An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors

Maxime Pigou, Jérôme Morchain, Pascal Fede, Marie-Isabelle Penet, Geoffrey Laronze

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*

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.

Keywords: Biological dynamics, Population Balance Model, Method of classes, Method of moments, (E)QMOM, Maximum entropy

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 themselves already require quite important computation power. The most advanced cell models, which result from community efforts in

*Corresponding author, Phone: +33561559275
Email address: maxime.pigou@insa-toulouse.fr (Maxime PIGOU)

integrating genome-scale reconstructions of a strain metabolic network, 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, 2013). These models describe state of the art knowledge of a cell functioning, 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 computation power increased significantly over the past few decades, coupling the CFD approach with a biological modelling that fully embraces the biological complexity –by solving the intracellular concentrations for each cell in a bioreactor with an Euler-Lagrange framework– is still not possible. Such an approach is numerically untractable. Usual solutions are either to neglect the spatial heterogeneity and solve a complex metabolic model in homogeneous batch or chemostat cultures (Meadows et al., 2010; Matsuoka and Shimizu, 2013) by assuming that all cells are identical; or to 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).

As concentration gradients are known to be responsible for metabolic dysfunctions in large-scale reactors (Enfors et al., 2001), 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) which constitutes transient conditions, while kinetic models are often 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, 2001) and in the metabolic description of *E. coli* by simplifying the key reactions of the central carbon metabolism into a 6 reactions model derived from 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., 2016).

Until now, we solved the PBM using a class method (also known as fixed pivot method, Kumar and Ramkrishna (1996a); Mantzaris et al. (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 focuses on improving the numerical tractability of the PBM, through the use of Methods of Moments (MOM), in order to increase back the allowed level of spatial accuracy. Instead of doing a direct resolution of the population balance equation, MOM describe 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 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 Maximum Entropy method (MaxEnt, 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 provided 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 a clever fusion of KDEM with the QMOM approach (Marchisio et al., 2003a,b,c). This method has proven to be stable and efficient, in particular near the moment-space frontier where MaxEnt is ill-conditioned, but requires to make assumptions over the shape of the reconstruction.

The current work is focused on assessing the methods QMOM, EQMOM and MaxEnt against the already used class method, in the perspective of running predictive, and numerically tractable, bioreactor simulations. All these methods will be implemented for the simulation of a homogeneous chemostat culture stressed with a dilution rate shift (Kätterer et al., 1986). After a first comparison of their numerical efficiency, their accuracy, and their stability when facing such a sudden perturbation, we will reproduce the results from Pigou and Morchain (2015) using these Methods of Moments, in a heterogeneous case based on the Compartment Model from Vrábel et al. (1999, 2000), and compare once again their results against the class method.

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 μ (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 μ 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 μ 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 μ. 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, μ; (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 μ^* 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:

$$\mu^* = \mu_{\max} \frac{C_G}{C_G + K_G} \frac{C_O}{C_O + K_O} \frac{K_{\mathrm{i},A}}{C_A + K_{\mathrm{i},A}} \tag{6}$$

with K_G and K_O the affinity constants ($kg.m^{-3}$) of the biomass toward glucose and oxygen, and $K_{\mathrm{i},A}$ 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 μ. 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, μ_a, as the minimum between its biological growth capability, μ, and the environment equilibrium growth rate μ^* (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, 1999). The detailed explanation for this change is given in Morchain et al. (2016).

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 micromixing time-scale, and depends on the fluid viscosity, ν (m^2/s), and the turbulent energy dissipation rate, ε (, whose value usually ranges from 0.5 to $10W/kg$ depending on the bioreactor stirring.

As R_G is an output of the metabolic calculation procedure (Eq. 2), which itself depends on μ_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. (2016), the actual growth rate is given by:

$$\mu_a = \Psi \ \mu \tag{9}$$

where the coefficient Ψ reads:

$$\Psi = 1 - e^{-C_G/C_{G_T}} \tag{10}$$

As a recall, the estimation of μ_a along with the calculation of μ^* 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 μ_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:

$$G + Y_{EG}\ E \xrightarrow{q^G_{ana}} Y_{XG}\ X \tag{R_1}$$

$$A + Y_{EA}\ E \xrightarrow{q^A_{ana}} Y_{XA}\ X \tag{R'_1}$$

$$G + Y_{OG}\ O \xrightarrow{q^G_{oxy}} Y^o_{EG}\ E \tag{R_2}$$

$$A + Y_{OA}\ O \xrightarrow{q^A_{oxy}} Y^o_{EA}\ E \tag{R'_2}$$

$$G \xrightarrow{q^G_{ferm}} Y^f_{EG}\ E + Y_{AG}\ A \tag{R_3}$$

$$G \xrightarrow{q^G_{over}} Y_{AG}\ A \tag{R_4}$$

G: Glucose, E: Energy, A: Acetate, O: Oxygen, X: Biomass. Y_{BA} is the stoichiometric molar coefficient in mol_B/mol_A. q_α^G and q_α^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 (μ), whose evolution is described by the Population Balance Model (see section 2.3),
- The environmental conditions (G, A and O concentrations) and the Ψ 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., 2016). 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:

$$\frac{\partial n(\mu)}{\partial t} = -\frac{\partial}{\partial \mu}\left[n(\mu)\zeta(\mu)\right] + \int_0^{+\infty} \beta(\mu,\mu')n(\mu')\Psi\mu'\mathrm{d}\mu' \tag{11}$$

