Application of the Direct Quadrature Method of Moments for the modelling of the enzymatic hydrolysis of cellulose: I. Case of soluble substrate

Noureddine Lebaz, Arnaud Cockx, Mathieu Sperandio, Alain Liné, Jérôme Morchain

Application of the Direct Quadrature Method of Moments for the modelling of the enzymatic hydrolysis of cellulose: I. Case of soluble substrate

Noureddine Lebaz[a,b,c,d], Arnaud Cockx[b,c,d], Mathieu Spérandio[b,c,d], Alain Liné[b,c,d], Jérôme Morchain[b,c,d,*]

[a] *Toulouse White Biotechnology (UMS INRA/INSA/CNRS), 3 rue Ariane, 31520 Ramonville Saint Agne, France*
[b] *Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077, Toulouse, France*
[c] *INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400, Toulouse, France*
[d] *CNRS, UMR5504, F-31400, Toulouse, France*

Abstract

The modelling of the enzymatic hydrolysis of cellulosic polymers is investigated through a population balance approach. Both Endoglucanase (EG) and Exoglucanase (CBH) activities are taken into account. EG achieves random attacks along cellulosic chains and cleaves the β-glycosidic bonds whereas CBH produces cellobiose molecules by chain-end scission mechanism. The EG activity is modelled as a pure breakage while the CBH activity is assimilated to an erosion process with a specific product (cellobiose). In the two cases, the inhibition of the cellulases activity by the end-product is incorporated. The population balance equation (PBE) accounting for breakage processes is solved using the Direct Quadrature Method of Moments (DQMOM) coupled to a distribution reconstruction technique based on the Maximum Entropy (ME) principle in order to track the time evolution of the chain length distribution (CLD) during the hydrolysis reaction. The β-glucosidase activity transforming the produced cellobiose into glucose is modelled as a Michaelis-Menten type kinetic with a competitive inhibition effect and solved simultaneously with the PBE. The numerical results show the time-evolution of the CLD during the hydrolysis reaction as well as the rate of conversion of the substrate into simple sugars. These results are in concordance with those predicted analytically. The synergistic action of the EG and CBH is highlighted and discussed and the inhibition effect is investigated. The approach is promising by its accuracy and fastness for the analysis of dynamic experimental data of the

*Corresponding author. Address: INSA, LISBP, 135, avenue de Rangueil, F-31077 Toulouse, France. Tel.: +33 56 155 9774; Fax: +33 56 155 9760

Email address: jerome.morchain@insa-toulouse.fr (Jérôme Morchain)

enzymatic hydrolysis reaction.

Keywords: Enzymatic hydrolysis, population balance, cellulose, Endoglucanase, Exoglucanase

1. Introduction

The growing demand for energy, the depletion of fossil fuels and the global warming have sparked a growth in research for renewable energy sources (Sun & Cheng, 2002). One of the promising alternatives is the bioconversion of biomass materials to bioethanol (Jorgensen et al., 2007). The biochemical conversion of lignocellulosic biomass to ethanol involves several processing steps (Gregg & Saddler, 1995). The reduction of cellulose, which is the major component of plant material, into fermentable sugars is carried out by means of enzymatic hydrolysis.

Cellulose is a homopolymer of β-1,4 linked D-glucose units (Mittal et al., 2011). Enzymes cocktails capable of degrading this polymer to sugars are a mixture of at least three different activities. The endoglucanase activity (EG) cleaves randomly the β-1,4-glycosidic linkages, the cellobiohydrolase activity (CBH) attacks the chain-ends and releases cellobiose (dimer of glucose) and the β-glucosidase activity hydrolyzes cellobiose into D-glucose (Andersen et al., 2008). A typical kinetic of the enzymatic hydrolysis of cellulose is characterised by a rapid initial rate followed by a gradual slowdown caused by a multitude of factors affecting the effectiveness of the biocatalysts, leading, finally, to a partial conversion of the substrate (Yang et al., 2006). These factors are related to the substrate structure (nature, pretreatments), the operating conditions (pH, temperature and mixing) and the enzymes (nature, deactivation during the reaction, inhibition by the end-products) (Van Dyk & Pletschke, 2012).

Understanding the complex interactions between the enzymes elementary mechanisms and the dynamic evolution of the substrate features during the hydrolysis reaction is of critical importance for the modelling, the design and the optimization of the transformation process. Michaelis-Menten based models are widely used for the modelling of this dynamic process however, since they have been developed initially for homoge-

neous reaction conditions, the parameters fitting the experimental data are only apparent and constants are sometimes lacking in physical or biochemical sense (Bansal et al., 2009). Mechanistic models incorporating one or several rate limiting factors such as the substrate accessibility, the cellulases inhibition/deactivation and the change in the substrate reactivity can be found in literature (Fan & Lee, 1983; South et al., 1995; Xu & Ding, 2007). A convincing critical analysis of the identifiability of these models has been published by Sin et al. (2010).

Recently, models based on the population balance approach have been proposed in order to better describe the heterogeneity of the substrate during the dynamic enzyme catalysis. Hosseini & Shah (2011a,b) assimilated the cellulose hydrolysis to soluble polymer degradation. The endoglucanase activity (EG) has been modelled as a pure random breakage of the cellulose chains and the exoglucanase activity (CBH) as a chain-end scission (Kostoglou, 2000). To be more realistic, Griggs et al. (2012a,b) developed a population balance model based on the depolymerisation mechanisms of EG and CBH assuming a population of monodisperse cylindrical cellulose particles since the reaction is heterogeneous. To solve the system of rate equations, Hosseini & Shah (2011a,b) considered all the possible chains in the system when Griggs et al. (2012a,b) reduced the number of equations by mapping the continuous particle size distribution (PSD) to a fixed discrete grid points. Lebaz et al. (2015) used the discretization method developed by Kumar & Ramkrishna (1996a) for the resolution of the population balance equation (PBE) with a fixed pivot technique for the EG activity and a moving pivot technique (Kumar & Ramkrishna, 1996b) for the CBH activity since it was assimilated to a continuous dissolution of the cellulosic chains. Ho et al. (2014), meanwhile, proposed a modified fixed pivot technique for the modelling of chain-end scission by CBH. All of these modelling approaches are computationaly expensive.

Cellulose particles can be characterised by their Degree of Polymerization (DP) which is the number of monomers (D-glucose) constituting each particle. If we note $\mathbf{N}$ the maximum DP in a given particulate

system, one needs to solve **N** equations to describe the time-evolution of all the possible sizes undergoing breakage processes. The discretization method (Kumar & Ramkrishna, 1996a) reduces considerably the number of equations by distributing the particles into a set of classes $M < \mathbf{N}$. Unfortunately, the accuracy of this method depends on the number of classes M. Increasing M implies more equations to solve and, thereby, substantial computational time. In this study, we develop a new approach for solving the PBE in the case of cellulose depolymerisation by EG and CBH. For the sake of accuracy and computational efficiency, the Direct Quadrature Method of Moments (DQMOM) is used as solution method (Marchisio & Fox, 2005). Since DQMOM provides information only on the time evolution of the moments of the particle size distribution (PSD), the recovering of the full distribution is needed. For this, the PSD is reconstructed from its first moments using the Maximum Entropy based method (Tagliani, 1999). This reconstruction method is preferred to other techniques (e.g. Spline based method, Beta Density Function based technique) for its fastness, accuracy and the small number of moments required (Lebaz et al., 2016). This is of critical importance when the PSD reconstruction and the PBE resolution are conducted simultaneously. Note that more recent moment methods are proposed in literature to solve the PBE, maintly the Conditional Quadrature Method of Moments (CQMOM) developed by Yuan & Fox (2011) which is more suitable for multivariate problems and the Extended Quadrature Method of Moments (EQMOM) (Yuan et al., 2012) allowing the reconstruction of the distribution all along the resolution using kernel density functions.

For sake of clarity, this contribution is organized in two parts. In this first part, the focus will be put on soluble substrates. The general theoretical framework including the DQMOM approach for solving the homogeneous monovariate PBE and the PSD reconstruction by the Maximum Entropy based method is developed and validated against analytical solutions. We show how the cellulolytic activities are modelled in the case of depolymerisation of soluble cellulose chains before testing and discussing the robustness of the model in the results section. The second part deals with the case of particulate systems.

In both cases, the β-glucosidase activity which converts cellobiose into D-glucose is modelled as a Michaelis-Menten type kinetic with competitive inhibition and resolved simultaneously with the PBE.

2. Theoretical framework

The PBE describes the evolution of a density function of a given population of chains/particles undergoing different transformation processes. In the case of enzymatic hydrolysis of cellulose chains, the substrate chains are degraded during the reaction by the synergistic action of an enzyme mixture. Thus, considering the chain length L as the internal coordinate, the PBE expressing the depolymerisation process in a spatially homogeneous system is written in its continuous form as:

$$\frac{\partial n(L,t)}{\partial t} = \int_L^\infty \beta(L,\lambda)\Gamma(\lambda)n(\lambda,t)d\lambda - \Gamma(L)n(L,t) \tag{1}$$

where $n(L,t)$ is the length-based number density function, $\Gamma(L)$ the breakage frequency for a chain of length L which is related to the enzyme activity, $\beta(L,\lambda)$ is the daughter distribution function giving the probability of obtaining a chain of length L from the breakup of a chain λ. The first term on the right hand side of equation (1) accounts for the formation (birth) of chains with length L resulting from the breakage of larger chains . The last term is the death term due to the loss of chains of length L because of their depolymerisation. Note that, in the particular case of soluble substrates, the internal coordinate L refers to the degree of polymerization (DP) of the cellulose chains.

Except for some cases where $\Gamma(L)$ and $\beta(L,\lambda)$ have simple mathematical expressions (Ziff & McGrady, 1985), equation (1) has no analytical solution. Numerical methods are used to solve it. We adopt in this contribution the Direct Quadrature Method of Moments (DQMOM). A brief description of this method is given in Appendix A. For further details, one can refer to the original paper by Marchisio & Fox (2005).

The k^{th} order moment of the chain length distribution (CLD) is defined as:

$$m_k(t) = \int_0^\infty L^k n(L,t) dL \tag{2}$$

The moments provide the main properties of the CLD. The zeroth order moment, $m_0(t)$, represents the molar concentration of the cellulose chains since $n(L,t)dL$ is the number of cellulose chains per volume having lengths between L and $L + dL$. The first order moment, $m_1(t)$, gives the total concentration of monomers present in cellulose chains. Average properties of the CLD can be derived from the moments such as the number-averaged chain length given by the ratio of the two first moments $m_1(t)/m_0(t)$ and the mass-averaged chain length $m_2(t)/m_1(t)$.

By applying the moment transformation of equation (1), the final moment transport equation is given as:

$$(1-k)\sum_{i=1}^{N} L_i^k a_i + k\sum_{i=1}^{N} L_i^{k-1} b_i = \sum_{i=1}^{N} \overline{b}_i^{(k)} \Gamma_i w_i - \sum_{i=1}^{N} L_i^k \Gamma_i w_i \tag{3}$$

where

$$\begin{cases} a_i &= \dfrac{\partial w_i}{\partial t} \\ b_i &= \dfrac{\partial c_i}{\partial t} \\ c_i &= w_i L_i \\ \overline{b}_i^{(k)} &= \displaystyle\int_0^\infty L^k \beta(L, L_i) dL \end{cases} \tag{4}$$

$(L_i(t), w_i(t))$ are the abscissa and weights of the N Gaussian quadrature nodes. $(L_i(0), w_i(0))$ can be computed using the Product-Difference (PD) algorithm (Gordon, 1968). The different steps of the transformation are explicited in Appendix A .

