], Marie-Isabelle Penet [c], Geoffrey Laronze [c]

[a] *LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France*
[b] *Institut de Mécanique des Fluides de Toulouse, Université de Toulouse, CNRS-INPT-UPS, Toulouse, France*
[c] *Sanofi Chimie, C&BD Biochemistry Vitry, 9 quai Jules Guesde, 94400 Vitry-sur-Seine, France*

## HIGHLIGHTS

- A biological population balance model is solved using class and moment methods.
- Homogeneous chemostat and heterogeneous fedbatch cultures are simulated.
- Methods are compared through accuracy, stability and computation time.
- The Maximum Entropy method is found to be unstable in the present test-cases.
- QMOM and EQMOM are well suited and have major advantages against class method.

## ARTICLE INFO

*Article history:*
Received 13 February 2017
Received in revised form 27 April 2017
Accepted 15 May 2017
Available online 20 May 2017

*Keywords:*
Biological dynamics
Population balance model
Method of classes
Method of moments
(E)QMOM
Maximum Entropy

## ABSTRACT

A predictive modelling for the simulation of bioreactors must account for both the biological and hydrodynamics complexities. Population balance models (PBM) are the best approach to conjointly describe these complexities, by accounting for the adaptation of inner metabolism for microorganisms that travel in a large-scale heterogeneous bioreactor. While being accurate for solving the PBM, the Class and Monte-Carlo methods are expensive in terms of calculation and memory use. Here, we apply Methods of Moments to solve a population balance equation describing the dynamic adaptation of a biological population to its environment. The use of quadrature methods (Maximum Entropy, QMOM or EQMOM) is required for a good integration of the metabolic behavior over the population. We then compare the accuracy provided by these methods against the class method which serves as a reference. We found that the use of 5 moments to describe a distribution of growth-rate over the population gives satisfactory accuracy against a simulation with a hundred classes. Thus, all methods of moments allow a significant decrease of memory usage in simulations. In terms of stability, QMOM and EQMOM performed far better than the Maximum Entropy method. The much lower memory impact of the methods of moments offers promising perspectives for the coupling of biological models with a fine hydrodynamics depiction.

## 1. Introduction

The large-scale simulation of bioreactors is currently a challenging issue. Such simulations must account for both (i) the (multiphase) hydrodynamics and (ii) the metabolic behaviour of the biological population carried by the fluid. The first can be achieved through the use of widespread CFD softwares which require significant computational power. The second can be addressed with advanced cell models which result from community efforts to integrate genome-scale reconstructions of a strain metabolic network and depict thousands of intracellular reactions and metabolite concentrations. Examples are the iJO1366 model for *Escherichia coli* (Orth et al., 2011) and the consensus YEAST model for *Saccharomyces cerevisiae* (Heavner et al., 2012; Heavner et al., 2013). These models describe state of the art knowledge of a cell metabolism, however their implementations require to solve either cumbersome optimization problems to access a steady-state cell-functioning, or to solve dynamically the metabolite concentrations in a cell that experiences exogeneous perturbations.

Even though the computational power increased significantly over the past few decades, it is still not possible to couple the CFD approach with a biological modelling that fully embraces the biological complexity. Such an approach is numerically untractable as it requires to solve dynamically the intracellular concentrations for each cell in a bioreactor with an Euler-Lagrange framework.

---

* Corresponding author.
*E-mail address:* maxime.pigou@insa-toulouse.fr (M. Pigou).

http://dx.doi.org/10.1016/j.ces.2017.05.026
0009-2509/

## Notation

*Roman*

| | |
|---|---|
| $C$ | concentration ($kg.m^{-3}$) |
| $H$ | Shannon entropy |
| $K$ | biological affinity constant ($kg.m^{-3}$) |
| $L$ | quadrature node abscissae ($h^{-1}$) |
| $m$ | moment of distribution n ($kg.m^{-3}.h^{-k}$) |
| $n$ | number density function ($h.kg_X.m^{-3}$) |
| $N$ | number of resolved moments |
| $NC$ | number of classes |
| $N_c$ | number of compartments |
| $P$ | order of moment methods |
| $q$ | specific reaction rate ($mol.kg_X^{-1}.h^{-1}$) |
| $Q$ | flow rate ($m^3.h^{-1}$) |
| $R$ | reaction rate ($kg.m^{-3}.h^{-1}$) |
| $T$ | time constant of adaptation (h) |
| $V$ | compartment volume ($m^{-3}$) |
| $w$ | quadrature node weight ($kg_X.m^{-3}$) |
| $Y$ | stoichiometric molar coefficient ($mol.mol^{-1}$) |

*Subscript and superscript*

| | |
|---|---|
| $\tilde{x}$ | population mean value |
| $x^*$ | equilibrium value |
| $x_a$ | achieved value |
| $x_A$ | acetate |
| $x_G$ | glucose |
| $x_{\mathrm{i}}$ | inhibition |
| $x_k$ | moment order |
| $x_m$ | compartment index |
| $x_n$ | compartment index |
| $x_O$ | oxygen |
| $x_T$ | threshold value |

*Greek symbols*

| | |
|---|---|
| $\varepsilon$ | turbulent energy dissipation rate ($W.kg^{-1}$) |
| $\kappa$ | PDF kernel |
| $\mu$ | growth rate ($g_X.g_X^{-1}.h^{-1}$) |
| $\nu$ | kinematic viscosity ($m^2.s^{-1}$) |
| $\varphi$ | polynomial coefficient |
| $\Phi$ | specific uptake rate ($g.g_X^{-1}.h^{-1}$) |
| $\Psi$ | environmental limitation coefficient |
| $\sigma$ | standard deviation ($h^{-1}$) |
| $\zeta$ | rate of change of specific growth rate ($h^{-2}$) |

Therefore, two simplified approaches are usually applied. On the one hand, one can neglect the spatial heterogeneity and solve a complex metabolic model in homogeneous batch or chemostat cultures (Meadows et al., 2010; Matsuoka and Shimizu, 2013). On the other hand, one will describe the hydrodynamic complexity jointly with a simplified biological approach such as either structured or unstructured kinetic models (Bezzo et al., 2003; Elqotbi et al., 2013; Lu et al., 2015).

Concentration gradients are known to be responsible for metabolic dysfunctions in large-scale reactors (Enfors et al., 2001), therefore we should avoid the first approach and describe the spatial heterogeneities. However, the use of kinetic models should be discarded too. Indeed, from the point of view of a cell travelling in these heterogeneous concentrations fields, the concentration signal is fluctuating (Linkès et al., 2014; Haringa et al., 2016). This make kinetic models inappropriate as they are usually based on the Monod kinetics law which reflects a steady-state equilibrium between a population and its environment. By making use of a Monod law, the kinetic models have "been over simplified by allowing instantaneous adaptation of the cell to the abiotic environment" (Silveston et al., 2008).

In previous work (Pigou and Morchain, 2015), we stepped back in both the hydrodynamic description by using a Compartment Model Approach (Cui et al., 1996; Mayr et al., 1993; Vrábel et al., 2000; Vrábel et al., 2001) and in the metabolic description of *E. coli* by simplifying the key reactions of the central carbon metabolism into a 6 reactions model inspired by the model proposed by Xu et al. (1999). More importantly, we introduced the use of a Population Balance Model (PBM) as a key modelling tool that allows describing simultaneously both (i) the concentration gradients, (ii) a dynamic adaptation of cells to the fluctuating conditions they experience along their trajectories and (iii) the metabolic impact of a disequilibrium between a cell and its local environment. This approach has been successfully challenged against experimental data in lab-scale batch culture and industrial-scale heterogeneous fedbatch culture. More recently, we improved the PBM to account for an experimentally observed stochastic diversity related to cell-division (Morchain et al., in press).

Until now, we solved the PBM using a class method (also known as fixed pivot method, Kumar and Ramkrishna (1996a, 2001)) with at least 60 classes to span the entire range of possible values for the chosen variable (i.e. the maximum growth-rate achievable by a cell provided enough nutrients are available). Each class represents a scalar that must be transported by the hydrodynamic framework. While transporting a hundred classes within a 70 compartments model (Pigou and Morchain, 2015) was perfectly feasible, doing the same in a CFD simulation would be prohibitively expensive.

The current paper thus makes the focus on improving the numerical tractability of the PBM, through the use of the Method of Moments (MOM), in order to increase the allowed level of spatial accuracy. Instead of performing a direct resolution of the population balance equation, the MOM describes the evolution of the first moments of a Number Density Function (NDF). However, it will be of interest to perform a reverse operation and to recover an approximation of the NDF from a finite set of its moments; this is known as a truncated moment problem (Abramov, 2007).

Many methods are available to tackle this problem. A review of such methods is available (John et al., 2007) though new methods or improvements are available since its publication. More recently, Lebaz et al. (2016) compared the most common approaches which are Kernel Density Element Method (KDEM), Spline-based method, and the Maximum Entropy (MaxEnt) method applied to the case of a depolymerization process. The KDEM approximates the unknown NDF as the sum of weighted Kernel Density Functions (KDF). The identification of the weights is performed through a constrained minimization procedure, which requires a high number of moments to prevent an underdetermined problem and the multiplicity of solutions. The spline method (John et al., 2007) leads to a piece-wise polynomial reconstruction, but the resulting reconstruction is highly dependent on numerical parameters, and can lead to negative values of the reconstructed NDF. For these reasons, the KDEM and spline methods will be discarded in the current work.

The MaxEnt method (Mead and Papanicolaou, 1984; Tagliani, 1999) was point out as efficient and accurate, even with a low number of moments, by Lebaz et al. (2016). It is however ill-conditioned at the boundaries of moment space (Massot et al., 2010), but this can be handled by providing some adjustments of the method (Vié et al., 2013). Finally, we consider the recent EQMOM method (Chalons et al., 2010; Yuan et al., 2012; Marchisio and Fox, 2013) which constitutes an interesting fusion of KDEM with the QMOM approach (Marchisio et al., 2003a; Marchisio et al., 2003b; Marchisio et al., 2003c). This method has proven to be stable and efficient –in particular near the frontier of the realizable moment space where MaxEnt is ill-conditioned– but requires to make assumptions on the shape of the reconstruction.

The current work is focused on assessing the methods QMOM, EQMOM and MaxEnt against the already used classes method, in the perspective of running predictive and numerically tractable bioreactor simulations. All these methods have been used to perform the simulation of a homogeneous chemostat culture stressed with a dilution rate shift (Kätterer et al., 1986). After, the methods are compared in terms of their numerical efficiency, their accuracy and their stability in this peculiar configuration. Finally, the moment and classes methods are compared on the configuration investigated by Pigou and Morchain (2015) of a heterogeneous fedbatch culture described by Vrábel et al. (1999, 2000). For this last, the heterogeneity is taken into account by a compartment model approach.

## 2. Models and methods

### 2.1. Local mass balance

The basis in the modelling of bioreactors is the formulation of local mass balances. They describe the evolution of local concentrations as a consequence of (i) transport by the carrying fluid and (ii) consumption or production by the biological phase. As in previous work (Pigou and Morchain, 2015), we will hereafter describe the hydrodynamics using Compartment Model Approach (CMA). Let $\mathbf{C}_n$ (kg.m$^{-3}$) be the vector of mass concentrations within the $n$-th compartment, $V_n$ the volume (m$^3$) of that compartment, and $Q_{n,m}$ (m$^3$.h$^{-1}$) the volume flow rate going from the $n$-th to the $m$-th compartment. The total number of compartments is $N_c$. Then, the mass balance equation in compartment $n$ is given as:

$$\frac{\partial V_n \mathbf{C}_n}{\partial t} + \mathbf{C}_n \sum_{m=1}^{N_c} (Q_{n,m}) - \sum_{m=1}^{N_c} (Q_{m,n}\mathbf{C}_m) = V_n \mathbf{R}(\mathbf{C}_n) \tag{1}$$

Our contribution is to express the vector of biological reaction rates $\mathbf{R}(\mathbf{C})$ (kg.m$^{-3}$.h$^{-1}$) as the sum of the substrate uptake rates, or product production rates, due to all cells considering their individual physiological states. Let $\mu$ (h$^{-1}$) be the biological growth capability of a cell (i.e. the growth rate they can achieve if permitted by the nutrient availability), we will distinguish each individual upon this value. Different cells, having different values of $\mu$ in a similar environment, will exhibit different metabolic behaviours. Then, in order to express the bioreaction rates at the scale of the biological population, one must know the statistical distribution of the property $\mu$ over that population, and integrate the uptake or production rates over that distribution:

$$R_i(\mathbf{C}_n) = \int_0^{+\infty} n(\mu)\Phi_i(\mu, \mathbf{C}_n)\mathrm{d}\mu \tag{2}$$

where $n(\mu)$ is the NDF defining the fraction of the biological phase whose specific growth rate is $\mu$. The first two moments of this NDF are defined as following:

$$\int_0^{+\infty} n(\mu)\mathrm{d}\mu = X \tag{3}$$

$$\int_0^{+\infty} \mu n(\mu)\mathrm{d}\mu = \tilde{\mu} X \tag{4}$$

with X the total biomass concentration (kg.m$^{-3}$) and $\tilde{\mu}$ the population mean growth rate (h$^{-1}$). In the current work, we consider the metabolic behaviour of *Escherichia coli* and the vector $\mathbf{C}$ actually consists in a vector of Glucose ($G$), Acetate ($A$) and Oxygen ($O$) concentrations.

$$\mathbf{C} = \begin{bmatrix} C_G \\ C_A \\ C_O \end{bmatrix} \tag{5}$$

We will also consider scalar variables to transport information about the distribution $n(\mu)$ as explained afterwards.

Therefore, the glucose uptake rate $\Phi_G(\mu, \mathbf{C})$, the oxygen uptake rate $\Phi_O(\mu, \mathbf{C})$ and the acetate uptake/production rate $\Phi_A(\mu, \mathbf{C})$ will be outcomes of the metabolic model calculation procedure. The later uses as inputs (i) the specific potential growth rate of individual, $\mu$; (ii) the vector of concentrations in the liquid phase $\mathbf{C}$, and (iii) the equilibrium law $\mu^* = f(\mathbf{C})$. The growth rate at equilibrium $\mu^*$ is the growth rate that cells would exhibit at steady state in an environment defined by the vector of concentrations $\mathbf{C}$. Such expressions are known from chemostat experiments and typically take the form of a multi-component Monod-Law, taking here into account the inhibitory effect of acetate:

$$\mu^* = \mu_{\max} \frac{C_G}{C_G + K_G} \frac{C_O}{C_O + K_O} \frac{K_{i,A}}{C_A + K_{i,A}} \tag{6}$$

with $K_G$ and $K_O$ the affinity constants (kg.m$^{-3}$) of the biomass toward glucose and oxygen, and $K_{i,A}$ the inhibitory constant of growth by acetate.

A noticeable point is that the substrate uptake rate is not algebraically related to the specific growth rate as it cannot be assumed in general that cells are at equilibrium with their environment (Ferenci, 1996). Therefore, our approach is consistent with theoretical considerations (Perret, 1960) and experimental observations (Abulesz and Lyberatos, 1989; Li, 1982; Silveston et al., 2008) indicating that the growth and uptake rates are decoupled in the dynamic regime whilst an algebraic relation exists between them at steady state.

The second point in terms of modelling resides in the calculation of the NDF $n(\mu)$ that defines the concentration of biomass whose potential growth rate is $\mu$. This calculation will be addressed in a dedicated paragraph.

### 2.2. Calculation procedure for the metabolic reaction rates

The procedure is almost identical to that presented in a previous paper (Pigou and Morchain, 2015), therefore, only the key features of the metabolic model, and the few differences of the calculation procedure are detailed here.

The first step of that procedure is to compute the actual growth rate of each cell, by taking into account its growth capabilities (see Section 2.3), and a potential limitation related to nutrient availability. In the previous work, we defined this actual growth rate of a cell, $\mu_a$, as the minimum between its biological growth capability, $\mu$, and the environment equilibrium growth rate $\mu^*$ (given by the Monod law, Eq. (6)): $\mu_a = \min(\mu^*, \mu)$. However, we recently shifted this formulation toward a more meaningful and physical one, based on a limitation by the micromixing, which proved to be consistent with experimental studies of membrane transporters at limiting nutrient concentrations (Ferenci, 1996; Ferenci, 1999). The detailed explanation for this change is given in Morchain et al. (in press).

We then defined a threshold glucose concentration, $C_{G_T}$, around which micromixing will start to be a limiting factor. As long as the bulk substrate concentration is significantly higher than this threshold concentration, cells will be fed enough by micromixing to be able to achieve their potential growth rate.

$$C_{G_T} = R_G \times 17\left(\frac{\nu}{\varepsilon}\right)^{0.5} \tag{7}$$

The term $17\sqrt{\nu/\varepsilon}$ is proposed by Baldyga and Bourne (1999) to evaluate the micro-mixing time-scale, and depends on the fluid viscosity, $\nu$ ($m^2/s$), and the turbulent energy dissipation rate, $\varepsilon$, whose value usually ranges from 0.5 to 10 W/kg depending on the bioreactor stirring.

As $R_G$ is an output of the metabolic calculation procedure (Eq. (2)), which itself depends on $\mu_a$, and considering that we only need the order of magnitude of the limiting concentration, we provide the following rough approximation of $R_G$ for the estimation of $C_{G_T}$:

$$R_G \approx \frac{M_G}{Y_{XG}M_X}\int_0^{+\infty} \mu n(\mu)\mathrm{d}\mu \tag{8}$$

With $Y_{XG}$ the molar yield of glucose to cell conversion ($mol_X.mol_G^{-1}$), $M_X$ the molar mass of biomass ($M_X = 113.1\ g_X.mol_X^{-1}$ considering the typical chemical formula $C_5H_7NO_2$) and $M_G$ the molar mass of glucose ($M_G = 180.2\ g_G.mol_G^{-1}$).

Now, following Morchain et al. (in press), the actual growth rate is given by:

$$\mu_a = \Psi\mu \tag{9}$$

where the coefficient $\Psi$ reads:

$$\Psi = 1 - e^{-C_G/C_{G_T}} \tag{10}$$

As a recall, the estimation of $\mu_a$ along with the calculation of $\mu^*$ is the very first step of the calculation procedure of the metabolic model, as detailed in Pigou and Morchain (2015) which explains why the choice of $\mu_a$ formulation is of importance. After estimating the actual -effectively achieved- growth rate of a cell, the calculation procedure of the metabolic model is exactly the one described in the previous work.

This metabolic model roughly describes the central metabolism of *Escherichia coli*, it accounts for:

- Anabolism based on either glucose or acetate as a carbon source, leading to the formation of new cells,
- Oxidative catabolism on both substrates for energy production,
- Fermentative catabolism of glucose, leading to the production of energy, and acetate as a by-product,
- Overflow metabolism, leading to production of acetate when glucose is over-consumed.

Each pathway is simplified into the following set of reactions:

$$G + Y_{EG}E \overset{q_{ana}^G}{\rightarrow} Y_{XG}X \tag{R1}$$

$$A + Y_{EA}E \overset{q_{ana}^A}{\rightarrow} Y_{XA}X \tag{R1$'$}$$

$$G + Y_{OG}O \overset{q_{oxy}^G}{\rightarrow} Y_{EG}^o E \tag{R2}$$

$$A + Y_{OA}O \overset{q_{oxy}^A}{\rightarrow} Y_{EA}^o E \tag{R2$'$}$$

$$G \overset{q_{ferm}^G}{\rightarrow} Y_{EG}^f E + Y_{AG}A \tag{R3}$$

$$G \overset{q_{over}^G}{\rightarrow} Y_{AG}A \tag{R4}$$

G: Glucose, E: Energy, A: Acetate, O: Oxygen, X: Biomass. $Y_{BA}$ is the stoichiometric molar coefficient in $mol_B/mol_A$. $q_\alpha^G$ and $q_\alpha^A$ are the specific reaction rates for reactions respectively based on glucose ($mol_G.g_X^{-1}.h^{-1}$) and on acetate ($mol_A.g_X^{-1}.h^{-1}$).

The calculation procedure gives access to the specific reaction rates, and is based on:

- The growth capability of a cell ($\mu$), whose evolution is described by the Population Balance Model (see Section 2.3),
- The environmental conditions (G, A and O concentrations) and the $\Psi$ coefficient,
- An assumption of non-accumulation within the cytoplasm. In particular, the rate of energy production is balanced by the rate of energy consumption.

### *2.3. Population balance model*

External and intrinsic perturbations are known to produce heterogeneity among the cell population. In order to track this diversity, the usual mathematical approach is to refer to a population balance model. The originality of our approach resides in that the discriminating factor is the specific growth rate of individuals. Recent observations have proved that this variable is actually distributed in a cell population (Yasuda, 2011). This formulation is advantageous since the relationship between the growth rate and the metabolic reaction rates is much more natural than when the size or mass of the cell is chosen as the discriminating parameter (Pigou and Morchain, 2015; Morchain et al., in press). The Population Balance Equation (PBE) for the specific growth rate distribution $n(\mu)$ is here given for a homogeneous case; terms accounting for the transport might be added on the left hand side depending on the hydrodynamic framework:

$$\frac{\partial n(\mu)}{\partial t} = -\frac{\partial}{\partial\mu}[n(\mu)\zeta(\mu)] + \int_0^{+\infty}\beta(\mu,\mu')n(\mu')\Psi\mu'\mathrm{d}\mu' \tag{11}$$

The first term in the right-hand side of Eq. (11) is a convection term in the $\mu$-space instead of the physical space. It describes the fact that individuals are able to adapt their specific growth rate in response to insufficient or excessive substrate concentrations. We refer to this term as the adaptation term. In the adaptation term, $\zeta(\mu)$ refers to a velocity in the $\mu$-space or equivalently to the rate of change of $\mu$ over time. This velocity can be either positive or negative depending on whether the environment is respectively rich or poor in nutrients, compared to what a cell is used to. In previous work, a general form for $\zeta(\mu)$ was proposed and validated against experimental data sets:

$$\zeta(\mu) = \left(\frac{1}{T} + \mu\right)(\mu^* - \mu) \tag{12}$$

The second term of Eq. (11) is often referred to as the birth and death term in PBM. $\beta(\mu,\mu')$ is a Probability Density function (PDF) which defines the probability that a mother cell having a specific growth rate $\mu'$ produces a daughter cell whose specific growth rate is $\mu$. The analysis of recent experimental data revealed that $\beta$ can be modeled using a skew-normal distribution (Yasuda, 2011; Morchain et al., in press) whose parameters are given in Appendix A.

Instead of looking for an analytic solution for the PBE (Eq. (11)), we will try to solve that equation numerically. The most straightforward method simply consists in a discretization of the $\mu$-space using either fixed (Kumar and Ramkrishna, 1996a) or moving (Kumar and Ramkrishna, 1996b) meshes. These methods tend to be expensive as soon as heterogeneous systems are considered. However, they accurately describe the solution distribution of the PBE and allow an easy coupling with the transport and reaction parts of the modelling (Eq. (1)). We used to apply these fixed mesh (also known as Methods Of Classes, MOC), as detailed in previous papers (Morchain et al., 2013; Pigou and Morchain, 2015). We will here focus on applying methods of moments and challenging their results against the already validated MOC. Knowing the law for the evolution of the distribution (Eq. (11)), the first step to apply moment methods is to transform the PBE so that it expresses the evolution of the distribution's moments. The $k$-th order moment of the distribution $n(\mu)$ is defined as:

$$m_k = \int_0^{+\infty} \mu^k n(\mu) \mathrm{d}\mu \tag{13}$$

The Appendix A details how this definition and the PBE lead to the following law of moments evolution:

$$\frac{\partial m_k}{\partial t} = k\left(\frac{\mu^*}{T} m_{k-1} + \left(\mu^* - \frac{1}{T}\right) m_k - m_{k+1}\right) + \Psi m_1 B_k(\tilde{\mu}) \tag{14}$$

- $B_k$ is the $k$-th order moment of the PDF $\beta(\mu, \mu')$ whose formulation is also given in Appendix A.
- $\tilde{\mu}$ is the population mean growth rate, defined in terms of moments by:

$$\tilde{\mu} = \frac{\int_0^{+\infty} \mu n(\mu) \mathrm{d}\mu}{\int_0^{+\infty} n(\mu) \mathrm{d}\mu} = \frac{m_1}{m_0} \tag{15}$$

$\frac{\partial m_k}{\partial t}$ depends on $m_{k+1}$ which leads to an unclosed formulation. To tackle this issue, McGraw (1997) introduced the Quadrature Method of Moments (QMOM) which is based on a Gaussian quadrature whose nodes and weights are chosen so that the $N$ first moments of the PDF are well computed by the quadrature:

$$m_k = \sum_{i=1}^{P} w_i L_i^k \qquad \forall k \in \{0, \ldots, N-1\} \tag{16}$$

$P$ is here the order of the method, which deals with $N = 2 * P$ number of moments. The core of the method lies in the identification of weights $w_i$ and abscissas $L_i$ of the Gaussian quadrature. These parameters allow an exact computation of moments of order ranging from 0 to $N-1$ and usually give satisfactory approximation of higher order moments. This method then allows closing the formulation given by Eq. (14).

We introduce here one refinement of the PBE compared to the one described in Pigou and Morchain (2015). The moment formulation of the PBE (Eq. (14)) is correct only if the time constant $T$ is not dependent on $\mu$. However, we used in Pigou and Morchain (2015) one time constant $T_u = 1.9h$ for individuals that are moving upward in the $\mu$-space, and a different time constant $T_d = 6.7h$ for individuals moving downward. This formulation implies that the decrease of growth capabilities in poor environments is slower than the increase of these capabilities in rich environments, and this fact is required to allow a good fitting of experimental data. In the current work, and in order to make use the moment formulation of the PBE given in Eq. (14), we define the time constant $T_m$ as the mean value of the functional $T(\mu)$ that we used previously:

$$T(\mu) = \begin{cases} T_u & \text{if } \mu < \mu^* \\ T_d & \text{otherwise} \end{cases} \tag{17}$$

$$T_m = m_0^{-1} \int_{\Omega\mu} T(\mu) n(\mu) \mathrm{d}\mu \tag{18}$$

$$= \alpha T_u + (1-\alpha) T_d \text{ with } \alpha = m_0^{-1} \int_0^{\mu^*} n(\mu) \mathrm{d}\mu \tag{19}$$

We then actually make use of the PBE given in Eq. (20) to describe the evolution of moments. Similarly, we use the time constant $T_m$ to describe the evolution of the distribution in the class method in order to have consistent formulations between methods.

$$\frac{\partial m_k}{\partial t} = k\left(\frac{\mu^*}{T_m} m_{k-1} + \left(\mu^* - \frac{1}{T_m}\right) m_k - m_{k+1}\right) + \Psi m_1 B_k(\tilde{\mu}) \tag{20}$$

### 2.4. Reconstruction methods

In the present case, and it seems very likely that this would extend to many biological applications, the calculation of the integral reaction term in Eq. (2) cannot be expressed in terms of moments of $n(\mu)$, at least because the uptake rates $\Phi_i(\mu, \mathbf{C})$ are not continuously differentiable with respect to $\mu$. To tackle this issue, we must construct a suited quadrature rule that will be used to approximate all integrals of the following form:

$$\int_{\Omega\mu} f(\mu) n(\mu) \mathrm{d}\mu \approx \sum_i w_i f(L_i) \tag{21}$$

where **w** and **L** are the weights and abscissas of the quadrature rule.

Different methods exist to provide a quadrature rule with the constraint that this rule does compute accurately the known $N$ moments of the distribution:

$$\int_{\Omega\mu} \mu^k n(\mu) \mathrm{d}\mu = \sum_i w_i L_i^k \qquad k \in \{0, \ldots, N-1\} \tag{22}$$

Each method formulates some assumptions about the properties of the NDF $n(\mu)$, and identify a unique NDF $\hat{n}(\mu)$ that matches the set of known moments and the formulated assumptions. We will refer to $\hat{n}(\mu)$ as a reconstruction – or approximation – of $n(\mu)$. Knowing the properties of $\hat{n}(\mu)$, obtaining the quadrature rule **w** and **L** will be quite straightforward.