The first term in the right-hand side of equation 11 is a convection term in the μ-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 μ-space or equivalently to the rate of change of μ 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 equation 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 μ' produces a daughter cell whose specific growth rate is μ. The analysis of recent experimental data revealed that β can be modeled using a skew-normal distribution (Yasuda, 2011; Morchain et al., 2016) 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 μ-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:

$$m_k = \sum_{i=1}^{O} w_i L_i^k \qquad \forall k \in \{0, \ldots, N-1\} \tag{16}$$

O is here the order of the method, which deals with $N = 2 * O$ 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 μ. However, we used in Pigou and Morchain (2015)

one time constant $T_u = 1.9h$ for individuals that are moving upward in the μ-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:

$$T(\mu) = \begin{cases} T_u & \text{if } \mu < \mu^* \\ T_d & \text{otherwise} \end{cases} \tag{17}$$

$$T_m = m_0^{-1} \int_{\Omega_\mu} T(\mu) n(\mu) \mathrm{d}\mu \tag{18}$$

$$= \alpha T_u + (1-\alpha) T_d \quad \text{with} \quad \alpha = m_0^{-1} \int_0^{\mu^*} n(\mu) \mathrm{d}\mu \tag{19}$$

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

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 $X_j, j \in \{0,\dots,NC\}$	Method of Moments $m_k, k \in \{0,\dots,N-1\}$		
		QMOM	EQMOM*	MaxEnt*
NDF approximation $\hat{n}(\mu)$	$\sum_{i=1}^{NC} \frac{X_j}{\delta\mu} \delta(\mu - \mu_j)$	$\sum_{i=1}^{N/2} w_i \delta(\mu - L_i)$	$\sum_{i=1}^{\lfloor N/2 \rfloor} w_i \kappa(\mu, L_i, \sigma)$	$\exp\left(\sum_{i=0}^{N-1} \varphi_i \mu^i\right)$
m_{N+1} 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 (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	N	N

Figure 1: Summary of applied methods to couple the population balance with transport and reaction. *Numerically expensive 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 equation 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 μ. 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 $\mathbf{w}$ and $\mathbf{L}$ 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 $\mathbf{w}$ and $\mathbf{L}$ 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 moment space, which implies that the distribution $\hat{n}(\mu)$ is a sum of $O = N/2$ weighted Dirac distributions. The reconstructed NDF is then given by:

$$\hat{n}(\mu) = \sum_{i=1}^{O} w_i \delta(\mu - L_i) \tag{23}$$

Thus the reaction term (Eq. 2) can be approximated by:

$$R(\mathbf{C}) \approx \sum_{i=1}^{O} w_i \mathbf{\Phi}(L_i, \mathbf{C}) \tag{24}$$

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.

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 O-nodes EQMOM reconstruction, has the following expression:

$$\hat{n}(\mu) = \sum_{i=1}^{O} w_i \kappa(\mu, L_i, \sigma) \tag{25}$$

This method then requires the first $N = 2O+1$ moments of the NDF, in order to identify uniquely the value of w_i, L_i and σ. The following kernels are known to be compatible with the EQMOM procedure: Gaussian κ_G (Chalons et al., 2010), Log-Normal κ_L (Madadi-Kandjani and Passalacqua, 2015), Beta κ_β of Gamma κ_Γ (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:

$$\kappa_G(\mu, L, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu-L)^2}{2\sigma^2}} \tag{26}$$

This method relies on the Wheeler algorithm (Marchisio and Fox, 2013), in order to identify the values of $\mathbf{w}$ and $\mathbf{L}$. On top of that, a non-linear solver must identify the unique value of σ 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 $O = 1$ and $O = 2$ 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).

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:

$$H[f] = -\int_{-\infty}^{+\infty} f(x) \ln(f(x)) \mathrm{d}x \tag{27}$$

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):

$$\hat{n}(\mu) = \exp\left(-\sum_{i=0}^{M} \varphi_i \mu^i\right) \tag{28}$$

With $M = N - 1$ the highest order of known moments.

The key issue is to identify the values of the polynomial coefficients φ_i, which is achieved through the minimization of the following function (Kapur, 1989; Mead and Papanicolaou, 1984):

$$\Gamma(\varphi_1, \dots \varphi_M) = \sum_{k=1}^{M} \varphi_k \frac{m_k}{m_0} + \ln \left(\int_0^1 \exp \left(- \sum_{k=1}^{M} \varphi_k \mu^k \right) \mathrm{d}\mu \right) \tag{29}$$

The Γ 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 of the frontier of the 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:

$$\hat{m}_k = \int_0^1 \mu^k \exp \left(- \sum_{i=0}^{M} \varphi_i \mu^i \right) \mathrm{d}\mu \qquad k \in \{0, \dots, 2M\} \tag{30}$$

These integrals must be evaluated numerically as no analytic form exists as soon as $M > 2$, 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 Figure 1. However as moments evolve in a continuous way over time, the φ_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.

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 bounds 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.