Although the intrinsic activities of EG and CBH are different, they lead to the depolymerisation of the substrate chains during the reaction. Thus, the two activities are modelled as breakage processes (equation 1), the distinction between them will appear in the formulation of the daughter distribution functions.

The validation of the DQMOM implementation in the case of breakage processes is given in Appendix B. This numerical validation is based on the analytical solution of equation (1) (Lebaz et al., 2016). Moreover, the CLD reconstruction from the moments using the maximum entropy technique is embedded with the DQMOM algorithm in order to access the CLD evolution simultaneously. The maximum entropy technique is briefly described in Appendix C and its numerical validation is fully discussed in Lebaz et al. (2016).

2.1. Daughter distribution functions formulation

The depolymerisation mechanism depends on the nature of the cellulases. In our case, two main activities (EG and CBH) are considered. This leads to the resolution of the global PBE below:

$$\frac{\partial n(L,t)}{\partial t} = S_L^{EG}(L,t) + S_L^{CBH}(L,t) \tag{5}$$

with

$$\begin{cases} S_L^{EG}(L,t) &= \displaystyle\int_L^\infty \beta_{EG}(L,\lambda)\Gamma_{EG}(\lambda)n(\lambda,t)d\lambda - \Gamma_{EG}(L)n(L,t) \\ S_L^{CBH}(L,t) &= \displaystyle\int_L^\infty \beta_{CBH}(L,\lambda)\Gamma_{CBH}(\lambda)n(\lambda,t)d\lambda - \Gamma_{CBH}(L)n(L,t) \end{cases} \tag{6}$$

where β_{EG} and β_{CBH} are the daughter distribution functions reflecting the mode of action of the enzymes. Γ_{EG} and Γ_{CBH} are the breakage frequencies directly related to the enzymatic activities (Lebaz et al., 2015).

The expression of the daughter distribution functions for both EG and CBH depends on the nature of the substrate : soluble polymer chains or particulate system. In the following paragraph, the formulation will be restricted to the case of polymer chains. The case of particulate substrates is treated in the second part of the contribution.

2.1.1. EG activity: random breakage

The EG activity cleaves randomly the β-1,4-glycosidic bonds along the cellulosic chains. Thus, in the case of soluble substrate, each EG attack leads to the scission of the initial chain λ producing thereby two

smaller chains $(\lambda - L)$ and L:

$$P(\lambda) \longrightarrow P(\lambda - L) + P(L) \tag{7}$$

The daughter distribution function $\beta(\lambda, L)$ defines the probability that a chain of length λ leads to a chain of size L when it breaks up. Under the assumption of a binary scission, two particles are always produced and the factor 2 takes into account that either of these particles can be of size $L < \lambda$. The time-evolution of the chain length number density function, $n(L,t)$ is given by the PBE :

$$S_L^{EG}(L,t) = \left.\frac{\partial n(L,t)}{\partial t}\right|_{EG} = 2\int_L^\infty \frac{1}{\lambda}\Gamma_{EG}(\lambda)n(\lambda,t)d\lambda - \Gamma_{EG}(L)n(L,t) \tag{8}$$

2.1.2. CBH activity: chain-end scission

The CBH activity has not a random aspect as the EG one, each effective attack produces one soluble cellobiose molecule and reduces, thereby, the DP of the attacked chain by 2. In the case of a soluble substrate, the CBH mechanism is described by:

$$P(L) \longrightarrow P(L - L_C) + P(L_C) \tag{9}$$

where L_C is the length of the cellobiose molecule.

Since the CBH activity leads to a specific product, the daughter distribution function is given by a Dirac delta function (McCoy & Wang, 1994; Wang et al., 1995) and the corresponding source term can be expressed as :

$$S_L^{CBH}(L,t) = \left.\frac{\partial n(L,t)}{\partial t}\right|_{CBH} = \int_L^\infty \Gamma_{CBH}(\lambda)\delta\left(L - (\lambda - L_C)\right) n(\lambda,t)d\lambda - \Gamma_{CBH}(L)n(L,t) \tag{10}$$

Additionally the cellobiose concentration $C_C(t)$ evolution during the reaction is given by :

$$\frac{\partial C_C(t)}{\partial t} = \int_L^\infty \Gamma_{CBH}(\lambda)\delta\left(L - L_C\right) n(\lambda, t) d\lambda + S_{L_C}^{EG}(L, t) \tag{11}$$

Where $S_{L_C}^{EG}(L, t)$ is the contribution of EG activity to cellobiose release which is negligible thus, the right hand side of equation (11) is reduced to the first term.

2.2. Moment transformation using the DQMOM formalism

The moment transformation of equations (8) and (10) is given by equation (3). We explicit hereafter the moment transformation of the equation (11) accounting for the cellobiose release after each CBH effective attack. Equation (11) can be written in a compact form as:

$$\frac{\partial C_C(t)}{\partial t} = Q_L(L, t) \tag{12}$$

where $Q_L(L, t)$ is the source term due to the erosion process.

One has to precise that $C_C(t)$ is a delta function centered on the cellobiose length L_C with a time increasing weight. Thus, $C_C(t)$ can be expressed as:

$$C_C(t) = w_C(t)\delta(L - L_C) \tag{13}$$

By introducing equation (13) in (12), we obtain:

$$\delta(L - L_C)\frac{\partial w_C}{\partial t} - \delta'(L - L_C) w_C \frac{\partial L_C}{\partial t} = Q_L(L, t) \tag{14}$$

Since L_C is constant and by applying the moment transformation, equation (14) is reduced to:

$$(1 - k) L_C^k \frac{\partial w_C}{\partial t} = \int_0^\infty L^k Q_L(L, t) dL \tag{15}$$

In the right hand side, the source term is transformed as follows :

$$\begin{aligned} \overline{Q}_k^{(N)} &= \int_0^\infty L^k Q_L(L,t)dL \\ &= \int_0^\infty \int_L^\infty L^k \Gamma(\lambda)\delta(L-L_C)n(\lambda,t)dLd\lambda \\ &= \sum_{i=1}^{N} \overline{b}_i^{(k)} \Gamma_i w_i \end{aligned} \tag{16}$$

In equation (16), w_i refers to the weight of the node i of the density $n(L,t)$. In the specific case of erosion process :

$$\overline{b}_i^{(k)} = \int_0^\infty L^k \delta(L-L_C)dL = L_C^k \tag{17}$$

Finally, equation 17 is tranformed to :

$$(1-k)L_C^k \frac{\partial w_C}{\partial t} = \sum_{i=1}^{N} L_C^k \Gamma_i w_i \tag{18}$$

Since $C_C(t)$ is fully defined by the concentration of cellobiose molecules produced at the time t, one needs only the zeroth order moment ($k=0$), thus:

$$\frac{\partial w_C}{\partial t} = \sum_{i=1}^{N} \Gamma_i w_i \tag{19}$$

Thus, equation (19) gives the time evolution of the cellobiose concentration.

2.3. Breakage frequencies formulation

The breakage frequencies for the two enzymatic activities give the rate of the depolymerisation process, in other words, the number of effective attacks per time unit. They are directly related to the cellulases concentration, more specifically to the active adsorbed enzymes concentration.

Different mathematical expressions can be used for the breakage frequencies. One can, for example, assume that Γ_{EG} is proportional to the chain length L since longer chains are more likely to be attacked

which is not necessarily the case for Γ_{CBH} because of the nature of the exoglucanase attacks. Different expressions for these frequencies are to be tested and confronted to the trends observed in experimental data sets (expressed as CLD at various times during hydrolysis, simple sugars concentrations, ...). The general form $\Gamma = \alpha L^p$ is often used (Nopens et al., 2002; Ding et al., 2006). This form is adopted and discussed in our case with different parameters (α and p).

2.4. The EG and CBH inhibition effect

The end-products inhibition effect is known as one of the most important factors affecting the hydrolysis rate (Xiao et al., 2004; Gruno et al., 2004) causing a progressive decrease of the cellulases activity as the end-product concentration increases with substrate conversion. To incorporate this critical effect in the model, we express the breakage frequencies of EG and CBH as functions of the end-product concentration which is time-dependent as follows (equations 20 and 21):

$$\Gamma_{EG} = \Gamma_{EG_0} \cdot \frac{K_{EG}}{C_C(t) + K_{EG}} \tag{20}$$

$$\Gamma_{CBH} = \Gamma_{CBH_0} \cdot \frac{K_{CBH}}{C_C(t) + K_{CBH}} \tag{21}$$

where $C_C(t)$ is the cellobiose molar concentration, K_{EG} and K_{CBH} are the inhibition constants for EG and CBH respectively, Γ_{EG_0} and Γ_{CBH_0} are the intrinsic EG and CBH activities respectively.

2.5. The β-glucosidase activity

Since the β-glucosidase activity takes place in the liquid phase and turns cellobiose into D-glucose, Michaelis-Menten type kinetic is used and competitive inhibition by the accumulation of glucose is incorpo-

rated as shown in equation (22).

$$\frac{dC_G}{dt} = \frac{2V_m C_C}{K_m\left(1+\frac{C_G}{K_P}\right)+C_C} \tag{22}$$

where C_G is the glucose molar concentration, K_m and V_m the Michaelis-Menten parameters, K_P the inhibition constant for the β-glucosidase activity. The factor 2 is due to the fact that one mole of cellobiose gives two moles of glucose.

3. Summary

Starting from a given chain length distribution (CLD), we solve simultaneously a set of equations (equation (23)) including a PBE accounting for a random breakage process for the EG activity, a PBE for an erosion process refering to the CBH activity and a Michaelis-Menten equation for the β-glucosidase activity.

$$\left\{\begin{array}{rcl} \dfrac{\partial n(L,t)}{\partial t} & = & 2\displaystyle\int_L^\infty \frac{1}{\lambda}\Gamma_{EG}(\lambda)n(\lambda,t)d\lambda - \Gamma_{EG}(L)n(L,t) \\ & + & \displaystyle\int_L^\infty \Gamma_{CBH}(\lambda)\delta\left(L-(\lambda-L_C)\right)n(\lambda,t)d\lambda - \Gamma_{CBH}(L)n(L,t) \\ \dfrac{\partial C_C(t)}{\partial t} & = & \displaystyle\int_L^\infty \Gamma_{CBH}(\lambda)\delta\left(L-L_C\right)n(\lambda,t)d\lambda - \frac{1}{2}\frac{dC_G(t)}{dt} \\ \dfrac{dC_G(t)}{dt} & = & \dfrac{2V_m C_C}{K_m\left(1+\frac{C_G}{K_P}\right)+C_C} \end{array}\right. \tag{23}$$

The integration of this set of equations using the DQMOM approach coupled with the reconstruction technique gives the time evolution of the CLD and the kinetic of the substrate conversion into simple sugars (cellobiose and glucose).

The initialisation of the resolution procedure requires the initial chain length distribution $n(L,0)$ and the model parameters. These parameters are the breakage frequencies for both EG and CBH activities (Γ_{EG} and Γ_{CBH}), the Michaelis-Menten parameters for the β-glucosidase activity (V_m and K_m) and the inhibition constants K_{EG}, K_{CBH} and K_P. One has to point out the low number of parameters required compared to the common kinetic-based models such as the model described by Kadam et al. (2004) in which 18 parameters

are to be identified. Thus, the low number of parameters in our case, clearly associated to the physics of the modelled phenomena, enhances their reliable identification from experimental data.

