Bioresource Technology 177 (2015) 255–265

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

Corn stover semi-mechanistic enzymatic hydrolysis model with tight parameter confidence intervals for model-based process design and optimization

Felipe Scott [a,b,*], Muyang Li [c,d], Daniel L. Williams [c,e], Raúl Conejeros [a,b], David B. Hodge [c,d,e,f], Germán Aroca [a,b]

[a] School of Biochemical Engineering, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2147, Valparaíso, Chile
[b] Bioenercel S.A. Barrio Universitario s/n, Ideaincuba building, Concepción, Chile
[c] DOE-Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA
[d] Department of Biosystems & Agricultural Engineering, Michigan State University, 48824 East Lansing, MI, USA
[e] Department of Chemical Engineering & Materials Science, Michigan State University, 48824 East Lansing, MI, USA
[f] Division of Chemical Engineering, Luleå University of Technology, Luleå, Sweden

HIGHLIGHTS

- A semi-mechanistic model is proposed for pretreated corn stover saccharification.
- The model considers high-solid saccharification and washed or unwashed solids.
- A subset of identifiable parameters was found showing tight confidence intervals.
- Uncertainty in parameters estimates was used to predict bands for glucose yield.
- The model reliably describes the saccharification kinetics of corn stover' glucan.

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form 13 November 2014
Accepted 14 November 2014
Available online 21 November 2014

Keywords:
Lignocellulose
Enzymatic hydrolysis
Kinetic model
High-solids saccharification
Biofuels

ABSTRACT

Uncertainty associated to the estimated values of the parameters in a model is a key piece of information for decision makers and model users. However, this information is typically not reported or the confidence intervals are too large to be useful. A semi-mechanistic model for the enzymatic saccharification of dilute acid pretreated corn stover is proposed in this work, the model is a modification of an existing one providing a statistically significant improved fit towards a set of experimental data that includes varying initial solid loadings (10–25% w/w) and the use of the pretreatment liquor and washed solids with or without supplementation of key inhibitors. A subset of 8 out of 17 parameters was identified, showing sufficiently tight confidence intervals to be used in uncertainty propagation and model analysis, without requiring interval truncation via expert judgment.

1. Introduction

The cell walls of plants comprising lignocellulosic biomass are a complex and heterogeneous matrix composed primarily of the biopolymers: cellulose, hemicelluloses, and lignin (Chundawat et al., 2011). These cell wall biopolymers offer the potential as feedstocks for the sustainable production of renewable fuels, chemicals, and biomaterials with a diverse range of biochemical, thermochemical, and catalytic routes. One promising conversion route involves the deconstruction of the cell wall polysaccharides into fermentable monosaccharides by a pretreatment and polysaccharide hydrolysis, followed by biological conversion of sugars to fuels such as ethanol (Galbe and Zacchi, 2012). Cellulose hydrolysis of pretreated lignocellulose can be performed using a cocktail of cooperative cellulase enzymes containing glycosyl hydrolases (Lynd et al., 2002) as well as a recently recognized class of lytic polysaccharide monooxygenases (Harris et al, 2014) that are responsible for

Abbreviation: CBH, cellobiohydrolases; EG, endoglucanases; FPU, Filter Paper Unit; PCS, pretreated corn stover.

* Corresponding author at: School of Biochemical Engineering, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2147, Valparaíso, Chile. Tel.: +56 32 2273755.
E-mail address: felipe.scott.c@mail.pucv.cl (F. Scott).

http://dx.doi.org/10.1016/j.biortech.2014.11.062
0960-8524/

Nomenclature

a_{iad}	adsorption decreasing factor (kg/g)
b_i	activity decreasing factor (kg/g)
$CI^{1-\alpha}$	confidence interval at α significance level
$\boldsymbol{COV}$	$m \times m$ covariance matrix
E_{1max}	maximum mass of enzyme that can be adsorbed onto a unit mass of substrate: 0.06 (g/g)
E_{2max}	maximum mass of enzyme that can be adsorbed onto a unit mass of substrate: 0.01 (g/g)
E_{1B}	bound concentration of CBH and EG (g/kg)
E_{2B}	bound concentration of β-glucosidase (g/kg)
E_{2F}	free concentration of β-glucosidase (g/kg)
E_T	total enzyme concentration (g/kg)
E_{1T}	concentration of CBH and EG (g/kg)
E_{2T}	concentration of β-glucosidase (g/kg)
f_2	fraction of β-glucosidase protein in Spezyme CP
$f_{\beta G}$	fraction of the maximum β-glucosidase activity
G	glucose concentration (g/kg)
G_2	cellobiose concentration (g/kg)
$J(\boldsymbol{\theta})$	cost function for parameter estimation
K_{3M}	cellobiose saturation constant (g/kg)
K_{1ad}	dissociation constant for the enzyme adsorption–desorption reaction: 0.4 (g/g)
K_{2ad}	dissociation constant for the enzyme adsorption–desorption reaction: 0.1 (g/g)
K_{ilA}	inhibition constant for acetic acid (g/kg)
K_{ilG}	inhibition constant for glucose (g/kg)
K_{ilG2}	inhibition constant for cellobiose (g/kg)
K_{ilX}	inhibition constant for five carbon sugars (g/kg)
k_{ir}	reaction rate (g· kg^{-1} · h^{-1})
R_S	substrate reactivity
$rCI^{1-\alpha}$	relative half confidence interval
S	cellulose concentration (g/kg)
$\boldsymbol{V}$	$n \times m$ derivative matrix
$\boldsymbol{W}$	$n \times n$ diagonal matrix of weights