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_M computed from the set of known moments $m_0, \dots, m_M$:

- If $p_M^{(n)} < 0.4 :\ K^{(n+1)} = 0.96 * K^{(n)}$
- If $p_M^{(n)} > 0.6 :\ K^{(n+1)} = 1.04 * K^{(n)}$

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 δt 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 $D = 0.1h^{-1}$ for $30h$ in order to reach a steady-state, we then apply a sudden shift in dilution rate toward $D' = 0.42h^{-1}$ in order to analyse 15 hours 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 Figure 2 shows simulation results for each method, with different orders or resolution. We applied QMOM with order ranging from $O = 2$ to $O = 5$ ($N = 2O$), EQMOM with order ranging from $O = 1$ to $O = 3$ ($N = 2O + 1$) and MaxEnt with N ranging from 3 to 7.

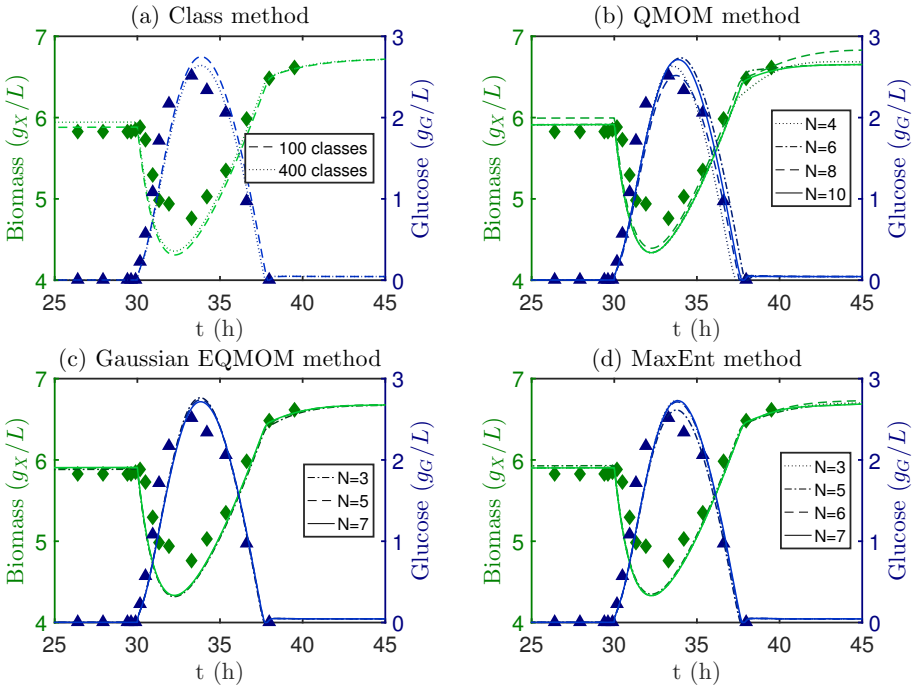

Figure 2: Simulation results for each method, compared with experimental data from Kätterer et al. (1986).

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 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 Figure 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 μ 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 (Figure 2b). However, as the resulting reconstructions are quite similar with EQMOM and MaxEnt (Figure 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 behaviour 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. Figure 4 illustrates these different acetate produc-

tions 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, due to the fact that they do not account for the narrow peak of the distribution (see Figure 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 Figure 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.

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. (2016) 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.

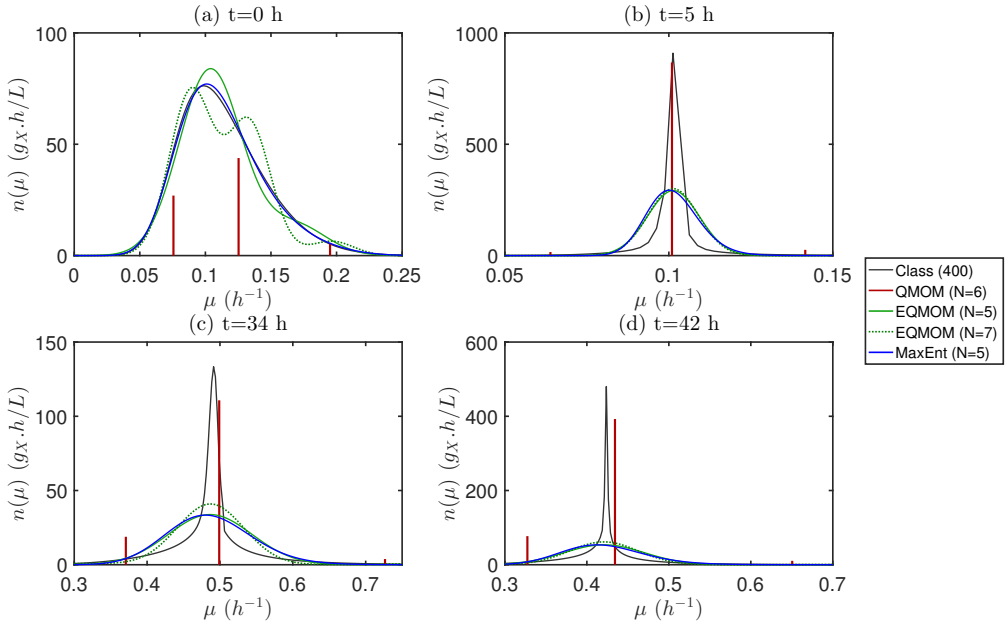

Figure 3: Comparison of reconstructed distributions against distribution resolved by class. An arbitrary scale is used for the Dirac distribution (QMOM).