Numerically, the *ode*45 integrator of MATLAB is used to solve the set of equations (23) under a workstation comprising of an Intel®Core^{TM}i7-3740QM CPU with a clock speed of 2.7 GHz and 16 GB of installed memory (RAM).

4. Results and discussion

For all the results presented in this article, the initial CLD is assimilated to a Gamma distribution given by the equation below:

$$n(L,0) = \frac{L^{\kappa-1}e^{-L/\theta}}{\mathbf{\Gamma}(\kappa)\theta^{\kappa}} \tag{24}$$

where κ and θ are the Gamma distribution parameters and $\mathbf{\Gamma}$ the classical Gamma function. The mode of the Gamma distribution is given by $(\kappa-1)\theta$ and its variance by $\kappa\theta^2$. In our case, the Gamma distribution parameters are $\kappa = 60$ and $\theta = 10$.

The two main activities EG and CBH are treated separately before showing the combination and the inhibition effect. The breakage frequencies are set proportional to the polymer chain length which means that longer chains are more likely to be attacked ($\Gamma = \alpha L^p$). This is the case especially for the EG random activity. Other breakage frequency expressions can be tested easily. The daughter distribution functions and frequencies for both EG and CBH are summerized in table 1. The system of ordinary differential equations (equation 23) is integrated over the time interval [0,10] hours.

For the erosion process, the formation of cellobiose is not integrated in the full CLD because it leads to a strong peak at the length $L_C = 2$ which constitutes a singularity and disturbs by the way the reconstruction procedure. Thus, the daughter distribution function is reduced to $\bar{b}_i^{(k)} = (L_i - L_C)^k$. This is consistent with

Mechanism	$\beta(L,\lambda)$	$\overline{b}_i^{(k)}$	Frequency (Γ)
EG: Uniform breakage	$2/\lambda$	$L_i^k \frac{2}{k+1}$	αL^p
CBH: Erosion process	$\begin{cases} 1 & if \quad L = L_C \\ 1 & if \quad L = \lambda - L_C \\ 0 & otherwise \end{cases}$	$L_C^k + (L_i - L_C)^k$	αL^p

Table 1: The daughter distribution functions and breakage frequencies for EG and CBH activities in the case of depolymerisation of soluble chains

the idea that cellobiose is no longer a chain per se but rather an elementary molecule. As a result, the total mass balance refers to the mass of polymer chains only. The cellobiose concentration is accessible via the mass loss in the system as well as by equation (11).

4.1. CBH + β-glucosidase activities

CBH attacks the chain-ends and releases cellobiose molecules as a product. The loss of two monomers by the attacked polymers is modelled as an erosion process leading to a progressive shortening of the chains length until their total transformation. The breakage frequency is set to $\Gamma_{CBH} = \alpha.L$ ($p = 1$). The Figure 1 gives the time evolution of the first four normalised moments of the CLD undergoing CBH attacks. The parameter α is set equal to 4.10^{-2} for practical considerations.

The total molar concentration (zeroth order moment) remains constant within the process which means that the number of polymer chains is constant. In the meantime, the total mass of the polymer chains (the first order moment) decreases due to the loss of cellobiose molecules. In other words, the initial number of chains is unchanged but the chains length is decreasing. This is consistent with the fact that the initial number-averaged chain length is far from the origin (much larger than the dimer of cellobiose). In order to observe the total transformation of the substrate (and the simultaneous diminution of the total molar concentration of soluble polymer) it would be necessary to increase the hydrolysis time and/or erosion frequency. Note that the set of equations (23) being conservative, the loss of mass due to the enzymatic action on the polymer

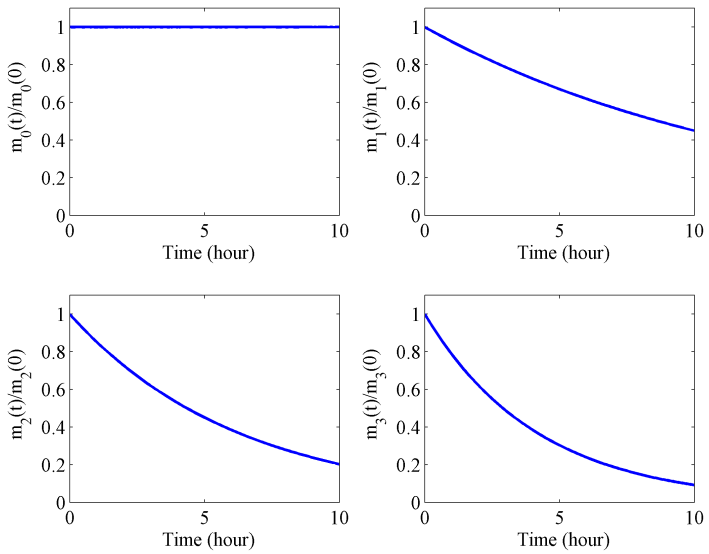

Figure 1: The time evolution of the first four normalised moments of the chain length distribution undergoing CBH attacks with a frequency proportional to the polymer length

chains (Figure 1) is identical to the accumulation of cellobiose and glucose. The interpretation of higher order moment is not direct. Indeed, in the present case, the mass is directly proportional to the chain length. However, the ratio $(\frac{m_2}{m_1})$ represents the mass averaged-chain length and is analogous to d_{43} for spherical particles. The results shows that this mass averaged-chain length decreases.

Figure 2 gives the evolution of the initial CLD reconstructed by the ME based method. The initial distribution is shrinking as it moves to the left since the erosion frequency is not equally distributed all along the CLD: longer chains are more frequently attacked than smaller ones.

These numerical simulations are validated against the analytical solution developed in Appendix D. The zeroth order moment is constant while the first order moment decreases exponentially with a time constant equal to $\tau_1 = \frac{1}{\alpha.L_C}$. Introducing non dimensional parameters will now help in analyzing the trends when different erosion laws are considered.

From the analytical solution, one can build a generic behaviour of the chains population undergoing erosion

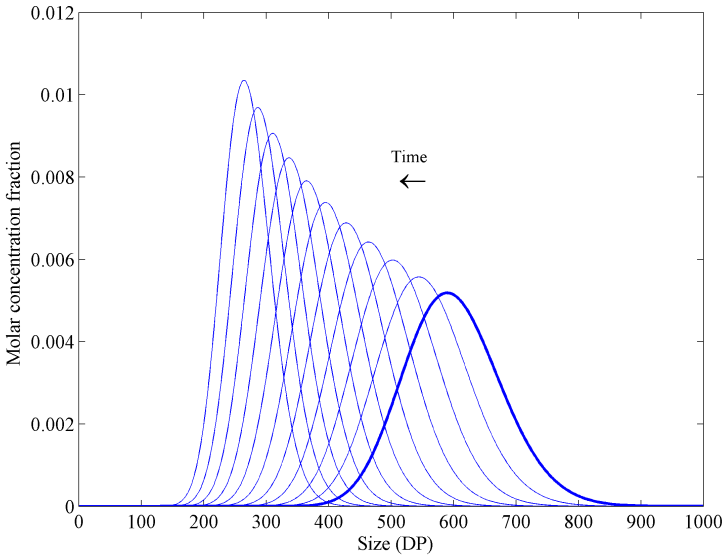

Figure 2: The time evolution of the CLD undergoing CBH attacks with a frequency proportional to the polymer length

process by introducing the adimensional time $\frac{t}{\tau_1}$ in the case of a length-dependent erosion frequency. In the case of a length-independent erosion frequency ($\Gamma_{CBH} = \alpha$), the first order moment of the CLD decreases linearly while the zeroth order moment remains constant. In this case, the time constant is $\tau_2 = \frac{m_1(0)}{\alpha L_C m_0(0)}$ and the adimensional time is $\frac{t}{\tau_2}$.

Figure 3 shows the evolution of the first normalised moment ($\frac{m_1(t)}{m_1(0)}$) reflecting the total mass in the system (without simple sugars: cellobiose and glucose). Two cases with length dependent/independent erosion frequencies, with the adimensional times $\frac{t}{\tau_1}$ and $\frac{t}{\tau_2}$ respectively are considered. The evolutions are independent from the absolute time and the numerical value of the factor α. They are caracteristic of the erosion frequency form.

The relative evolution of the first moment of the CLD considering the CBH attacks shown in Figure 3 is of critical importance to determine the CBH elementary mechanism by fitting experimental data. A linear fit of the mass versus time curve reveals a length independent frequency while an exponential fit is the footprint of a length dependent frequency.

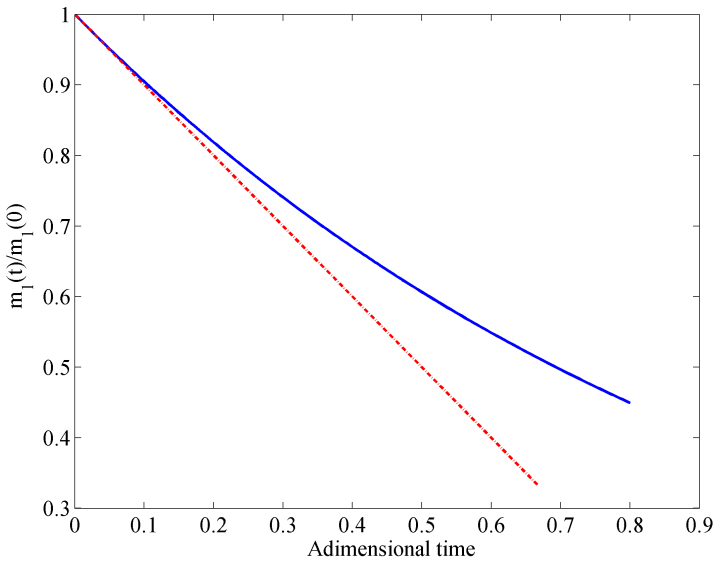

Figure 3: Evolution of the first normalised moment of the chain length distribution undergoing erosion process with a length-dependent (continuous line) and a length-independent (dashed line) frequencies

4.2. EG activity

EG attacks randomly the cellulose chains and breaks up the glucosidic bonds leading to two daughter smaller chains after each attack. The fragmentation process increases the total molar concentration of the cellulose chains in the system as shown by the evolution of the zeroth order normalised moment in Figure 4. In this case, the breakage frequency is length-dependent $\Gamma_{EG} = \alpha L$ and the factor α is set equal to 10^{-4}. The total number of chains increases while the total mass represented by the first order moment remains constant (batch reactor). The number-averaged chain length given by the ratio of the first two moments decreases from a DP of 600 to 375 at the end of the process.