We will then use this rule to perform the estimation of higher order unknown moments (Eq. (14)) as well as the numerical computation of unclosed integral terms (Eq. (2) and (19)).

#### 2.4.1. The QMOM method

The QMOM method is the easiest method to implement. It makes the assumption that the moment set is at the frontier of the realizable moment space. This implies that the distribution $\hat{n}(\mu)$ is a discrete distribution, written as the sum of $P = N/2$ weighted Dirac distributions. The reconstructed NDF is then given by:

$$\hat{n}(\mu) = \sum_{i=1}^{P} w_i \delta(\mu - L_i) \tag{23}$$

Thus the reaction term (Eq. (2)) can be approximated by:

$$R(\mathbf{C}) \approx \sum_{i=1}^{P} w_i \mathbf{\Phi}(L_i, \mathbf{C}) \tag{24}$$

Due to the complexity of the function $\mathbf{\Phi}(\mu, \mathbf{C})$, a high order quadrature will be required, which implies the need of a high number of resolved moments to correctly approximate the integral term in Eq. (2).

The computation of the weights $w_i$ and abscissas $L_i$ of the quadrature nodes is performed using either the Product-Difference Algorithm (PDA) or the Wheeler Algorithm (WA) as implemented by Marchisio and Fox (2013), with some code tuning to improve efficiency.

With this method, a generic quadrature is given by:

$$\int_{\Omega\mu} f(\mu) n(\mu) \mathrm{d}\mu \approx \sum_{i=1}^{P} w_i f(L_i) \tag{25}$$

#### 2.4.2. The EQMOM method

Yuan et al. (2012) introduced the Extended Quadrature Method of Moments (EQMOM) which consists in coupling QMOM with the

Kernel Density Element Method (KDEM) in which the NDF is reconstructed as the weighted sum of Kernel Density Functions.

The reconstructed NDF, using a P-nodes EQMOM reconstruction, has the following expression:

$$\hat{n}(\mu) = \sum_{i=1}^{P} w_i \kappa(\mu, L_i, \sigma) \tag{26}$$

This method then requires the first $N = 2P + 1$ moments of the NDF, in order to identify uniquely the value of $w_i, L_i$ and $\sigma$. The following kernels are known to be compatible with the EQMOM procedure: Gaussian $\kappa_G$ (Chalons et al., 2010), Log-Normal $\kappa_L$ (Madadi-Kandjani and Passalacqua, 2015), Beta $\kappa_\beta$ and Gamma $\kappa_\Gamma$ (Athanassoulis and Gavriliadis, 2002; Yuan et al., 2012) kernels. We tested each of these kernels but we will only focus on the Gaussian kernel in this paper. Its expression is given hereafter:

$$\kappa_G(\mu, L, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu-L)^2}{2\sigma^2}} \tag{27}$$

This method relies on the Wheeler algorithm (Marchisio and Fox, 2013), in order to identify the values of **w** and **L**. On top of that, a non-linear solver must identify the unique value of $\sigma$ which leads to a reconstructed distribution whose moments match the expected values. We implemented a bisection method to find numerically the root of the objective function that quantify the good agreement of the reconstruction with the set of known moments. We also implemented analytical solutions for $P = 1$ and $P = 2$ as described by Marchisio and Fox (2013).

The integration of the metabolic behaviour over the population (Eq. (2)) is performed by using a 10-nodes Gauss-Hermite quadrature for each node of the Gaussian EQMOM reconstruction as suggested by Yuan et al. (2012).

With this method, a generic quadrature is given by:

$$\int_{\Omega\mu} f(\mu) n(\mu) \mathrm{d}\mu \approx \sum_{i=1}^{P} w_i \sum_{j=1}^{P'} \frac{b_j}{\sqrt{\pi}} f\left(L_i + a_j \sigma \sqrt{2}\right) \tag{28}$$

$w_i$ and $L_i$ and $\sigma$ are the weights and nodes returned by the EQMOM procedure. $a_j$ and $b_j$ are the nodes and weights of a Gauss-Hermite quadrature of order $P'$.

### 2.4.3. The Maximum Entropy method

Given a finite realizable set of $N$ moments, there exists an infinite set of NDF with the same set of first $N$ moments (Mead and Papanicolaou, 1984). Therefore, the goal of any reconstruction method is to choose one plausible NDF out of this infinite set of possibilities. While the EQMOM method enforces the expected shape of the reconstruction by choosing arbitrarily a specific kernel, the Maximum Entropy method aims to find, out of all possible reconstructions, the one that maximizes the Shannon Entropy defined for any PDF $f$ as:

$$H[f] = -\int_{-\infty}^{+\infty} f(x) \ln(f(x)) \mathrm{d}x \tag{29}$$

Tagliani (1999) describes the application of this method for the specific case of a positive PDF defined on the closed support $x \in [0, 1]$. This method can be extended to any finite support $[a, b]$ without loss of generality by a mere linear change of variable.

The reconstructed distribution whose Shannon entropy is the highest takes the following form (Mead and Papanicolaou, 1984; Tagliani, 1999):

$$\hat{n}(\mu) = \exp\left(-\sum_{i=0}^{P} \varphi_i \mu^i\right) \tag{30}$$

With $P = N - 1$ the highest order of known moments.

The key issue is to identify the values of the polynomial coefficients $\varphi_i$, which is achieved through the minimization of the following function (Kapur, 1989; Mead and Papanicolaou, 1984):

$$\Gamma(\varphi_1, \ldots \varphi_P) = \sum_{k=1}^{P} \varphi_k \frac{m_k}{m_0} + \ln\left(\int_0^1 \exp\left(-\sum_{k=1}^{P} \varphi_k \mu^k\right) \mathrm{d}\mu\right) \tag{31}$$

The $\Gamma$ function is both convex and smooth which makes its minimization possible through an iterative Newton-Raphson procedure, with the necessary and sufficient condition that the moment sequence is realizable and not too close from the frontier of the realizable moment space, otherwise the Hessian matrix will be ill-conditioned.

The Jacobian and Hessian matrices of this function are easily expressed, but they require the numerical computation of the following integrals:

$$\hat{m}_k = \int_0^1 \mu^k \exp\left(-\sum_{i=0}^{P} \varphi_i \mu^i\right) \mathrm{d}\mu \qquad k \in \{0, \ldots, 2P\} \tag{32}$$

These integrals must be evaluated numerically as no analytic form exists as soon as $P > 2$, which is done using the adaptative support quadrature proposed by Vié et al. (2013). The fact that such integrals must be numerically computed, at each step of the Newton-Raphson procedure, which itself is called at each time-step, explains why we marked that method as computationally intensive on Fig. 1. However as moments evolve in a continuous way over time, the $\varphi_i$ will also evolve continuously, and the initial guess of the Newton-Raphson procedure is set as the solution of the previous time-step, leading to a fast convergence.

The number of nodes for the resulting quadrature rule actually depends on the results of the procedure described by Vié et al. (2013). We used a 15 nodes Gauss-Legendre quadrature for each sub-intervals identified by their procedure. The number of sub-interval, $s$, is variable depending on the $\varphi_i$ values: $s \in \{1, \ldots, N\}$. Thus, a generic quadrature is given by:

$$\int_{\Omega\mu} f(\mu) n(\mu) \mathrm{d}\mu \approx \sum_{i=1}^{s} \sum_{j=1}^{P'} b_j \frac{x_{\max,i} - x_{\min,i}}{2} f\left(\frac{a_j + 1}{2}(x_{\max,i} - x_{\min,i}) + x_{\min,i}\right) \tag{33}$$

with

- $s$ the number of sub-intervals returned by the procedure described by Vié et al. (2013) ($s \leqslant N$),
- $x_{\min,i}$ and $x_{\max,i}$ the minimum and maximum limits of the $i$-th sub-interval,
- $a_j$ and $b_j$ the nodes and weights of a Gauss-Legendre quadrature of order $P'$.

Finally, in our following simulations, we did encounter cases where the moment set was too close from the frontier of the moment space which led to ill-conditioned Hessian matrices. We first performed the reconstruction on the support $[0; K * \mu_{\max}]$ with $K = 1.5$, however we observed that our distributions only span a tiny fraction of this interval at each time. This often led to moment sets whose last moment were close to their upper or lower bound in the moment space (we underlined this by calculating the canonical moments using the QD algorithm from Dette and Studden (1997)). We then decided to adapt dynamically the value of $K$ between 0 and 2 in order to stretch the support of the reconstruction so that the moments of the distribution are always far enough from the frontier of the moment space, which then allows a fast and accurate convergence of the MaxEnt method.

Population balance equation:

$$\frac{\partial n(\mu)}{\partial t} = -\frac{\partial}{\partial \mu}[n(\mu)\zeta(\mu)] + \int_0^{+\infty} \beta(\mu,\mu')n(\mu')\psi\mu' d\mu'$$

| | Method of classes<br>$X_j, j \in \{0,\dots,NC\}$ | Method of Moments<br>$m_k, k \in \{0,\dots,N-1\}$ | | |
|---|---|---|---|---|
| | | QMOM | EQMOM* | MaxEnt* |
| NDF approximation<br>$\hat{n}(\mu)$ | $\sum_{i=1}^{NC} \frac{X_j}{\Delta\mu}\delta(\mu-\mu_j)$ | $\sum_{i=1}^{N/2} w_i\delta(\mu-L_i)$ | $\sum_{i=1}^{\lfloor N/2 \rfloor} w_i\kappa(\mu,L_i,\sigma)$ | $\exp\left(\sum_{i=0}^{N-1}\varphi_i\mu^i\right)$ |
| $m_{N+1}$<br>formulation | $\emptyset$ | $\sum_{i=1}^{N/2} w_i L_i^{N+1}$ | $\int \hat{n}(\mu)\mu^{N+1}d\mu$ | $\int \hat{n}(\mu)\mu^{N+1}d\mu$ |
| Source term<br>(eq. 2) | $\Sigma_j X_j\Phi(\mu_j)$ | $\sum_{i=1}^{N/2} w_i\Phi(L_i)$ | $\int \hat{n}(\mu)\Phi(\mu)d\mu$ | $\int \hat{n}(\mu)\Phi(\mu)d\mu$ |
| Number ODE | $NC$ | $N = 2P$ | $N = 2P+1$ | $N = P+1$ |
| Number of $f(\mu)$ evaluations for approximating $\int f(\mu)n(\mu)d\mu$ | $NC$ | $P$ | $P \cdot P'$ | $\leq N\,P'$ |

**Fig. 1.** Summary of applied methods to couple the population balance with transport and reaction. *Numerically expensive methods. $P$: Order of moment method (positive integer). $P'$ : Order of nested quadrature (we use $P' = 10$ (EQMOM) or $P' = 15$ (MaxEnt)).

The rules for the evolution of $K$, from time step $(n)$ to timestep $(n+1)$ are based on the value of the last canonical moments $p_P \in [0;1]$ computed from the set of known moments $m_0,\dots,m_P$:

- If $p_P^{(n)} < 0.4 : K^{(n+1)} = 0.96 * K^{(n)}$.
- If $p_P^{(n)} > 0.6 : K^{(n+1)} = 1.04 * K^{(n)}$.

This proposition is most probably not universal and might only work in our specific application cases.

### 2.5. Simulation software

All following simulations are performed using ADENON, a user-friendly simulation software we developed using the environment provided by MATLAB R2016a. This software is mainly focused on the simulation of bioreactors, by applying our PBM/Metabolic biological models within a hydrodynamic framework (compartment models, plug-flow reactors, batch or fedbatch cultures as well as accelerostat cultures). Population balances can be solved using either class or moment methods, with all core routines –for moment quadrature or distribution reconstruction– built into this software.

Following the case configuration provided by the user, this tool formulates the corresponding ODE in terms of mass and volume balances. This set of ODE is then solved using an explicit scheme for time integration, either the Runge-Kutta 2,3 pair of Bogacki and Shampine (1989) or a simple first-order Euler scheme. The specificities of our solver compared to the built-in “ode23” function are (i) its capability of running in parallel (multi-core) mode by distributing the resolved variables and the reconstruction computing across CPU cores, and (ii) the fact that it enforces the consistency of resolved variables (mainly their non-negativity) in a more stringent way.

We used the simple explicit Euler scheme for all simulations, and choose a timestep $\delta t$ tiny enough to make the solution independent from this timestep.

## 3. Results

### 3.1. Stressed chemostat culture

In a first attempt of applying the method of moments with reconstruction of the NDF, we chose to reproduce numerically the experimental results from Kätterer et al. (1986). We simulate a homogeneous chemostat culture with a constant initial dilution rate $D = 0.1\ \text{h}^{-1}$ for 30 h in order to reach a steady-state, we then apply a sudden shift in dilution rate toward $D' = 0.42\ \text{h}^{-1}$ in order to analyse 15 h of the transient-state.

As the original experiments were conducted using *Candida tropicalis* instead of *E. coli*, we adjusted the parameters of our metabolic model to fit quantitatively the biomass and substrate curves provided by Kätterer et al. (1986). It is however obvious that the metabolic behaviours of the yeast *C. tropicalis* and the bacteria *E. coli* are quite different and a mere parameter adjustment of a *E. coli* metabolic model will not produce a model exhibiting the metabolic behaviour of *C. tropicalis*. Here, we are only interested in the analysis of the population balance part of the model. We shall investigate each reconstruction method in terms of stability, computation time, and accuracy of the reconstruction. The shape of the reconstruction will have a metabolic impact in terms of acetate production, and we will only compare these productions between class and moment methods, not against experimental results.

The Fig. 2 shows simulation results for each method, with different orders or resolution. We applied QMOM with order ranging from $P = 2$ to $P = 5$ $(N = 2P)$, EQMOM with order ranging from $P = 1$ to $P = 3$ $(N = 2P + 1)$ and MaxEnt with $P$ ranging from 2 to 6 $(N = P + 1)$.

The overall dynamics are well reproduced by each method, even with as few as two nodes with QMOM, even though that last method gives noticeably different results depending on its order. As explained before, the overall dynamics does not depend directly on the redistribution term of the PBE (Eq. (11)) but mainly on the

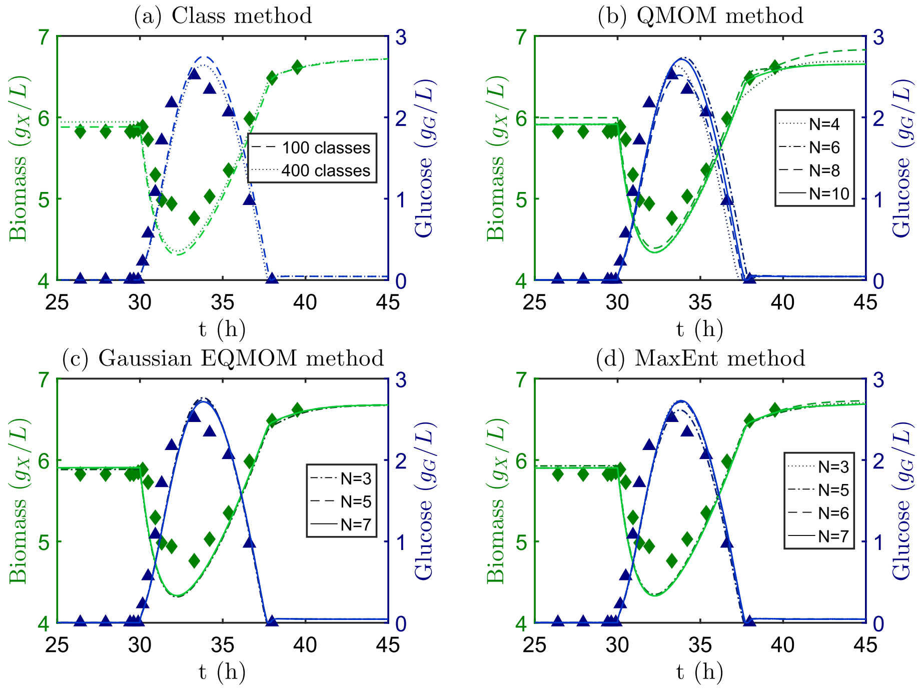


**Fig. 2.** Simulation results for each method, compared with experimental data from Kätterer et al. (1986).

adaptation term. The moment formulation of this term needs a closure method to estimate the next unsolved moment, so as long as this estimation is reasonably accurate, the dynamics should be well reproduced.

We then assessed the error on the estimation of the next unknown moment for each method and order by comparing them to the moments calculated with 400 classes. Full data set is provided as supplementary data. It is shown that the error is mainly kept under 0.2%.

In terms of shape of the reconstruction, we can use the same data set to compare the original distribution solved with the class method to the reconstructions as illustrated on Fig. 3.

The shape of the reconstruction has two main effects. It affects the biomass concentration at steady-state due to the Pirt law which changes the yield of substrate conversion to biomass depending on the property $\mu$ of each individual. The population mean conversion yield will then depend on that shape, which explains why steady-state biomass concentrations are order-dependent for QMOM (Fig. 2b). However, as the resulting reconstructions are quite similar with EQMOM and MaxEnt (Fig. 3), no matter the order of the method, they always predict similar steady-state biomass concentrations.

The second effect is the metabolic behaviour. As stated before, our metabolic model does not represent the actual metabolic beha-

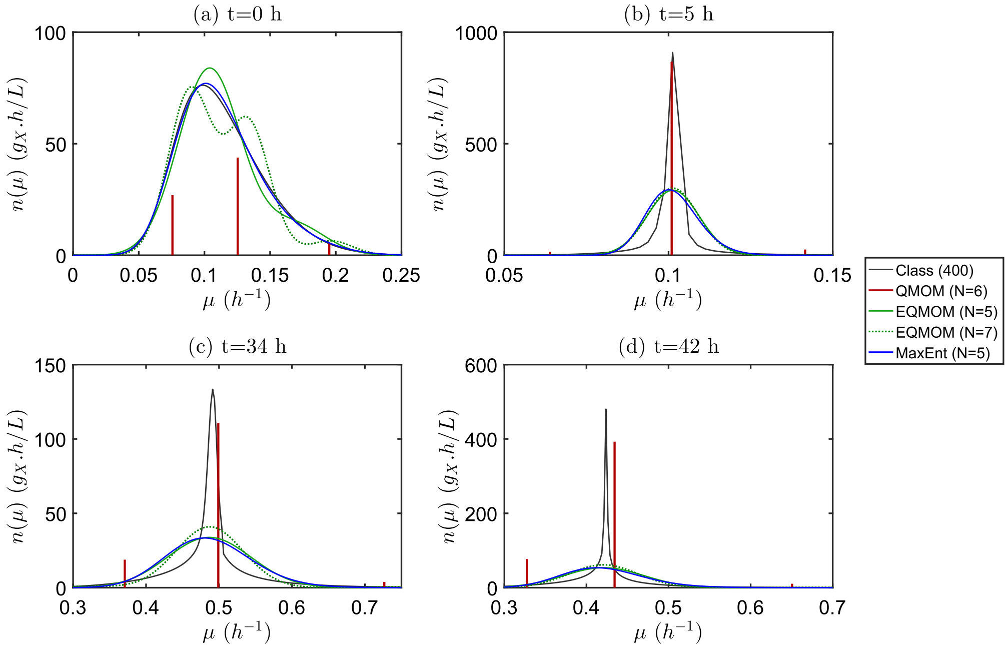
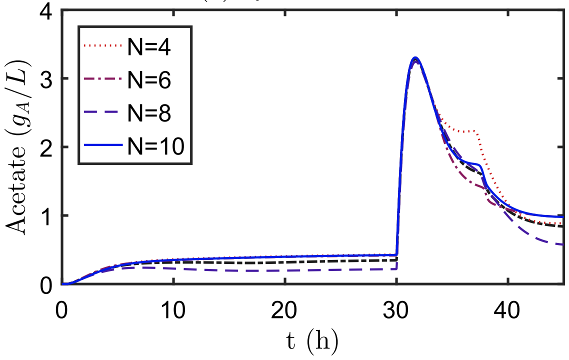
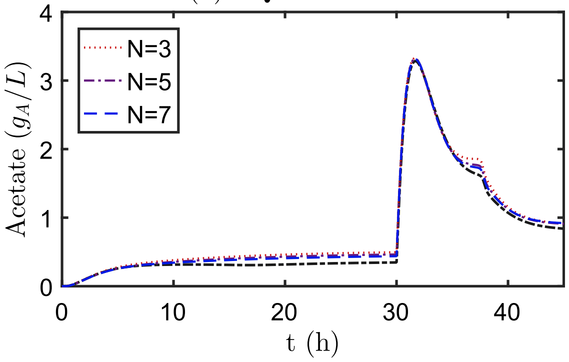


**Fig. 3.** Comparison of reconstructed distributions against distribution resolved by class. An arbitrary scale is used for the Dirac distribution (QMOM).

(a) QMOM method



(b) EQMOM method



(c) MaxEnt method

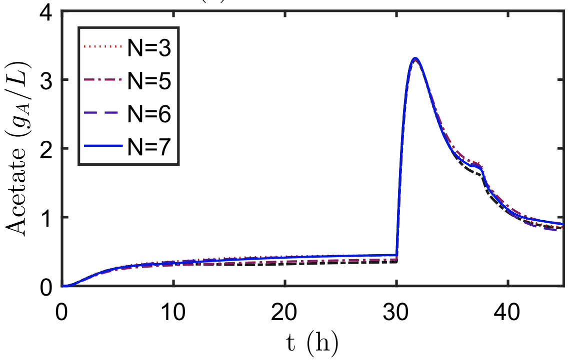


**Fig. 4.** Evolution of acetate concentrations as predicted by the *E. coli* metabolic model. Black dash-dotted line: results from class method.

viour of *Candida tropicalis*, however, it describe the overflow metabolism existing in *E. coli* which leads to acetate production in a way that depends on the shape of the distribution. Fig. 4 illustrates these different acetate productions depending on the chosen method. Once again, QMOM exhibits different behaviours depending on the order of the method, while EQMOM and MaxEnt lead to predictions close to the class method.

Acetate production is slightly overestimated by all moment methods (Fig. 4), due to the fact that they do not account for the narrow peak of the distribution (Fig. 3b,c,d). This slightly overestimates the disequilibrium between the individuals and their environment, which is a key point in our modelling: the disequilibrium between the cell uptake of substrate and its requirements for growth determines the intensity of the overflow metabolism (i.e. the production of by-products, here the acetate).

In terms of simulation performances, the Fig. 5 details the mean computation time spent on each time-step of the simulation. A blank simulation –ran without computing the terms related to bioreaction or population balance– is shown in order to estimate and distinguish the actual models computation time from the time spent on other tasks in the software.

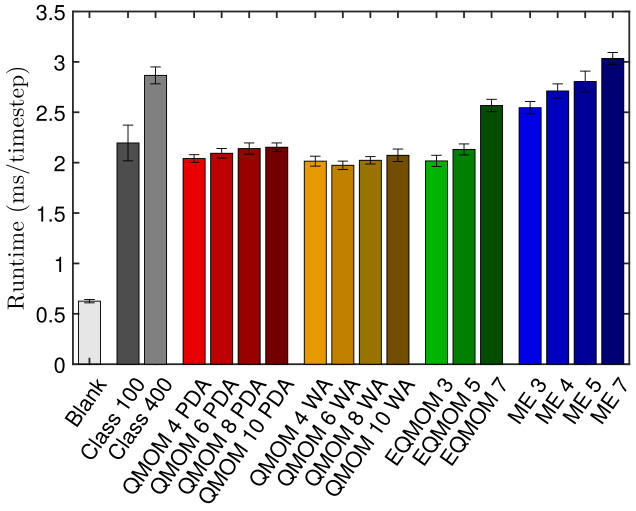


**Fig. 5.** Mean run-time per timestep for each method and different orders (ms/ts) (± standard deviation measured on 20 simulations per method and order).

The class method is a direct one, the computation time is mainly spent on (i) computing the metabolic model from Pigou and Morchain (2015) for each class and (ii) computing the redistribution term of the PBE as detailed in Morchain et al. (in press) for each class, which implies computing $NC + 1$ values of the Owen's T function using 10-nodes Gauss-Legendre quadratures (with $NC$ the number of classes).

All method of moments must compute the first N moments of the skewnormal distribution which is not expensive considering that their expressions are available. The major computational cost then comes from (i) establishing the quadrature rule and (ii) computing the metabolic model for each node of the quadrature.

In this regard, QMOM is the least expensive method: the quadrature rule is computed using directly either the Product-Difference Algorithm (PDA) or the Wheeler Algorithm (WA), both consisting in computing the eigenvalues and eigenvectors of a particular $N/2 \times N/2$ matrix, and computing the metabolic model for $N/2$ nodes. The WA seems to be slightly faster than the PDA.

In order to establish a quadrature rule with the EQMOM method, Marchisio and Fox (2013) detail the analytical solution for $N = 3$ and a solution whose cost is hardly higher than a 2 nodes QMOM for $N = 5$, which explains the low computation times for these two orders of resolution. The case $N = 7$ needs an iterative algorithm to find the suited quadrature rule, based on a dichotomic method. We speed-up that method by making use of the result from the previous timestep, the dichotomic algorithm then converges most of the time in 3 to 6 evaluation of the objective function, each of which requiring a single call to the WA.

Finally, MaxEnt is the most expensive method. It is actually as fast as QMOM and EQMOM when the moment set is far from the frontier of the moment space, but our model often produces moment sets near the frontier. Then, we slow down the method by using different tweaks in order to stabilise it: (i) the adaptive quadrature proposed by (Vié et al., 2013), (ii) the dynamic adaptation of the distribution's support and (iii) the computation of canonical moments to check realizability of the moment set. The underlying Newton-Raphson procedure often converges in a single iteration, but this number increases up to 10 for many time-steps after the dilution rate shift, so simulating the next few hours following this shift is actually as long as computing the rest of the time range.

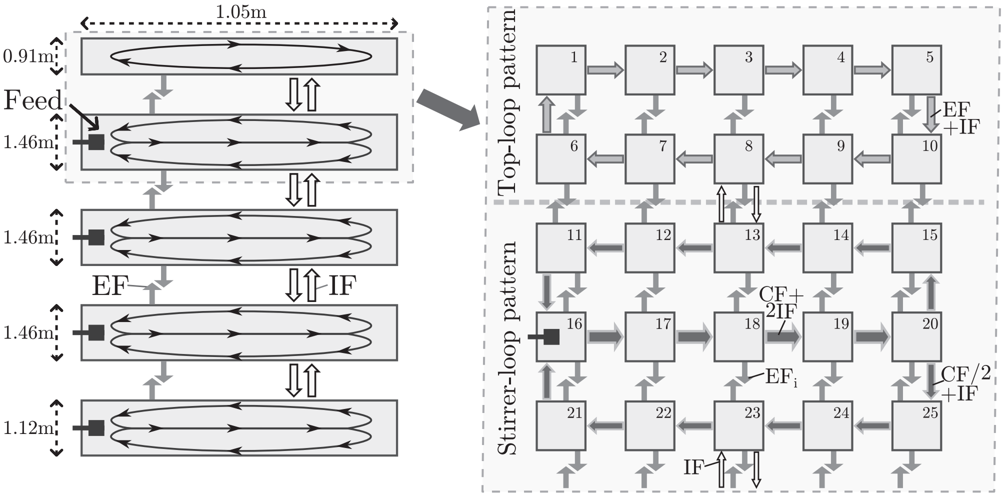


**Fig. 6.** Representation of the macroscopic flow patterns in the fedbatch reactor (left) and details of its compartmentalization and specific flows for the top of the reactor (right). Values for the flow rates CF, IF and EF are given in the appendix B of Pigou and Morchain (2015).