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

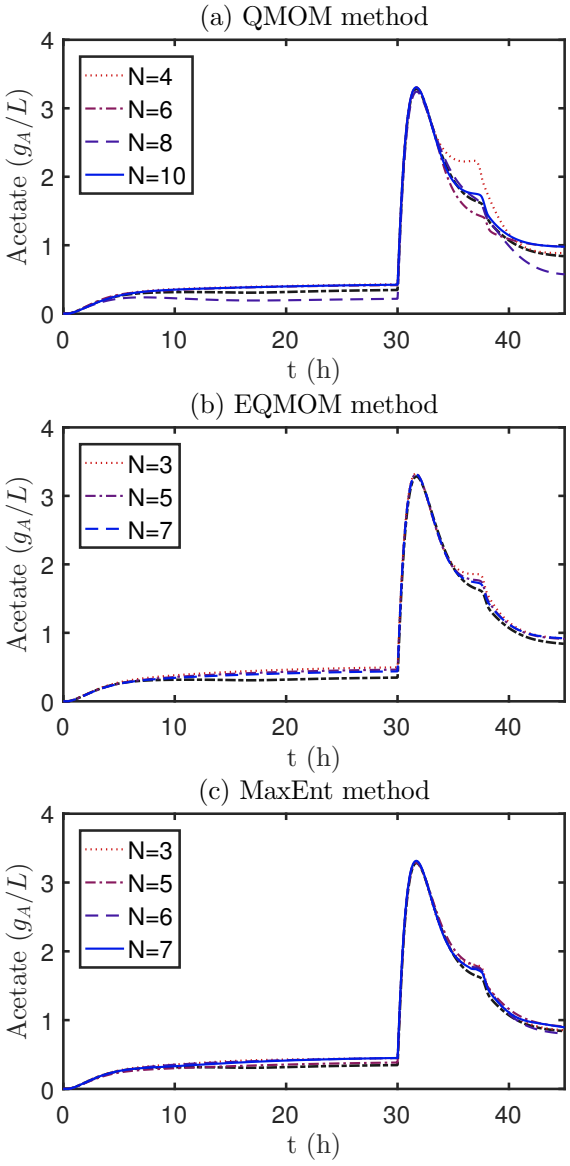

Figure 4: Evolution of acetate concentrations as predicted by the *E. coli* metabolic model. Black dash-dotted line: results from class method.

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 0 to 1 iterations, 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.

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 (Figures 2 and 4). The MaxEnt method is used with 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 Figure 4c). Finally, EQMOM was also applied with N=5 moments as going up to 7 moments did not increase the precision drastically (Figure 4b) but did increase significantly the computation time (Figure 5). This will also make the comparison between MaxEnt and EQMOM more relevant.

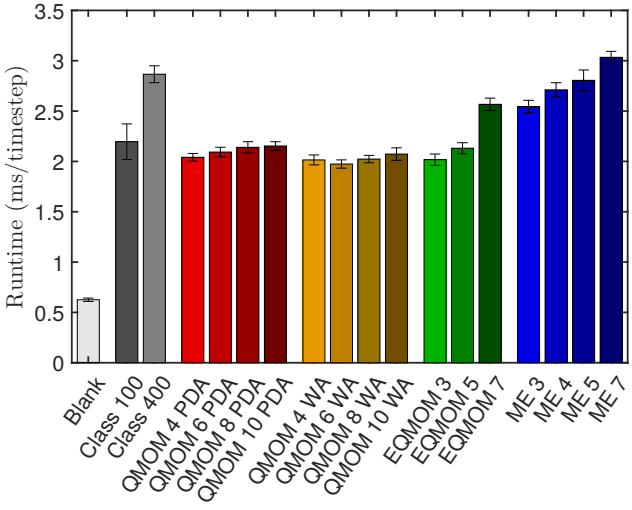

Figure 5: Mean run-time per timestep for each method and different orders (ms/ts) (± standard deviation measured on 20 simulations per method and order)

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:

$$\delta t_{\max} = \min_{n=1}^{N_c} \left(\frac{V_n}{\sum_{m=1}^{N_c} Q_{n,m}} \right) \tag{31}$$

The value of the maximum time-step for the compartment model shown in Figure 6 is $\delta t_{\max} = 1.47\ 10^{-5}\ h$.