Figure 5 gives the time-evolution of the CLD undergoing fragmentation process. The molar concentration of the longest chains decreases while small chains appear in the system. The initial CLD shifts toward the smaller lengths with a consequent shape transformation (bimodal distributions appear). The CLDs are reconstructed over the interval $[0, +\infty[$. The action of EG on small chains with $L \in [0, 2]$ is therefore

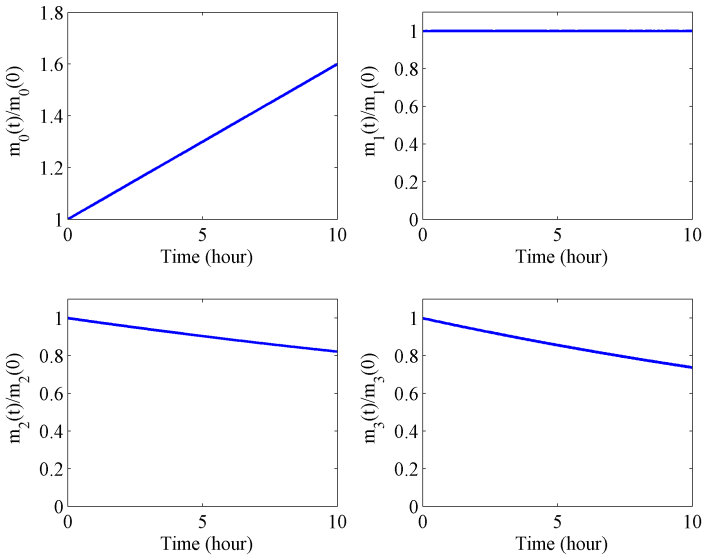

Figure 4: The time evolution of the first four normalised moments of the chain lenght distribution undergoing EG attacks with a frequency proportional to the polymer length

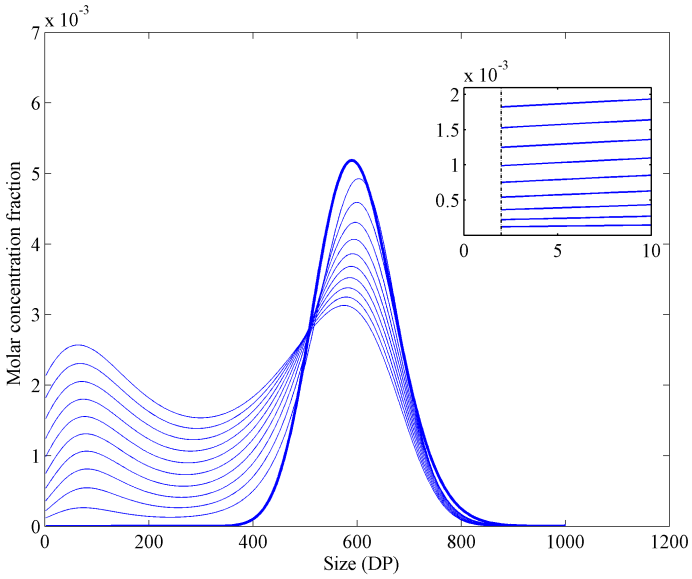

Figure 5: The time evolution of the CLD undergoing EG attacks with a frequency proportional to the polymer length

considered in the model in order to formulate a continuous daughter distribution function expression and thus ensure low computational cost for the PBE resolution. This is conceptually erroneous but there is no consequence from a quantitative point of view as mentionned before (equation (11)). The probability to

produce cellobiose when cutting a chain of length L is very small, the breakage frequency goes to zero when $L \to 2$. As a result, the amount of cellobiose produced by EG is several order of magnitude smaller than that resulting from CBH activity. This supports the neglection of the term $S_{L_C}^{EG}$ in equation (11).

The analytical results relative to the random breakage given in Appendix D are in concordance with the numerical ones. The total mass remains constant while the molar concentration of the polymers (zeroth order moment) increases linearly with a slope equal to $\alpha m_1(0)$. In this case the adimensional time is $\frac{t}{\tau_1}$ with $\tau_1 = \frac{m_0(0)}{\alpha m_1(0)}$.

In the case where the breakage frequency is constant ($\Gamma_{EG} = \alpha$), the zeroth order moment increases exponentially with a time-constant equal to $\tau_2 = \frac{1}{\alpha}$. Figure 6 shows the relative evolution of the normalised zeroth order moment of the CLD when pure breakage process is considered (EG activity) with length-dependent/independent frequencies.

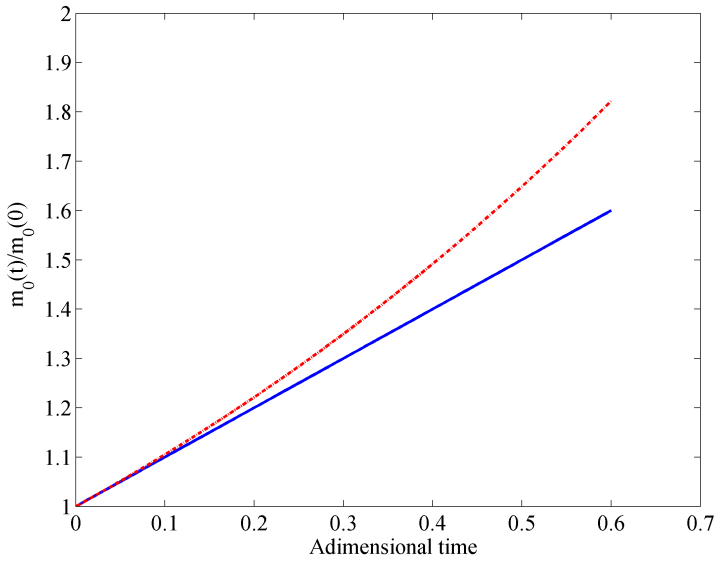

Figure 6: Evolution of the zeroth normalised moment of the chain length distribution undergoing pure breakage process with a length-dependent (continuous line) and a length-independent (dashed line) frequencies

As mentioned previously in the case of CBH activity, the relative evolution of the zeroth order moment

when considering EG activity (pure breakage process) shown in Figure 6 is time and parameter independent. It depends only on the mathematical form of the breakage frequency and here again the fitting of experimental data (by either linear or exponential functions) can be used to identify the breakage law.

4.3. EG-CBH + β-glucosidase combined activity

The combination of the three different enzymatic activities results typically in a CLD evolution presented in Figure 7. In this example, the breakage frequencies those used in the previous cases ($\Gamma = \alpha L$).

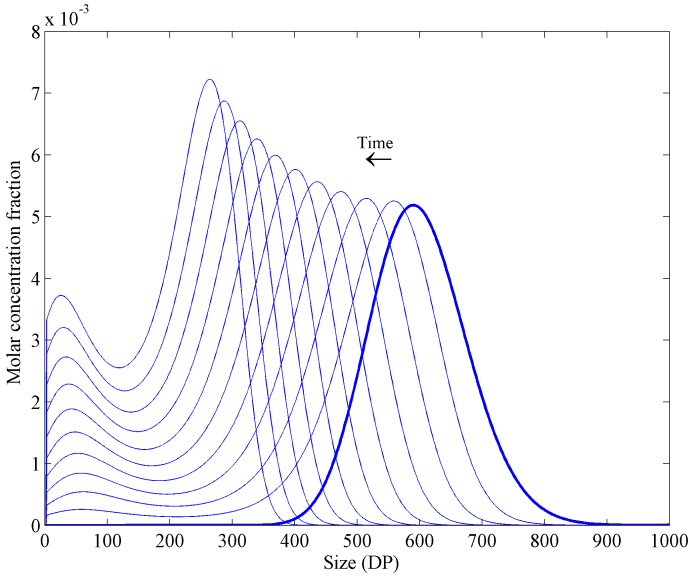

Figure 7: The time evolution of the CLD undergoing EG and CBH attacks with a frequency proportional to the polymer length for the two activities

The initial CLD in Figure 7 shifts toward small sizes and new small chains are continuously formed by the breakage of the longest ones because of the EG activity. This is reflected by the increase of the zeroth order moment and the exponential decrease of the first order moment as predicted via the analytical solution. Clearly the trends depends on the relative values of α_1 and α_2 i.e. the composition of the enzymatic cocktail.

In order to get access to the general behaviour of the first two moments of the CLD depending on the form of the breakage/erosion frequencies, three different combination forms are tested. The first one is a length-dependent frequency ($\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2.L$), the second one is a length-independent

frequency ($\Gamma_{EG} = \alpha_1$ and $\Gamma_{CBH} = \alpha_2$) and the last one combines a length-dependent breakage frequency and a length-independent erosion frequency ($\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2$). The time constants for these three cases are respectively $\tau_1 = \frac{1}{\alpha_2 L}$, $\tau_2 = \frac{1}{\alpha_1}$ and $\tau_3 = \frac{1}{\sqrt{\alpha_1 \alpha_2 L_C}}$. The results are shown in Figure 8.

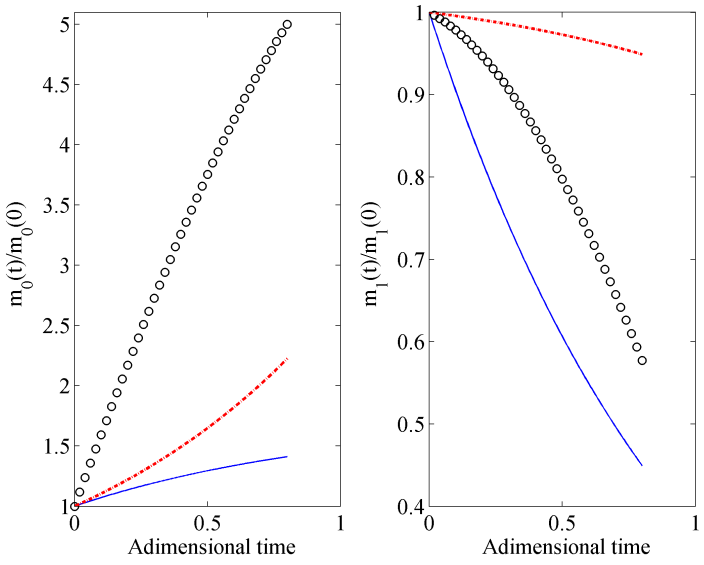

Figure 8: Relative evolution of the two first normalised moments when the two activities EG and CBH are combined with different breakage/erosion frequency forms: $\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2.L$ by the continuous line, $\Gamma_{EG} = \alpha_1$ and $\Gamma_{CBH} = \alpha_2$ represented by the dashed line, $\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2$ by the circles.

Figure 8 shows that the relative evolution of each moment depends on the form of the breakage/erosion frequencies. Thus, we have three generic trends for each moment. The combination of these informations (relative to the total molar concentration of the substrate and its total mass in the system) allows the determination of the breakage/erosion frequencies forms without any need of separation of the different activities at the experimental scale, step which can be critical.

4.4. Influence of the frequency formulation on the prediction of simple sugars

The synergistic action of the different activities increases the rate of degradation of the substrate. Additionally this can be quantified by comparing the time-evolution of the number-averaged chain length when using the enzymes separately and in cocktail as shown in Figure 9 (the right subplot). This chosen criterion

is the ratio of the first two moments ($\frac{m_1(t)}{m_0(t)}$). The total amount of substrate converted into soluble sugars is reflected by the diminution of the first moment (left subplot in Figure 9). The lower the curve the faster the conversion.

In the case presented in Figure 9 where $\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2$, the number-averaged chain length decreases from an initial degree of polymerization (DP) of 600 to 520 through the sole action of CBH. The decrease reaches a DP of 375 when one considers the EG activity only. The combination of the two activities affects substantially the number-averaged chain length. At the end of the time-process, the mean DP is only of 320. The repercussion of the combination of the two different activities on the production of simple sugars is given in the first subplot (Figure 9). As shown, the conversion of the substrate is boosted when the two enzymes are combined. This Endo-Exo synergistic effect is explained by the appearance of new chains in the system due to the EG activity and, since all chains are equiprobably attacked by CBH, the conversion is accelerated.