### 3.2. Fedbatch culture - Vrabel et al.

We simulated the very same fedbatch culture described with a 70 compartments hydrodynamic model by Vrábel et al. (2001) but using our own biological modelling as detailed in Pigou and Morchain (2015). Here, we reproduce these simulations, by using the methods of moments to solve the population balance model. The QMOM method is applied with 5 nodes ($N = 10$) as this seemed to be required to produce the same results than the EQMOM and MaxEnt methods (Figs. 2 and 4). The MaxEnt method is used with

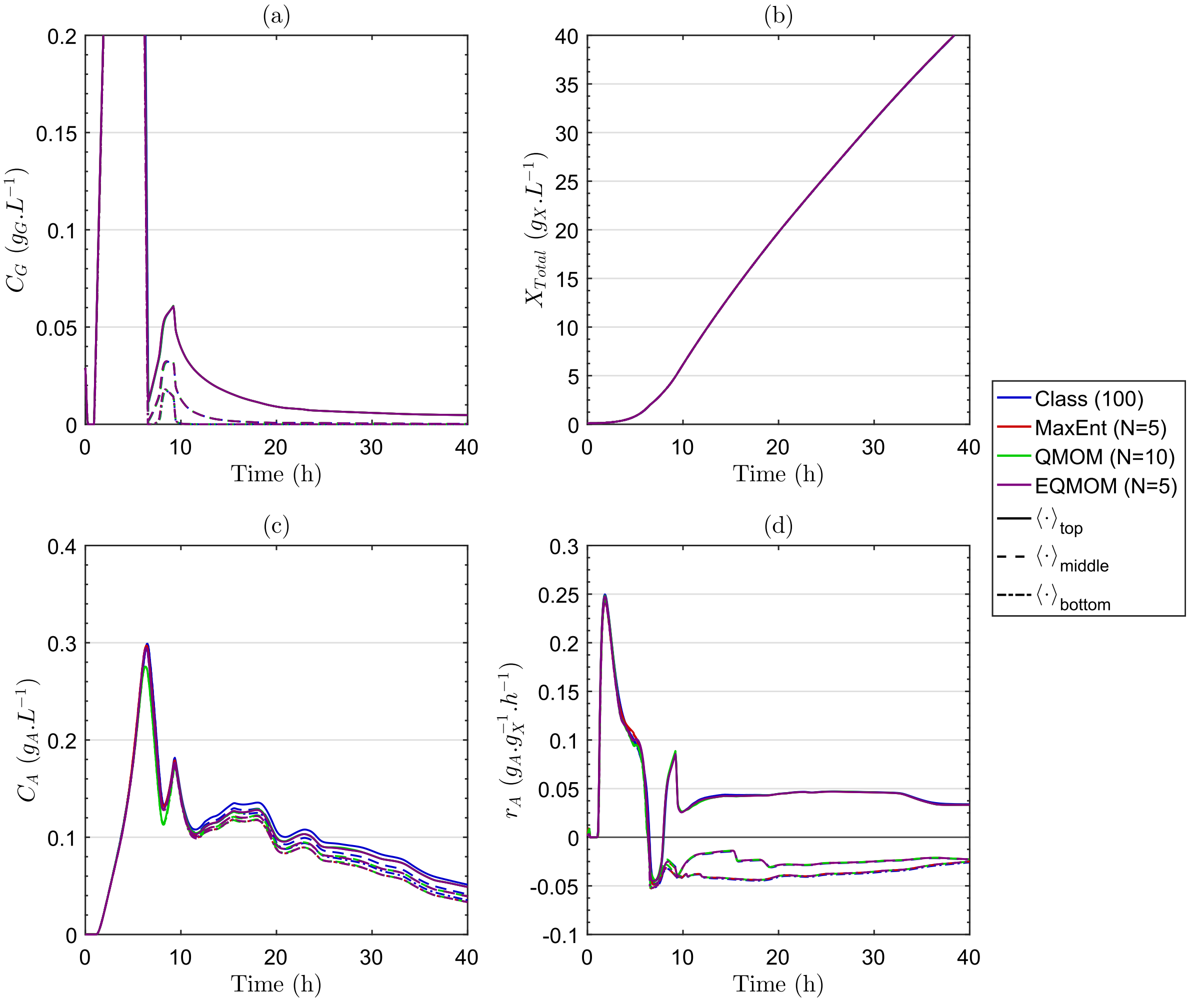


**Fig. 7.** Simulation results for the different population balance methods in the heterogeneous fedbatch culture. (a) Glucose concentration, (b) Total biomass concentration, (c) Acetate concentration and (d) Acetate specific production rate.

$N = 5$ moments as we did not manage to increase the number of moments up to 7 is this setup due to stability issues and also because, surprisingly, the prediction of acetate production was actually better with 5 moments than with 7 moments (see Fig. 4c). Finally, EQMOM was also applied with $N = 5$ moments as going up to 7 moments did not increase the precision drastically (Fig. 4b) but did increase significantly the computation time (Fig. 5). This will also make the comparison between MaxEnt and EQMOM more relevant.

In order to enforce the consistency of numerical results, we limit the maximum value of the time-step to the minimum compartment mean-residence-time:

$$\delta t_{\max} = \min_{n=1}^{N_c}\left(\frac{V_n}{\sum_{m=1}^{N_c} Q_{n,m}}\right) \tag{34}$$

The value of the maximum time-step for the compartment model shown in Fig. 6 is $\delta t_{\max} = 1.4710^{-5}$ h.

The Fig. 7 gathers the results in term of glucose, total biomass and acetate concentrations as well as mean population reaction rates. The plotted values are mean values at different heights (volumetric mean value over compartments of the same row). The three heights (top, middle and bottom) correspond to the following compartments (see Fig. 6 for numbering):

- top: compartments 11–15,
- middle: compartments 36–40,
- bottom: compartments 61–65.

The good agreements between the methods is related to the fact that, in the heterogeneous large-scale reactor, the distribution is continuously perturbed by external fluctuations which prevent the apparition of the narrow distribution seen previously (Fig. 3). The expected distribution has a smoother shape which is well reconstructed by MaxEnt and EQMOM as shown in Fig. 8.

Finally, we ran 5 times the first hour of simulation in order to gather statistics about simulation runtimes in the heterogeneous case with different orders of resolution. The results are shown in Fig. 9.

Each CPU core had to perform calculations for 14 compartments in Fig. 9 while a single compartment was considered in Fig. 5 which explains the overall higher computation times. However, the previous analysis about the comparison of the complexity of each method remains the same, and the observations on the heterogeneous case are the same than in the homogeneous case: QMOM maintains a constant computation time, EQMOM is as fast as QMOM as long as $N \leq 5$ and MaxEnt is slower than other methods due to the stabilisation of the method.

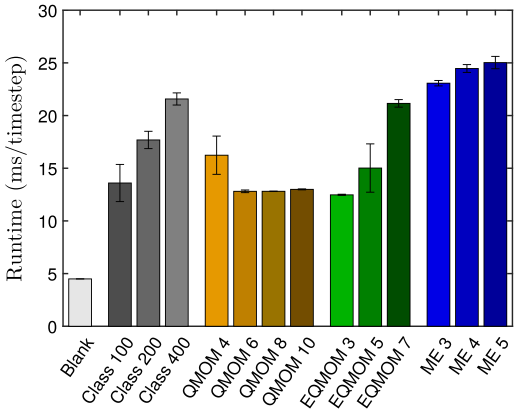


**Fig. 9.** Comparison of simulation runtimes for each method ($ms/ts$) (± standard deviation measured on 5 simulations per method and order). Simulations performed using 5 CPU cores.

## 4. Discussion

Dealing with a biological phase naturally leads to the use of the "population" semantic field due to the individual nature of cells, each of which having its own set of properties and its own "memory". Hence, the use of a Population Balance Model to describe a biological population seems to be obvious, almost axiomatic.

The most natural way to solve a PBM is the class method, which constitutes a direct resolution of the equations. However, its accuracy comes with the price of a high memory cost. Unpublished data show that for simple batch and chemostat simulations, the results are dependent on the number of classes up to 60 classes, and we can even notice differences between 100 and 400 classes in Fig. 2a. This number of classes is needed to span the entire property space with sufficient accuracy, however simulations clearly show that most of the time a large fraction of classes are nearly empty. This means that we allocate memory for variables that most of the time carry almost no information, but still happen sometime to be used, depending on the state of the population.

This explains why we are shifting toward methods of moments. They resolve basic properties of the distribution (total number, mean, variance, skewness, flatness, ...) which all contain useful information no matter the state of the population. Moments gather higher entropy about the distribution than classes, in the sense of

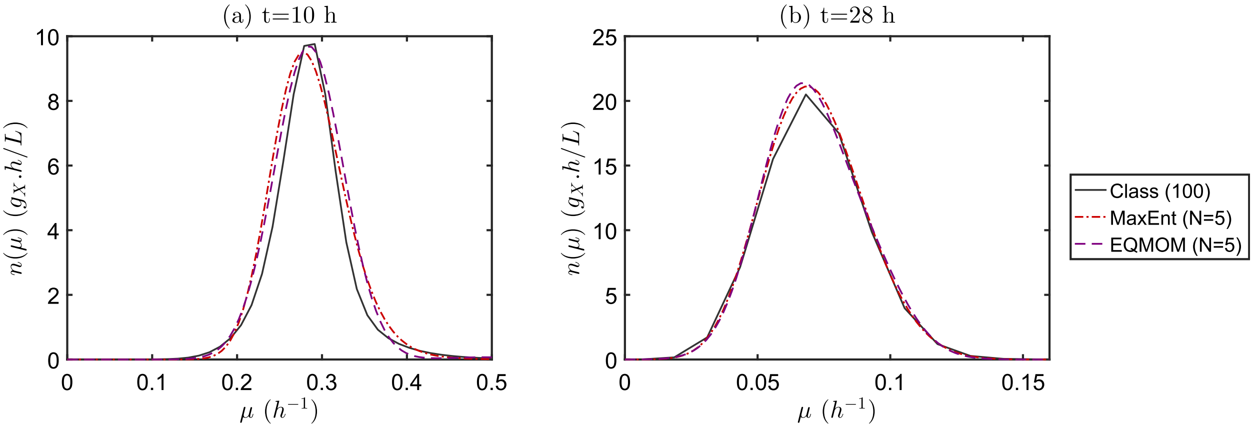


**Fig. 8.** Comparison of distribution shapes as resolved by the class method and reconstructed by the EQMOM and MaxEnt methods both with $N = 5$ moments.

information theory. This significantly reduces the required number of resolved variables, from more than 60, to half a dozen.

For some applications, each equation of the model can be formulated in terms of moments (Hulburt and Katz, 1964; McGraw, 1997) leading to closed formulations or easy closure through the used of quadrature based methods. In these cases, the accuracy of the methods of moments poses no question –and is even better than a class method considering that the latter induces numerical diffusion in the parameter space– for a smaller memory and computational cost.

Unfortunately, we deal here with a metabolic complexity that offers no model formulation in terms of moments. We tackle this issue by using reconstruction methods, namely QMOM, EQMOM and MaxEnt methods; but these methods introduce quadrature inaccuracy as well as extra computational cost, which we tried to quantify in our simulations.

The results in terms of concentration fields are really promising, as the biomass, glucose and acetate concentrations were well reproduced both in the homogeneous and the heterogeneous cases. We mainly noticed slight errors consisting in an overproduction of acetate with moment methods in the Katterer case (see Fig. 4) due to difficulties in reconstructing a narrow distribution with a wide base (similar to a Laplace distribution). However, we are shifting toward moment methods only to perform large-scale simulations at low memory cost: the class method performs just fine enough for homogeneous cases. The crucial comparison must then be made on the large-scale simulation.

In Fig. 7, we observe a surprisingly good agreement between the class method and the moment methods. The only noticeable difference is seen on acetate curves for the moment methods which slightly underestimate the acetate production for $t \approx 10$ h which induces a persistent shift along time when compared to the class method. On this regard, the accuracy for variables whose value is of importance (substrate and product concentrations) is satisfactory for all methods.

In terms of computational cost, which we evaluate through the simulation time, Yuan et al. (2012) already performed a comparison of EQMOM and the MaxEnt methods. They observed that EQMOM was a hundred times faster than MaxEnt for the reconstruction of two NDF. However, we do not feel that their comparison is fair: the slow convergence of MaxEnt is only due to a bad initialization of the Newton-Raphson procedure, similar to what we observe in our simulations for the very first time-step. We want to supplement their observations by pointing out that when used in time resolved simulations, which happens to be our specific application for these methods, the MaxEnt method performs only slightly slower than EQMOM (see Figs. 5 and 9) due to different adjustments made in order to improve stability.

We do not develop nor try to promote a specific reconstruction method, but only want to draw general guidelines about which method should be used for the simulation of large-scale bioreactors. We then did our best to keep the comparison of the methods as fair as possible. On the basis of our results, here are the key observations we made about the different methods.

EQMOM is a stable method: it behaves well near the moment space frontier and we did not notice any particular difficulty when increasing the number of resolved moments. When used on mono-modal distribution, it can be applied at a low computational cost with 3 or 5 moments, thanks to analytical solutions (Marchisio and Fox, 2013). The possibility to increase the number of resolved moments, without loss of stability, means that this method is also well suited to reconstruct multi-modal distributions, this will be useful when tracking fast population dynamics in heterogeneous systems. Moreover, the method naturally embeds the feature of nested quadratures: a relevant Gauss-Hermite quadrature can be constructed on each Gaussian node of the EQMOM reconstruction, which helps performing efficiently the integration of Eq. (2), compared to MaxEnt. This can be an important advantage of this method if the metabolic model (computation of $\boldsymbol{\Phi}$) is computationally expensive. Here, the metabolic model was easy enough to compute, which hide this salient feature of EQMOM reconstructions.

The same level of accuracy than EQMOM with 5 moments could be reached with QMOM using 10 moments with a similar computation time, but with less calls to the metabolic model. Both methods can then be used based on whether we try to reduce the memory usage or the number of calls to a potentially complex metabolic model. Similarly, as shown in A, if the redistribution law $\beta(\mu, \mu')$ did not allow such a simple moment formulation (Eq. (20)), the computational cost of EQMOM or MaxEnt to compute the redistribution term would be significantly higher than QMOM. This is directly related to the number of nodes returned by each method for a generic integral approximation (see Fig. 1, Eqs. (25), (28) and (33)).

MaxEnt is known to be ill-conditioned near the boundaries of the moment space (Massot et al., 2010) or when the number of resolved moments increases (Gzyl and Tagliani, 2010), in particular we did not manage to perform simulations using this method and more than 7 resolved moments, or even 5 moments in the heterogeneous case. This comes from the fact that we describe quite narrow distributions on a large support, which naturally correspond to moments near the frontier of the moment space. When the method is working, it might give a better reconstruction with 5 moments than EQMOM with 7 moments (Fig. 3a), however, using this reconstruction to construct a relevant quadrature rule is more difficult than for EQMOM. We then recommend its use when assessing slow dynamics (preserving mono-modal distribution in heterogeneous systems) for simulations where the time-step is negligible compared to the characteristic time of moment evolution (to ensure a good initialization), and only if a method can be designed to form a set of moment far enough from the frontier of the moment space. These are quite restrictive conditions which do not make this method the more advisable.

Finally, in terms of memory footprint, we managed to reduce the number of resolved variables to describe the population from about a hundred (class method) to about only 5 variables which will be significant when moving toward CFD simulations of bioreactors. However, it should be noted that for MaxEnt, memory registers must be allocated both for the transported moments and for the vector of polynomial coefficients $\boldsymbol{\varphi}$ which serves as initial value for the Newton-Raphson procedure. MaxEnt then requires twice as much memory space than QMOM and EQMOM methods for equivalent number of resolved moments.

The significant improvements in terms of memory usage will be even more significant when we will shift toward multivariate population balance models. For a bivariate distribution, the number of classes or moments will roughly be squared leading to about $10^4$ classes opposed to 25 moments in each geometrical node.

## 5. Conclusions

The point of applying population balance based modelling for the predictive simulation of heterogeneous bioreactors is now well established (Morchain et al., 2014; Morchain et al., in press; Pigou and Morchain, 2015; Heins et al., 2015; Bertucco et al., 2015; Fredrickson and Mantzaris, 2002). This paper is more focused on numerical methods to solve the population balance model, in order to shift from a class method to moment based methods. In our modelling, the reconstruction methods of a NDF from a finite set of moments is required for the computation of the population metabolic behaviour. We then implemented QMOM, EQMOM and

Maximum Entropy methods, and challenged them in terms of stability, memory footprint, computational cost and accuracy against class method results.

At equivalent number of resolved moments, QMOM is noticeably less accurate than EQMOM and MaxEnt. However, increasing the number of moments for QMOM does not increase significantly the computation time, which make this method competitive with the others when looking only at accuracy and simulation runtime.

If reducing the memory footprint is the main concern, EQMOM actually reaches the same accuracy than QMOM with half the number of resolved moments. However, its computational cost increases significantly between 5 and 7 resolved moments, due to the need of an iterative procedure rather than an analytical or direct solution.

Depending on the use case, MaxEnt has often been reported as an interesting method (Massot et al., 2010; Vié et al., 2013; Lebaz et al., 2016), however, we will tend to discard it for our future works. Indeed, even when the method is well-conditioned, it is not particularly competitive with EQMOM in terms of computation time and accuracy, but comes with twice the memory usage of EQMOM. Moreover, the method tends to be quickly unstable if moments are near the limit of realizability. This poses problem for our modelling as the lack of experimental data about the dynamics of internal biological properties will make us formulate models which tend toward narrow distributions until data are available. An example of that is the PBM from Pigou and Morchain (2015) which led to a Dirac distribution in steady-state homogeneous systems until experimental data from Nobs and Maerkl (2014, 2011) allowed us to improve the PBM and add an experimentally justified redistribution term as explained in Morchain et al. (in press).

Overall, the QMOM and EQMOM methods have shown to be accurate and stable enough for the simulation of a large scale bioreactor with a significantly reduced memory impact and a simulation time of the same order of magnitude than the class method.

## Acknowledgments

The authors thank Sanofi Chimie - C&BD Biochemistry Vitry for its financial support. The authors declare no conflict of interest.

## Appendix A. Moment formulation of the PBE

As a recall, the population balance equation is defined as

$$\frac{\partial n(\mu)}{\partial t} = -\frac{\partial}{\partial \mu}[n(\mu)\zeta(\mu)] + \int_0^{+\infty} \beta(\mu,\mu')n(\mu')\Psi\mu'\mathrm{d}\mu' \tag{A.1}$$

and the $k$-th order moment of the distribution $n(\mu)$ is defined by

$$m_k = \int_0^{+\infty} \mu^k n(\mu)\mathrm{d}\mu \tag{A.2}$$

We want to formulate the law of moments evolution, as the sum of contributions from an adaptation term $\left(\frac{\partial m_{a,k}(t)}{\partial t}\right)$, and a growth term $\left(\frac{\partial m_{g,k}(t)}{\partial t}\right)$:

$$\frac{\partial m_k(t)}{\partial t} = \frac{\partial m_{a,k}(t)}{\partial t} + \frac{\partial m_{g,k}(t)}{\partial t} \tag{A.3}$$

The formulation of $\frac{\partial m_{a,k}(t)}{\partial t}$ comes by multiplying the first RHS term of Eq. (A.1) by $\mu^k$ and integrating by part with respect to $\mu$:

$$\frac{\partial m_{a,k}(t)}{\partial t} = -\int_{\Omega\mu} \mu^k \frac{\partial}{\partial \mu}(n(\mu,t)\cdot\zeta(\mu))\mathrm{d}\mu \tag{A.4}$$

$$= \int_{\Omega\mu} k\mu^{k-1} n(\mu,t)\zeta(\mu)\mathrm{d}\mu - \left[\mu^k n(\mu,t)\zeta(\mu)\right]_{\partial\Omega\mu} \tag{A.5}$$

Considering that the adaptation will not allow individuals to cross the frontier of the $\mu$-space ($\partial\Omega\mu$), the second term of Eq. (A.5) is necessarily null. By expanding the formulation of $\zeta(\mu) = \left(T^{-1} + \mu\right)(\mu^* - \mu)$, the formulation of $\frac{\partial m_{a,k}(t)}{\partial t}$ in terms of moments of the distribution is trivial:

$$\frac{\partial m_{a,k}(t)}{\partial t} = k\left(\frac{\mu^*}{T} m_{k-1}(t) + \left(\mu^* - \frac{1}{T}\right) m_k(t) - m_{k+1}(t)\right) \tag{A.6}$$

The contribution of the growth term to the moment evolution depends both on the used quadrature method and on the propability function modeling the redistribution phenomena related to cell division. We have

$$\frac{\partial m_{g,k}(t)}{\partial t} = \int_{\Omega\mu} \mu^k \left[\int_{\Omega\mu} \beta(\mu,\tilde{\mu})\Psi\mu' n(\mu')\mathrm{d}\mu'\right]\mathrm{d}\mu \tag{A.7}$$

The approximations of $n(\mu)$ by the methods of moments lead to

$$\int_{\Omega\mu} f(\mu)n(\mu)\mathrm{d}\mu \approx \sum_{i=1}^{I} w_i * f(L_i) \tag{A.8}$$

where $I, w_i$ and $L_i$ depend on the methods used to perform the quadrature of moments (Eqs. (25), (28) and (33)).

Using these quadratures, we reach the following expression:

$$\frac{\partial m_{g,k}(t)}{\partial t} \approx \Psi \int_{\Omega\mu} \mu^k \left(\sum_{i=1}^{I} w_i L_i \beta(\mu, L_i)\right)\mathrm{d}\mu \tag{A.9}$$

$$\approx \Psi \sum_{i=1}^{I} w_i L_i \int_{\Omega\mu} \mu^k \beta(\mu, L_i)\mathrm{d}\mu \tag{A.10}$$

This last formulation is generic and can be used for any redistribution law $\beta(\mu,\mu')$. In the case of current simulations, we base this term on the previous work from Morchain et al. (in press) were we identified the following probability density function as a good model for experimental data from the literature (Nobs and Maerkl, 2014; Yasuda, 2011):

$$\beta(\mu,\mu') = \frac{2}{\sigma}\phi\left(\frac{\mu - l}{\sigma}\right)\Phi\left(\alpha \times \frac{\mu - l}{\sigma}\right) \tag{A.11}$$

with:

- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,
- $\Phi(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$.

The redistribution law is a skew-normal distribution whose parameters depend on the population mean growth rate $\tilde{\mu}$ but not on the growth rate of the mother cell (hence, $\mu'$ is not used in the expression):

$$\tilde{\mu} = \frac{m_1(t)}{m_0(t)} \tag{A.12}$$

$$l = k_l \tilde{\mu} \tag{A.13}$$

$$\sigma = k_\sigma \tilde{\mu} \tag{A.14}$$

$$\alpha = 3.65 \tag{A.15}$$

The constants $k_l$ and $k_\sigma$ were chosen so that the PDF $\beta(\mu,\mu')$ fits experimental data, but also with the constraint that the first moment of this PDF is equal to $\tilde{\mu}$ so that the redistribution term will have no impact on the population mean growth rate. The used values are:

Then, the growth related evolution of the distribution moments is easily expressed in terms of moments of $\beta(\mu,\mu') = \beta(\mu)$ and does not require the moment quadrature:

$$\frac{\partial m_{g,k}(t)}{\partial t} = \int_{\Omega\mu} \left[ \mu^k \int_{\Omega\mu} \beta(\mu)\Psi\mu' n(\mu')\mathrm{d}\mu' \right] \mathrm{d}\mu \tag{A.16}$$

$$= \Psi \int_{\Omega\mu} \mu^k \beta(\mu) \left[ \int_{\Omega\mu} \mu' n(\mu')\mathrm{d}\mu' \right] \mathrm{d}\mu \tag{A.17}$$

$$= \Psi \int_{\Omega\mu} \mu^k \beta(\mu) m_1(t)\mathrm{d}\mu \tag{A.18}$$

$$= \Psi m_1(t) B_k \tag{A.19}$$

$B_k$ is the $k$-th order moment of the PDF $\beta(\mu)$ which happens to only depend on $\tilde{\mu}$ whose value is accessible using the first two moments of the distribution (Eq. (A.12)). The moments $B_k$ can be determined analytically using the Moment Generating Function of the skew-normal distribution:

$$B_k = \frac{\partial^k M}{\partial t^k}(0) \tag{A.23}$$

$$M(t) = \exp\left(lt + \frac{\sigma^2 t^2}{2}\right)\left(1 + \mathrm{erf}\left(\frac{\sigma \alpha t}{\sqrt{2(1+\alpha^2)}}\right)\right) \tag{A.24}$$

We used the MATLAB Symbolic Toolbox to pre-compute the expressions of $B_k, \forall k \in \{0, \ldots, 9\}$.

Note that in order to respect the constraint $B_1 = \tilde{\mu}, k_\sigma$ and $k_l$ must satisfy the following relationship:

$$k_l + \sqrt{\frac{2\alpha^2}{\pi(1+\alpha^2)}} k_\sigma = 1 \tag{A.25}$$

## Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ces.2017.05.026.

## References

Abramov, R.V., 2007. An improved algorithm for the multidimensional moment-constrained maximum entropy problem. J. Comput. Phys. 226, 621–644. http://dx.doi.org/10.1016/j.jcp.2007.04.026.

Abulesz, E.M., Lyberatos, G., 1989. Periodic operation of a continuous culture of baker's yeast. Biotechnol. Bioeng. 34, 741–749. http://dx.doi.org/10.1002/bit.260340603.

Athanassoulis, G., Gavriliadis, P., 2002. The truncated hausdorff moment problem solved by using kernel density functions. Probab. Eng. Mech. 17, 273–291. http://dx.doi.org/10.1016/S0266-8920(02)00012-7.

Baldyga, J., Bourne, J.R., 1999. Turbulent Mixing and Chemical Reactions. John Wiley and Sons.

Bertucco, A., Sforza, E., Fiorenzato, V., Strumendo, M., 2015. Population balance modeling of a microalgal culture in photobioreactors: comparison between experiments and simulations. AIChE J. 61, 2702–2710. http://dx.doi.org/10.1002/aic.14893.

Bezzo, F., Macchietto, S., Pantelides, C.C., 2003. General hybrid multizonal/CFD approach for bioreactor modeling. AIChE J. 49, 2133–2148. http://dx.doi.org/10.1002/aic.690490821.

Bogacki, P., Shampine, L., 1989. A 3(2) pair of Runge-Kutta formulas. Appl. Math. Lett. 2, 321–325. http://dx.doi.org/10.1016/0893-9659(89)90079-7.

Chalons, C., Fox, R.O., Massot, M., 2010. A multi-Gaussian quadrature method of moments for gas-particle flows in a LES framework. In: Proceedings of the 2010 Summer Program, Center for turbulence Research. Stanford University, pp. 347–358.

Cui, Y.Q., van der Lans, R.G., Noorman, H.J., Luyben, K.C.A.M., 1996. Compartment mixing model for stirred reactors with multiple impellers. Chem. Eng. Res. Des. 74, 261–271.

Dette, H., Studden, W.J., 1997. The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis. John Wiley & Sons, New York; Chichester.

Elqotbi, M., Vlaev, S., Montastruc, L., Nikov, I., 2013. CFD modelling of two-phase stirred bioreaction systems by segregated solution of the euler-euler model. Comput. Chem. Eng. 48, 113–120. http://dx.doi.org/10.1016/j.compchemeng.2012.08.005.

Enfors, S.O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H.Y., Neubauer, P., van der Lans, R., Luyben, K., Vrábel, P., Manelius, Å., 2001. Physiological responses to mixing in large scale bioreactors. J. Biotechnol. 85, 175–185. http://dx.doi.org/10.1016/S0168-1656(00)00365-5. twenty years of the European Federation of Biotechnology.

Ferenci, T., 1996. Adaptation to life at micromolar nutrient levels: the regulation of Escherichia coli glucose transport by endoinduction and cAMP. FEMS Microbiol. Rev. 18, 301–317. http://dx.doi.org/10.1111/j.1574-6976.1996.tb00246.x.

Ferenci, T., 1999. Regulation by nutrient limitation. Curr. Opin. Microbiol. 2, 208–213. http://dx.doi.org/10.1016/S1369-5274(99)80036-8.

Fredrickson, A., Mantzaris, N.V., 2002. A new set of population balance equations for microbial and cell cultures. Chem. Eng. Sci. 57, 2265–2278. http://dx.doi.org/10.1016/S0009-2509(02)00116-1.

Gzyl, H., Tagliani, A., 2010. Hausdorff moment problem and fractional moments. Appl. Math. Comput. 216, 3319–3328. http://dx.doi.org/10.1016/j.amc.2010.04.059.