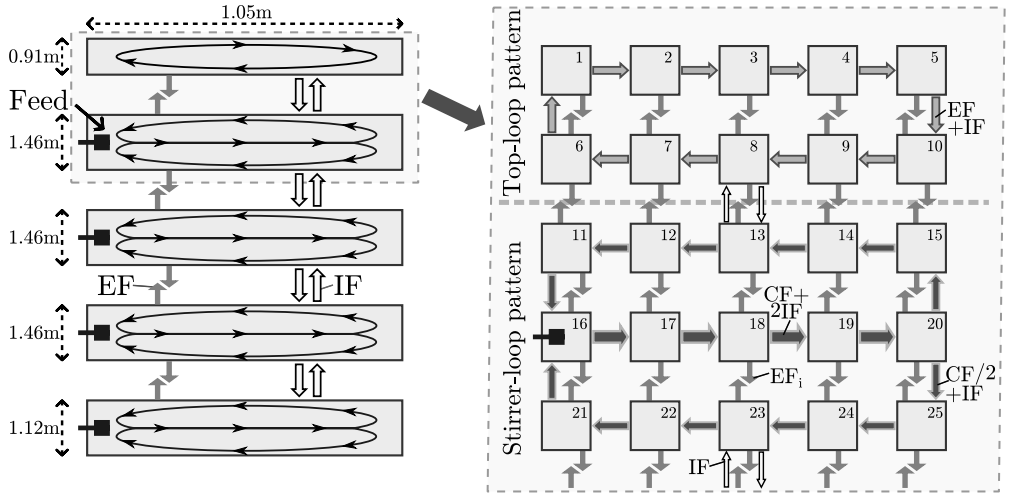

Figure 6: Representation of the macroscopic flow patterns in the fedbatch reactor (left) and details of its compartmentalization and specific flows for the top of the reactor (right). Values for the flow rates CF, IF and EF are given in Pigou and Morchain (2015), appendix B.

The Figure 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 Figure 6 for numbering):

- top: compartments 11 to 15
- middle: compartments 36 to 40
- bottom: compartments 61 to 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 (Figure 3). The expected distribution has a smoother shape which is well reconstructed by MaxEnt and EQMOM as shown in Figure 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 Figure 9.

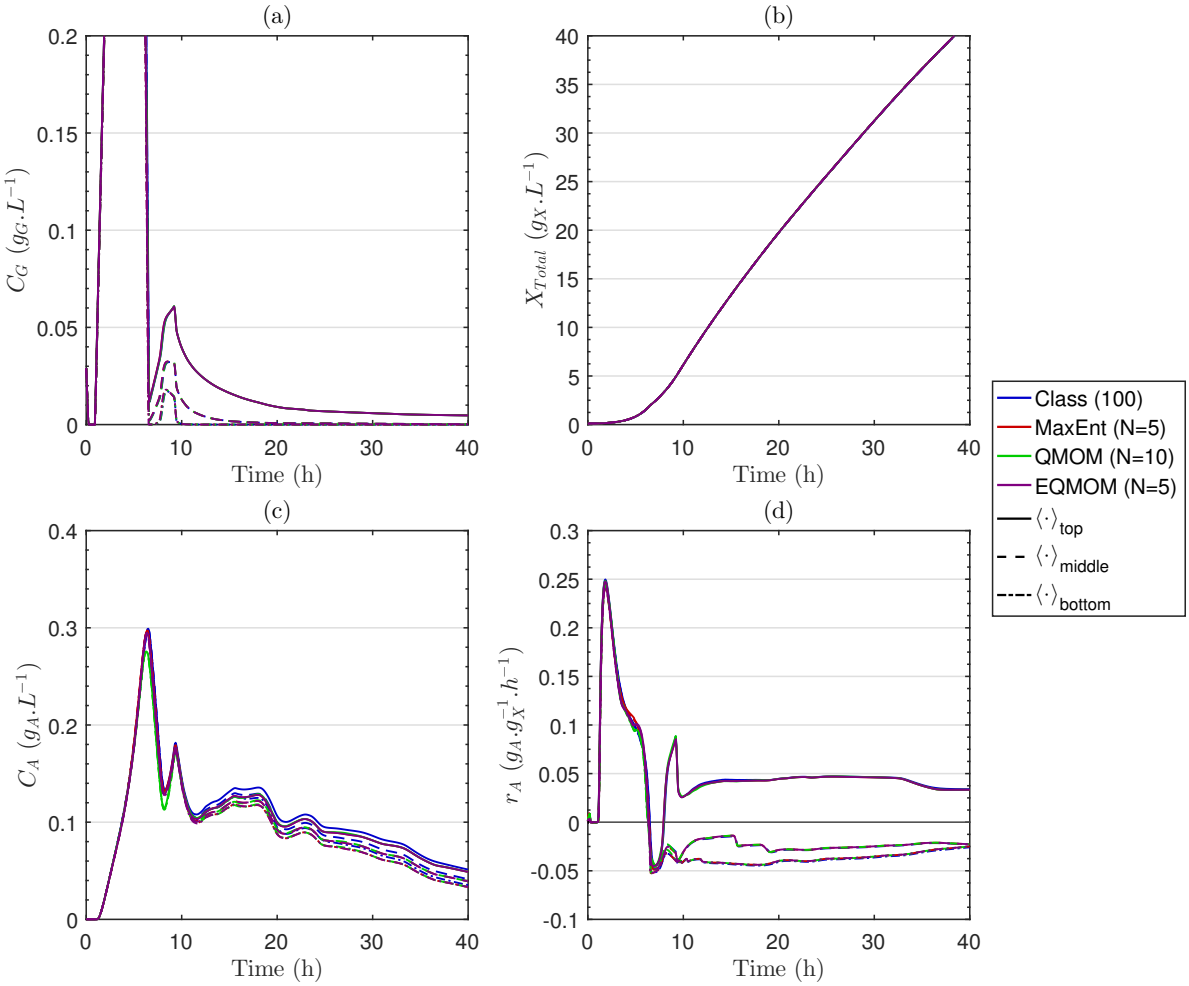

Figure 7: Simulation results for the different population balance methods in the heterogeneous fedbatch culture. (a) Glucose concentration, (b) Total biomass concentration, (c) Acetate concentration and (d) Acetate specific production rate.

Each CPU core had to perform calculations for 14 compartments in figure 9 while a single compartment was considered in figure 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.