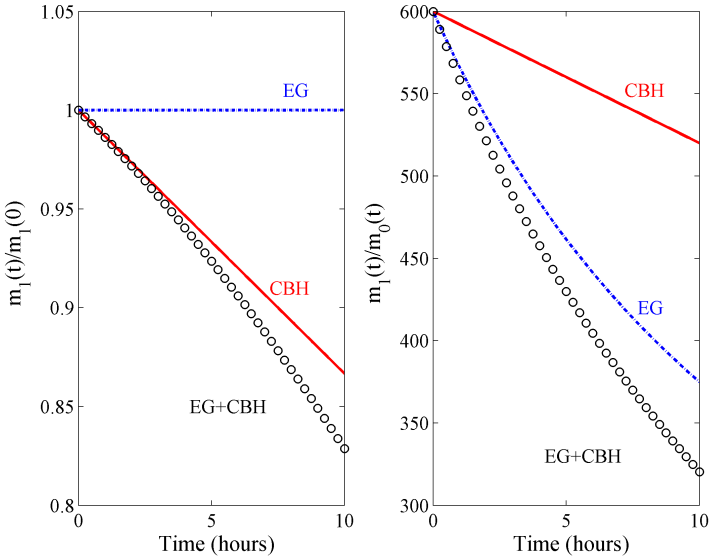

Figure 9: Evolution of the first normalised moment (left subplot) and the number-averaged chain length (right subplot) for the three different cases: EG (dashed line), CBH (continuous line) and EG-CBH (circles). The breakage/erosion frequencies are $\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2$.

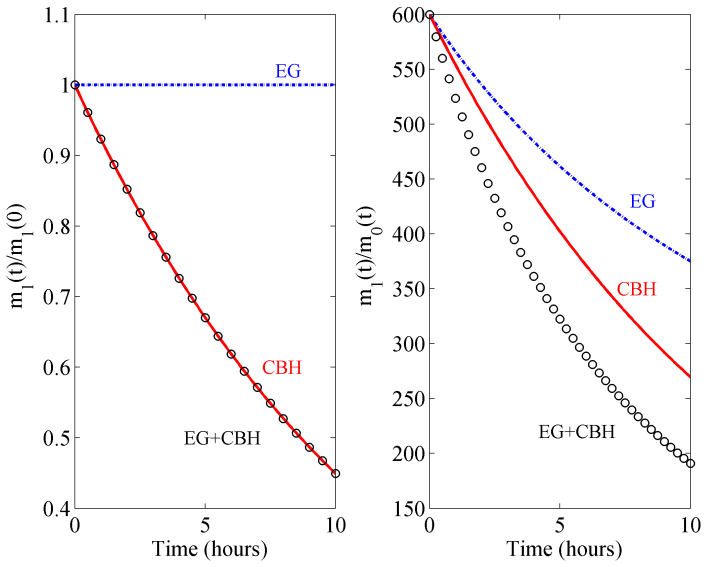

Figure 10: Evolution of the first normalised moment (left subplot) and the number-averaged chain length (right subplot) for the three different cases: EG (dashed line), CBH (continuous line) and EG-CBH (circles). The breakage/erosion frequencies are $\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{CBH} = \alpha_2.L$.

In the case presented in Figure 10 ($\Gamma_{EG} = \alpha_1.L$ and $\Gamma_{EG} = \alpha_2.L$), the combined action of the two different activities has no incidence on the substrate conversion curves (in terms of simple sugars). The fractions of glucose and cellobiose produced by EG and CBH independently and in cocktail are compared and there is no noticeable difference in the two cases. This is confirmed by the analytical solution given in Appendix D in which the evolution of the first order moment of the CLD is similar when considering CBH activity only and both EG and CBH activities. This is not the case for the number-averaged chain length (right subplot in Figure 10). So one would conclude that there is no synergistic effect.

These two numerical examples ($[\Gamma_{EG} = \alpha_1.L, \Gamma_{CBH} = \alpha_2]$ and $[\Gamma_{EG} = \alpha_1.L, \Gamma_{CBH} = \alpha_2.L]$) reveal that the erosion frequency can not be length dependent. In this case, no synergistic effect is possible observed in terms of simple sugars production. This is in contradiction with experimental results that reveal an increase of the conversion rate (and hence sugar production) when the two types of enzymes are combined (Sun & Cheng, 2002; Van Dyk & Pletschke, 2012). It appears that a size independent erosion process is necessary to

observe synergism. From a physical point of view it is consistent to assume that the CBH activity does not depend on the chain length since such enzymes operate at both ends of the chain irrespective of its length. We considered synergism as a phenomenon leading to a faster conversion of the substrate. With this definition, a size dependent erosion process does not lead to a synergetic effect (left subplot figure 10). Yet there is actually a consequence of the combined action of the two types of enzymes on the chain length. This could also be regarded as a form of synergism (but the information is rarely accessed experimentally). Comparing the right plots in figure 9 and 10 we can claim that there is always a synergetic effect (in terms of average chain length) and that the contribution of CBH to the average chain length reduction is more pronounced in the case of a linear dependency.

4.5. EG-CBH + β-glucosidase inhibition effect

In a batch reactor, the disintegration of the polymer chains by the cooperative action of the different cellulolytic activities leads to the accumulation of the end-products in the system. This causes the so-called inhibition effect since the enzymes are sensitive to their own products, inactive enzyme-product complexes are formed. As a consequence, the enzymes activity slows down as reported in Figure 11 where no inhibition and inhibition resultas are compared in terms of cellobiose and glucose fractions (last two subplots) and the first two normalised moments ot the CLD (first two subplots).

The introduction of the inhibition effect in the population balance model leads to the slowdown of the conversion rate of the substrate as shown in Figure 11. The total molar concentration ($m_0(t)$) as well as the overall conversion rate ($m_1(t)$, cellobiose and glucose fractions) are affected. The parameters K_{EG}, K_{CBH} and K_P corresponding to the three main activities control the significance of the inhibition effect.

By introducing the inhibition effect, the formulation of the model is more complete. The model inhibition parameters can be reached experimentally by fitting typical saccharification data in the specific case of depolymerisation process for example. The model is built considering the elementary mechanisms of the

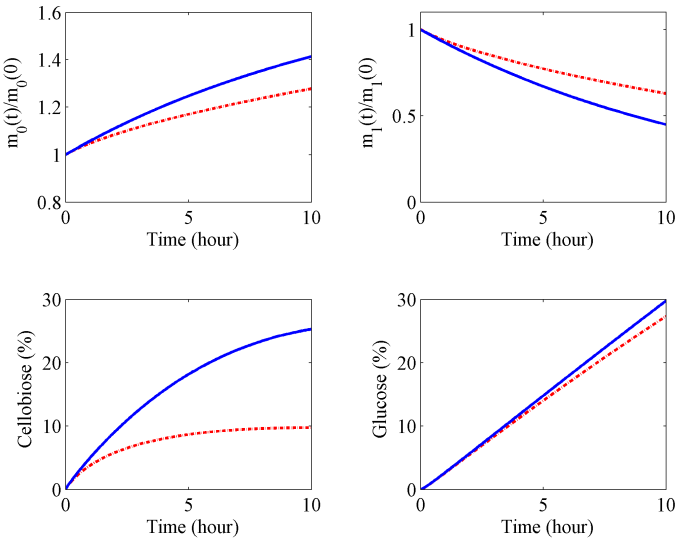

Figure 11: Comprison between the evolution of the two first normalised moments of the CLD and the cellobiose and glucose fractions when considering EG-CBH combined action without inhibition effect (continuous line) and with the inhibition effect (dashed line)

three main cellulase activities on soluble polymers.

5. Conclusion

In this first part, a dynamic population balance model has been developed in the case of cellulose polymers degradation by the synergistic action of cellulases. The three main cellulolytic actions are taken into account since they act differently with specific mechanisms. First, the general theoretical framework is developed in the case of enzymatic hydrolysis process. The random action of the endoglucanase activity (EG) is assimilated to a pure breakage process when the cellobiohydrolase activity (CBH) is modelled as an erosion phenomenon since it proceeds with a chain-end scission and produces cellobiose molecules.

Since the resolution of the population balance equation is computationaly intensive, the Direct Quadrature Method of Moments (DQMOM) is adopted. This resolution method based on the time tracking of a finite number of moments of the chain length distribution (CLD) has been validated against an analytical solution in the case of a specific breakage process. With a view to future comparison with transient experimental

data, the Maximum Entropy based method is coupled to DQMOM for a simultaneous reconstruction of the full CLDs during the PBE resolution. The β-glucosidase activity is modelled as a Michaelis-Menten type kinetic and incorporated to the PBE model.

The actions of EG and CBH activities are tested independently and in combination. Different breakage/erosion frequency laws are used. The numerical results are consistent with the analytical predictions. Global trends of the two first CLD moments (related to the molar concentration and to the mass of the substrate) are shown in the different cases. This guides the experimental determination of the breakage/erosion frequency laws. The EG/CBH synergistic action is discussed. This confirms the discriminatory character of the model since it proves numerically the fact that the erosion frequency can not be length dependent. The incorporation of the inhibition effect is investigated in the last section since the accumulation of cellobiose and glucose affects substantially the enzymes activity.

The use of DQMOM resolution method coupled with the ME-based recontruction technique reduces the computational time drastically. It doesn't exceed few seconds in all the considered cases offering by the fact the possibility to perform a global optimization loop for the determination of the model parameters based on experimental data.

Aknowledgments

The authors would like to thank *Toulouse White Biotechnology* and *Région Midi-Pyrénées* for their financial support.

Nomenclature

A	Square matrix of the linear system in DQMOM approach
a_i	First time derivative of the weights of the quadrature nodes
b_i	First time derivative of the product of the weighted abscissas
C_C	Cellobiose molar concentration
C_G	Glucose molar concentration
c_i	Weighted abscissas
d	Term source vector of the linear system in DQMOM approach
$f(x)$	Target probability density function
$f_M(x)$	Reconstructed probability density function
$H[f]$	Shannon entropy
k	Moments order
K_{CBH}	Inhibition constant of the CBH
K_{EG}	Inhibition constant of the EG
K_m	Parameter of the Michaelis-Menten type kinetic
K_P	Inhibition constant of the β-glucosidase
L	Polymer chain length
L_C	Cellobiose length
L_i	Abscissas of the quadrature nodes
$m_k(t)$	k^{th} order moment
$n(L,t)$	Number based chain length distribution
N	Number of quadrature nodes
$\mathbf{N}$	Maximum degree of polymerization
$P(L)$	Polymer chain
$S_L(L,t)$	Source term of the PBE accounting for the breakage/erosion process
t	Time
V_m	Parameter of the Michaelis-Menten type kinetic
w_i	Weights of the quadrature nodes
x	Generic variable
x_{min}, x_{max}	Lower and upper bourdaries of the reconstruction interval

Greek Symbols

α	Coefficient
$\beta(L,\lambda)$	Daughter distribution function
$\beta_{EG}(L,\lambda)$, $\beta_{CBH}(L,\lambda)$	Daughter distribution function of EG an CBH respectively
$\Gamma(L)$	Breakage frequency
$\Gamma_{EG}(L)$, $\Gamma_{CBH}(L)$	Breakage frequency of EG an CBH respectively
$\mathbf{\Gamma}$	Gamma function
$\delta(L-L_i)$	Dirac delta function centered on L_i
θ	Second parameter of the Gamma distribution
κ	First parameter of the Gamma distribution
λ	Chain length longer than L
μ	Mean of the normal law
ξ_j	Lagrange's multipliers
σ	Standard deviation of the normal law
τ	Time constant

Abreviations

CBH	Cellobiohydrolase
CLD	Chain Length Distribution
DP	Degree of Polymerization
DQMOM	Direct Quadrature Method of Moments
EG	Endoglucanase
ME	Maximum Entropy
PBE	Population Balance Equation
PD	Product-Difference algorithm
PSD	Particle Size Distribution

Appendix A: The Monovariate Direct Quadrature Method of Moments (DQMOM)

We give hereafter a succinct description of the Direct Quadrature Method of Moments (DQMOM) applied for the population balance equation accounting for breakage processes, for more detail one can refer to the original work by Marchisio & Fox (2005).