Haringa, C., Tang, W., Deshmukh, A.T., Xia, J., Reuss, M., Heijnen, J.J., Mudde, R.F., Noorman, H.J., 2016. Euler-Lagrange computational fluid dynamics for (bio) reactor scale down: an analysis of organism lifelines. Eng. Life Sci. http://dx.doi.org/10.1002/elsc.201600061.

Heavner, B.D., Smallbone, K., Barker, B., Mendes, P., Walker, L.P., 2012. Yeast 5 – an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. BMC Syst. Biol. 6. http://dx.doi.org/10.1186/1752-0509-6-55.

Heavner, B.D., Smallbone, K., Price, N.D., Walker, L.P., 2013. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. Database, 2013. http://dx.doi.org/10.1093/database/bat059.

Heins, A.L., Fernandes, R.L., Gernaey, K.V., Lantz, A.E., 2015. Experimental and in silico investigation of population heterogeneity in continuous Saccharomyces cerevisiae scale-down fermentation in a two-compartment setup. J. Chem. Technol. Biotechnol. 90, 324–340. http://dx.doi.org/10.1002/jctb.4532.

Hulburt, H., Katz, S., 1964. Some problems in particle technology: a statistical mechanical formulation. Chem. Eng. Sci. 19, 555–574. http://dx.doi.org/10.1016/0009-2509(64)85047-8.

John, V., Angelov, I., Öncül, A., Thévenin, D., 2007. Techniques for the reconstruction of a distribution from a finite number of its moments. Chem. Eng. Sci. 62, 2890–2904. http://dx.doi.org/10.1016/j.ces.2007.02.041.

Kapur, J., 1989. Maximum-entropy Models in Science and Engineering. Wiley.

Kätterer, L., Allemann, H., Käppeli, O., Fiechter, A., 1986. Transient responses of continuously growing yeast cultures to dilution rate shifts: a sensitive means to analyze biology and the performance of equipment. Biotechnol. Bioeng. 28, 146–150. http://dx.doi.org/10.1002/bit.260280126.

Kumar, S., Ramkrishna, D., 1996a. On the solution of population balance equations by discretization – I. A fixed pivot technique. Chem. Eng. Sci. 51, 1311–1332. http://dx.doi.org/10.1016/0009-2509(96)88489-2.

Kumar, S., Ramkrishna, D., 1996b. On the solution of population balance equations by discretization – II. A moving pivot technique. Chem. Eng. Sci. 51, 1333–1342. http://dx.doi.org/10.1016/0009-2509(95)00355-X.

Lebaz, N., Cockx, A., Spérandio, M., Morchain, J., 2016. Reconstruction of a distribution from a finite number of its moments: a comparative study in the case of depolymerization process. Comput. Chem. Eng. 84, 326–337. http://dx.doi.org/10.1016/j.compchemeng.2015.09.008.

Li, W., 1982. Estimating heterotrophic bacterial productivity by inorganic radiocarbon uptake: importance of establishing time courses of uptake. Deep Sea Res. Part B. Oceanogr. Literature Rev. 29, 167–172. http://dx.doi.org/10.1016/0198-0254(82)90325-9.

Linkès, M., Fede, P., Morchain, J., Schmitz, P., 2014. Numerical investigation of subgrid mixing effects on the calculation of biological reaction rates. Chem. Eng. Sci. 116, 473–485. http://dx.doi.org/10.1016/j.ces.2014.05.005.

Lu, H., Li, C., Tang, W., Wang, Z., Xia, J., Zhang, S., Zhuang, Y., Chu, J., Noorman, H., 2015. Dependence of fungal characteristics on seed morphology and shear stress in bioreactors. Bioprocess Biosyst. Eng. 38, 917–928. http://dx.doi.org/10.1007/s00449-014-1337-8.

Madadi-Kandjani, E., Passalacqua, A., 2015. An extended quadrature-based moment method with log-normal kernel density functions. Chem. Eng. Sci. 131, 323–339. http://dx.doi.org/10.1016/j.ces.2015.04.005.

Mantzaris, N.V., Daoutidis, P., Srienc, F., 2001. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. Comput. Chem. Eng. 25, 1411–1440. http://dx.doi.org/10.1016/S0098-1354(01)00709-8.

Marchisio, D., Fox, R., 2013. Computational Models for Polydisperse Particulate and Multiphase Systems. Cambridge Series in Chemical Engineering. Cambridge University Press.

Marchisio, D.L., Pikturna, J.T., Fox, R.O., Vigil, R.D., Barresi, A.A., 2003a. Quadrature method of moments for population-balance equations. AIChE J. 49, 1266–1276. http://dx.doi.org/10.1002/aic.690490517.

Marchisio, D.L., Vigil, R., Fox, R.O., 2003b. Quadrature method of moments for aggregation-breakage processes. J. Colloid Interface Sci. 258, 322–334. http://dx.doi.org/10.1016/S0021-9797(02)00054-1.

Marchisio, D.L., Vigil, R.D., Fox, R.O., 2003c. Implementation of the quadrature method of moments in CFD codes for aggregation-breakage problems. Chem. Eng. Sci. 58, 3337–3351. http://dx.doi.org/10.1016/S0009-2509(03)00211-2.

Massot, M., Laurent, F., Kah, D., De Chaisemartin, S., 2010. A robust moment method for evaluation of the disappearance rate of evaporating sprays. SIAM J. Appl. Math. 70, 3203–3234. http://dx.doi.org/10.1137/080740027.

Matsuoka, Y., Shimizu, K., 2013. Catabolite regulation analysis of Escherichia coli for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation. J. Biotechnol. 168, 155–173. http://dx.doi.org/10.1016/j.jbiotec.2013.06.023. special issue: Biotechnology for a healthy and green world.

Mayr, B., Horvat, P., Nagy, E., Moser, A., 1993. Mixing-models applied to industrial batch bioreactors. Bioprocess Eng. 9, 1–12. http://dx.doi.org/10.1007/BF00389534.

McGraw, R., 1997. Description of aerosol dynamics by the quadrature method of moments. Aerosol Sci. Technol. 27, 255–265. http://dx.doi.org/10.1080/02786829708965471.

Mead, L.R., Papanicolaou, N., 1984. Maximum entropy in the problem of moments. J. Math. Phys. 25, 2404–2417.

Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metabolic Eng. 12, 150–160. http://dx.doi.org/10.1016/j.ymben.2009.07.006. metabolic Flux Analysis for Pharmaceutical Production Special Issue.

Morchain, J., Gabelle, J.C., Cockx, A., 2013. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. AIChE J. 59, 369–379. http://dx.doi.org/10.1002/aic.13820.

Morchain, J., Gabelle, J.C., Cockx, A., 2014. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AIChE J. 60, 27–40. http://dx.doi.org/10.1002/aic.14238.

Morchain, J., Pigou, M., Lebaz, N., in press. A population balance model for bioreactors combining interdivision time distributions and micromixing concepts. Biochem. Eng. J. http://dx.doi.org/10.1016/j.bej.2016.09.005. in press.

Nobs, J.B., Maerkl, S.J., 2014. Long-term single cell analysis of S. pombe on a microfluidic microchemostat array. PLoS ONE 9. http://dx.doi.org/10.1371/journal.pone.0093466.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2011. A comprehensive genome-scale reconstruction of Escherichia coli metabolism. Molec. Syst. Biol. 7. http://dx.doi.org/10.1186/1752-0509-6-55.

Perret, C.J., 1960. A new kinetic model of a growing bacterial population. Microbiology 22, 589–617. http://dx.doi.org/10.1099/00221287-22-3-589.

Pigou, M., Morchain, J., 2015. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. Chem. Eng. Sci. 126, 267–282. http://dx.doi.org/10.1016/j.ces.2014.11.035.

Silveston, P., Budman, H., Jervis, E., 2008. Forced modulation of biological processes: a review. Chem. Eng. Sci. 63, 5089–5105. http://dx.doi.org/10.1016/j.ces.2008.06.017. 5TH Unsteady-State Processes in Catalysis: a Special Issue of Chemical Engineering Science.

Tagliani, A., 1999. Hausdorff moment problem and maximum entropy: a unified approach. Appl. Math. Comput. 105, 291–305. http://dx.doi.org/10.1016/S0096-3003(98)10084-X.

Vié, A., Laurent, F., Massot, M., 2013. Size-velocity correlations in hybrid high order moment/multi-fluid methods for polydisperse evaporating sprays: modeling and numerical issues. J. Comput. Phys. 237, 177–210. http://dx.doi.org/10.1016/j.jcp.2012.11.043.

Vrábel, P., van der Lans, R.G., Cui, Y., Luyben, K., 1999. Compartment model approach: mixing in large scale aerated reactors with multiple impellers. Chem. Eng. Res. Des. 77, 291–302. http://dx.doi.org/10.1205/026387699526223.

Vrábel, P., van der Lans, R.G., Luyben, K.C., Boon, L., Nienow, A.W., 2000. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements. Chem. Eng. Sci. 55, 5881–5896. http://dx.doi.org/10.1016/S0009-2509(00)00175-5.

Vrábel, P., van der Lans, R.G., van der Schot, F.N., Luyben, K.C., Xu, B., Enfors, S.O., 2001. CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. Chem. Eng. J. 84, 463–474. http://dx.doi.org/10.1016/S1385-8947(00)00271-0.

Xu, B., Jahic, M., Enfors, S.O., 1999. Modeling of overflow metabolism in batch and fed-batch cultures of Escherichia coli. Biotechnol. Prog. 15, 81–90. http://dx.doi.org/10.1021/bp9801087.

Yasuda, K., 2011. Algebraic and geometric understanding of cells: epigenetic inheritance of phenotypes between generations. In: Müller, S., Bley, T. (Eds.), High Resolution Microbial Single Cell Analytics, Advances in Biochemical Engineering/Biotechnology, vol. 124. Springer, Berlin, Heidelberg, pp. 55–81. http://dx.doi.org/10.1007/10_2010_97.

Yuan, C., Laurent, F., Fox, R., 2012. An extended quadrature method of moments for population balance equations. J. Aerosol Sci. 51, 1–23. http://dx.doi.org/10.1016/j.jaerosci.2012.04.003.

# C

# NEW DEVELOPMENTS OF THE EXTENDED QUADRATURE METHOD OF MOMENTS TO SOLVE POPULATION BALANCE EQUATIONS

Journal of Computational Physics 365 (2018) 243–268

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

# New developments of the Extended Quadrature Method of Moments to solve Population Balance Equations

Maxime Pigou [a,b,*], Jérôme Morchain [a], Pascal Fede [b], Marie-Isabelle Penet [c], Geoffrey Laronze [c]

[a] LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France
[b] Institut de Mécanique des Fluides de Toulouse – Université de Toulouse, CNRS-INPT-UPS, Toulouse, France
[c] Sanofi Chimie – C&BD Biochemistry Vitry - 9 quai Jules Guesde, 94400 Vitry-sur-Seine, France

ARTICLE INFO

*Article history:*
Received 17 October 2017
Received in revised form 15 February 2018
Accepted 16 March 2018
Available online 21 March 2018

*Keywords:*
Extended Quadrature Method of Moments (EQMOM)
Quadrature Based Method of Moments (QBMM)
Population Balance
Mathematical modelling
Gauss quadrature

ABSTRACT

Population Balance Models have a wide range of applications in many industrial fields as they allow accounting for heterogeneity among properties which are crucial for some system modelling. They actually describe the evolution of a Number Density Function (NDF) using a Population Balance Equation (PBE). For instance, they are applied to gas–liquid columns or stirred reactors, aerosol technology, crystallisation processes, fine particles or biological systems. There is a significant interest for fast, stable and accurate numerical methods in order to solve for PBEs, a class of such methods actually does not solve directly the NDF but resolves their moments. These methods of moments, and in particular quadrature-based methods of moments, have been successfully applied to a variety of systems. Point-wise values of the NDF are sometimes required but are not directly accessible from the moments. To address these issues, the Extended Quadrature Method of Moments (EQMOM) has been developed in the past few years and approximates the NDF, from its moments, as a convex mixture of Kernel Density Functions (KDFs) of the same parametric family. In the present work EQMOM is further developed on two aspects. The main one is a significant improvement of the core iterative procedure of that method, the corresponding reduction of its computational cost is estimated to range from 60% up to 95%. The second aspect is an extension of EQMOM to two new KDFs used for the approximation, the Weibull and the Laplace kernels. All MATLAB source codes used for this article are provided with this article.

## 1. Introduction

Population Balance Equations (PBEs) are particular formalisms that allows describing the evolution of properties among heterogeneous populations. They are used to track the size distribution of fine particles [1]; the bubble size distribution in gas–liquid stirred-tank reactors or bubble columns [2,3]; the crystal-size distribution in crystallizers; the distribution of biological cell properties in bioreactors [4,5]; the volume and/or surface distribution of soot particles in flames [6,7] or the formation of nano-particles [8], among other examples.

---

* Corresponding author at: LISBP-INSA Toulouse, 135 Avenue de Rangueil, 31077 Toulouse, France.
*E-mail address:* maxime.pigou@insa-toulouse.fr (M. Pigou).

https://doi.org/10.1016/j.jcp.2018.03.027

## Nomenclature

### Greek symbols

| | |
|---|---|
| $\varepsilon$ | relative tolerance |
| $\lambda_j$ | $j$-th nested quadrature node |
| $\mu$ | positive measure |
| $\omega_j$ | $j$-th nested quadrature weight |
| $\Omega_\xi$ | NDF support |
| $\pi_k$ | $k$ order orthogonal polynomial |
| $\sigma$ | shape parameter |
| $\xi$ | random variable |
| $\xi_i$ | $i$-th main quadrature node |
| $\zeta$ | realisability criteria on $]0, +\infty[$ |

### Roman

| | |
|---|---|
| $a$ | orthogonal polynomials recurrence coefficient |
| $\boldsymbol{A}$ | transition matrix to degenerated moments |
| $b$ | orthogonal polynomials recurrence coefficient |
| $\mathcal{H}$ | Hankel determinant |
| $\boldsymbol{J}_n$ | $n$ order Jacobi matrix |
| $m_k$ | moment of order $k$ |
| $\mathcal{M}$ | realisable moment space |
| $n$ | number density function |
| $\tilde{n}$ | approximation of $n$ |
| $N$ | order of moment set |
| $\mathcal{N}$ | order of realisability |
| $p_k$ | canonical moment of order $k$ |
| $P$ | number of main quadrature nodes |
| $Q$ | number of nested quadrature nodes |
| $w_i$ | $i$-th main quadrature weight |

A PBE describes the evolution and transport of a Number Density Function (NDF), under the influence of multiple processes which modify the tracked property distribution (e.g. erosion, dissolution, aggregation, breakage, coalescence, nucleation, adaptation, etc.).

One often requires low-cost numerical methods to solve PBEs, for instance when coupling with a flow solver (e.g. Computational Fluid Dynamics software). Monte-Carlo methods constitute a stochastic resolution of the population balance and can be applied to such PBE–CFD simulations [9]. Similarly, sectional methods allow direct numerical resolutions of the PBE through the discretisation of the property space [10,11]. They respectively require a high number of parcels or sections in order to reach high accuracy and are thus often discarded for large-scale simulations.

An interesting alternative approach lies in the field of methods of moments. A PBE, which describes the evolution of a NDF, is transformed in a set of equations which describes the evolution of the moments of that distribution. Moments are integral properties of NDFs, the first low order integer moments are related to the mean, variance, skewness and flatness of the statistical distributions described by NDFs. This approach then reduces the number of resolved variables to a finite set of NDF moments. It also comes with some difficulties when one must compute non-moment integral properties, or point-wise evaluations, of the distribution [12].

To tackle these issues, one can try to recover a NDF from a finite set of its moments. In most cases, this reverse problem has an infinite number of solutions and different approaches exist to identify one or an other out of them. The simplest is probably to assume that the NDF is a standard distribution (Gaussian, Log-normal, ...) whose parameters will be deduced from its first few moments. Other methods that lead to continuous approximations, and which preserve a higher number of moments, are the Spline method [13], the Maximum-Entropy approach [12,14,15] or the Kernel Density Element Method (KDEM) [16].

More recently, the Extended Quadrature Method of Moments (EQMOM) was proposed as a new approach which is more stable than the previous ones, and yields either continuous or discrete NDFs depending on the moments [1,17,18]. EQMOM has been implemented in OpenFOAM [19] for the purpose of PBE–CFD coupling. The core of this method relies on an iterative procedure that is a computational bottleneck.

The current work focuses on EQMOM and develops a new core procedure whose computational cost is significantly lower than previous implementations by reducing both (i) the cost of each iteration and (ii) the total number of required iterations.

The previous core procedure [1] will be recalled before describing how it can be shifted toward the new – cheaper – approach. Both implementations will be compared in terms of computational cost (number of required floating-point operations) and run-time.

Multiple variations of EQMOM exist, the Gauss EQMOM [17,20], Log-normal EQMOM [21] as well as Gamma and Beta EQMOM [18]. Two new variations, namely Laplace EQMOM and Weibull EQMOM, are proposed along with a unified formalism among all six variations.

The whole source code used to write this article (figures and data generation) is provided as supplementary data, as well as our implementations of EQMOM in the form of a MATLAB functions library [22].

## 2. Quadrature Based Methods of Moments: QMOM and EQMOM

### 2.1. Definitions

Let $\mathrm{d}\mu(\xi)$ be a positive measure, induced by a non-decreasing function $\mu(\xi)$ defined on a support $\Omega_\xi$. This measure is associated to a Number Density Function $n(\xi)$ such that $\mathrm{d}\mu(\xi) = n(\xi)\mathrm{d}\xi$. Let $\boldsymbol{m}_N$ be the vector of the first $N+1$ integer moments of this measure:

$$\boldsymbol{m}_N = \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_N \end{bmatrix}, \qquad m_k = \int\limits_{\Omega_\xi} \xi^k n(\xi)\mathrm{d}\xi \tag{1}$$

Three actual supports will be considered: (i) $\Omega_\xi = ]-\infty, +\infty[$, (ii) $\Omega_\xi = ]0, +\infty[$ and (iii) $\Omega_\xi = ]0, 1[$. For each support, one can define the associated realisable moment space, $\mathcal{M}_N(\Omega_\xi)$, as the set of all vectors of finite moments $\boldsymbol{m}_N$ induced by all possible positive measures defined on $\Omega_\xi$.

A moment set is said to be “weakly realisable” if located on the boundary of the realisable moment space ($\boldsymbol{m}_N \in \partial\mathcal{M}_N(\Omega_\xi)$). Otherwise, if located within the realisable moment space, $\boldsymbol{m}_N$ is said to be “strictly realisable”.

### 2.2. Quadrature method of moments

EQMOM is based on the Quadrature Method of Moments (QMOM) that was first introduced by McGraw [23]. It is used to approximate integral properties of a distribution where only a finite number of its moments is known. By making use of an even number of moments $2P$, one can compute a Gauss quadrature rule characterised by its weights $\boldsymbol{w}_P = [w_1, \ldots, w_P]^\mathrm{T}$ and nodes $\boldsymbol{\xi}_P = [\xi_1, \ldots, \xi_P]^\mathrm{T}$ such that:

$$\int\limits_{\Omega_\xi} f(\xi)\mathrm{d}\mu(\xi) = \sum_{i=1}^{P} w_i f(\xi_i) \tag{2}$$

holds true if $f(\xi) = \xi^k, \forall k \in \{0, \ldots, 2P-1\}$. Otherwise, this quadrature rule will produce an approximation of the integral property. The computation of the quadrature rule (i.e. the vectors $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$) is of special interest for following developments, which is why its two main steps will be detailed.

Any positive measure $\mathrm{d}\mu(\xi)$ is associated with a sequence of monic polynomials (i.e. polynomial whose leading coefficient equals 1) denoted $\pi_k$ – with $k$ the order of the polynomial – such that:

$$\int\limits_{\Omega_\xi} \pi_i(\xi)\pi_j(\xi)\mathrm{d}\mu(\xi) = 0, \qquad \text{for } i \neq j \tag{3}$$

These polynomials are said orthogonal with respect to the measure $\mathrm{d}\mu(\xi)$ and are defined by:

$$\pi_k(\xi) = \frac{1}{c_k} \begin{vmatrix} m_0 & m_1 & \cdots & m_{k-1} & m_k \\ m_1 & m_2 & \cdots & m_k & m_{k+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_{k-1} & m_k & \cdots & m_{2k-2} & m_{2k-1} \\ 1 & \xi & \cdots & \xi^{k-1} & \xi^k \end{vmatrix} \tag{4}$$

with $c_k$ a constant chosen so that the leading coefficient (of order $k$) of $\pi_k$ equals 1, hence making $\pi_k$ a monic polynomial.

It is known that monic orthogonal polynomials satisfy a three-term recurrence relation [24]:

$$\pi_{k+1}(\xi) = (\xi - a_k)\pi_k(\xi) - b_k\pi_{k-1}(\xi) \tag{5}$$

with $a_k$ and $b_k$ being the recurrence coefficients specific to the measure $\mathrm{d}\mu(\xi)$, $\pi_{-1}(\xi) = 0$ and $\pi_0(\xi) = 1$.

Let $\boldsymbol{J}_n(\mathrm{d}\mu)$ be the $n \times n$ Jacobi matrix associated to the measure $\mathrm{d}\mu$. This is a tridiagonal symmetric matrix defined as:

$$\boldsymbol{J}_n(\mathrm{d}\mu) = \begin{pmatrix} a_0 & \sqrt{b_1} & & 0 \\ \sqrt{b_1} & a_1 & \ddots & \\ & \ddots & \ddots & \sqrt{b_{n-1}} \\ 0 & & \sqrt{b_{n-1}} & a_{n-1} \end{pmatrix} \tag{6}$$

The weights and nodes of the quadrature rule from Eq. (2) are given by spectral properties of $\boldsymbol{J}_P(\mathrm{d}\mu)$. The nodes $\boldsymbol{\xi}_P$ of the rule are the eigenvalues of $\boldsymbol{J}_P(\mathrm{d}\mu)$. The weights of the rule are given by:

$$w_i = m_0 v_{1,i}^2 \tag{7}$$

where $v_{1,i}$ is the first component of the normalised eigenvector belonging to the eigenvalue $\xi_i$. The computation of the quadrature rule (Eq. (2)) then relies on two steps:

1. The computation of the recurrence coefficients $\boldsymbol{a}_{P-1} = [a_0, \ldots, a_{P-1}]^\mathsf{T}$ and $\boldsymbol{b}_{P-1} = [b_1, \ldots, b_{P-1}]^\mathsf{T}$.
2. The computation of the eigenvalues and the normalised eigenvectors of $\boldsymbol{J}_P(\mathrm{d}\mu)$.

Multiple algorithms are available in the literature to compute the recurrence coefficients:

- The Quotient-Difference algorithm [25,26]
- The Product-Difference algorithm [27]
- The Chebyshev algorithm [28]

The Chebyshev algorithm was found to be the stablest one of the three [1,28], its description is given in Appendix A.

### *2.3. Extended Quadrature Method of Moments*

The QMOM method is well suited for the approximation of integral properties of the NDF, which is actually the main purpose of Gauss quadratures. However, in many applications such as evaporation [12] or dissolution [29] processes, point-wise values of the NDF $n(\xi)$ are required but not directly accessible from the moments. For that purpose, a method is needed to produce an approximation $\widetilde{n}(\xi)$ of the original distribution $n(\xi)$, by knowing only a finite set of its moments.

In a sense, one can consider that the Gaussian quadrature computed with QMOM approximates $n(\xi)$ as a weighted sum of Dirac distributions:

$$\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta(\xi, \xi_i) \tag{8}$$

with the Dirac $\delta$ distribution defined by its sifting property

$$\int_{-\infty}^{+\infty} f(\xi)\, \delta(\xi, \xi_m) \mathrm{d}\xi = f(\xi_m) \tag{9}$$

For most applications, $n(\xi)$ is expected to be a continuous distribution whilst QMOM yields monodisperse or discrete polydisperse reconstructions of $n(\xi)$, with $\widetilde{n}(\xi) = 0$ for all values of $\xi$ except some finite number of these values.

Many methods were suggested to tackle this problem and to propose a continuous reconstruction $\widetilde{n}(\xi)$ from a finite number of moments $\boldsymbol{m}_N$. Some of them are the Spline method [13], the Maximum-Entropy approach [14,15,12] or the Kernel Density Element Method [16]. Their properties will not be discussed here but one only underlines that they tend to be unstable, ill-conditioned, or have a high sensitivity to numerical parameters [13,29,30]. In particular, none of them can handle the case of a weakly realisable moment set. Such a moment set is associated to a discrete (or degenerated) distribution and, in this specific case, the distribution provided by QMOM is the only possible reconstruction (see Eq. (8)).

Note that a failure – or instabilities – in a numerical method can compromise the integrity of large-scale simulations. For this reason, Chalons et al. [17], Yuan et al. [18] and Nguyen et al. [1] proposed a robust and stable method to tackle this reconstruction problem by handling both continuous approximations and discrete solutions. Their approach, the Extended Quadrature Method of Moments, approximates $n(\xi)$ as a convex mixture of Kernel Density Functions (KDFs) of the same parametric family:

$$\widetilde{n}(\mu) = \sum_{i=1}^{P} w_i \delta_\sigma(\xi, \xi_i) \tag{10}$$

with

- $w_i$: the weight of the i-th node, $w_i \geq 0, \forall i \in \{1, \ldots, P\}$
- $\xi_i$: the location parameter of the i-th node, $\xi_i \in \Omega_\xi, \forall i \in \{1, \ldots, P\}$
- $\delta_\sigma$: a KDF chosen to perform the approximation, referred later to as the reconstruction kernel. $\sigma$ is the shape parameter of the approximation.

The computation of the weights $\boldsymbol{w}_P = [w_1, \dots, w_P]^\mathsf{T}$, the nodes $\boldsymbol{\xi}_P = [\xi_1, \dots, \xi_P]^\mathsf{T}$ and the shape parameter $\sigma$ from the moment set $\boldsymbol{m}_{2P}$ is performed by the EQMOM moment-inversion procedure. The improvement of this procedure constitutes the core of this article and is detailed in section 3.

Multiple standard normalized distribution functions can be used as the reconstruction kernel $\delta_\sigma$ (e.g. Gaussian, Log-normal, etc.). A list of them is given in Appendix B. All of these kernels degenerate into Dirac distribution if their shape parameters are sufficiently small:

$$\lim_{\sigma\to 0} \delta_\sigma(\xi, \xi_m) = \delta(\xi, \xi_m) \tag{11}$$

This allows EQMOM to be numerically stable in the case of a moment set $\boldsymbol{m}_{2P}$ being on the boundary of the realisable moment space $\partial\mathcal{M}_{2P}(\Omega_\xi)$. Indeed, in such cases, the EQMOM approximation simply degenerates in a weighted sum of Dirac distribution and the definition given in Eq. (10) still holds true, with $\sigma = 0$.

EQMOM can also be used to compute integral properties of the NDF with high accuracy. This comes with the introduction of nested quadratures. The main quadrature proposes the following approximation of integral terms:

$$\int_{\Omega_\xi} f(\xi) n(\xi) \mathrm{d}\xi \approx \sum_{i=1}^{P} w_i \left[ \int_{\Omega_\xi} f(\xi) \delta_\sigma(\xi, \xi_i) \mathrm{d}\xi \right] \tag{12}$$

Moreover, a quadrature rule can be used to approximate the bracketed integral in Eq. (12). This will be the nested quadrature that actually depends on the kernel $\delta_\sigma(\xi, \xi_m)$. For instance, Gauss–Hermite quadratures can be used to approximate integrals over a Gaussian kernel (see Appendix B.1). Nested quadratures then give the following approximation:

$$\int_{\Omega_\xi} f(\xi) n(\xi) \mathrm{d}\xi \approx \sum_{i=1}^{P} w_i \sum_{j=1}^{Q} \omega_j f\left(g(\sigma, \xi_i, \lambda_j)\right) \tag{13}$$

with $Q$ the order, $\boldsymbol{\omega}_Q = [\omega_1, \dots, \omega_Q]^\mathsf{T}$ the weights and $\boldsymbol{\lambda}_Q = [\lambda_1, \dots, \lambda_Q]^\mathsf{T}$ the nodes of the sub-quadrature. $g$ defines the nodes of the nested quadrature from $\sigma$, $\xi_i$ and $\lambda_j$. These nested quadratures are detailed for all KDFs in Appendix B and Appendix C.