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

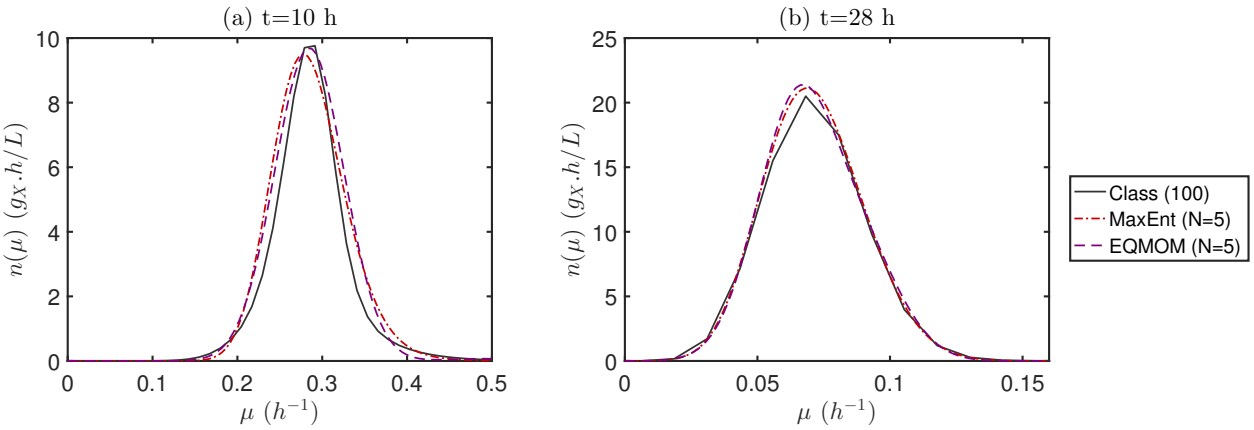

Figure 8: Comparison of distribution shapes as resolved by the class method and reconstructed by the EQMOM and MaxEnt methods both with $N = 5$ moments.

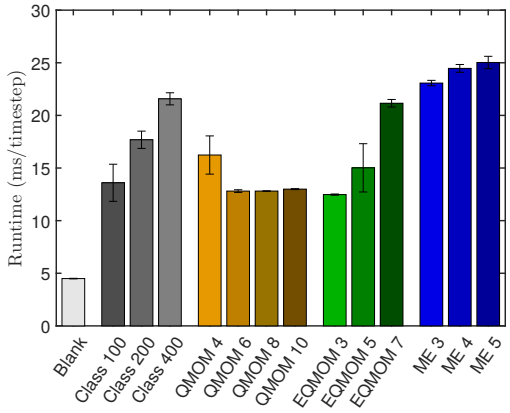

Figure 9: Comparison of simulation runtimes for each method (ms/ts) ($\pm$ standard deviation measured on 5 simulations per method and order). Simulations performed using 5 CPU cores.

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 figure 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 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 over-production of acetate with moment methods in the Katterer case (see Figure 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 Figure 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 $t \approx 10h$ 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 Figures 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 monomodal 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 of increasing the number of resolved moments 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 embeds the principle 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. This can be an important advantage of this method if the metabolic model (computation of $\mathbf{\Phi}$) 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 similarly, based on whether we try to reduce the memory usage or the number of calls to a potentially complex metabolic model.

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 (Figure 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 monomodal 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 mesh.

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, 2016; 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); Yasuda (2011) allowed us to improve the PBM and add an experimentally justified redistribution term as explained in Morchain et al. (2016).

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.

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
O : Order of (E)QMOM method
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_{i} : Inhibition
x_k : Moment order
x_m : Compartment index
x_n : Compartment index
x_O : Oxygen
x_T : Threshold value

Greek

ε : Turbulent energy dissipation rate ($W.kg^{-1}$)
κ : PDF Kernel
μ : Growth rate ($g_X.g_X^{-1}.h^{-1}$)
ν : Kinematic viscosity ($m^2.s^{-1}$)
φ : Polynomial coefficient
Φ : Specific uptake rate ($g.g_X^{-1}.h^{-1}$)
Ψ : Environmental limitation coefficient
σ : Standard deviation (h^{-1})
ζ : rate of change of specific growth rate (h^{-2})

Literature Cited

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.

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.

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 Journal 61, 2702–2710. doi:10.1002/aic.14893.

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.

Bogacki, P., Shampine, L., 1989. A 3(2) pair of Runge-Kutta formulas. Applied Mathematics Letters 2, 321–325. doi: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. Chemical Engineering Research and Design 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. 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.

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. doi:10.1111/j.1574-6976.1996.tb00246.x.

Ferenci, T., 1999. Regulation by nutrient limitation. Curr Opin Microbiol 2, 208–213. doi: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. Chemical Engineering Science 57, 2265 – 2278. doi:10.1016/S0009-2509(02)00116-1.

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.

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 *Saccharomyces cerevisiae* scale-down fermentation in a two-compartment setup. Journal of Chemical Technology and Biotechnology 90, 324–340. doi:10.1002/jctb.4532.

Hulburt, H., Katz, S., 1964. Some problems in particle technology: A statistical mechanical formulation. Chemical Engineering Science 19, 555–574. doi: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. Chemical Engineering Science 62, 2890–2904. doi: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. Biotechnology and Bioengineering 28, 146–150. doi:10.1002/bit.260280126.