The homogeneous monovariate PBE for breakage processes can be written in a compact form as:

$$\frac{\partial n(L,t)}{\partial t} = S_L(L,t) \tag{A.1}$$

where $S_L(L,t)$ is the source term due to the breakage, given in its explicit form in equation (1).

In the monovariate DQMOM approach, the continuous distribution function $n(L,t)$ is expressed as a discrete summation of Dirac delta functions (Marchisio & Fox, 2005):

$$n(L,t) = \sum_{i=1}^{N} w_i(t)\delta\left[L(t) - L_i(t)\right] \tag{A.2}$$

where N is the number of delta functions (nodes i), L_i is the property of the node and w_i its weight. By substituting $n(L,t)$ by its discrete decomposition (equation A.2 in equation A.1), one obtains:

$$\sum_{i=1}^{N} \delta\left(L - L_i\right) \frac{\partial w_i}{\partial t} - \sum_{i=1}^{N} \delta'\left(L - L_i\right) w_i \frac{\partial L_i}{\partial t} = S_L(L,t) \tag{A.3}$$

where $\delta'(L - L_i)$ is the first derivative of the delta function $\delta(L - L_i)$. By introducing the weighted abscissa c_i instead of the abscissa L_i (equation A.4),

$$c_i = w_i L_i \tag{A.4}$$

and by setting :

$$\frac{\partial w_i}{\partial t} = a_i \qquad \frac{\partial c_i}{\partial t} = b_i \tag{A.5}$$

Marchisio & Fox (2005) reformulate the equation A.3 as :

$$\sum_{i=1}^{N} \left[\delta\left(L - L_i\right) + \delta'(L - L_i)L_i\right] a_i - \sum_{i=1}^{N} \delta'\left(L - L_i\right) b_i = S_L(L,t) \tag{A.6}$$

The unknowns a_i and b_i are only time-dependent and are reachable by applying moment transformation. Hulburt & Katz (1964) defined the k^{th} integer moment of the distribution function $n(L,t)$ as:

$$m_k(t) = \int_0^\infty L^k n(L,t) dL = \sum_{i=1}^{N} w_i L_i^k \tag{A.7}$$

The moment transformation (multiplying by L^k and integrating over L) is applied to equation A.6. The final equation is given as:

$$(1-k)\sum_{i=1}^{N} L_i^k a_i + k\sum_{i=1}^{N} L_i^{k-1} b_i = \int_0^\infty L^k S_L(L,t) dL \tag{A.8}$$

Considering the definition of S_L given in equation (1), the right hand side term of equation (A.8) now writes:

$$\int L^k S_L(L,t) dL = \int_0^\infty \int_0^\infty L^k \beta(L,\lambda)\Gamma(\lambda) n(\lambda,t) dL d\lambda - \int_0^\infty L^k \Gamma(L) n(L,t) dL \tag{A.9}$$

Thus, equation (A.8) becomes :

$$(1-k)\sum_{i=1}^{N} L_i^k a_i + k\sum_{i=1}^{N} L_i^{k-1} b_i = \sum_{i=1}^{N} \bar{b}_i^{(k)} \Gamma_i w_i - \sum_{i=1}^{N} L_i^k \Gamma_i w_i \tag{A.10}$$

where

$$\bar{b}_i^{(k)} = \int_0^\infty L^k \beta(L, L_i) dL \tag{A.11}$$

The system in equation A.10 can be written in a matrix form as :

$$Ax = d \tag{A.12}$$

where A is a square matrix $(2N, 2N)$, x is the unknowns vector $(2N$: $a_1, \ldots, a_N, b_1, \ldots, b_N)$ and d is the term source vector $(2N)$. Thus, $2N$ moments $(k = 0, \ldots, 2N-1)$ are needed to solve the system. Generally, DQMOM requires at least a three nodes quadrature (N=3) for an accurate tracking of the time evolution of the moments (Marchisio & Fox, 2005)). The initial abscissas $L_i(0)$ and weights $w_i(0)$ are obtained using the Product-Difference algorithm (Gordon, 1968).

Appendix B: DQMOM validation in the case of breakage process

The DQMOM implementation was validated considering a normal law as initial CLD in the case where the PBE has analytical solution (Ziff & McGrady, 1985). In one hand, the time evolution of the CLD is known (see the analytical solution of the PBE for pure breakage process given in Lebaz et al. (2016)), thus the moments $m_k(t)$ of the distribution can be easily derived (equation 2). In the other hand, DQMOM gives directly the time evolution of the CLD moments. The comparison between the two methods is given in the figure bellow:

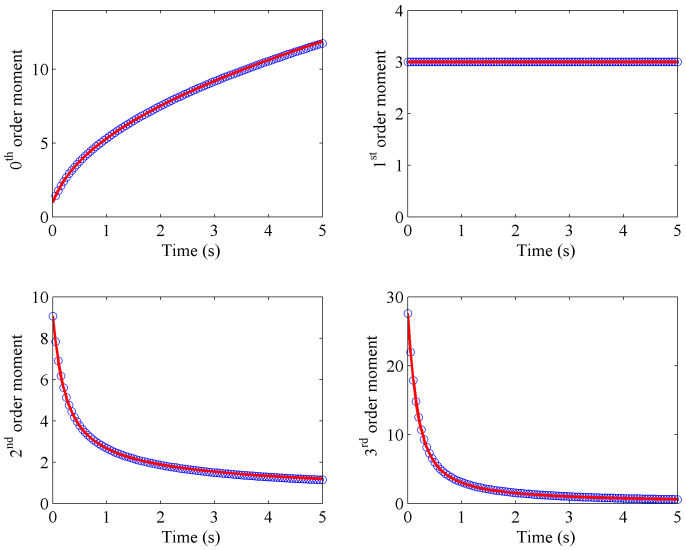

Figure B.12: The four first moments time-evolution of a normal size distribution calculated by DQMOM (circles) and analytical solution (continuous line). The parameters of the normal distribution are $\sigma = 0.25$ and $\mu = 3$.

In order to assess the error induced by DQMOM, we define the global error function as:

$$E(k) = \frac{|m_k - m_k^*|}{m_k^*} \tag{B.1}$$

where m_k is the k^{th} order moment estimated via DQMOM and m_k^* the k^{th} order moment calculated via the analytical solution.

As we can see in figure B.12 , there is a good agreement between the moments calculated from the analytical solution and those estimated via DQMOM. The first order moment refers to the total chains length and remains constant within the time since the polymers are broken up but their total mass correlated with the total length is constant.

The figure B.13 represents the time evolution of the global error function $E(k)$ for the four first moments when using a quadrature with four nodes.

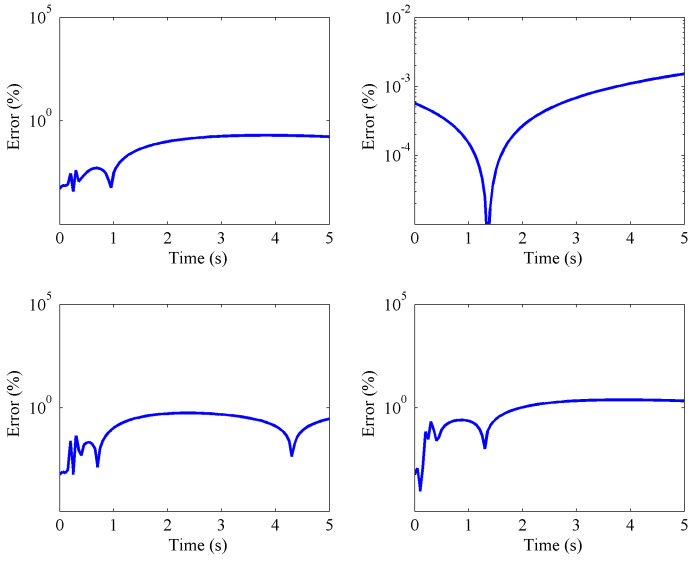

Figure B.13: The time evolution of the global error function $E(k)$ using a four nodes quadrature (eight first moments)

Appendix C: Number density distribution reconstruction from the moments

The recovery of a distribution knowing only a finite number of its moments is known as the moment problem in mathematical analysis and arises in many scientific applications (Gavriliadis & Athanassoulis, 2009). Different numerical methods have been developed for the reconstruction of distributions from the moments reviewed by John et al. (2007). For its simplicity and accuracy, the Maximum Entropy (ME) approach (Tagliani, 1999) is used in this work. We give hereafter a synthetic description of this technique.

We note by $f(x)$ the probability density function (PDF) and by $f_M(x)$ its approximation. The ME method is based on the maximization of the Shannon entropy $H[f]$ given by:

$$H[f] = -\int_0^\infty f(x) ln f(x) dx \tag{C.1}$$

Under the moments constraints:

$$m_k = \int_0^\infty x^k f(x) dx \qquad k = 0, 1, \ldots \tag{C.2}$$

The explicit representation of the ME approximate is given by:

$$f_M(x) = exp\left[-\sum_{j=0}^{2N-1} \xi_j x^j \right] \tag{C.3}$$

To be supplemented by the $2N$ constraints:

$$m_i = \int_0^\infty x^i f_M(x) dx \qquad i = 0, 1, \ldots, 2N-1 \tag{C.4}$$

The $(2N)$ Lagrange's multipliers $\xi_0, \ldots, \xi_{2N-1}$ are obtained by solving the following set of $2N$ nonlinear

equations:

$$\int_0^\infty x^i exp\left[-\sum_{j=0}^{2N-1}\xi_j x^j\right]dx = m_i \qquad i = 0, 1, \ldots, 2N-1 \tag{C.5}$$

Numerically, iterative methods are used to solve equation (C.5) (Mead & Papanicolaou, 1984). In our case, we use the standard Newton method starting from an initial choice of the Lagrange's multipliers as $\xi = (-ln(m_0)/(x_{max} - x_{min}), 0, \ldots, 0)$ with an accuracy of 10^{-6}. The set of moments are normalised before such as $m_0 = 1$. For more details as well as for the validation, one can refer to Lebaz et al. (2016).

Appendix D: Analytical solution of the PBE in the case of breakage processes

The evolution of the CLD moments when considering breakage processes (EG+CBH) can be written in a discret form as :

$$\frac{\partial m_k(t)}{\partial t} = \left[\sum_{i=1}^{N}\overline{b}_i^{(k)}\Gamma_i w_i - \sum_{i=1}^{N} L_i^k \Gamma_i w_i\right]_{EG} + \left[\sum_{i=1}^{N}\overline{b}_i^{(k)}\Gamma_i w_i - \sum_{i=1}^{N} L_i^k \Gamma_i w_i\right]_{CBH} \tag{D.1}$$

By substituting the breakage and erosion daughter distribution functions and frequencies in equation (D.1) as reported in table 1, we obtain the general equation below :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_1\left(\frac{2}{k+1} - 1\right)m_{k+p_1} + \alpha_2\left(\sum_i (L_i - L_C)^k L_i^{p_2} w_i - m_{k+p_2}\right) \tag{D.2}$$

with (α_1, p_1) the parameters of the breakage frequency of EG and (α_2, p_2) the parameters of the erosion frequency of CBH.

We explicit hereafter the expected evolution of the two first moments considering length-dependent ($p_1 = p_2 = 1$) and length-independent ($p_1 = p_2 = 0$) frequencies.