## 3. Moment inversion procedure

The EQMOM moment-inversion procedure comes with analytical solutions for some kernels in the case of low-order quadratures. The one-node analytical solutions are detailed for all kernels in Appendix B. When they exist, the two-nodes analytical solutions are implemented in MATLAB code (see supplementary data) but are not detailed in this article. The current section is focusing on the numerical procedure used to compute the reconstruction parameters in absence of an analytical solution.

The procedure proposed by Yuan et al. [18] and Nguyen et al. [1] is first recalled in section 3.1. The section 3.2 details how their approach can be shifted toward a new convergence criteria that will be applied to the specific cases of

- the *Hamburger* moment problem (section 3.3): NDF defined on the whole phase space $\Omega_\xi = ]-\infty, +\infty[$
- the *Stieltjes* moment problem (section 3.4): NDF defined on the positive phase space $\Omega_\xi = ]0, +\infty[$
- the *Hausdorff* moment problem (section 3.5): NDF defined on the closed support $\Omega_\xi = ]0, 1[$

Some moment sets lead to ill-conditioned situations that need to be specifically handled by EQMOM implementations. These are addressed in section 3.6.

### 3.1. Standard procedure

Let $\boldsymbol{m}_N$ be the vector of the first $N+1$ integer moments of the measure $\mathrm{d}\mu(\xi) = n(\xi)\mathrm{d}\xi$, with $N = 2P$ an even integer:

$$\boldsymbol{m}_N = \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_N \end{bmatrix}, \qquad m_k = \int_{\Omega_\xi} \xi^k n(\xi) \mathrm{d}\xi \tag{14}$$

The EQMOM moment-inversion procedure aims to identify the parameters $\sigma$, $\boldsymbol{w}_P = [w_1, \dots, w_P]^\mathsf{T}$ and $\boldsymbol{\xi}_P = [\xi_1, \dots, \xi_P]^\mathsf{T}$ such that $\boldsymbol{m}_N = \widetilde{\boldsymbol{m}}_N$ with:

$$\widetilde{\boldsymbol{m}}_N = \begin{bmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \vdots \\ \widetilde{m}_N \end{bmatrix}, \qquad \widetilde{m}_k = \int_{\Omega_\xi} \xi^k \widetilde{n}(\xi)\mathrm{d}\xi, \qquad \widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma(\xi, \xi_i) \tag{15}$$

For any value of $\sigma$, Yuan et al. [18] identified a procedure which leads to the parameters $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ such that $\boldsymbol{m}_{N-1} = \widetilde{\boldsymbol{m}}_{N-1}$. The EQMOM moment-inversion problem has then been reduced to solving a scalar non-linear equation by looking for a root of the function $D_N(\sigma) = m_N - \widetilde{m}_N(\sigma)$.

The approach developed by Yuan et al. [18] and then improved by Nguyen et al. [1] is based on the fact that, for the KDFs used in EQMOM, it is possible to write the following linear system:

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n(\sigma) \cdot \boldsymbol{m}_n^* \tag{16}$$

where $\boldsymbol{A}_n(\sigma)$ is a lower-triangular $(n+1) \times (n+1)$ matrix whose elements depend only on the chosen KDF and on the value $\sigma$, whereas $\boldsymbol{m}_n^*$ is defined as:

$$\boldsymbol{m}_n^* = \begin{bmatrix} m_0^* \\ m_1^* \\ \vdots \\ m_n^* \end{bmatrix}, \qquad m_k^* = \sum_{i=1}^{P} w_i \xi_i^k \tag{17}$$

By their definition, the moments $\boldsymbol{m}_n^*$ correspond to the moments of a degenerated distribution (i.e. a finite sum of Dirac distributions), hence these moments will be referred as the *degenerated moments of the approximation*. Degenerated moments are defined in such a way that the vectors $\boldsymbol{w}_P$ and $\boldsymbol{\xi}_P$ can be computed from $\boldsymbol{m}_{2P-1}^*$ using a Gauss Quadrature (see 2.2).

At this point, one has the basis required to compute the objective function $D_N(\sigma)$ and to search for its root. The computation of $D_N(\sigma)$ from a vector $\boldsymbol{m}_N$ is as follow (see also Fig. 1a):

1. Compute $\boldsymbol{m}_{N-1}^*(\sigma) = \boldsymbol{A}_{N-1}^{-1}(\sigma) \cdot \boldsymbol{m}_{N-1}$.
2. Compute the recurrence coefficients $\boldsymbol{a}_{P-1}^*(\sigma)$ and $\boldsymbol{b}_{P-1}^*(\sigma)$ by applying the Chebyshev algorithm to $\boldsymbol{m}_{N-1}^*(\sigma)$.
3. Use the recurrence coefficients to compute the Gaussian quadrature rule $\boldsymbol{w}_P(\sigma)$ and $\boldsymbol{\xi}_P(\sigma)$.
4. Knowing the parameters $\sigma$, $\boldsymbol{w}_P(\sigma)$ and $\boldsymbol{\xi}_P(\sigma)$ of the reconstruction, compute $\widetilde{m}_N(\sigma)$, this can be done easily by:
   - Computing the N-th order degenerated moment of the approximated NDF: $\overline{m}_N^*(\sigma) = \sum_{i=1}^P w_i(\sigma)\xi_i(\sigma)^N$.
   - Multiplying the last line of $\boldsymbol{A}_N(\sigma)$ and the vector of degenerated moments: $\widetilde{m}_N(\sigma) = [0, 0, \ldots, 1] \cdot \boldsymbol{A}_N(\sigma) \cdot \left[m_0^*(\sigma), \ldots, m_{N-1}^*(\sigma), \overline{m}_N^*(\sigma)\right]^{\mathrm{T}}$.
5. Compute $D_N(\sigma) = m_N - \widetilde{m}_N(\sigma)$.

For each compatible KDF, it is possible to use low order moments to compute an upper bound $\sigma_{max}$ so that the search of a root of $D_N$ is restricted to the interval $\sigma \in [0, \sigma_{max}]$. Then a bounded non-linear equation solver such as Ridder's method can be applied to actually find the root of the function.

Two specific cases were discarded in the previous description of the method. First, it happens that the function $D_N$ does not admit any root, in such a case the procedure is switched toward the minimisation of this function in order to reduce the error on the last moment of the approximation.

Second, during the computation of $D_N(\sigma)$, one must compute degenerated moments from which weights and nodes are extracted. If degenerated moments $\boldsymbol{m}_{N-1}^*(\sigma)$ turn out not to be realisable on the support $\Omega_\xi$ of the NDF, the quadrature performed on this vector will lead to nodes outside $\Omega_\xi$, or even to negative/complex weights. Nguyen et al. [1] then suggest to check for the realisability of degenerated moments, and if these are not realisable, to set $\widetilde{m}_N(\sigma)$ to a arbitrarily high value such as $10^{100}$. This will force the non-linear equation solver to test a lower value of $\sigma$ in order to bring back the vector $\boldsymbol{m}_{N-1}^*(\sigma)$ within the realisable moment space. However note that this is only a numerical trick to converge toward the actual root, but $D_N(\sigma)$ is actually undefined as soon as $\boldsymbol{m}_{N-1}^*(\sigma)$ is not realisable.

### 3.2. *A new procedure based on moment realisability*

The reversible linear system linking raw moments of the approximation $\widetilde{\boldsymbol{m}}_N$ to its degenerated moments $\boldsymbol{m}_N^*$ is such that a new objective function $D_N^*(\sigma)$ – whose root is the same as that of $D_N(\sigma)$ – can be formulated. Its computation is as follow (see also Fig. 1b):

1. Compute $\boldsymbol{m}_N^*(\sigma) = \boldsymbol{A}_N^{-1}(\sigma) \cdot \boldsymbol{m}_N$.
2. Compute a quadrature on the vector $\boldsymbol{m}_{N-1}^*(\sigma)$ to obtain the vectors $\boldsymbol{w}_P(\sigma)$ and $\boldsymbol{\xi}_P(\sigma)$.
3. Compute $\overline{m}_N^*(\sigma) = \sum_{i=1}^P w_i(\sigma)\xi_i(\sigma)^N$.
4. Compute $D_N^*(\sigma) = m_N^*(\sigma) - \overline{m}_N^*(\sigma)$.

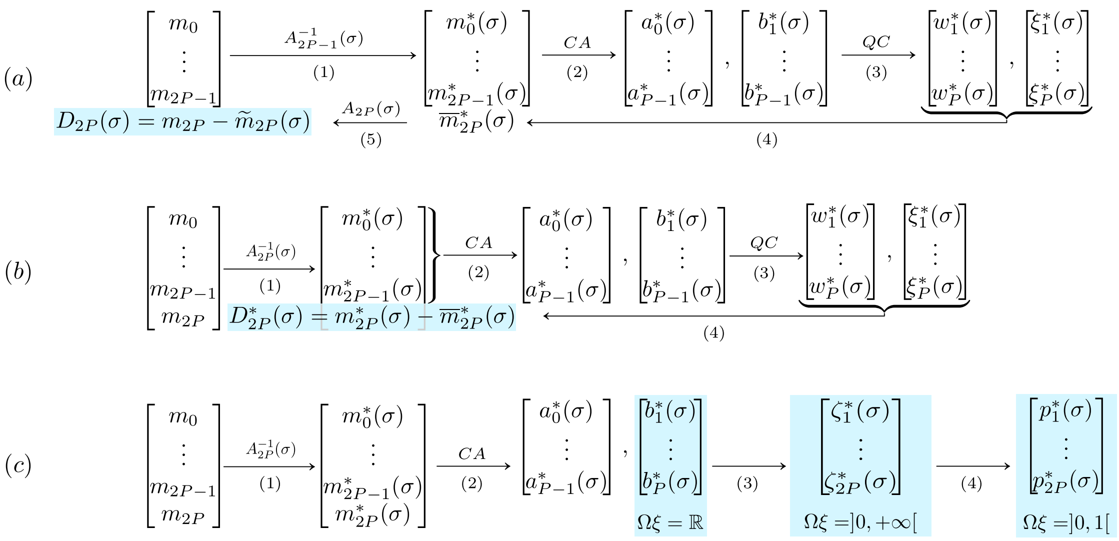


**Fig. 1.** Comparison of the computation of convergence criteria based on (a) $D_N(\sigma)$, (b) $D_N^*(\sigma)$ and (c) the realisability criteria of the support $\Omega_\xi$. CA: Chebyshev Algorithm. QC: Quadrature Computation. The convergence criteria are highlighted in light blue. Inspired by Fig. 1 from Nguyen et al. [1]. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

Note that $D_N(\sigma) = D_N^*(\sigma) \times A_{N,N}(\sigma)$. As shown in Appendix B for all kernels, diagonal elements of $\boldsymbol{A}_n(\sigma)$ are always strictly positive, therefore the two objective functions do share the same roots.

The benefit of this new objective function is that it only requires the matrix $\boldsymbol{A}_N^{-1}(\sigma)$ instead of both the matrix $\boldsymbol{A}_{N-1}^{-1}(\sigma)$ and the last line of $\boldsymbol{A}_N(\sigma)$. This only increases the clarity of the method, but has hardly no effect on its numerical cost.

The point of this alternative approach is however to underline a crucial element for the new EQMOM implementation: we actually look for a value of $\sigma$ for which $m_{2P}^*(\sigma) = \overline{m}_{2P}^*(\sigma)$. This implies that, for this specific searched $\sigma$ value, the vector $\boldsymbol{m}_{2P}^*(\sigma)$ reads

$$\boldsymbol{m}_{2P}^*(\sigma) = \begin{bmatrix} \sum_{i=1}^P w_i \xi_i^0 \\ \sum_{i=1}^P w_i \xi_i^1 \\ \vdots \\ \sum_{i=1}^P w_i \xi_i^{2P} \end{bmatrix} \tag{18}$$

which is, by construction, the vector of the first $2P+1$ moments of the sum of $P$ Dirac distributions.

Under the condition that a $P$-node EQMOM reconstruction exists for the moment set $\boldsymbol{m}_{2P}$ with $\sigma > 0$, $w_i > 0$, $\xi_i \neq 0, i \in \{1, \dots, P\}$, the vector $\boldsymbol{m}_{2P}^*(\sigma)$ will have the following specific properties:

1. The vector $\boldsymbol{m}_{2P-1}^*(\sigma)$ must be strictly within the realisable moment space $\mathcal{M}_{N-1}(\Omega_\xi)$.
2. The vector $\boldsymbol{m}_{2P}^*(\sigma)$ must be on the boundary of the realisable moment space $\mathcal{M}_N(\Omega_\xi)$.

EQMOM procedure will then rely on the realisability of the vector $\boldsymbol{m}_{2P}^*(\sigma)$ instead of the computation of the error on the last moment, this will be a cheaper approach.

Situations were the EQMOM reconstruction exist but with $\sigma = 0$, or $\exists i \in \{1, \dots, P\}, w_i = 0$ or $\xi_i = 0$ are tackled in section 3.6 but are always based on checking the realisability of $\boldsymbol{m}_{2P}^*(\sigma)$.

The actual definition of the realisable moment space of order $n$, $\mathcal{M}_n$, depends on the support $\Omega_\xi$ of the NDF. The three classical supports, corresponding to the *Hamburger*, *Stieltjes* and *Hausdorff* moment problems, come with different constraints on a moment set to ensure its realisability. The realisability criteria for each of these supports will then be detailed.

Fig. 1 sums up the “standard approach” based on $D_N(\sigma)$, the shifted approach, based on $D_N^*(\sigma)$, as well as the new approach based on the realisability criteria of $\boldsymbol{m}_{2P}^*(\sigma)$ for all three supports.

### 3.3. *Application to the* Hamburger *problem*

As stated in 2.2, it is known that monic polynomials which are orthogonal to a measure $\mathrm{d}\mu(\xi) = n(\xi)\mathrm{d}\xi$ satisfy a three-term recurrence relation (Eq. (5)) with $a_k$ and $b_k, k \in \mathbb{N}$, the recurrence coefficients specific to the measure $\mathrm{d}\mu(\xi)$. The Favard’s theorem [31] and its converse [32] imply that the measure $\mathrm{d}\mu(\xi)$ is realisable on $\Omega_\xi = ]-\infty, +\infty[$ if and only if $a_k \in \mathbb{R}$ and $b_k > 0$, $\forall k \in \mathbb{N}$.

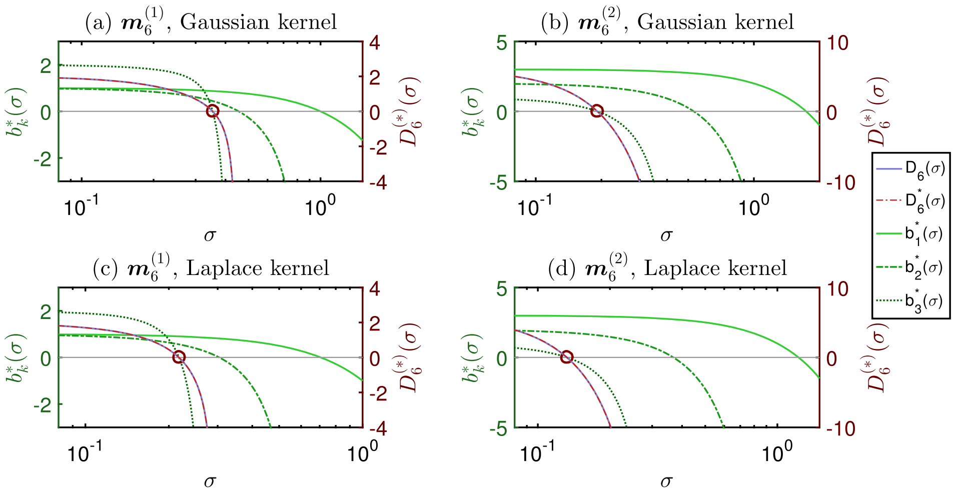


**Fig. 2.** Evolution of the different convergence criteria for both Gaussian (a and b) and Laplace (c and d) kernels depending on $\sigma$ value. The two initial moment sets are $\boldsymbol{m}_6^{(1)} = [1\ 1\ 2\ 5\ 12\ 42\ 133]^{\mathrm{T}}$ and $\boldsymbol{m}_6^{(2)} = [1\ 2\ 7\ 17\ 58\ 149\ 493]^{\mathrm{T}}$.

One looks for a value of $\sigma$ such that the associated degenerated moments $\boldsymbol{m}^*_{2P-1}(\sigma)$ are strictly realisable (i.e. within the moment space), and the moments $\boldsymbol{m}^*_{2P}(\sigma)$ are weakly realisable (i.e. on the frontier of realisability). Then, if the Chebyshev algorithm is used to compute the recurrence coefficients $\boldsymbol{a}^*_{P-1}(\sigma) = [a^*_0(\sigma), \dots, a^*_{P-1}(\sigma)]^{\mathrm{T}}$ and $\boldsymbol{b}^*_P(\sigma) = [b^*_1(\sigma), \dots, b^*_P(\sigma)]^{\mathrm{T}}$ from the vector $\boldsymbol{m}^*_{2P}(\sigma)$, the condition of realisability can be written in terms of values of $\boldsymbol{b}^*_P(\sigma)$: looking for the EQMOM reconstruction parameters with the Gaussian and Laplace kernels is equivalent to looking for a value of $\sigma$ such as:

- $b^*_k(\sigma) > 0,\ \forall k \in \{1, \dots, P-1\}$
- $b^*_P(\sigma) = 0$

Fig. 2 makes use of the developments from Appendix B.1 and Appendix B.2, about the Gaussian and Laplace kernels respectively, to show the evolution of $D_6(\sigma)$, $D^*_6(\sigma)$ and $b^*_k(\sigma)$, $k \in \{1, 2, 3\}$ for two sets of 7 moments ($P = 3$). This figure illustrates the fact that indeed the approaches based on $D_N(\sigma)$, $D^*_N(\sigma)$ and $b^*_P(\sigma)$ are equivalent as they share the same circled root.

Let denote $\sigma_k$ the root of $b_k(\sigma)$. One can notice that the root $\sigma_k$ lies within the interval $[0, \sigma_{k-1}]$. We actually observed the existence of all roots $\sigma_k$, $k \in \{1, \dots, P\}$ on numerous (about $10^6$) randomly selected moment sets of $N + 1 = 13$ moments, and never observed an undefined root. The generality of this observation has not been mathematically proved, but it seems that indeed $\sigma_k$ is always defined and always lies in $\sigma_k \in [0, \sigma_{k-1}]$, $k \in \{2, \dots, P\}$. $\sigma_1$ is defined analytically.

The previous observations were used to design a simple algorithm which allows identifying the root $\sigma_P$. This algorithm is based on the fact that it is possible to check whether a value $\sigma_t$ is higher or lower than $\sigma_P$ at low cost and with no prior knowledge of $\sigma_P$ value:

- If $b^*_k(\sigma_t) > 0,\ \forall k \in \{1, \dots, P\}$, then $\sigma_t < \sigma_P$.
- Otherwise, that is if $\exists k \in \{1, \dots, P\}$, $b^*_k(\sigma_t) < 0$, then $\sigma_t > \sigma_P$.

One can then use an iterative approach that will

1. Check the realisability of the raw moments $\boldsymbol{m}_{2P} = \boldsymbol{m}^*_{2P}(0)$ by computing $\boldsymbol{b}^*_P(0)$ and checking the positivity of all elements.
2. Initialise an interval $\left[\sigma_l^{(0)},\ \sigma_r^{(0)}\right]$ such that $\sigma_l^{(0)} < \sigma_P$ and $\sigma_r^{(0)} > \sigma_P$, and then update these bounds to shrink the search interval. These initial values will be $\sigma_l^{(0)} = 0$ and $\sigma_r^{(0)} = \sigma_1$ with $\sigma_1$ the analytical solution of $b^*_1(\sigma) = 0$.
3. Iterate over $k$
   (a) Choose $\sigma_t \in \left[\sigma_l^{(k-1)}, \sigma_r^{(k-1)}\right]$.
   (b) Compute $\boldsymbol{b}^*_P(\sigma_t)$.
   (c) If all elements of $\boldsymbol{b}^*_P(\sigma_t)$ are positive, set $\sigma_l^{(k)} = \sigma_t$ and $\sigma_r^{(k)} = \sigma_r^{(k-1)}$.
   (d) Otherwise, set $\sigma_l^{(k)} = \sigma_l^{(k-1)}$ and $\sigma_r^{(k)} = \sigma_t$.

The choice of $\sigma_t$ at step 3a will be made by trying to locate the root $\sigma_j$ of $b^*_j(\sigma)$ with $j$ the index of the first negative element of $\boldsymbol{b}^*_P\left(\sigma_r^{(k)}\right)$. Following Nguyen et al. [1] developments, the use of Ridder's method is advised to select $\sigma_t$. This method actually tests two $\sigma$ values per iteration. Consequently, the step 3 of the previous algorithm becomes:

3. Iterate over $k$
    (a) Identify $j$ the index of the first negative element of $\boldsymbol{b}_P^*\left(\sigma_r^{(k-1)}\right)$.
    (b) Compute $\sigma_{t_1} = \frac{1}{2}\left(\sigma_l^{(k-1)} + \sigma_r^{(k-1)}\right)$ and $\boldsymbol{b}_P^*(\sigma_{t_1})$.
    (c) Compute $\sigma_{t_2} = \sigma_{t_1} + \left(\sigma_{t_1} - \sigma_l^{(k-1)}\right) \frac{b_j^*(\sigma_{t_1})}{\sqrt{b_j^*(\sigma_{t_1})^2 - b_j^*\left(\sigma_l^{(k-1)}\right) * b_j^*\left(\sigma_r^{(k-1)}\right)}}$ and $\boldsymbol{b}_P^*(\sigma_{t_2})$.
    (d) Set $\sigma_l^{(k)}$ as the highest value between $\sigma_l^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{b}_P^*$ contains only positive values.
    (e) Set $\sigma_r^{(k)}$ as the lowest value between $\sigma_r^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{b}_P^*$ contains at least one negative value.

Stop the computation if $\sigma_r^{(k)} - \sigma_l^{(k)} < \varepsilon\ \sigma_1$ or if $b_P^*\left(\sigma_l^{(k)}\right) < \varepsilon\ b_P^*(0)$, with $\varepsilon$ a relative tolerance (e.g. $\varepsilon = 10^{-10}$). Then compute the weights $\boldsymbol{w}_P$ and nodes $\boldsymbol{\xi}_P$ of the EQMOM reconstruction by computing a Gauss quadrature based on the recurrence coefficients $\boldsymbol{a}_{P-1}^*\left(\sigma_l^{(k)}\right)$ and $\boldsymbol{b}_{P-1}^*\left(\sigma_l^{(k)}\right)$.

Actual implementations of this algorithm for both kernels are provided as supplementary data.

### 3.4. *Application to the* Stieltjes *problem*

It is well known that the realisability of a moment set $\boldsymbol{m}_N$ on the support $\Omega_\xi = ]0, +\infty[$ is strictly equivalent to the positivity of the Hankel determinants $\underline{\mathcal{H}}_{2n+d}$ [33] defined as:

$$\underline{\mathcal{H}}_{2n+d} = \begin{vmatrix} m_d & \cdots & m_{n+d} \\ \vdots & \ddots & \vdots \\ m_{n+d} & \cdots & m_{2n+d} \end{vmatrix} \tag{19}$$

with $d \in \{0, 1\}$ and $n \in \mathbb{N},\ 2n + d \leq N$.

This condition on the positivity of Hankel determinants can be translated into a condition on the positivity of the numbers $\zeta_k$ [32] defined by:

$$\zeta_k = \frac{\underline{\mathcal{H}}_{k-3}\underline{\mathcal{H}}_k}{\underline{\mathcal{H}}_{k-2}\underline{\mathcal{H}}_{k-1}}, \qquad \underline{\mathcal{H}}_j = 1 \text{ if } j < 0 \tag{20}$$

These numbers can be directly computed from the recurrence coefficients $\boldsymbol{a}_P$ and $\boldsymbol{b}_P$ defined in 2.2 through the following relations:

$$\zeta_{2k} = \frac{b_k}{\zeta_{2k-1}}, \qquad \zeta_{2k+1} = a_k - \zeta_{2k} \tag{21}$$

with $\zeta_1 = a_0 = m_1/m_0$.

The goal here is to use these realisability criteria to compute the parameters of EQMOM quadrature with either the Log-normal, the Gamma or the Weibull kernel (see Appendix B.3, Appendix B.4 and Appendix B.5 respectively). In these cases, one must

1. Compute $\boldsymbol{m}_N^*(\sigma) = \boldsymbol{A}_N^{-1}(\sigma) \cdot \boldsymbol{m}_N$ with $\boldsymbol{A}_N(\sigma)$ the matrix associated to the chosen kernel (see Appendix B.3, Appendix B.4, Appendix B.5).
2. Apply the Chebyshev algorithm to $\boldsymbol{m}_N^*(\sigma)$ to access the recurrence coefficients $\boldsymbol{a}_P^*(\sigma)$ and $\boldsymbol{b}_P^*(\sigma)$.
3. Compute $\boldsymbol{\zeta}_N^*(\sigma) = [\zeta_1^*(\sigma), \ldots, \zeta_N^*(\sigma)]^{\mathrm{T}}$ using relations in Eq. (21).

One actually looks for $\sigma$ such that

- $\zeta_k^*(\sigma) > 0,\ \forall k \in \{1, \ldots, N-1\}$
- $\zeta_N^*(\sigma) = 0$

Let $\sigma_k$ be the root of $\zeta_k^*(\sigma)$. In all cases, the root $\sigma_2$ is defined, analytically for the Log-normal and Gamma kernels, and numerically for the Weibull kernel. Fig. 3 shows the evolution of $D_6(\sigma)$, $D_6^*(\sigma)$ and $\boldsymbol{\zeta}_6^*(\sigma)$ for three moment sets when the developments relative to the Weibull (see Appendix B.5) kernel are used. Three situations can be observed on that figure:

1. All roots $\sigma_k$, $k \in \{2, \ldots, N\}$ are defined (Fig. 3a).
2. Some intermediary roots $\sigma_k$, $k \in \{3, \ldots, N-1\}$, are not defined but the root $\sigma_N$ still exists (Fig. 3b).
3. The root $\sigma_N$ is not defined (Fig. 3c).

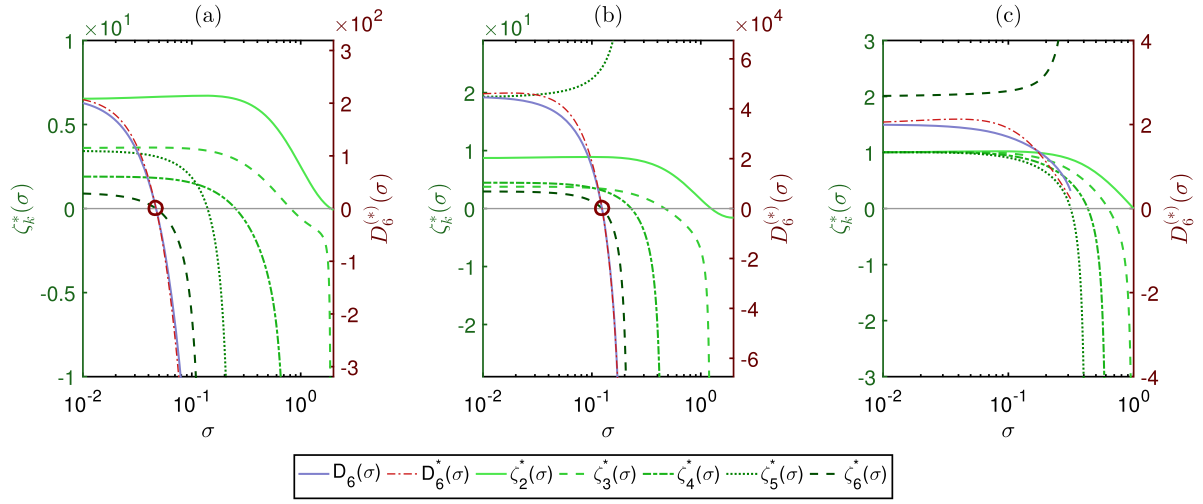


**Fig. 3.** Evolution of the different convergence criteria for the Weibull kernel depending on $\sigma$ value. The initial moment sets are $\boldsymbol{m}_6^{(a)} = [1\ 1.5\ 12\ 131\ 15200\ 18033\ 2.16e5]^{\mathsf{T}}$, $\boldsymbol{m}_6^{(b)} = [1\ 5.5\ 78\ 1285\ 22225\ 4.05e5\ 7.88e6]^{\mathsf{T}}$ and $\boldsymbol{m}_6^{(c)} = [1\ 1\ 2\ 5\ 14\ 42\ 133]^{\mathsf{T}}$.