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.

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.

Li, W., 1982. Estimating heterotrophic bacterial productivity by inorganic radiocarbon uptake: importance of establishing time courses of uptake. Deep Sea Research Part B. Oceanographic Literature Review 29, 167–172. doi: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. Chemical Engineering Science 116, 473–485. doi: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 and Biosystems Engineering 38, 917–928. doi:10.1007/s00449-014-1337-8.

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.

730 Marchisio, D., Fox, R., 2013. Computational Models for Polydisperse Particulate and Multi-
731 phase Systems. Cambridge Series in Chemical Engineering, Cambridge University Press.

732 Marchisio, D.L., Pikturna, J.T., Fox, R.O., Vigil, R.D., Barresi, A.A., 2003a. Quadrature
733 method of moments for population-balance equations. AIChE Journal 49, 1266–1276.
734 doi:10.1002/aic.690490517.

735 Marchisio, D.L., Vigil, R., Fox, R.O., 2003b. Quadrature method of moments for aggregation-
736 breakage processes. Journal of Colloid and Interface Science 258, 322–334. doi:10.1016/
737 S0021-9797(02)00054-1.

738 Marchisio, D.L., Vigil, R.D., Fox, R.O., 2003c. Implementation of the quadrature method of
739 moments in CFD codes for aggregation-breakage problems. Chemical Engineering Science
740 58, 3337–3351. doi:10.1016/S0009-2509(03)00211-2.

741 Massot, M., Laurent, F., Kah, D., De Chaisemartin, S., 2010. A robust moment method
742 for evaluation of the disappearance rate of evaporating sprays. SIAM Journal on Applied
743 Mathematics 70, 3203–3234. doi:10.1137/080740027.

744 Matsuoka, Y., Shimizu, K., 2013. Catabolite regulation analysis of *Escherichia coli* for
745 acetate overflow mechanism and co-consumption of multiple sugars based on systems
746 biology approach using computer simulation. Journal of Biotechnology 168, 155–173.
747 doi:10.1016/j.jbiotec.2013.06.023. special issue: Biotechnology for a healthy and
748 green world.

749 Mayr, B., Horvat, P., Nagy, E., Moser, A., 1993. Mixing-models applied to industrial batch
750 bioreactors. Bioprocess Engineering 9, 1–12. doi:10.1007/BF00389534.

751 McGraw, R., 1997. Description of aerosol dynamics by the quadrature method of moments.
752 Aerosol Science and Technology 27, 255–265. doi:10.1080/02786829708965471.

753 Mead, L.R., Papanicolaou, N., 1984. Maximum entropy in the problem of moments. Journal
754 of Mathematical Physics 25, 2404–2417.

755 Meadows, A.L., Karnik, R., Lam, H., Forestell, S., Snedecor, B., 2010. Application of
756 dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metabolic
757 Engineering 12, 150–160. doi:10.1016/j.ymben.2009.07.006. metabolic Flux Analysis
758 for Pharmaceutical Production Special Issue.

759 Morchain, J., Gabelle, J.C., Cockx, A., 2013. Coupling of biokinetic and population balance
760 models to account for biological heterogeneity in bioreactors. AIChE Journal 59, 369–379.
761 doi:10.1002/aic.13820.

762 Morchain, J., Gabelle, J.C., Cockx, A., 2014. A coupled population balance model and CFD
763 approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AIChE
764 Journal 60, 27–40. doi:10.1002/aic.14238.

765 Morchain, J., Pigou, M., Lebaz, N., 2016. A population balance model for bioreactors com-
766 bining interdivision time distributions and micromixing concepts. Biochemical Engineering
767 Journal , In pressdoi:10.1016/j.bej.2016.09.005.

768 Nobs, J.B., Maerkl, S.J., 2014. Long-term single cell analysis of *S. pombe* on a microfluidic
769 microchemostat array. PLoS ONE 9. doi:10.1371/journal.pone.0093466.

770 Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B.O., 2011.
771 A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. Molecular
772 Systems Biology 7. doi:10.1186/1752-0509-6-55.

773 Perret, C.J., 1960. A new kinetic model of a growing bacterial population. Microbiology 22,
774 589–617. doi:10.1099/00221287-22-3-589.

775 Pigou, M., Morchain, J., 2015. Investigating the interactions between physical and biolog-
776 ical heterogeneities in bioreactors using compartment, population balance and metabolic
777 models. Chemical Engineering Science 126, 267–282. doi:10.1016/j.ces.2014.11.035.

778 Silveston, P., Budman, H., Jervis, E., 2008. Forced modulation of biological processes:
779 A review. Chemical Engineering Science 63, 5089–5105. doi:10.1016/j.ces.2008.06.
780 017. 5TH Unsteady-State Processes in Catalysis: a Special Issue of Chemical Engineering
781 Science.

782 Tagliani, A., 1999. Hausdorff moment problem and maximum entropy: A unified approach.
783 Applied Mathematics and Computation 105, 291–305. doi:10.1016/S0096-3003(98)
784 10084-X.

785 Vié, A., Laurent, F., Massot, M., 2013. Size-velocity correlations in hybrid high order
786 moment/multi-fluid methods for polydisperse evaporating sprays: Modeling and numerical
787 issues. Journal of Computational Physics 237, 177–210. doi:10.1016/j.jcp.2012.11.043.