- **Case 1 :** $p_1 = p_2 = 1$

- *Random breakge* :

In this case, equation (D.2) is reduced to :

$$\frac{\partial m_k(t)}{\partial t} = \left(\frac{2}{k+1} - 1\right) \alpha_1 m_{k+1} \tag{D.3}$$

The evolution of the two first moments is given by :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= \alpha_1 m_1(t) \\ \frac{\partial m_1(t)}{\partial t} &= 0 \end{cases} \tag{D.4}$$

The integration of the two differential equations in (D.4) leads to :

$$\begin{cases} m_0(t) &= \alpha_1.m_1(0).t + m_0(0) \\ m_1(t) &= m_1(0) \end{cases} \tag{D.5}$$

- *Erosion process* :

Equation (D.2) is written as :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_2 \left(\sum_i (L_i - L_C)^k L_i w_i - m_{k+1} \right) \tag{D.6}$$

The evolution of the two first moments is :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= 0 \\ \frac{\partial m_1(t)}{\partial t} &= -\alpha_2 L_C m_1(t) \end{cases} \tag{D.7}$$

This leads to the two first explicit moments :

$$\begin{cases} m_0(t) &= m_0(0) \\ m_1(t) &= m_1(0).e^{-\alpha_2 L_C t} \end{cases} \tag{D.8}$$

- *Random breakage + Erosion processes* :

The combination of the two processes is traduced by the equation below :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_1 \left(\frac{2}{k+1} - 1 \right) m_{k+1} + \alpha_2 \left(\sum_i (L_i - L_C)^k L_i w_i - m_{k+1} \right) \tag{D.9}$$

The evolution of the two first moments is :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= \alpha_1 m_1(t) \\ \frac{\partial m_1(t)}{\partial t} &= -\alpha_2 L_C m_1(t) \end{cases} \tag{D.10}$$

The two first moments in this case are :

$$\begin{cases} m_0(t) &= \frac{\alpha_1 m_1(0)}{\alpha_2 L_C} \left(1 - e^{-\alpha_2 L_C t}\right) + m_0(0) \\ m_1(t) &= m_1(0).e^{-\alpha_2 L_C t} \end{cases} \tag{D.11}$$

- **Case 2 :** $p_1 = p_2 = 0$

- *Random breakge* :

In this case, equation (D.2) is reduced to :

$$\frac{\partial m_k(t)}{\partial t} = \left(\frac{2}{k+1} - 1 \right) \alpha_1 m_k \tag{D.12}$$

The evolution of the two first moments is given by :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= \alpha_1 m_0(t) \\ \frac{\partial m_1(t)}{\partial t} &= 0 \end{cases} \tag{D.13}$$

The integration of the two differential equations in (D.13) leads to :

$$\begin{cases} m_0(t) &= m_0(0).e^{\alpha_1 t} \\ m_1(t) &= m_1(0) \end{cases} \tag{D.14}$$

- *Erosion process* :

Equation (D.2) is written as :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_2 \left(\sum_i (L_i - L_C)^k w_i - m_k \right) \tag{D.15}$$

The evolution of the two first moments is :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= 0 \\ \frac{\partial m_1(t)}{\partial t} &= -\alpha_2 L_C m_0(t) \end{cases} \tag{D.16}$$

This leads to the two first explicit moments :

$$\begin{cases} m_0(t) &= m_0(0) \\ m_1(t) &= -\alpha_2 L_C m_0(0) t + m_1(0) \end{cases} \tag{D.17}$$

- *Random breakage + Erosion processes* :

The combination of the two processes is traduced by the equation below :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_1 \left(\frac{2}{k+1} - 1 \right) m_k + \alpha_2 \left(\sum_i (L_i - L_C)^k w_i - m_k \right) \tag{D.18}$$

The evolution of the two first moments is :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= \alpha_1 m_0(t) \\ \frac{\partial m_1(t)}{\partial t} &= -\alpha_2 L_C m_0(t) \end{cases} \tag{D.19}$$

The two first moments in this case are :

$$\begin{cases} m_0(t) &= m_0(0).e^{\alpha_1 t} \\ m_1(t) &= \frac{\alpha_2 L_C m_0(0)}{\alpha_1} \left(1 - e^{\alpha_1 t}\right) + m_1(0) \end{cases} \tag{D.20}$$

- **Case 3 :** $p_1 = 1$ and $p_2 = 0$

We explore the combination of a random breakage with a length-dependent frequency and an erosion process with length-independent frequency. Equation (D.2) is written as :

$$\frac{\partial m_k(t)}{\partial t} = \alpha_1 \left(\frac{2}{k+1} - 1 \right) m_{k+1} + \alpha_2 \left(\sum_i (L_i - L_C)^k w_i - m_k \right) \tag{D.21}$$

The evolution of the two first moments is :

$$\begin{cases} \frac{\partial m_0(t)}{\partial t} &= \alpha_1 m_1(t) \\ \frac{\partial m_1(t)}{\partial t} &= -\alpha_2 L_C m_0(t) \end{cases} \tag{D.22}$$

The analytical resolution of the set of equations in (D.22) is given below :

$$\begin{cases} m_0(t) &= m_0(0) \cos\left(\sqrt{\alpha_1 \alpha_2 L_C}.t\right) + \sqrt{\frac{\alpha_1}{\alpha_2 L_C}}.m_1(0) \sin\left(\sqrt{\alpha_1 \alpha_2 L_C}.t\right) \\ m_1(t) &= m_1(0) \cos\left(\sqrt{\alpha_1 \alpha_2 L_C}.t\right) - \sqrt{\frac{\alpha_2 L_C}{\alpha_1}}.m_0(0) \sin\left(\sqrt{\alpha_1 \alpha_2 L_C}.t\right) \end{cases} \tag{D.23}$$

Appendix E: Supplementary data

Figures E.14, E.15 and E.16 summarize the numerical results compared to those predicted analytically for the first, second and third cases respectively.

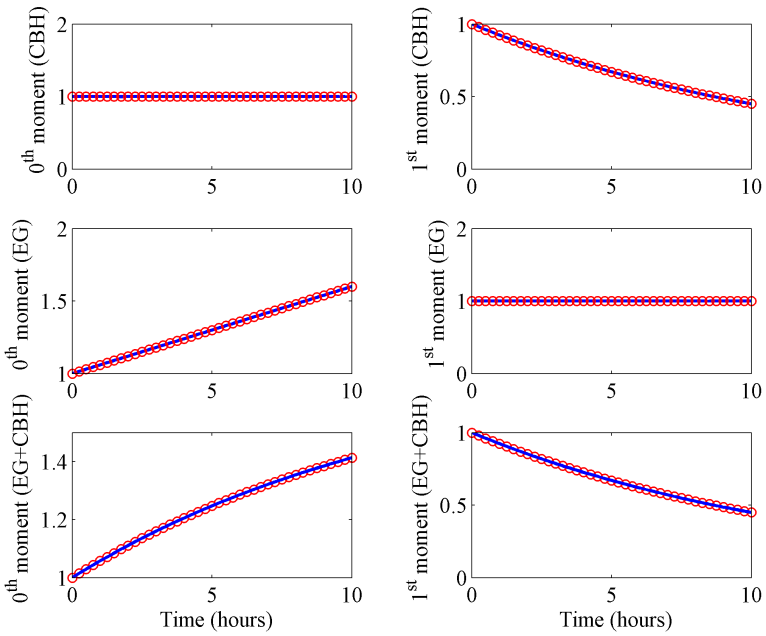

Figure E.14: Comparison between the analytical solution (continuous line) and the numerical results (circles) for the three different cases with $\Gamma_{EG} = \alpha_1 L$ and $\Gamma_{CBH} = \alpha_2 L$.

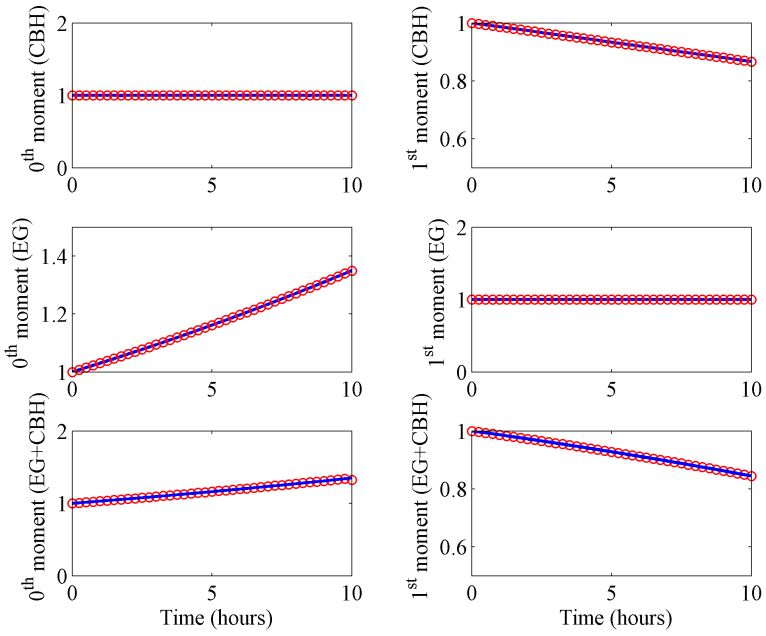

Figure E.15: Comparison between the analytical solution (continuous line) and the numerical results (circles) in the case of EG+CBH with $\Gamma_{EG} = \alpha_1$ and $\Gamma_{CBH} = \alpha_2$.

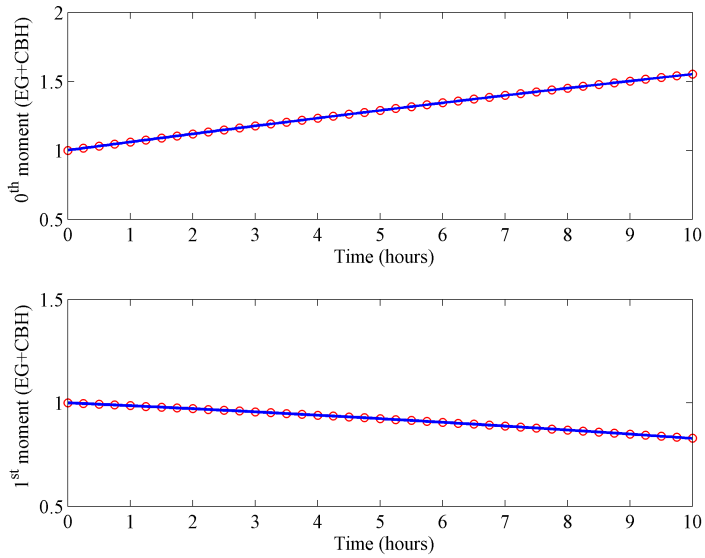

Figure E.16: Comparison between the analytical solution (continuous line) and the numerical results (circles) in the case of EG+CBH with $\Gamma_{EG} = \alpha_1 L$ and $\Gamma_{CBH} = \alpha_2$.

References

Andersen, N., Johansen, K. S., Michelsen, M., Stenby, E. H., Krogh, K. B. R. M., & Olsson, L. (2008). Hydrolysis of cellulose using mono-component enzymes shows synergy during hydrolysis of phosphoric acid swollen cellulose (PASC), but competition on Avicel. *Enzyme and Microbial Technology*, *42*, 362–370. URL: http://www.sciencedirect.com/science/article/pii/S0141022907003456. doi:10.1016/j.enzmictec.2007.11.018.

Bansal, P., Hall, M., Realff, M. J., Lee, J. H., & Bommarius, A. S. (2009). Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances*, *27*, 833–848. URL: http://www.sciencedirect.com/science/article/pii/S0734975009001402. doi:10.1016/j.biotechadv.2009.06.005.