These three cases can be observed for the Gamma and Log-normal kernels too.

In the first two cases, when $\sigma_N$ exists, the EQMOM approximation is well defined. The last case – where $\zeta_N^*(\sigma)$ admits no root in $[0, \sigma_{N-1}]$ – actually corresponds to the case described by Nguyen et al. [1] where $D_N(\sigma)$ did not admit any root either. In this case, it was suggested to minimise $D_N(\sigma)$ in order to reduce the difference between $m_N$ and $\widetilde{m}_N(\sigma)$ as much as possible.

$D_N(\sigma)$ tends to be a decreasing function, but is undefined as soon as any element of $\boldsymbol{\zeta}_{N-1}^*(\sigma)$ is negative. The minimum of $D_N(\sigma)$ is then usually located at the highest order defined root. For instance, in the case shown in Fig. 3c, the minimum of $D_6(\sigma)$ is located at the root $\sigma_5$ of $\zeta_5^*(\sigma)$.

The moment-inversion procedure for reconstruction kernels defined on $\Omega_\xi = ]0, +\infty[$ is then reduced to the identification of the defined root $\sigma_k$, $k \in \{2, \ldots, N\}$, of highest index. The algorithm proposed in section 3.3 already converges toward this root and only requires little adjustments:

1. Check the realisability of the raw moments $\boldsymbol{m}_{2P} = \boldsymbol{m}_{2P}^*(0)$ by computing $\boldsymbol{\zeta}_N^*(0)$ and checking the positivity of all elements.
2. Initialise an interval $\left[\sigma_l^{(0)},\ \sigma_r^{(0)}\right]$ with $\sigma_l^{(0)} = 0$ and $\sigma_r^{(0)} = \sigma_2$ with $\sigma_2$ the solution of $\zeta_2^*(\sigma) = 0$.
3. Iterate over $k$
   (a) Identify $j$ the index of the first negative element of $\zeta_N^*\left(\sigma_r^{(k-1)}\right)$.
   (b) Compute $\sigma_{t_1} = \frac{1}{2}\left(\sigma_l^{(k-1)} + \sigma_r^{(k-1)}\right)$ and $\boldsymbol{\zeta}_N^*(\sigma_{t_1})$.
   (c) Compute $\sigma_{t_2} = \sigma_{t_1} + \left(\sigma_{t_1} - \sigma_l^{(k-1)}\right) \dfrac{\zeta_j^*(\sigma_{t_1})}{\sqrt{\zeta_j^*(\sigma_{t_1})^2 - \zeta_j^*\left(\sigma_l^{(k-1)}\right) * \zeta_j^*\left(\sigma_r^{(k-1)}\right)}}$ and $\boldsymbol{\zeta}_N^*(\sigma_{t_2})$.
   (d) Set $\sigma_l^{(k)}$ as the highest value between $\sigma_l^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{\zeta}_N^*$ contains only positive values.
   (e) Set $\sigma_r^{(k)}$ as the lowest value between $\sigma_r^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{\zeta}_N^*$ contains at least one negative value.

Stop the computation if $\sigma_r^{(k)} - \sigma_l^{(k)} < \varepsilon \sigma_1$ or if $\zeta_N^*\left(\sigma_l^{(k)}\right) < \varepsilon \zeta_N^*(0)$, with $\varepsilon$ a relative tolerance (e.g. $\varepsilon = 10^{-10}$). Then compute the weights $\boldsymbol{w}_P$ and nodes $\boldsymbol{\xi}_P$ of the EQMOM reconstruction by computing a Gaussian-quadrature based on recurrence coefficients $\boldsymbol{a}_{P-1}^*\left(\sigma_l^{(k)}\right)$ and $\boldsymbol{b}_{P-1}^*\left(\sigma_l^{(k)}\right)$.

### 3.5. *Application to the* Hausdorff *problem*

Moments of a distribution defined on the closed support $\Omega_\xi = ]0, 1[$ must obey two sets of conditions in order to be within the realisable moment space [15,26]. The moment set $\boldsymbol{m}_N$ is interior to the realisable moment space associated to the support $\Omega_\xi = ]0, 1[$ if and only if:

- $\underline{\mathcal{H}}_k > 0$, $\forall k \in \{0, \ldots, N\}$
- $\overline{\mathcal{H}}_k > 0$, $\forall k \in \{1, \ldots, N\}$

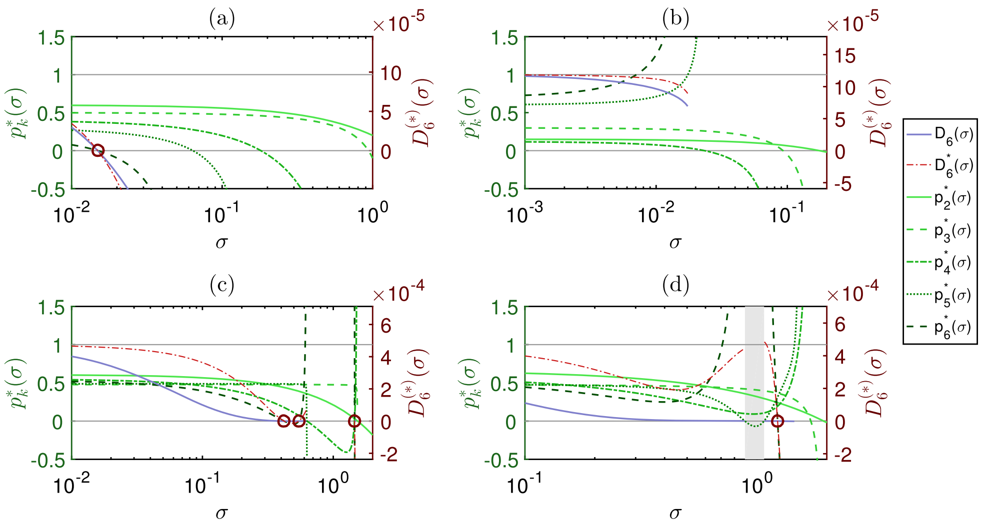


**Fig. 4.** Evolution of the different convergence criteria for the Beta reconstruction kernel and four initial moment sets. These sets can be found in the figure source code provided as supplementary data.

with $\underline{\mathcal{H}}_k$ defined in Eq. (19) and $\overline{\mathcal{H}}_k$ defined by

$$\overline{\mathcal{H}}_{2n+d} = \begin{vmatrix} m_{d-1} - m_d & \cdots & m_{n+d-1} - m_{n+d} \\ \vdots & \ddots & \vdots \\ m_{n+d-1} - m_{n+d} & \cdots & m_{2n+d-1} - m_{2n+d} \end{vmatrix} \tag{22}$$

Leaving aside the obvious condition $\underline{\mathcal{H}}_0 = m_0 > 0$, the conditions $\overline{\mathcal{H}}_k > 0$ and $\underline{\mathcal{H}}_k > 0$ induce a lower bound $m_k^-$ and an upper bound $m_k^+$ for the values of $m_k$, $k \in \{1, \dots, N\}$. Consequently, one can define the canonical moments of the distribution $\boldsymbol{p}_N = [p_1, \dots, p_N]^\mathrm{T}$ as

$$p_k = \frac{m_k - m_k^-}{m_k^+ - m_k^-} \tag{23}$$

A moment set $\boldsymbol{m}_N$ is strictly realisable if and only if the associated canonical moment set $\boldsymbol{p}_N$ lies in the hypercube $]0, 1[^N$. Canonical moments can be computed through the recurrence relation [34]:

$$p_k = \frac{\zeta_k}{1 - p_{k-1}} \tag{24}$$

with $\zeta_k$ defined in Eq. (20) and $p_1 = m_1$.

In the case of the Beta kernel (see B.6), one is looking for a value of $\sigma$ such that the vector $\boldsymbol{p}_N^*(\sigma)$ has the following properties:

- $p_k^*(\sigma) \in ]0, 1[\,, \qquad \forall k \in \{1, \dots, N-1\}$
- $p_N^*(\sigma) = 0$

$\boldsymbol{p}_N^*(\sigma)$ is computed from the vector $\boldsymbol{\zeta}_N^*(\sigma)$ which is deduced from the recurrence coefficients $\boldsymbol{a}_{P-1}^*(\sigma)$ and $\boldsymbol{b}_P^*(\sigma)$. These are computed –like previously– through the Chebyshev algorithm applied to the vector $\boldsymbol{m}_N^*(\sigma) = \boldsymbol{A}_N^{-1}(\sigma) \cdot \boldsymbol{m}_N$.

Fig. 4 shows the evolution of the canonical moments and the convergence criteria $D_6(\sigma)$ and $D_6^*(\sigma)$ for four different sets of 7 moments with the developments relative to the Beta kernel (see Appendix B.6). Each of these sets corresponds to one of the four situations encountered when dealing with Beta EQMOM:

- Fig. 4a: the root $\sigma_N$ of $D_N(\sigma)$, $D_N^*(\sigma)$ and $p_N^*(\sigma)$ exists and can be identified through a similar procedure than that described in sections 3.3 and 3.4.
- Fig. 4b: the root $\sigma_N$ is not defined but the minimum of $D_N(\sigma)$ is located at the $\sigma$ value for which $\boldsymbol{p}_{N-1}^*(\sigma)$ is on the boundary of the hypercube $]0, 1[^{N-1}$.
- Fig. 4c: $D_N(\sigma)$, $D_N^*(\sigma)$ and $p_N^*(\sigma)$ admit multiple roots.
- Fig. 4d: the root $\sigma_N$ is defined, but there is a range $]\sigma_{v_1}, \sigma_{v_2}[$ with $\sigma_{v_2} < \sigma_N$, highlighted in light grey, such that in this interval the convergence criteria are undefined because $\forall \sigma \in ]\sigma_{v_1}, \sigma_{v_2}[, \ \boldsymbol{p}_{N-1}^*(\sigma) \notin ]0, 1[^{N-1}$.

The algorithm proposed in sections 3.3 and 3.4 can still be applied here by replacing the convergence criteria by the canonical moments, and by checking that the values of $\boldsymbol{p}_N^*(\sigma)$ all lie in the interval $]0,1[$ instead of checking only for positivity:

1. Check the realisability of the raw moments $\boldsymbol{m}_{2P} = \boldsymbol{m}_{2P}^*(0)$ by computing $\boldsymbol{p}_N^*(0)$ and checking that all elements lie in $]0,1[$.
2. Initialise an interval $\left[\sigma_l^{(0)},\ \sigma_r^{(0)}\right]$ with $\sigma_l^{(0)} = 0$ and $\sigma_r^{(0)} = \sigma_2$ with $\sigma_2$ the analytical solution of $p_2^*(\sigma) = 0$.
3. Iterate over $k$
   - (a) Identify $j$ the index of the first element of $\boldsymbol{p}_N^*\left(\sigma_r^{(k-1)}\right)$ that is either negative or higher than 1.
   - (b) Compute $\sigma_{t_1} = \frac{1}{2}\left(\sigma_l^{(k-1)} + \sigma_r^{(k-1)}\right)$ and $\boldsymbol{p}_N^*(\sigma_{t_1})$.
   - (c) If $j < N$ and $p_j^*\left(\sigma_r^{(k-1)}\right) > 1$
     - Compute $\sigma_{t_2} = \sigma_{t_1} + \left(\sigma_{t_1} - \sigma_l^{(k-1)}\right) \dfrac{q_j^*(\sigma_{t_1})}{\sqrt{q_j^*(\sigma_{t_1})^2 - q_j^*\left(\sigma_l^{(k-1)}\right) * q_j^*\left(\sigma_r^{(k-1)}\right)}}$ and $\boldsymbol{p}_N^*(\sigma_{t_2})$, with $q_j^*(\sigma) = 1 - p_j^*(\sigma)$.
   - (d) Else, that is if $j = N$ or $p_j^*\left(\sigma_r^{(k-1)}\right) < 0$
     - Compute $\sigma_{t_2} = \sigma_{t_1} + \left(\sigma_{t_1} - \sigma_l^{(k-1)}\right) \dfrac{p_j^*(\sigma_{t_1})}{\sqrt{p_j^*(\sigma_{t_1})^2 - p_j^*\left(\sigma_l^{(k-1)}\right) * p_j^*\left(\sigma_r^{(k-1)}\right)}}$ and $\boldsymbol{p}_N^*(\sigma_{t_2})$.
   - (e) Set $\sigma_l^{(k)}$ as the highest value between $\sigma_l^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{p}_N^*$ lies in $]0,1[^N$.
   - (f) Set $\sigma_r^{(k)}$ as the lowest value between $\sigma_r^{(k-1)}$, $\sigma_{t_1}$ and $\sigma_{t_2}$ such that the corresponding vector $\boldsymbol{p}_N^*$ does not lie in $]0,1[^N$.

Stop the computation if $\sigma_r^{(k)} - \sigma_l^{(k)} < \varepsilon\sigma_2$ or if $p_N^*\left(\sigma_l^{(k)}\right) < \varepsilon p_N^*(0)$, with $\varepsilon$ a relative tolerance (e.g. $\varepsilon = 10^{-10}$). As previously, once convergence is achieved, the weights $\boldsymbol{w}_P$ and nodes $\boldsymbol{\xi}_P$ of the reconstruction can be obtained by computing a Gaussian quadrature rule based on the recurrence coefficients $\boldsymbol{a}_{P-1}^*\left(\sigma_l^{(k)}\right)$ and $\boldsymbol{b}_{P-1}^*\left(\sigma_l^{(k)}\right)$.

This algorithm will converge to the root $\sigma_N$ for cases similar to Fig. 4a; to the minimum of $D_N(\sigma)$ for cases similar to Fig. 4b; to one of the multiple roots for cases similar to Fig. 4c. In the case illustrated in Fig. 4d, the algorithm may or may not identify the existing root, depending on whether one of the intermediate tested $\sigma$ values lies in the greyed area.

One could try to develop a more robust algorithm, that will always find the root if it is defined, even in the case shown in Fig. 4d. An other improvement would be to ensure a consistent result when multiple roots exist, for instance by converging toward the lowest root, so that a small perturbation in the raw moments will only cause a small change on the resulting $\sigma$ value. Nothing prevents the current algorithm from converging toward one root for a moment set and toward another one after a small perturbation of this set which could induce instabilities in large-scale simulations. Note that these limitations already existed in previous EQMOM implementations and do not result from the new approach developed in this article.

### *3.6. Handling weakly realisable and ill-conditioned moment sets*

The EQMOM moment-inversion procedure attempts to identify a NDF defined by

$$\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma(\xi, \xi_i) \tag{25}$$

whose first $2P+1$ integer moments are given by $\boldsymbol{m}_{2P}$.

This approximation is not always possible as shown in sections 3.4 and 3.5. When the EQMOM approximation exists, it may be ill-conditioned if at least one of the followings holds true:

- $\sigma = 0$
- $\exists i,\ w_i = 0$
- $\exists i,\ \xi_i = 0$

The first situation is that of $\boldsymbol{m}_{2P}$ being weakly realisable. The second situation occurs if $\boldsymbol{m}_{2P}$ is the moment set of a convex mixture of the reconstruction kernel with less than $P$ nodes. These situations are not mutually exclusive, a vector $\boldsymbol{m}_6$ could be the vector of the 7 first moments of a bi-Dirac distribution, one of which could be located in $\xi = 0$.

Accounting for these situations requires introducing the order of realisability of a moment set, $\mathcal{N}(\boldsymbol{m}_N)$. This notation was introduced by Nguyen et al. [1] but was only defined on $\Omega_\xi = ]0,+\infty[$ in terms of Hankel determinants. The following definition is broader as it encompasses theirs but extends it to other supports. $\mathcal{N}(\boldsymbol{m}_N)$ is the number of moments in the largest strictly realisable subset of $\boldsymbol{m}_N$. For each support, the order of realisability is defined in terms of the realisability criterion:

- For $\Omega_\xi = ]-\infty, +\infty[$, compute $\boldsymbol{b}_P$ from $\boldsymbol{m}_{2P}$;
  - if all elements are positive, $\mathcal{N}(\boldsymbol{m}_{2P}) = 2P + 1$;
  - else, if there is $n$ such that $b_n = 0$, $\mathcal{N}(\boldsymbol{m}_{2P}) = 2n$;
  - else, if there is $n$ such that $b_n < 0$, $\mathcal{N}(\boldsymbol{m}_{2P}) = 2n - 1$.
- For $\Omega_\xi = ]0, +\infty[$, compute $\boldsymbol{\zeta}_{2P}$ from $\boldsymbol{m}_{2P}$;
  - if all elements are positive, $\mathcal{N}(\boldsymbol{m}_{2P}) = 2P + 1$;
  - else identify $n$ such that $\zeta_n \leq 0$, $\mathcal{N}(\boldsymbol{m}_{2P}) = n$.
- For $\Omega_\xi = ]0, 1[$, compute $\boldsymbol{p}_{2P}$ from $\boldsymbol{m}_{2P}$;
  - if all elements are included on $]0, 1[$, $\mathcal{N}(\boldsymbol{m}_{2P}) = 2P + 1$;
  - else identify $n$ such that $p_n \notin ]0, 1[$, $\mathcal{N}(\boldsymbol{m}_{2P}) = n$.

Detecting situations where $\sigma = 0$ requires to check the order of realisability of raw moments. If $\mathcal{N}(\boldsymbol{m}_{2P})$ is even, set $\sigma = 0$; otherwise apply the iterative procedure to $\boldsymbol{m}_{2P'}$ with $\mathcal{N}(\boldsymbol{m}_{2P}) = 2P' - 1$ to identify $\sigma$ [1].

The actual number of nodes required by the EQMOM approximation, *i.e.* the number of non-zero weights $P''$, is determined from $\mathcal{N}(\boldsymbol{m}^*_{2P}(\sigma))$. If it is even, $P'' = \mathcal{N}(\boldsymbol{m}^*_{2P}(\sigma))/2$; otherwise, $P'' = (\mathcal{N}(\boldsymbol{m}^*_{2P}(\sigma)) + 1)/2$ but one node will be located in $\xi = 0$ which might be an issue for KDFs defined on $\Omega_\xi = ]0, +\infty[$ or $\Omega_\xi = ]0, 1[$. The weights and nodes will be computed from the recurrence coefficients $\boldsymbol{a}^*_{P''-1}(\sigma)$ and $\boldsymbol{b}^*_{P''-1}(\sigma)$. If $P'' < P$, let $w_k = 0, \xi_k = 1/2, \ \forall k \in \{P'' + 1, \dots, P\}$.

These adjustments of the first and last steps of algorithms described in sections 3.3, 3.4 and 3.5 give great stability to the moment-inversion procedure at low cost.

In the situation where $\mathcal{N}(\boldsymbol{m}^*_{2P}(\sigma)) = 2P$, the EQMOM approximation is guaranteed to preserve the whole moment set $\boldsymbol{m}_{2P}$. However, if $\mathcal{N}(\boldsymbol{m}^*_{2P}(\sigma)) < 2P$, the approximation may, or may not, preserve all moments with no simple method to check for this. One should compute the moments of the EQMOM approximation and measure the relative error from original moments.

## 4. Comparison of EQMOM approaches

### 4.1. Method

The new EQMOM moment-inversion procedure only requires computation of the realisability criteria of the vector of degenerated moments $\boldsymbol{m}^*_{2P}(\sigma)$ in order to identify $\sigma$. These computations were already performed in the original approach [1] to ensure the realisability of the vector $\boldsymbol{m}^*_{2P-1}(\sigma)$ prior to the quadrature computation and ulterior steps.

It is therefore obvious that the new approach will always require a lower number of floating point operations (FLOP). In order to quantify this reduction on FLOP number, and the actual performance gain, different implementations of EQMOM are compared, they are based either on the realizability criteria, or on a quadrature-based objective function.

#### 4.1.1. Tested EQMOM implementations

Comparison are performed for kernels defined on $\Omega_\xi = ]-\infty, +\infty[$ (i.e. Gauss and Laplace kernels), and on $\Omega_\xi = ]0, +\infty[$ (i.e. Log-Normal, Gamma and Weibull kernels), using MATLAB [22] implementations.

Implementations that are based on the realizability criteria of $\boldsymbol{m}^*_{2P}(\sigma)$ use algorithms that were fully described in sections 3.3 and 3.4 and adjustments from section 3.6.

For quadrature-based moment-inversion implementations, we optimized codes from Marchisio and Fox [20] and the OpenQBMM project [19] by implementing optimizations suggested by Nguyen et al. [1] and adjustments from section 3.6. Instead of searching for the root of $D_{2P}(\sigma)$ (see Fig. 1a), these implementations directly search the root of $D^*_{2P}(\sigma)$ (Fig. 1b). Doing so, all compared implementations only require the matrix $\boldsymbol{A}^{-1}_{2P}(\sigma)$ and can benefit from the same code optimization when computing $\boldsymbol{m}^*_{2P}(\sigma) = \boldsymbol{A}^{-1}_{2P}(\sigma) \cdot \boldsymbol{m}_{2P}$.

For kernels defined on $\Omega_\xi = ]0, +\infty[$, if Ridder's method fails to identify a root of $D^*_{2P}(\sigma)$, the golden-ratio method is used to minimize $D_{2P}(\sigma)^2 = \left(D^*_{2P}(\sigma) \cdot A_{2P,2P}(\sigma)\right)^2$. The golden-ratio minimization method was already used in OpenQBMM [19].

#### 4.1.2. Performance measurements

The main element of comparison is the number of floating-point operations required for the whole moment-inversion procedure. The MATLAB implementations embed a simple FLOP counter that distinguishes each operation ($+$, $-$, $*$, $/$, exp, $\sqrt{\cdot}$, $\Gamma(\cdot)$, ...) and counts them for each step of the moment-inversion procedure (linear system, Chebyshev algorithm, quadrature computation and others).

In order to evaluate the number of operations used in the computation of the eigenvalues and eigenvectors of the Jacobi matrix (Eq. (6)), the Jacobi and the Francis algorithms which are suited for symmetric matrices [35] are used in place of the MATLAB built-in "eig" function [22]. The Jacobi algorithm is used for matrices of size up to 3 × 3 and the Francis algorithm for larger matrices in order to always use the fastest method.

Two others metrics are measured for each call to the moment-inversion procedure: the number of tested $\sigma$ values and the wall-time of function calls.

**Table 1**
Comparison of Gauss EQMOM implementations corresponding to Fig. 1b and 1c for moment sets far from the frontier of realisability. The count of FLOP details the operations related to (i) the matrix–vector product $\boldsymbol{A}_{2P}^{-1}(\sigma) \cdot \boldsymbol{m}_{2P}$, (ii) the Chebyshev Algorithm (CA), (iii) the Quadrature Computation (QC) and (iv) a miscellaneous category. Results are given as mean±standard-deviation among $10^4$ moment sets.

| | | | $P=2$ | $P=3$ | $P=4$ | $P=5$ |
|---|---|---|---|---|---|---|
| New approach | FLOP | $\boldsymbol{A}_{2P}^{-1}(\sigma)$ | $237_{\pm 59}$ | $767_{\pm 141}$ | $1709_{\pm 253}$ | $3201_{\pm 476}$ |
| | | CA | $177_{\pm 40}$ | $477_{\pm 83}$ | $979_{\pm 139}$ | $1751_{\pm 251}$ |
| | | QC | $52_{\pm 0}$ | $474_{\pm 42}$ | $995_{\pm 120}$ | $1746_{\pm 188}$ |
| | | Misc. | $54_{\pm 12}$ | $65_{\pm 11}$ | $75_{\pm 11}$ | $86_{\pm 12}$ |
| | | Total | $519_{\pm 112}$ | $1783_{\pm 242}$ | $3759_{\pm 441}$ | $6784_{\pm 830}$ |
| | Evaluations | | $12_{\pm 3}$ | $14_{\pm 2}$ | $17_{\pm 2}$ | $19_{\pm 3}$ |
| | Run-time (ms) | | $1_{\pm 0}$ | $2_{\pm 0}$ | $3_{\pm 0}$ | $4_{\pm 1}$ |
| Former approach | FLOP | $\boldsymbol{A}_{2P}^{-1}(\sigma)$ | $295_{\pm 161}$ | $1433_{\pm 423}$ | $4060_{\pm 869}$ | $8516_{\pm 1870}$ |
| | | CA | $202_{\pm 102}$ | $853_{\pm 241}$ | $2246_{\pm 467}$ | $4509_{\pm 967}$ |
| | | QC | $742_{\pm 377}$ | $9171_{\pm 2910}$ | $24997_{\pm 9966}$ | $52312_{\pm 14096}$ |
| | | Misc. | $191_{\pm 99}$ | $430_{\pm 129}$ | $804_{\pm 156}$ | $1298_{\pm 251}$ |
| | | Total | $1429_{\pm 739}$ | $11887_{\pm 3603}$ | $32108_{\pm 10645}$ | $66635_{\pm 16085}$ |
| | Evaluations | | $14_{\pm 7}$ | $26_{\pm 7}$ | $39_{\pm 8}$ | $50_{\pm 11}$ |
| | Run-time (ms) | | $1_{\pm 1}$ | $9_{\pm 3}$ | $17_{\pm 5}$ | $31_{\pm 7}$ |
| Gain in | FLOP | | $59.1\%_{\pm 12.3\%}$ | $84.2\%_{\pm 3.5\%}$ | $87.9\%_{\pm 2.5\%}$ | $88.0\%_{\pm 13.1\%}$ |
| | Evaluations | | $8.6\%_{\pm 27.7\%}$ | $40.9\%_{\pm 17.7\%}$ | $54.2\%_{\pm 12.8\%}$ | $53.0\%_{\pm 55.2\%}$ |
| | Run-time | | $53.2\%_{\pm 13.2\%}$ | $81.9\%_{\pm 4.2\%}$ | $84.0\%_{\pm 3.6\%}$ | $83.3\%_{\pm 18.0\%}$ |

### *4.1.3. Tested moment sets*

Each comparison was performed on $10^4$ randomly generated moment sets. These have varying size $2P+1 \in \{5, 7, 9, 11\}$ and were either far from, or close to, the boundary of the realisable moment space.

Moments sets for kernels defined on $\Omega_\xi = ]-\infty, +\infty[$ were computed from random vectors $\boldsymbol{a}_{P-1}$ and $\boldsymbol{b}_P$ using a reversed Chebyshev algorithm. Distribution laws for the elements of these vectors are

- $a_k \sim \mathcal{N}(0, 25), k \in \{0, \ldots, P-1\}$.
- $b_k \sim 1 + \text{Exp}(4), k \in \{1, \ldots, P\}$.
- $b_P \sim \text{Exp}(0.5)$ for moment sets close from the frontier of realisability.

Similarly, moments sets for kernels defined on $\Omega_\xi = ]0, +\infty[$ were computed from random vectors $\boldsymbol{\zeta}_{2P}$ using a reversed $\zeta$-Chebyshev algorithm [1]. Elements of these vectors are generated using following distribution laws:

- $\zeta_k \sim 1 + \text{Exp}(4), k \in \{1, \ldots, 2P\}$.
- $\zeta_{2P} \sim \text{Exp}(0.5)$ for moment sets close from the frontier of realisability.

### *4.1.4. Reproducibility*

To allow reproducibility of results described hereafter, every source codes previously described, and randomly generated data, are available as supplementary data.