788 Vrábel, P., van der Lans, R.G., Luyben, K.C., Boon, L., Nienow, A.W., 2000. Mixing in
789 large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers:
790 modelling and measurements. Chemical Engineering Science 55, 5881–5896. doi:10.1016/
791 S0009-2509(00)00175-5.

792 Vrábel, P., van der Lans, R.G., van der Schot, F.N., Luyben, K.C., Xu, B., Enfors, S.O.,
793 2001. CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale
794 fermentations. Chemical Engineering Journal 84, 463–474. doi:10.1016/S1385-8947(00)
795 00271-0.

796 Vrábel, P., der Lans, R.V., Cui, Y., Luyben, K., 1999. Compartment model approach:
797 Mixing in large scale aerated reactors with multiple impellers. Chemical Engineering
798 Research and Design 77, 291–302. doi:10.1205/026387699526223.

799 Xu, B., Jahic, M., Enfors, S.O., 1999. Modeling of overflow metabolism in batch and
800 fed-batch cultures of *Escherichia coli*. Biotechnology Progress 15, 81–90. doi:10.1021/
801 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. Springer Berlin Heidelberg. volume 124 of *Advances in Biochemical Engineering / Biotechnology*, pp. 55–81. doi:10.1007/10_2010_97.

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.

Appendix A. Moment formulation of the PBE

As a recall, the population balance equation is defined as

$$\frac{\partial n(\mu)}{\partial t} = -\frac{\partial}{\partial \mu}\left[n(\mu)\zeta(\mu)\right] + \int_0^{+\infty} \beta(\mu,\mu')n(\mu')\Psi\mu'\mathrm{d}\mu' \tag{A.1}$$

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 μ^k and integrating by part with respect to μ:

$$\frac{\partial m_{a,k}(t)}{\partial t} = -\int_{\Omega\mu} \mu^k \frac{\partial}{\partial \mu}\left(n(\mu,t)\cdot\zeta(\mu)\right)\mathrm{d}\mu \tag{A.4}$$

$$= \int_{\Omega\mu} k\mu^{k-1} n(\mu,t)\zeta(\mu)\mathrm{d}\mu - \left[\mu^k n(\mu,t)\zeta(\mu)\right]_{\partial\Omega\mu} \tag{A.5}$$

Considering that the adaptation will not allow individuals to cross the frontier of the μ-space ($\partial\Omega\mu$), the second term of Eq. A.5 is necessarily null. By expanding the formulation of $\zeta(\mu) = \left(T^{-1}+\mu\right)\left(\mu^*-\mu\right)$, 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 strongly depends on the choice of a closure for the redistribution model. In Morchain et al. (2016), we identified the following probability density function as a good model for experimental data from the literature (Nobs and Maerkl, 2014; Yasuda, 2011):

$$\beta(\mu,\mu') = \beta(\mu,\tilde{\mu}) = \frac{2}{\sigma}\phi\left(\frac{\mu-l}{\sigma}\right)\Phi\left(\alpha\times\frac{\mu-l}{\sigma}\right) \tag{A.7}$$

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:

$$\tilde{\mu} = \frac{m_1(t)}{m_0(t)} \tag{A.8}$$

$$l = k_l \ \tilde{\mu} \tag{A.9}$$

$$\sigma = k_\sigma \ \tilde{\mu} \tag{A.10}$$

$$\alpha = 3.65 \tag{A.11}$$

The constants k_l and k_σ 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:

$$k_l = 0.651836690282903 \tag{A.12}$$

$$k_\sigma = 0.452438435252710 \tag{A.13}$$

Then, the growth related evolution of the distribution moments is easily expressed in terms of moments of $\beta(\mu)$:

$$\frac{\partial m_{g,k}(t)}{\partial t} = \int_{\Omega_\mu} \left[\mu^k \int_{\Omega_\mu} \beta(\mu, \tilde{\mu}) \Psi \mu' n(\mu') \mathrm{d}\mu' \right] \mathrm{d}\mu \tag{A.14}$$

$$= \Psi \int_{\Omega_\mu} \mu^k \beta(\mu, \tilde{\mu}) \left[\int_{\Omega_\mu} \mu' n(\mu') \mathrm{d}\mu' \right] \mathrm{d}\mu \tag{A.15}$$

$$= \Psi \int_{\Omega_\mu} \mu^k \beta(\mu, \tilde{\mu}) m_1(t) \mathrm{d}\mu \tag{A.16}$$

$$= \Psi m_1(t) B_k \tag{A.17}$$

$$\tag{A.18}$$

B_k is the k-th order moment of the PDF $\beta(\mu, \tilde{\mu})$ which happens to only depend on $\tilde{\mu}$ whose value is accessible using the first two moments of the distribution (Eq. A.8). The moments B_k can be determined analytically using the Moment Generating Function of the skew-normal distribution:

$$B_k = \frac{\partial^k M}{\partial t^k}(0) \tag{A.19}$$

$$M(t) = \exp\left(lt + \frac{\sigma^2 t^2}{2} \right) \left(1 + \mathrm{erf}\left(\frac{\sigma \alpha t}{\sqrt{1 + \alpha^2}} \right) \right) \tag{A.20}$$

We used the MATLAB Symbolic Toolbox to pre-compute the expressions of $B_k, \forall k \in \{0, \ldots, 9\}$.