Ding, A., Hounslow, M. J., & Biggs, C. A. (2006). Population balance modelling of activated sludge flocculation: Investigating the size dependence of aggregation, breakage and collision efficiency. *Chem-*

ical engineering science, *61*, 63–74. URL: http://www.sciencedirect.com/science/article/pii/S0009250905004410.

Fan, L. T., & Lee, Y.-h. (1983). Kinetic studies of enzymatic hydrolysis of insoluble cellulose: Derivation of a mechanistic kinetic model. *Biotechnology and Bioengineering*, *25*, 2707–2733. URL: http://onlinelibrary.wiley.com/doi/10.1002/bit.260251115/abstract. doi:10.1002/bit.260251115.

Gavriliadis, P. N., & Athanassoulis, G. A. (2009). Moment information for probability distributions, without solving the moment problem, II: Main-mass, tails and shape approximation. *Journal of Computational and Applied Mathematics*, *229*, 7–15. URL: http://www.sciencedirect.com/science/article/pii/S037704270800513X. doi:10.1016/j.cam.2008.10.011.

Gordon, R. G. (1968). Error Bounds in Equilibrium Statistical Mechanics. *Journal of Mathematical Physics*, *9*, 655–663. URL: http://scitation.aip.org/content/aip/journal/jmp/9/5/10.1063/1.1664624. doi:10.1063/1.1664624.

Gregg, D., & Saddler, J. (1995). Bioconversion of lignocellulosic residue to ethanol: Process flowsheet development. *Biomass and Bioenergy*, *9*, 287–302. URL: http://www.sciencedirect.com/science/article/pii/0961953495000976. doi:10.1016/0961-9534(95)00097-6.

Griggs, A. J., Stickel, J. J., & Lischeske, J. J. (2012a). A mechanistic model for enzymatic saccharification of cellulose using continuous distribution kinetics I: Depolymerization by EGI and CBHI. *Biotechnology and Bioengineering*, *109*, 665–675. URL: http://onlinelibrary.wiley.com/doi/10.1002/bit.23355/abstract. doi:10.1002/bit.23355.

Griggs, A. J., Stickel, J. J., & Lischeske, J. J. (2012b). A mechanistic model for enzymatic saccharification of cellulose using continuous distribution kinetics II: Cooperative enzyme action, solution kinetics, and prod-

uct inhibition. *Biotechnology and Bioengineering*, *109*, 676–685. URL: http://onlinelibrary.wiley.com/doi/10.1002/bit.23354/abstract. doi:10.1002/bit.23354.

Gruno, M., Vljame, P., Pettersson, G., & Johansson, G. (2004). Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnology and Bioengineering*, *86*, 503–511. URL: http://onlinelibrary.wiley.com/doi/10.1002/bit.10838/abstract. doi:10.1002/bit.10838.

Ho, Y. K., Doshi, P., Yeoh, H. K., & Ngoh, G. C. (2014). Modeling chain-end scission using the Fixed Pivot technique. *Chemical Engineering Science*, *116*, 601–610. URL: http://www.sciencedirect.com/science/article/pii/S0009250914002528. doi:10.1016/j.ces.2014.05.035.

Hosseini, S. A., & Shah, N. (2011a). Enzymatic hydrolysis of cellulose part II: Population balance modelling of hydrolysis by exoglucanase and universal kinetic model. *biomass and bioenergy*, *35*, 3830–3840. URL: http://www.sciencedirect.com/science/article/pii/S0961953411002315.

Hosseini, S. A., & Shah, N. (2011b). Modelling enzymatic hydrolysis of cellulose part I: Population balance modelling of hydrolysis by endoglucanase. *Biomass and Bioenergy*, *35*, 3841–3848. URL: http://www.sciencedirect.com/science/article/pii/S0961953411002285. doi:10.1016/j.biombioe.2011.04.026.

Hulburt, H. M., & Katz, S. (1964). Some problems in particle technology: A statistical mechanical formulation. *Chemical Engineering Science*, *19*, 555–574. URL: http://www.sciencedirect.com/science/article/pii/0009250964850478. doi:10.1016/0009-2509(64)85047-8.

John, V., Angelov, I., ncl, A. A., & Thvenin, D. (2007). Techniques for the reconstruction of a distribution from a finite number of its moments. *Chemical Engineering Science*, *62*, 2890–2904. URL: http://www.sciencedirect.com/science/article/pii/S0009250907002072. doi:10.1016/j.ces.2007.02.041.

Jorgensen, H., Kristensen, J. B., & Felby, C. (2007). Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining*, *1*, 119–134. URL: http://onlinelibrary.wiley.com/doi/10.1002/bbb.4/abstract. doi:10.1002/bbb.4.

Kadam, K. L., Rydholm, E. C., & McMillan, J. D. (2004). Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. *Biotechnology progress*, *20*, 698–705. URL: http://onlinelibrary.wiley.com/doi/10.1021/bp034316x/pdf.

Kostoglou, M. (2000). Mathematical analysis of polymer degradation with chain-end scission. *Chemical Engineering Science*, *55*, 2507–2513. URL: http://www.sciencedirect.com/science/article/pii/S0009250999004716. doi:10.1016/S0009-2509(99)00471-6.

Kumar, S., & Ramkrishna, D. (1996a). On the solution of population balance equations by discretizationI. A fixed pivot technique. *Chemical Engineering Science*, *51*, 1311–1332. URL: http://www.sciencedirect.com/science/article/pii/0009250996884892. doi:10.1016/0009-2509(96)88489-2.

Kumar, S., & Ramkrishna, D. (1996b). On the solution of population balance equations by discretizationII. A moving pivot technique. *Chemical Engineering Science*, *51*, 1333–1342. URL: http://www.sciencedirect.com/science/article/pii/000925099500355X. doi:10.1016/0009-2509(95)00355-X.

Lebaz, N., Cockx, A., Sprandio, M., & Morchain, J. (2015). Population balance approach for the modelling of enzymatic hydrolysis of cellulose. *The Canadian Journal of Chemical Engineering*, *93*, 276–284. URL: http://onlinelibrary.wiley.com/doi/10.1002/cjce.22088/abstract. doi:10.1002/cjce.22088.

Lebaz, N., Cockx, A., Sprandio, 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. URL: http://www.sciencedirect.com/science/article/pii/S0098135415003038. doi:10.1016/j.compchemeng.2015.09.008.

Marchisio, D. L., & Fox, R. O. (2005). Solution of population balance equations using the direct quadrature method of moments. *Journal of Aerosol Science*, *36*, 43–73. URL: http://www.sciencedirect.com/science/article/pii/S0021850204003052. doi:10.1016/j.jaerosci.2004.07.009.

McCoy, B. J., & Wang, M. (1994). Continuous-mixture fragmentation kinetics: particle size reduction and molecular cracking. *Chemical Engineering Science*, *49*, 3773–3785. URL: http://www.sciencedirect.com/science/article/pii/0009250994001723. doi:10.1016/0009-2509(94)00172-3.

Mead, L. R., & Papanicolaou, N. (1984). Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, *25*, 2404–2417. URL: http://scitation.aip.org/content/aip/journal/jmp/25/8/10.1063/1.526446. doi:10.1063/1.526446.

Mittal, A., Katahira, R., Himmel, M. E., Johnson, D. K., & others (2011). Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility. *Biotechnol Biofuels*, *4*, 41. URL: http://www.biomedcentral.com/content/pdf/1754-6834-4-41.pdf.

Nopens, I., Biggs, C. A., De Clercq, B., Govoreanu, R., Wilen, B. M., Lant, P., & Vanrolleghem, P. A. (2002). Modelling the activated sludge flocculation process combining laser light diffraction particle sizing and population balance modelling(PBM). *Water Science & Technology*, *45*, 41–49. URL: http://modeleau.fsg.ulaval.ca/fileadmin/modeleau/documents/Publications/pvr317.pdf.

Sin, G., Meyer, A. S., & Gernaey, K. V. (2010). Assessing reliability of cellulose hydrolysis models to support biofuel process designIdentifiability and uncertainty analysis. *Computers & chemical engineering*, *34*, 1385–1392. URL: http://www.sciencedirect.com/science/article/pii/S0098135410000554.

South, C. R., Hogsett, D. A. L., & Lynd, L. R. (1995). Modeling simultaneous saccharification and fermentation of lignocellulose to ethanol in batch and continuous reactors. *Enzyme and Microbial Tech-*

nology, *17*, 797–803. URL: http://www.sciencedirect.com/science/article/pii/014102299400016K. doi:10.1016/0141-0229(94)00016-K.

Sun, Y., & Cheng, J. (2002). Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresource Technology*, *83*, 1–11. URL: http://www.sciencedirect.com/science/article/pii/S0960852401002127. doi:10.1016/S0960-8524(01)00212-7.

Tagliani, A. (1999). Hausdorff moment problem and maximum entropy: A unified approach. *Applied Mathematics and Computation*, *105*, 291–305. URL: http://www.sciencedirect.com/science/article/pii/S009630039810084X. doi:10.1016/S0096-3003(98)10084-X.

Van Dyk, J. S., & Pletschke, B. I. (2012). A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymesFactors affecting enzymes, conversion and synergy. *Biotechnology Advances*, *30*, 1458–1480. URL: http://www.sciencedirect.com/science/article/pii/S0734975012000687. doi:10.1016/j.biotechadv.2012.03.002.

Wang, M., Smith, J. M., & McCoy, B. J. (1995). Continuous kinetics for thermal degradation of polymer in solution. *AIChE Journal*, *41*, 1521–1533. URL: http://onlinelibrary.wiley.com/doi/10.1002/aic.690410616/abstract. doi:10.1002/aic.690410616.

Xiao, Z., Zhang, X., Gregg, D. J., & Saddler, J. N. (2004). Effects of sugar inhibition on cellulases and β-glucosidase during enzymatic hydrolysis of softwood substrates. *Applied Biochemistry and Biotechnology*, *115*, 1115–1126. URL: http://link.springer.com/article/10.1385/ABAB%3A115%3A1-3%3A1115. doi:10.1385/ABAB:115:1-3:1115.

Xu, F., & Ding, H. (2007). A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions from the fractal and jamming (overcrowding) effects. *Applied Catalysis A: General*, *317*, 70–

81. URL: http://www.sciencedirect.com/science/article/pii/S0926860X06007150. doi:10.1016/j.apcata.2006.10.014.

Yang, B., Willies, D. M., & Wyman, C. E. (2006). Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion. *Biotechnology and Bioengineering*, *94*, 1122–1128. URL: http://onlinelibrary.wiley.com/doi/10.1002/bit.20942/abstract. doi:10.1002/bit.20942.

Yuan, C., & Fox, R. O. (2011). Conditional quadrature method of moments for kinetic equations. *Journal of Computational Physics*, *230*, 8216–8246. URL: http://www.sciencedirect.com/science/article/pii/S0021999111004396. doi:10.1016/j.jcp.2011.07.020.

Yuan, C., Laurent, F., & Fox, R. O. (2012). An extended quadrature method of moments for population balance equations. *Journal of Aerosol Science*, *51*, 1–23. URL: http://www.sciencedirect.com/science/article/pii/S0021850212000699. doi:10.1016/j.jaerosci.2012.04.003.

Ziff, R. M., & McGrady, E. D. (1985). The kinetics of cluster fragmentation and depolymerisation. *Journal of Physics A: Mathematical and General*, *18*, 3027. URL: http://iopscience.iop.org/0305-4470/18/15/026. doi:10.1088/0305-4470/18/15/026.