## *4.2. Results*

Results of the comparison performed on Gauss-EQMOM for moment sets far from the boundary of the realisable moment space are given in Table 1. Similar tables are available as supplementary data for all kernels and moment sets.

Table 2 underlines a decrease in the number of tested $\sigma$ values, in particular for high order reconstructions. This decrease is mainly due to the fact that in the former approach, if $\boldsymbol{m}_{N-1}^*(\sigma)$ turns out not to be realisable, the objective function is set to a arbitrarily high negative value. The use of such an arbitrary value slows down the convergence of the non-linear equation solver. Meanwhile, the new approach never makes use of arbitrary values, all the elements of the vectors of realisability criteria ($\boldsymbol{b}_P^*(\sigma)$, $\boldsymbol{\zeta}_{2P}^*(\sigma)$ or $\boldsymbol{p}_{2P}^*(\sigma)$) are used one after the other which yields a better choice of the next tested $\sigma$ value.

Moreover, for kernels defined on $\Omega_\xi = ]0, +\infty[$ and in situations illustrated in Fig. 3c, the former approach may switch from a root search to a minimization process if no root is found. This induces numerous supplementary tested $\sigma$ values before convergence is reached while this situation never occurs in the new approach.

A significant drop in the total number of FLOP can be observed in Table 3. This was expected and is mainly justified by the fact that the quadrature computation is only called once in the new approach whilst it is called for most tested $\sigma$ values in the former moment-inversion procedure. This quadrature, which consists in the computation of the eigenvalues

**Table 2**
Gain in number of tested $\sigma$ values.

| | | $P=2$ | $P=3$ | $P=4$ | $P=5$ |
|---|---|---|---|---|---|
| Gauss | Strict | $8.6\%_{\pm 27.7\%}$ | $40.9\%_{\pm 17.7\%}$ | $54.2\%_{\pm 12.8\%}$ | $53.0\%_{\pm 55.2\%}$ |
| | Weak | $10.2\%_{\pm 25.5\%}$ | $43.0\%_{\pm 17.4\%}$ | $54.7\%_{\pm 20.0\%}$ | $42.4\%_{\pm 86.3\%}$ |
| Laplace | Strict | $8.6\%_{\pm 28.3\%}$ | $41.1\%_{\pm 17.6\%}$ | $54.3\%_{\pm 12.7\%}$ | $53.2\%_{\pm 54.4\%}$ |
| | Weak | $9.9\%_{\pm 23.7\%}$ | $43.3\%_{\pm 17.4\%}$ | $54.8\%_{\pm 19.9\%}$ | $42.5\%_{\pm 85.7\%}$ |
| Log-normal | Strict | $8.8\%_{\pm 45.9\%}$ | $21.9\%_{\pm 32.8\%}$ | $30.5\%_{\pm 27.0\%}$ | $49.5\%_{\pm 23.8\%}$ |
| | Weak | $4.0\%_{\pm 38.1\%}$ | $17.8\%_{\pm 26.4\%}$ | $40.0\%_{\pm 27.0\%}$ | $59.1\%_{\pm 16.1\%}$ |
| Gamma | Strict | $15.5\%_{\pm 38.2\%}$ | $24.7\%_{\pm 31.1\%}$ | $34.9\%_{\pm 30.1\%}$ | $57.2\%_{\pm 24.9\%}$ |
| | Weak | $7.5\%_{\pm 28.3\%}$ | $20.7\%_{\pm 26.0\%}$ | $47.2\%_{\pm 29.6\%}$ | $65.4\%_{\pm 16.0\%}$ |
| Weibull | Strict | $26.3\%_{\pm 35.0\%}$ | $27.2\%_{\pm 30.2\%}$ | $32.4\%_{\pm 28.1\%}$ | $54.0\%_{\pm 25.8\%}$ |
| | Weak | $9.4\%_{\pm 15.7\%}$ | $19.4\%_{\pm 17.7\%}$ | $41.8\%_{\pm 25.3\%}$ | $63.3\%_{\pm 16.4\%}$ |

**Table 3**
Gain in FLOP for all tested kernels and moment sets.

| | | $P=2$ | $P=3$ | $P=4$ | $P=5$ |
|---|---|---|---|---|---|
| Gauss | Strict | $59.1\%_{\pm 12.3\%}$ | $84.2\%_{\pm 3.5\%}$ | $87.9\%_{\pm 2.5\%}$ | $88.0\%_{\pm 13.1\%}$ |
| | Weak | $59.3\%_{\pm 11.3\%}$ | $84.4\%_{\pm 3.7\%}$ | $87.8\%_{\pm 4.8\%}$ | $85.3\%_{\pm 20.4\%}$ |
| Laplace | Strict | $64.2\%_{\pm 10.7\%}$ | $87.5\%_{\pm 2.7\%}$ | $91.0\%_{\pm 1.7\%}$ | $91.6\%_{\pm 8.9\%}$ |
| | Weak | $64.1\%_{\pm 9.1\%}$ | $87.5\%_{\pm 2.9\%}$ | $90.8\%_{\pm 3.5\%}$ | $89.6\%_{\pm 14.1\%}$ |
| Log-normal | Strict | $58.9\%_{\pm 20.0\%}$ | $85.6\%_{\pm 5.8\%}$ | $89.1\%_{\pm 4.3\%}$ | $93.3\%_{\pm 3.5\%}$ |
| | Weak | $56.7\%_{\pm 16.7\%}$ | $84.7\%_{\pm 4.6\%}$ | $90.8\%_{\pm 4.4\%}$ | $94.7\%_{\pm 2.4\%}$ |
| Gamma | Strict | $58.2\%_{\pm 18.8\%}$ | $82.1\%_{\pm 7.3\%}$ | $85.7\%_{\pm 6.8\%}$ | $91.6\%_{\pm 5.4\%}$ |
| | Weak | $54.2\%_{\pm 14.1\%}$ | $81.1\%_{\pm 6.1\%}$ | $88.7\%_{\pm 6.9\%}$ | $93.4\%_{\pm 3.5\%}$ |
| Weibull | Strict | $67.7\%_{\pm 15.1\%}$ | $87.4\%_{\pm 5.1\%}$ | $90.0\%_{\pm 4.2\%}$ | $94.2\%_{\pm 3.6\%}$ |
| | Weak | $61.0\%_{\pm 6.5\%}$ | $86.7\%_{\pm 3.0\%}$ | $91.8\%_{\pm 3.9\%}$ | $95.7\%_{\pm 2.3\%}$ |

and eigenvectors of a tridiagonal symmetric matrix, is the most expensive operation used in the EQMOM moment-inversion procedure.

Concerning the impact of whether the moment sets are close or far from the boundary of the moment space, no significant difference appears in Tables 2 and 3. This implies that there are no preferential situations where the former approach would have been more interesting. The new EQMOM core procedure should be used against all moment sets.

Overall, one observes a net decrease in the number of floating-point operations and in the computation run-times of 60% up to 95% for these implementations of EQMOM and the tested moment sets.

One final interesting observation is the evolution of variability in the computational cost of each EQMOM reconstruction. This is illustrated by the ratios standard-deviation/mean shown in Table 4. The high ratios occurring for the former approach, ranging from 24% up to 794%, show that the distributions of required FLOP by EQMOM reconstruction are highly skewed. This leads to high probabilities of significantly different computational costs between different moment sets.

On the other hand, that ratio never goes higher than 26% for the new approach if $P=2$, and 16% for higher order reconstructions. It implies that the distribution of numerical cost is more narrow and that this new approach will induce more consistent numerical costs among different moment sets. This is a salient feature of this new moment-inversion algorithm as it allows better load-balancing in high performance computing, in particular in highly parallelized CFD softwares.

## 5. Conclusion

The first developments relative to the Extended Quadrature Method of Moments are quite recent [17]. Most of these developments were dedicated to widening the use of this method to new application cases, in particular by adding new reconstruction kernels to the EQMOM formalism, and to demonstrate its stability and accuracy compared to other methods. This article summarised all of these developments, relative to the Gaussian kernel [17], to the Log-normal kernel [21] and to the Gamma and Beta kernels [18]. It was also shown that at least two other kernels are perfectly compatible with the EQMOM formalism: the Laplace and Weibull kernels.

The youth of EQMOM explains that there is still room left for improvements. The core of this method – the moment-inversion procedure – is an iterative process which is its computational bottleneck. Nguyen et al. [1] proposed some modifications, compared to previous implementations, in order to stabilise the method and to speed-up its resolution, namely the use of Ridder's method instead of bounded-secant or dichotomic methods to solve the non-linear problem, and the realisability checks performed prior to the quadrature computation.

**Table 4**
Ratio standard-deviation/mean of the distribution of total number of FLOP required by EQMOM reconstructions. Only moment sets generated far from the frontier of realisability are considered.

| | | $P=2$ | $P=3$ | $P=4$ | $P=5$ |
|---|---|---|---|---|---|
| Gauss | New approach | 0.22 | 0.14 | 0.12 | 0.12 |
| | Former approach | 0.52 | 0.30 | 0.33 | 0.24 |
| Laplace | New approach | 0.21 | 0.12 | 0.11 | 0.11 |
| | Former approach | 0.67 | 0.30 | 0.43 | 0.43 |
| Log-normal | New approach | 0.26 | 0.15 | 0.12 | 0.11 |
| | Former approach | 1.10 | 0.97 | 7.94 | 6.48 |
| Gamma | New approach | 0.25 | 0.16 | 0.15 | 0.15 |
| | Former approach | 1.26 | 1.14 | 0.87 | 3.23 |
| Weibul | New approach | 0.21 | 0.13 | 0.12 | 0.11 |
| | Former approach | 1.05 | 1.46 | 7.21 | 6.41 |

Further improvements were proposed by shifting the resolution toward a new paradigm. This results in a significant decrease in computational cost of about 60%–95% in terms of required floating-point operations. This resulted in our MATLAB implementations in a similar gain in terms of computation wall-time. Moreover, the new approach offers more consistent numerical costs which will be beneficial to load-balancing in parallelized software.

In multiple works [1,18,30], EQMOM has been compared to other methods (Maximum Entropy approach or sectional methods) and exhibited (i) similar accuracy even with a lower number of resolved variables, and (ii) faster or comparable computation times. The new improvements of EQMOM will make it even more competitive as its stability and accuracy are kept while reducing the gap in terms of numerical cost between EQMOM and other cheaper methods such as Gauss or Gauss–Radau quadratures.

We strongly believe that transparency about these developments will help further refinements of EQMOM. For that reason, all sources used to generate figures and data in this article are provided as supplementary data. We also release all our EQMOM source codes both with this article and in an open-access GIT repository (url: https://gitlab.com/open-eqmom). It will be updated as well as supplemented with implementations of EQMOM in languages other than MATLAB. In the case of the Beta reconstruction kernel, some suggestions for further improvements in terms of accuracy and stability were listed in section 3.5. These will be tackled in ulterior work.

## Acknowledgements

The authors thank Sanofi Chimie – C&BD Biochemistry Vitry for its financial support. The authors would also like to thank Prof. Frédérique Laurent and Prof. Rodney O. Fox for the precious discussions and insights about EQMOM and the developments presented in this paper. The authors declare no conflict of interest.

## Appendix A. Chebyshev algorithm

The Chebyshev algorithm allows to compute the three-term recurrence coefficients of the monic polynomials orthogonal to a measure $\mathrm{d}\mu(\xi)$ whose moments are given by the vector $\boldsymbol{m}_N = [m_0, \dots, m_N]$. This version of the algorithm fills column-wise a $N+1 \times \lceil \frac{N+1}{2} \rceil$ matrix denoted $\boldsymbol{S}$.

First, fill the first column with the moments $S_{i,0} = m_i$, compute $a_0 = m_1/m_0$ and fill the second column with $S_{i,1} = S_{i+1,0} - a_0 S_{i,0}, \quad \forall i \in \{1, \dots, N-1\}$.

Then iterate for $j \in \left\{2, \dots, \lceil \frac{N-1}{2} \rceil\right\}$:

$$a_{j-1} = \frac{S_{j,j-1}}{S_{j-1,j-1}} - \frac{S_{j-1,j-2}}{S_{j-2,j-2}}$$

$$b_{j-1} = \frac{S_{j-1,j-1}}{S_{j-2,j-2}}$$

$$S_{i,j} = S_{i+1,j-1} - a_{j-1} S_{i,j-1} - b_{j-1} S_{i,j-2}, \qquad i \in \{j, \dots, N-j\}$$

## Appendix B. Kernels for EQMOM

There exists multiple variations of the EQMOM method depending on the Kernel Density Function that is used for the reconstruction in Eq. (15). This section details the specificities of multiple KDF that were found to be compatible with the EQMOM procedure. It details for each kernel

1. the actual expression of that kernel $\delta_\sigma(\xi, \xi_m)$;
2. the expression of its moments;
3. the matrix $\boldsymbol{A}_n(\sigma)$ that allows the transfer between the raw moments of the reconstruction $\widetilde{\boldsymbol{m}}_n$ and its degenerated moments $\boldsymbol{m}_n^*$;
4. the nested quadrature rules suiting this kernel;
5. the analytical solutions available for one-node EQMOM ($P = 1$).

Two-nodes analytical solutions exist for the Gaussian, Gamma, Laplace and Log-normal kernels and are accessible using the same methodology than that used by Chalons et al. [17] for the Gaussian kernel. These solutions are not detailed here but are implemented in the MATLAB code given in supplementary data.

All definitions of matrices $\boldsymbol{A}_n(\sigma)$ are given using zero-offset. The element of the first line and column of this matrix then reads $A_{0,0}(\sigma)$.

### *B.1. Gaussian kernel*

#### *B.1.1. Definition*

The Gaussian kernel $\delta_\sigma^{(G)}(\xi, \xi_m)$ was first used in EQMOM by Chalons et al. [17]. It is defined on $\Omega_\xi = ]-\infty, +\infty[$ by

$$\delta_\sigma^{(G)}(\xi, \xi_m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\xi - \xi_m)^2}{2\sigma^2}\right) \tag{B.1}$$

#### *B.1.2. Moments and linear system*

Moments of the Gaussian kernel are given by:

$$\int_{-\infty}^{+\infty} \xi^k \delta_\sigma^{(G)}(\xi, \xi_m) \mathrm{d}\xi = \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{k!}{j!(k-2j)!} \left(\frac{\sigma^2}{2}\right)^j \xi_m^{k-2j} \tag{B.2}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma^{(G)}(\xi, \xi_i)$ are given by the linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(G)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.3}$$

with

$$A_{i,j}^{(G)}(\sigma) = \begin{cases} 0 & \text{if } j > i \text{ or } (i-j \mod 2) = 1 \\ \frac{i!}{\left(\frac{i-j}{2}\right)! j!} \left(\frac{\sigma^2}{2}\right)^{\frac{i-j}{2}} & \text{otherwise} \end{cases} \tag{B.4}$$

The inverse of this matrix is given by:

$$A_{i,j}^{(G)-1}(\sigma) = \begin{cases} 0 & \text{if } j > i \text{ or } (i-j \mod 2) = 1 \\ \frac{i!}{\left(\frac{i-j}{2}\right)! j!} \left(-\frac{\sigma^2}{2}\right)^{\frac{i-j}{2}} & \text{otherwise} \end{cases} \tag{B.5}$$

which translates, for the case $n = 4$, into:

$$\begin{bmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \end{bmatrix} = \begin{pmatrix} 1 & & & & 0 \\ 0 & 1 & & & \\ \sigma^2 & 0 & 1 & & \\ 0 & 3\sigma^2 & 0 & 1 & \\ 3\sigma^4 & 0 & 6\sigma^2 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} m_0^* \\ m_1^* \\ m_2^* \\ m_3^* \\ m_4^* \end{bmatrix} \tag{B.6}$$

$$\begin{bmatrix} m_0^* \\ m_1^* \\ m_2^* \\ m_3^* \\ m_4^* \end{bmatrix} = \begin{pmatrix} 1 & & & & 0 \\ 0 & 1 & & & \\ -\sigma^2 & 0 & 1 & & \\ 0 & -3\sigma^2 & 0 & 1 & \\ 3\sigma^4 & 0 & -6\sigma^2 & 0 & 1 \end{pmatrix} \cdot \begin{bmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \end{bmatrix} \tag{B.7}$$

#### B.1.3. Moment preserving nested quadrature

The approximation of integral properties using Gauss EQMOM is performed through the following nested quadrature:

$$\int_{-\infty}^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \frac{1}{\sqrt{\pi}}\sum_{i=1}^{P} w_i \sum_{j=1}^{Q} \omega_j f\left(\xi_i + \sigma\lambda_j\sqrt{2}\right) \tag{B.8}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes Gauss–Hermite quadrature rule (see Appendix C).

#### B.1.4. Single node analytical solution

The case $P = 1$ has the following analytical solution:

$$\begin{aligned} w_1 &= m_0 \\ \xi_1 &= \frac{m_1}{m_0} \\ \sigma &= \frac{\sqrt{m_2 m_0 - m_1^2}}{m_0} \end{aligned}$$

### B.2. Laplace kernel

#### B.2.1. Definition

The Laplace kernel $\delta_\sigma^{(\lambda)}(\xi, \xi_m)$ is defined on $\Omega_\xi = ]-\infty, +\infty[$ by

$$\delta_\sigma^{(\lambda)}(\xi, \xi_m) = \frac{1}{2\sigma}\exp\left(-\frac{|\xi - \xi_m|}{\sigma}\right) \tag{B.9}$$

#### B.2.2. Moments and linear system

Moments of the Laplace kernel are given by

$$\int_{-\infty}^{+\infty} \xi^k \delta_\sigma^{(\lambda)}(\xi, \xi_m)\mathrm{d}\xi = \sum_{j=0}^{k} \frac{k!}{(k-j)!}\frac{1+(-1)^j}{2}\xi_m^{k-j}\sigma^j \tag{B.10}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma^{(\lambda)}(\xi, \xi_i)$ are given by the linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(\lambda)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.11}$$

with

$$A_{i,j}^{(\lambda)}(\sigma) = \begin{cases} 0 & \text{if } j > i \text{ or } (i - j \mod 2) = 1 \\ \frac{i!}{j!}\sigma^{i-j} & \text{otherwise} \end{cases} \tag{B.12}$$

The inverse matrix is defined by

$$A_{i,j}^{(\lambda)\ -1}(\sigma) = \begin{cases} 1 & \text{if } i = j \\ -(j+1)(j+2)\sigma^2 & \text{if } i = j+2 \\ 0 & \text{otherwise} \end{cases} \tag{B.13}$$

which translates for $n = 6$ into

$$\begin{pmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \\ \widetilde{m}_5 \\ \widetilde{m}_6 \end{pmatrix} = \begin{pmatrix} 1 & & & & & & 0 \\ 0 & 1 & & & & & \\ \frac{2!\sigma^2}{0!} & 0 & 1 & & & & \\ 0 & \frac{3!\sigma^2}{1!} & 0 & 1 & & & \\ \frac{4!\sigma^4}{0!} & 0 & \frac{4!\sigma^2}{2!} & 0 & 1 & & \\ 0 & \frac{5!\sigma^4}{1!} & 0 & \frac{5!\sigma^2}{3!} & 0 & 1 & \\ \frac{6!\sigma^6}{0!} & 0 & \frac{6!\sigma^4}{2!} & 0 & \frac{6!\sigma^2}{4!} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} m_0^* \\ m_1^* \\ m_2^* \\ m_3^* \\ m_4^* \\ m_5^* \\ m_6^* \end{pmatrix} \tag{B.14}$$

$$\begin{pmatrix} m_0^* \\ m_1^* \\ m_2^* \\ m_3^* \\ m_4^* \\ m_5^* \\ m_6^* \end{pmatrix} = \begin{pmatrix} 1 & & & & & & 0 \\ 0 & 1 & & & & & \\ -2\sigma^2 & 0 & 1 & & & & \\ & -6\sigma^2 & 0 & 1 & & & \\ & & -12\sigma^2 & 0 & 1 & & \\ & & & -20\sigma^2 & 0 & 1 & \\ 0 & & & & -30\sigma^2 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \\ \widetilde{m}_5 \\ \widetilde{m}_6 \end{pmatrix} \tag{B.15}$$

### *B.2.3. Moment preserving nested quadrature*

The approximation of integral properties using Laplace EQMOM is performed through the following nested quadrature:

$$\int_{-\infty}^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \sum_{i=1}^{P} w_i \sum_{j=1}^{Q} \omega_j f\left(\xi_i + \sigma\lambda_j\right) \tag{B.16}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes "Gauss-Laplace" quadrature rule (see Appendix C).

### *B.2.4. Single node analytical solution*

The case $P = 1$ has the following analytical solution:

$$\begin{aligned} w_1 &= m_0 \\ \xi_1 &= \frac{m_1}{m_0} \\ \sigma &= \sqrt{\frac{m_2 m_0 - m_1^2}{2m_0^2}} \end{aligned}$$

## *B.3. Log-normal kernel*

### *B.3.1. Definition*

The Log-normal kernel $\delta_\sigma^{(L)}(\xi, \xi_m)$ was first used in EQMOM by Madadi-Kandjani and Passalacqua [21]. It is defined on $\Omega_\xi = ]0, +\infty[$ by

$$\delta_\sigma^{(L)}(\xi, \xi_m) = \frac{1}{\sigma\xi\sqrt{2\pi}} \exp\left(-\frac{(\log(\xi) - \log(\xi_m))^2}{2\sigma^2}\right) \tag{B.17}$$

### *B.3.2. Moments and linear system*

Moments of the Log-normal kernel are given by

$$\int_{0}^{+\infty} \xi^k \delta_\sigma^{(L)}(\xi, \xi_m)\mathrm{d}\xi = \xi_m^k z^{k^2} \qquad \text{with } z = e^{\sigma^2/2} \tag{B.18}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma^{(L)}(\xi, \xi_i)$ are given by

$$\widetilde{m}_k = m_k^* z^{k^2} \tag{B.19}$$

This can be translated into a linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(L)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.20}$$

with $\boldsymbol{A}_n^{(L)}(\sigma)$ a diagonal matrix:

$$A_{i,j}^{(L)}(\sigma) = \begin{cases} z^{i^2} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{B.21}$$

whose inverse matrix is directly given by

$$A_{i,j}^{(L)-1}(\sigma) = \begin{cases} z^{-i^2} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{B.22}$$

### B.3.3. Low cost nested quadrature

A variable change allows approximating integral properties over a LogN EQMOM reconstruction using Gauss–Hermite quadratures [21]:

$$\int_0^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \frac{1}{\sqrt{\pi}}\sum_{i=1}^{P} w_i \sum_{j=1}^{Q} \omega_j f\left(\xi_i \exp\left(\sigma\lambda_j\sqrt{2}\right)\right) \tag{B.23}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes Gauss–Hermite quadrature rule (see Appendix C).

Parameters of this nested quadrature do not depend on $\sigma$ of the main quadrature nodes $\boldsymbol{\xi}_P$. Consequently, $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ only need to be computed once. It is worth noting that this quadrature does not preserve the moments of the distribution and only yields exact results for $f(\xi) = \log(\xi)^k,\ k \in \{0, \ldots, 2\min(P, Q) - 1\}$.

### B.3.4. Moment preserving nested quadrature

Passalacqua et al. [19] suggested the use of Gauss–Wigert quadratures [36] to preserve the moments of a LogN EQMOM reconstruction:

$$\int_0^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \sum_{i=1}^{P} w_i \sum_{j=1}^{Q} \omega_j^{(\sigma)} f\left(\xi_i \lambda_j^{(\sigma)}\right) \tag{B.24}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q^{(\sigma)}$ and $\boldsymbol{\lambda}_Q^{(\sigma)}$ are the weights and nodes of a $Q$-nodes Gauss–Wigert quadrature rule of parameter $\sigma$ (see Appendix C). This quadrature rule must be computed for each value of $\sigma$, i.e. for each LogN EQMOM reconstruction.

### B.3.5. Single node analytical solution

The case $P = 1$ has the following analytical solution:

$$w_1 = m_0$$

$$\xi_1 = \sqrt{\frac{m_1^4}{m_2 m_0^3}}$$

$$\sigma = \sqrt{\log\left(\frac{m_2 m_0}{m_1^2}\right)}$$

## B.4. Gamma kernel

### B.4.1. Definition

The Gamma kernel $\delta_\sigma^{(\Gamma)}(\xi, \xi_m)$ was first used in EQMOM by Yuan et al. [18]. It is defined on $\Omega_\xi = ]0, +\infty[$ by

$$\delta_\sigma^{(\Gamma)}(\xi, \xi_m) = \frac{\xi^{(l-1)}\exp(-\xi/\sigma)}{\Gamma(l)\sigma^l} \qquad \text{with } l = \frac{\xi_m}{\sigma} \text{ and } \Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t \tag{B.25}$$

### B.4.2. Moments and linear system

Moments of the Gamma kernel are given by

$$\int_0^{+\infty} \xi^k \delta_\sigma^{(\Gamma)}(\xi, \xi_m)\mathrm{d}\xi = G_k(\xi_m, \sigma) = \begin{cases} 1 & \text{if } k = 0 \\ \prod_{j=0}^{k-1}(\xi_m + j\sigma) & \text{otherwise} \end{cases} \tag{B.26}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma^{(\Gamma)}(\xi, \xi_i)$ are given by the linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(\Gamma)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.27}$$

with

$$A_{i,j}^{(\Gamma)}(\sigma) = \begin{cases} 0 & \text{if } j > i \text{ or } i = 0 \text{ or } j = 0 \\ 1 & \text{if } i = 0 \text{ and } j = 0 \\ A_{i-1,j-1}^{(\Gamma)}(\sigma) + (i-1)\sigma A_{i-1,j}^{(\Gamma)}(\sigma) & \text{otherwise} \end{cases} \tag{B.28}$$

The inverse of this matrix is given by

$$A^{(\Gamma)-1}_{i,j}(\sigma) = \begin{cases} 0 & \text{if } j > i \text{ or } i = 0 \text{ or } j = 0 \\ 1 & \text{if } i = 0 \text{ and } j = 0 \\ A^{(\Gamma)-1}_{i-1,j-1}(\sigma) - j\sigma A^{(\Gamma)-1}_{i-1,j}(\sigma) & \text{otherwise} \end{cases} \tag{B.29}$$

which translates, for $n = 6$ into

$$\begin{bmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \\ \widetilde{m}_5 \\ \widetilde{m}_6 \end{bmatrix} = \begin{pmatrix} 1 & & & & & & 0 \\ 0 & 1 & & & & & \\ 0 & 1\sigma & 1 & & & & \\ 0 & 2\sigma^2 & 3\sigma & 1 & & & \\ 0 & 6\sigma^3 & 11\sigma^2 & 6\sigma & 1 & & \\ 0 & 24\sigma^4 & 50\sigma^3 & 35\sigma^2 & 10\sigma & 1 & \\ 0 & 120\sigma^5 & 274\sigma^4 & 225\sigma^3 & 85\sigma^2 & 15\sigma & 1 \end{pmatrix} \cdot \begin{bmatrix} m^*_0 \\ m^*_1 \\ m^*_2 \\ m^*_3 \\ m^*_4 \\ m^*_5 \\ m^*_6 \end{bmatrix} \tag{B.30}$$

$$\begin{bmatrix} m^*_0 \\ m^*_1 \\ m^*_2 \\ m^*_3 \\ m^*_4 \\ m^*_5 \\ m^*_6 \end{bmatrix} = \begin{pmatrix} 1 & & & & & & 0 \\ 0 & 1 & & & & & \\ 0 & -\sigma & 1 & & & & \\ 0 & \sigma^2 & -3\sigma & 1 & & & \\ 0 & -\sigma^3 & 7\sigma^2 & -6\sigma & 1 & & \\ 0 & \sigma^4 & -15\sigma^3 & 25\sigma^2 & -10\sigma & 1 & \\ 0 & -\sigma^5 & 31\sigma^4 & -90\sigma^3 & 65\sigma^2 & -15\sigma & 1 \end{pmatrix} \cdot \begin{bmatrix} \widetilde{m}_0 \\ \widetilde{m}_1 \\ \widetilde{m}_2 \\ \widetilde{m}_3 \\ \widetilde{m}_4 \\ \widetilde{m}_5 \\ \widetilde{m}_6 \end{bmatrix} \tag{B.31}$$

### *B.4.3. Low cost nested quadrature*

A Gauss–Laguerre quadrature can be used to approximate integral properties over a Gamma EQMOM reconstruction:

$$\int_0^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \sum_{j=1}^{Q} \omega_j f(\sigma\lambda_j) \sum_{i=1}^{P} \frac{w_i}{\Gamma\left(\frac{\xi_i}{\sigma}\right)} \lambda_j^{\frac{\xi_i}{\sigma}-1} \tag{B.32}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes Gauss–Laguerre quadrature rule of parameter $\alpha = 0$ (see Appendix C). The advantage of this quadrature is that it only requires $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ to be computed once. However, this quadrature will not preserve the moments of the distribution.

### *B.4.4. Moment preserving nested quadrature*

A generalized Gauss–Laguerre quadrature preserves the moments of a Gamma EQMOM reconstruction:

$$\int_0^{+\infty} f(\xi)n(\xi)\mathrm{d}\xi \approx \sum_{i=1}^{P} \frac{w_i}{\Gamma\left(\frac{\xi_i}{\sigma}\right)} \sum_{j=1}^{Q} \omega_j^{(\alpha_i)} f\left(\sigma\lambda_j^{(\alpha_i)}\right) \tag{B.33}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q^{(\alpha_i)}$ and $\boldsymbol{\lambda}_Q^{(\alpha_i)}$ are the weights and nodes of a $Q$-nodes Gauss–Laguerre quadrature rule of parameter $\alpha_i = \frac{\xi_i}{\sigma} - 1$ (see Appendix C).

The accuracy of this quadrature comes with a cost related to the computation of $\boldsymbol{\omega}_Q^{(\alpha_i)}$ and $\boldsymbol{\lambda}_Q^{(\alpha_i)}$ for each value of $\alpha_i$.

### *B.4.5. Single node analytical solution*

The case $P = 1$ has the following analytical solution:

$$\begin{aligned} w_1 &= m_0 \\ \xi_1 &= \frac{m_1}{m_0} \\ \sigma &= \frac{m_2}{m_1} - \frac{m_1}{m_0} \end{aligned}$$

### B.5. Weibull kernel

#### B.5.1. Definition

The Weibull kernel $\delta_\sigma^{(W)}(\xi, \xi_m)$ is defined on $\Omega_\xi = ]0, +\infty[$ by

$$\delta_\sigma^{(W)}(\xi, \xi_m) = \frac{1}{\sigma \xi_m}\left(\frac{\xi}{\xi_m}\right)^{\frac{1-\sigma}{\sigma}} \exp\left(-\left(\frac{\xi}{\xi_m}\right)^{1/\sigma}\right) \tag{B.34}$$

#### B.5.2. Moments and linear system

Moments of the Weibull kernel are given by

$$\int_0^{+\infty} \xi^k \delta_\sigma^{(W)}(\xi, \xi_m)\mathrm{d}\xi = \xi_m^k \Gamma(1 + k\sigma) \tag{B.35}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^P w_i \delta_\sigma^{(W)}(\xi, \xi_i)$ are given by

$$\widetilde{m}_k = m_k^* \Gamma(1 + k\sigma) \tag{B.36}$$

This can be translated into a linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(W)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.37}$$

with $\boldsymbol{A}_n^{(W)}(\sigma)$ a diagonal matrix:

$$A_{i,j}^{(W)}(\sigma) = \begin{cases} \Gamma(1 + i\sigma) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{B.38}$$

whose inverse matrix is directly given by

$$A_{i,j}^{(W)-1}(\sigma) = \begin{cases} \frac{1}{\Gamma(1+i\sigma)} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{B.39}$$

#### B.5.3. Low cost nested quadrature

A Gauss–Laguerre quadrature can be used to approximate integral properties over a Weibull EQMOM reconstruction:

$$\int_0^{+\infty} f(\xi) n(\xi)\mathrm{d}\xi \approx \sum_{i=1}^P w_i \sum_{j=1}^Q \omega_j f\left(\xi_i \lambda_j^\sigma\right) \tag{B.40}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes Gauss–Laguerre quadrature rule of parameter $\alpha = 0$ (see Appendix C). The advantage of this quadrature is that it only requires $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ to be computed once. However, this quadrature will not preserve the moments of the distribution and only yields exact results for $f(\xi) = \xi^{k/\sigma},\ k \in \{0, \dots, 2\min(P, Q) - 1\}$

#### B.5.4. Moment preserving nested quadrature

One can produce a Gauss quadrature that preserves the moments of Weibull EQMOM approximations:

$$\int_0^{+\infty} f(\xi) n(\xi)\mathrm{d}\xi \approx \sum_{i=1}^P w_i \sum_{j=1}^Q \omega_j^{(\sigma)} f\left(\xi_i \lambda_j^{(\sigma)}\right) \tag{B.41}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q^{(\sigma)}$ and $\boldsymbol{\lambda}_Q^{(\sigma)}$ are the weights and nodes of a $Q$-nodes "Gauss-Weibull" quadrature rule of parameter $\sigma$ (see Appendix C). The weights and nodes of the nested quadrature need to be computed for each value of $\sigma$, i.e. for each Weibull EQMOM approximation of the NDF.

### *B.5.5. Single node numerical solution*

The parameters $w_1$, $\xi_1$ and $\sigma$ of the one-node Weibull EQMOM must be solution of the following system:

$$\begin{aligned} m_0 &= w_1 \\ \frac{m_1}{\Gamma(1+\sigma)} &= w_1\xi_1 \\ \frac{m_2}{\Gamma(1+2\sigma)} &= w_1\xi_1^2 \end{aligned}$$

The first equation gives $w_1 = m_0$ but no explicit solution exists for the two other equations. One can however notice that $s = \frac{\sigma}{1+\sigma}$ must be a root of

$$G(s) = \frac{m_2 m_0}{m_1^2} - \frac{\Gamma(\frac{1+s}{1-s})}{\Gamma(\frac{1}{1-s})^2} \tag{B.42}$$

which is monotonous, defined on $s \in [0,\ 1[$ and has the following limits

$$\begin{aligned} &G(0) = \frac{m_2 m_0}{m_1^2} > 0 \\ &\lim_{s\to 1^-} G(s) < 0 \end{aligned}$$

$G(s)$ then admits a single root that can be computed numerically with the Ridder's method. One can also narrow down, at a very low cost, the search interval $[0,\ 1[$ by using the property

$$g_n = G\left(\frac{n}{n+1}\right) = \frac{m_2 m_0}{m_1^2} - \frac{(2n)!}{(n!)^2} \tag{B.43}$$

with $n$ an integer, which induces the following recurrence relation:

$$g_n = c - h_n \tag{B.44}$$

$$h_{n+1} = \left(4 - \frac{2}{n+1}\right) h_n \tag{B.45}$$

with $c = \frac{m_2 m_0}{m_1^2}$ and $h_1 = 2$.

The proposed algorithm to identify the root of $G(s)$ is

1. Compute $c = \frac{m_2 m_0}{m_1^2}$
    - if $c < 1$, cancel the operation as the moments are not realisable;
    - if $c = 1$, $s = 0$ is the root of $G(s)$;
    - if $c < 2$, set $s_l = 0$, $v_l = c - 1$, $s_r = \frac{1}{2}$ and $v_r = c - 2$ and go to step 3.
    - otherwise, set $s_l = 0$, $v_l = c - 1$ and go to step 2.
2. Initialise $i = 1$, $h = 2$ and iterate
    - (a) increment $i$ by 1;
    - (a) compute $h = h * \left(4 - \frac{2}{i}\right)$
        - if $h = c$, then $s = \frac{i}{i+1}$ is a root of $G(s)$;
        - if $h < c$, set $s_l = \frac{i}{i+1}$ and $v_l = c - h$;
        - if $h > c$, set $s_r = \frac{i}{i+1}$, $v_r = c - h$ and go to step 3.
3. Apply the Ridder's method to $G(s)$ on the interval $[s_l,\ s_r]$
    - (a) compute $s_{t_1} = \frac{1}{2}(s_l + s_r)$ and $v_{t_1} = G(s_{t_1})$;
    - (b) compute $s_{t_2} = s_{t_1} + (s_{t_1} - s_l)\frac{v_{t_1}}{\sqrt{v_{t_1}^2 - v_l v_r}}$ and $v_{t_2} = G(s_{t_2})$;
    - (c) set $s_l$ the highest value between $s_l$, $s_{t_1}$ and $s_{t_2}$ whose image by $G$ is positive;
    - (d) set $s_r$ the lowest value between $s_r$, $s_{t_1}$ and $s_{t_2}$ whose image by $G$ is negative;
    - (e) stop the computation if $v_l < \varepsilon(c-1)$ with $\varepsilon$ a relative tolerance (e.g. $\varepsilon = 10^{-10}$) and consider $s_l$ as a root of $G(s)$.

Once the root of $G(s)$ is identified, compute

$$\begin{aligned} \sigma &= \frac{s}{1-s} \\ \xi_1 &= \frac{m_1}{m_0\Gamma(1+\sigma)} \end{aligned}$$

Note that each iteration of the Ridder's method requires two computations of $G(s)$, that implies four computations of the Gamma function – which is quite expensive – by iteration. This explains the interest of the second step which allows to narrow down the research interval at hardly no cost.

### B.6. Beta kernel

#### B.6.1. Definition

The Beta kernel $\delta_\sigma^{(\beta)}(\xi, \xi_m)$ was first used in EQMOM by Yuan et al. [18]. It is defined on $\Omega_\xi = ]0, 1[$ by

$$\delta_\sigma^{(\beta)}(\xi, \xi_m) = \frac{\xi^{(l-1)}(1-\xi)^{(m-1)}}{B(l, m)} \qquad \text{with } l = \frac{\xi_m}{\sigma} \text{ and } m = \frac{1-\xi_m}{\sigma} \tag{B.46}$$

with $B(l, m) = \int_0^1 x^{(l-1)}(1-x)^{(m-1)}\mathrm{d}x$ the beta function.

#### B.6.2. Moments and linear system

Moments of the Beta kernel are given by

$$\int_0^1 \xi^k \delta_\sigma^{(\beta)}(\xi, \xi_m)\mathrm{d}\xi = H_k(\xi_m, \sigma) = \begin{cases} 1 & \text{if } k = 0 \\ \prod_{j=0}^{k-1}\left(\frac{\xi_m + j\sigma}{1 + j\sigma}\right) & \text{otherwise} \end{cases} \tag{B.47}$$

Moments of the distribution $\widetilde{n}(\xi) = \sum_{i=1}^{P} w_i \delta_\sigma^{(\beta)}(\xi, \xi_i)$ are given by the linear system

$$\widetilde{\boldsymbol{m}}_n = \boldsymbol{A}_n^{(\beta)}(\sigma) \cdot \boldsymbol{m}_n^* \tag{B.48}$$

with the elements of $\boldsymbol{A}_n^{(\beta)}(\sigma)$ being computed from the elements of the matrix relative to Gamma EQMOM, $\boldsymbol{A}_n^{(\Gamma)}(\sigma)$:

$$A_{i,j}^{(\beta)}(\sigma) = \frac{A_{i,j}^{(\Gamma)}(\sigma)}{F_i(\sigma)} \tag{B.49}$$

$$F_i(\sigma) = \begin{cases} 1 & \text{if } i \le 1 \\ (1 + (i-1)\sigma)F_{i-1}(\sigma) & \text{otherwise} \end{cases} \tag{B.50}$$

The inverse of this matrix is also easily defined from $\boldsymbol{A}_n^{(\Gamma)-1}(\sigma)$:

$$A_{i,j}^{(\beta)-1}(\sigma) = A_{i,j}^{(\Gamma)-1}(\sigma)F_j(\sigma) \tag{B.51}$$

#### B.6.3. Low cost nested quadrature

A Gauss–Legendre quadrature can be used to approximate integral properties over a Beta EQMOM reconstruction:

$$\int_0^1 f(\xi)n(\xi)\mathrm{d}\xi \approx \frac{1}{2}\sum_{i=1}^{P} \frac{w_i}{B(\alpha_i+1, \beta_i+1)} \sum_{j=1}^{Q} \omega_j f\left(\frac{1-\lambda_j}{2}\right)\left(\frac{1-\lambda_j}{2}\right)^{\alpha_i}\left(\frac{1+\lambda_j}{2}\right)^{\beta_i} \tag{B.52}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ are the weights and nodes of a $Q$-nodes Gauss–Legendre quadrature rule (see Appendix C); $\alpha_i = \frac{\xi_i - \sigma}{\sigma}$ and $\beta_i = \frac{1 - \xi_i - \sigma}{\sigma}$. This nested quadrature only requires $\boldsymbol{\omega}_Q$ and $\boldsymbol{\lambda}_Q$ to be computed once, but will not preserve the moments of the distribution.

#### B.6.4. Moment preserving nested quadrature

A Gauss–Jacobi quadrature will preserve the moments of the distribution:

$$\int_0^1 f(\xi)n(\xi)\mathrm{d}\xi \approx 2^{\frac{\sigma-1}{\sigma}} \sum_{i=1}^{P} \frac{w_i}{B(\alpha_i+1, \beta_i+1)} \sum_{j=1}^{Q} \omega_j^{(\alpha_i, \beta_i)} f\left(\frac{1-\lambda_j^{(\alpha_i, \beta_i)}}{2}\right) \tag{B.53}$$

with $\boldsymbol{w}_P$, $\boldsymbol{\xi}_P$ and $\sigma$ the EQMOM reconstruction parameters computed from $\boldsymbol{m}_{2P}$; $\boldsymbol{\omega}_Q^{(\alpha_i, \beta_i)}$ and $\boldsymbol{\lambda}_Q^{(\alpha_i, \beta_i)}$ are the weights and nodes of a $Q$-nodes Gauss–Jacobi quadrature rule of parameters $\alpha_i = \frac{\xi_i - \sigma}{\sigma}$ and $\beta_i = \frac{1 - \xi_i - \sigma}{\sigma}$ (see Appendix C). The moment-preserving property of this quadrature comes with the need to compute $\boldsymbol{\omega}_Q^{(\alpha_i, \beta_i)}$ and $\boldsymbol{\lambda}_Q^{(\alpha_i, \beta_i)}$ for each node of the main Beta EQMOM quadrature.

### *B.6.5. Single node analytical solution*

The case $P = 1$ has the following analytical solution:

$$\begin{aligned} w_1 &= m_0 \\ \xi_1 &= \frac{m_1}{m_0} \\ \sigma &= \frac{m_1^2 - m_0 m_2}{m_0(m_2 - m_1)} \end{aligned}$$

## Appendix C. Gaussian quadratures

A Q-node Gaussian quadrature allows to approximate a function integral as a weighted sum of point wise values of this function over an interval $I$:

$$\int_I f(x) p(x) \mathrm{d}x \approx \sum_{j=1}^{Q} \omega_j f\left(\lambda_j\right) \tag{C.1}$$

$p(x)$ is a weight function, and the quadrature rule yields accurate integral evaluations if $f(x) = x^k,\ k \in \{0, \ldots, 2Q-1\}$. The computation of the weights $\boldsymbol{\omega}_Q$ and nodes $\boldsymbol{\lambda}_Q$ is performed as detailed in 2.2 by considering polynomials that are orthogonal with respect to the weight function $p(x)$.

Table C.1 details for each Gauss quadrature:

- the weight function $p(x)$;
- the integration support $I$;
- the computation of recurrence coefficients $\boldsymbol{a}_{Q-1}$ and $\boldsymbol{b}_{Q-1}$;
- the zero-th order moment $P_0$ of $p(x)$.

The recurrence coefficients are used to construct the Jacobi matrix $\boldsymbol{J}_Q$ associated with $p(x)$ on $I$ (see Eq. (6)). The nodes $\boldsymbol{\lambda}_Q$ are the eigenvalues of $\boldsymbol{J}_Q$, and the weights $\boldsymbol{\omega}_Q$ are given by $\omega_j = P_0 v_{1,j}^2$ with $v_{1,j}$ the first component of the normalised eigenvector belonging to the eigenvalue $\lambda_j$.

**Table C.1**
Specifics of Gauss quadratures used for EQMOM nested quadratures.

| Gauss- | $I$ | $p(x)$ | $\boldsymbol{a}_Q$ and $\boldsymbol{b}_Q$ | $P_0$ |
|---|---|---|---|---|
| Hermite | $\mathbb{R}$ | $\exp\left(-x^2\right)$ | $a_k = 0$<br>$b_k = k/2$ | $\sqrt{\pi}$ |
| Laplace[c] | $\mathbb{R}$ | $\exp\left(-\lvert x\rvert\right)/2$ | Apply Chebyshev algorithm to<br>$\boldsymbol{P}_{2Q-1}$ with $P_k = \begin{cases} 0 & \text{if } k \text{ odd} \\ k! & \text{if } k \text{ even} \end{cases}$ | 1 |
| Laguerre[f] | $\mathbb{R}^+$ | $x^\alpha \exp(-x)$ | $a_0 = 1 + \alpha$<br>$a_k = 2 + a_{k-1}$<br>$b_k = k(k+\alpha)$ | $\Gamma(1+\alpha)$[d] |
| Wigert[a,f] | $\mathbb{R}^+$ | $\frac{1}{\gamma x\sqrt{2\pi}} \exp\left(\frac{\log^2(x)}{2\gamma^2}\right)$ | $a_k = \left(\left(z^2+1\right)z^{2k} - 1\right)z^{2k-1}$<br>$b_k = \left(z^{2k} - 1\right)z^{6k-4}$<br>$z = \exp(\gamma^2/2)$ | 1 |
| Weibull[c,f] | $\mathbb{R}^+$ | $\gamma x^{\gamma-1} \exp\left(-x^\gamma\right)$ | Apply Chebyshev algorithm to<br>$\boldsymbol{P}_{2Q-1}$ with $P_k = \Gamma(1 + k/\gamma)$ | 1 |
| Legendre[b] | $]-1, 1[$ | 1 | $a_k = 0$<br>$b_k = \dfrac{k^2}{4k^2 - 1}$ | 2 |
| Jacobi[b,f] | $]-1, 1[$ | $(1-x)^\alpha\ (1+x)^\beta$ | $a_k = \dfrac{\beta^2-\alpha^2}{\delta_k(\delta_k+2)}$<br>$b_k = \dfrac{4k(k+\alpha)(k+\beta)(k+\alpha+\beta)}{\delta_k^2(\delta_k^2-1)}$<br>$\delta_k = 2k+\alpha+\beta$ | $2^{\alpha+\beta+1} \times$<br>$B(\alpha+1, \beta+1)$[e] |

---

[a] Wilck [36].

[b] Shen et al. [37].

[c] Not standard Gauss-quadrature.

[d] $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t$.

[e] $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

[f] $\alpha > -1,\ \beta > -1,\ \gamma > 0$.

## Appendix D. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcp.2018.03.027.

## References

[1] T. Nguyen, F. Laurent, R. Fox, M. Massot, Solution of population balance equations in applications with fine particles: mathematical modeling and numerical schemes, J. Comput. Phys. (ISSN 0021-9991) 325 (2016) 129–156, https://doi.org/10.1016/j.jcp.2016.08.017.

[2] P. Moilanen, M. Laakkonen, O. Visuri, V. Alopaeus, J. Aittamaa, Modelling mass transfer in an aerated 0.2 $m^3$ vessel agitated by Rushton, Phasejet and Combijet impellers, Chem. Eng. J. (ISSN 1385-8947) 142 (1) (2008) 95–108, https://doi.org/10.1016/j.cej.2008.01.033.

[3] A. Buffo, V. Alopaeus, A novel simplified multivariate PBE solution method for mass transfer problems, Chem. Eng. Sci. (ISSN 0009-2509) 172 (Supplement C) (2017) 463–475, https://doi.org/10.1016/j.ces.2017.06.036.

[4] J. Morchain, J. Gabelle, A. Cockx, A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors, AIChE J. (ISSN 1547-5905) 60 (1) (2014) 27–40, https://doi.org/10.1002/aic.14238.

[5] A. Heins, R.L. Fernandes, K.V. Gernaey, A.E. Lantz, Experimental and in silico investigation of population heterogeneity in continuous Sachharomyces cerevisiae scale-down fermentation in a two-compartment setup, J. Chem. Technol. Biotechnol. (ISSN 1097-4660) 90 (2) (2015) 324–340, https://doi.org/10.1002/jctb.4532.

[6] S. Salenbauch, A. Cuoci, A. Frassoldati, C. Saggese, T. Faravelli, C. Hasse, Modeling soot formation in premixed flames using an extended conditional quadrature method of moments, Combust. Flame (ISSN 0010-2180) 162 (6) (2015) 2529–2543, https://doi.org/10.1016/j.combustflame.2015.03.002.

[7] A. Wick, T.-T. Nguyen, F. Laurent, R.O. Fox, H. Pitsch, Modeling soot oxidation with the extended quadrature method of moments, Proc. Combust. Inst. (ISSN 1540-7489) 36 (1) (2017) 789–797, https://doi.org/10.1016/j.proci.2016.08.004.

[8] J. Akroyd, A.J. Smith, R. Shirley, L.R. McGlashan, M. Kraft, A coupled CFD-population balance approach for nanoparticle synthesis in turbulent reacting flows, Chem. Eng. Sci. (ISSN 0009-2509) 66 (17) (2011) 3792–3805, https://doi.org/10.1016/j.ces.2011.05.006.

[9] P. Fede, O. Simonin, P. Villedieu, Monte-Carlo simulation of colliding particles or coalescing droplets transported by a turbulent flow in the framework of a joint fluid–particle pdf approach, Int. J. Multiph. Flow (ISSN 0301-9322) 74 (Supplement C) (2015) 165–183, https://doi.org/10.1016/j.ijmultiphaseflow.2015.04.006.

[10] S. Kumar, D. Ramkrishna, On the solution of population balance equations by discretization, I: a fixed pivot technique, Chem. Eng. Sci. (ISSN 0009-2509) 51 (8) (1996) 1311–1332, https://doi.org/10.1016/0009-2509(96)88489-2.

[11] S. Kumar, D. Ramkrishna, On the solution of population balance equations by discretization, II: moving pivot technique, Chem. Eng. Sci. (ISSN 0009-2509) 51 (8) (1996) 1333–1342, https://doi.org/10.1016/0009-2509(95)00355-X.

[12] M. Massot, F. Laurent, D.K.S. de Chaisemartin, A robust moment method for evaluation of the disappearance rate of evaporating sprays, SIAM J. Appl. Math. 70 (8) (2010) 3203–3234, https://doi.org/10.1137/080740027.

[13] V. John, I. Angelov, A. Öncül, D. Thévenin, Techniques for the reconstruction of a distribution from a finite number of its moments, Chem. Eng. Sci. (ISSN 0009-2509) 62 (11) (2007) 2890–2904, https://doi.org/10.1016/j.ces.2007.02.041.

[14] L. Mead, N. Papanicolaou, Maximum entropy in the problem of moments, J. Math. Phys. 25 (8) (1984) 2404–2417.

[15] A. Tagliani, Hausdorff moment problem and maximum entropy: a unified approach, Appl. Math. Comput. (ISSN 0096-3003) 105 (2–3) (1999) 291–305, https://doi.org/10.1016/S0096-3003(98)10084-X.

[16] G. Athanassoulis, P. Gavriliadis, The truncated Hausdorff moment problem solved by using kernel density functions, Probab. Eng. Mech. (ISSN 0266-8920) 17 (3) (2002) 273–291, https://doi.org/10.1016/S0266-8920(02)00012-7.

[17] C. Chalons, R. Fox, M. Massot, A multi-Gaussian quadrature method of moments for gas–particle flows in a LES framework, in: Proceedings of the 2010 Summer Program, 2010, pp. 347–358, Center for Turbulence Research, Stanford University.

[18] C. Yuan, F. Laurent, R. Fox, An extended quadrature method of moments for population balance equations, J. Aerosol Sci. (ISSN 0021-8502) 51 (2012) 1–23, https://doi.org/10.1016/j.jaerosci.2012.04.003.

[19] A. Passalacqua, F. Laurent, E. Madadi-Kandjani, J. Heylmun, R. Fox, An open-source quadrature-based population balance solver for OpenFOAM, Chem. Eng. Sci. (ISSN 0009-2509) 176 (2018) 306–3018, https://doi.org/10.1016/j.ces.2017.10.043.

[20] D. Marchisio, R. Fox, Computational Models for Polydisperse Particulate and Multiphase Systems, Cambridge Series in Chemical Engineering, Cambridge University Press, ISBN 9781107328174, 2013.

[21] E. Madadi-Kandjani, A. Passalacqua, An extended quadrature-based moment method with log-normal kernel density functions, Chem. Eng. Sci. (ISSN 0009-2509) 131 (2015) 323–339, https://doi.org/10.1016/j.ces.2015.04.005.

[22] MATLAB, version 9.0 (R2016a), The MathWorks, Inc., Natick, Massachusetts, United States, 2016.

[23] R. McGraw, Description of aerosol dynamics by the quadrature method of moments, Aerosol Sci. Technol. 27 (2) (1997) 255–265, https://doi.org/10.1080/02786829708965471.

[24] W. Gautschi, Orthogonal Polynomials: Computation and Approximation, Numerical Mathematics and Scientific Computation, Oxford University Press, ISBN 9780198506720, 2004.

[25] P. Henrici, The quotient-difference algorithm, Natl. Bur. Stand., Appl. Math. Ser. 49 (1958) 23–46.

[26] H. Dette, W. Studden, The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis, John Wiley & Sons, New York, Chichester, 1997.

[27] R. Gordon, Error Bounds in Equilibrium Statistical Mechanics, J. Math. Phys. 9 (1968) 655, https://doi.org/10.1063/1.1664624.

[28] J. Wheeler, Modified moments and Gaussian quadratures, Rocky Mt. J. Math. 4 (2) (1974) 287–296, RMJ-1974-4-2-287.

[29] N. Lebaz, A. Cockx, M. Spérandio, J. Morchain, Reconstruction of a distribution from a finite number of its moments: a comparative study in the case of depolymerization process, Comput. Chem. Eng. (ISSN 0098-1354) 84 (2016) 326–337, https://doi.org/10.1016/j.compchemeng.2015.09.008.

[30] M. Pigou, J. Morchain, P. Fede, M. Penet, G. Laronze, An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors, Chem. Eng. Sci. (ISSN 0009-2509) 171 (2017) 218–232, https://doi.org/10.1016/j.ces.2017.05.026.

[31] J. Favard, Sur les polynomes de Tchebicheff, C. R. Hebd. Séances Acad. Sci. 200 (1935) 2052–2053.

[32] T. Chihara, An Introduction to Orthogonal Polynomials, vol. 13, Mathematics and Its Applications, Gordon and Breach, ISBN 9780677041506, 1978.

[33] J. Shohat, J. Tamarkin, The Problem of Moments, Mathematical Surveys and Monographs, vol. 4, American Mathematical Society, 1943.

[34] H. Wall, Analytic Theory of Continued Fractions, The University Series in Higher Mathematics, Van Nostrand, 1948.

[35] W. Ford, Chapter 19 – The symmetric eigenvalue problem, in: Numerical Linear Algebra with Applications, Academic Press, Boston, ISBN 978-0-12-394435-1, 2015, pp. 439–468.

[36] M. Wilck, A general approximation method for solving integrals containing a lognormal weighting function, J. Aerosol Sci. (ISSN 0021-8502) 32 (9) (2001) 1111–1116, https://doi.org/10.1016/S0021-8502(01)00044-1.

[37] J. Shen, T. Tang, L.-L. Wang, Orthogonal Polynomials and Related Approximation Results, Springer, Berlin, Heidelberg, ISBN 978-3-540-71041-7, 2011, pp. 47–140.

# D

# COUPLED SIMULATION RESULTS

This Appendix details all simulations results for coupled simulations described in Chapter V.

For confidentiality reasons, the complete thesis cannot be made fully available. In the current version, this appendix has been edited out.