Indices and sets

m	number of parameters
n	number of experimental measures
$\boldsymbol{\theta}$	set of parameters
$\boldsymbol{\theta}_f$	set of parameters with fixed values
$\boldsymbol{K}$	set of combinations of m parameters taken k at a time, each row (K) represents a particular combination of k parameters

Greek symbols

α	significance level for t-test and F-test
δ_{min}	minimum acceptable parameter sensitivity
δ^{msqr}	sensitivity measure
γ_K	collinearity index of parameter subset K
γ_K^C	collinearity index of complement of parameter subset K
γ_K^{max}	maximum allowable collinearity index
$\boldsymbol{\Lambda}'$	matrix of elements of $\boldsymbol{K}$ showing $\gamma_K \leqslant \gamma_K^{max}$

depolymerizing cellulose. Some hemicellulose-retaining pretreatments also require hemicellulose-depolymerizing enzymes to maximize cellulose hydrolysis (Decker et al., 2008). The process is complex due to the number of enzymes that take part and the fact that reactions take place on the surface of a water-insoluble crystalline polymer (i.e. cellulose hydrolysis) as well as reactions in the liquid phase (i.e. cellobiose hydrolysis).

An extensive collection of kinetic models of cellulose enzymatic hydrolysis for model cellulosic substrates and pretreated biomass can be found in literature and have been recently reviewed (Bansal et al., 2009). Models range from simple empirical or black-box models to complex mechanistic models, which attempt to use the current understanding of how the process works to derive causal hypotheses that are incorporated in the mathematical model. While empirical models may have a small number of parameters to adjust (although there are exceptions such as neural network models), mechanistic models can involve a large number of parameters, which need to be found by fitting the model to a large number of purposely generated experimental data (Brun et al., 2001). From a model-based process design point of view; the use of either empirical or mechanistic models depends on whether the user is interested in testing conditions within the experimental data (interpolation), or in testing conditions that lie outside the experimental conditions (extrapolation) where the mechanistic model provide a rational basis for predicting the behavior of the system.

Due to the complexity of the enzymatic hydrolysis process, the changing enzyme formulations made available by the major commercial enzyme producers, as well as the strong influence that pretreatment and feedstock have over the outcome of enzymatic hydrolysis, semi-mechanistic models with the smallest possible number of parameters may be the most adequate choice from a model-based development point of view, thereby reducing the amount of experimental data required to estimate the parameters values. Among the existing semi-mechanistic models, the one developed by NREL researchers in 2004 (Kadam et al., 2004) has been used in a number of biofuel production processes flowsheets evaluation and alternatives comparison (Scott et al., 2013; Morales-Rodriguez et al., 2011; Hodge et al., 2009) and it has been subjected to an identifiability and uncertainty analysis (Sin et al., 2010). Results indicate that only 6 out of 26 parameters are identifiable from the original data, and any attempt to identify a higher number of parameters results in significant errors on their estimates. This is evidenced by the wide confidence intervals for the values of the parameters.

Uncertainty in parameter estimates can arise from a number of sources, including insufficiently informative experimental data, i.e. the model is not sensitive to some of the parameters to be estimated over the experimental data set (Raue et al., 2009) and parameters that are correlated, i.e. parameters are mathematically related to each other through some implicit function (Li and Vu, 2013; Raue et al., 2009). Furthermore, to be used in model-based process development, a cellulose hydrolysis kinetic model should incorporate as many aspects controlling the process behavior as possible. Among them, initial solids loading is a key factor since the higher the concentration of substrate the higher the ethanol titer in the fermentation stage, which decreases the energy needs in the recovery stage. However, it is known that high solid loadings affect the final glucan conversion (Wang et al., 2011; Kristensen et al., 2009; Hodge et al., 2008). Usage of the liquor generated during pretreatment is desirable from a technical and economical point of view, since no capital-intensive separation equipment is required and the sugar oligomers released during pretreatment can be hydrolyzed by the action of enzymes in the saccharification stage, however, enzymes are known to be inhibited by soluble sugar monomers, dimers, and oligomers (Teugjas and Väljamäe, 2013; Qing et al., 2010), organic acids (Hodge et al., 2008), and phenolic compounds (Tejirian and Xu, 2011) contained in the pretreatment liquor. Despite these inherent difficulties and

complexities in modeling cellulose enzymatic saccharification, process engineers and decision-makers do not only require a single point estimate of model parameters, but also a measure of the uncertainty associated with them. If available, information regarding parameter uncertainty (in the form of confidence intervals) can be approximately translated into prediction confidence intervals using suitable methods (Kreutz et al., 2012; Sin et al., 2010), or it can be directly used for optimal process design under uncertainty (Rooney and Biegler, 2001). However, to be useful, confidence regions and confidence intervals need to be physically realizable and as tight as possible.

This works aims to identify the parameters of a semi-mechanistic model allowing prediction of enzyme hydrolysis of lignocellulosic materials at several initial solid loadings, which is capable of simulating the effect of the pretreatment liquor and with feasible (whitout negative or zero elements) conventional parameters confidence intervals. Additionally, the bounds of the confidence region are calculated based on a more appropriate method for non-linear models. To our knowledge, this is the first work addressing kinetic modeling of enzymatic hydrolysis of lignocellulosic material when the pretreatment liquor is present in the reaction mixture, at high solid loadings and focusing on obtaining tight confidence intervals.

2. Methods

2.1. Experimental data

The experimental data used in this study has been reported previously (Hodge, 2005; Hodge et al., 2008) along with a thorough description of the experimental setup, hence, only a brief description will be given here. Additionally, the experimental measurements are reported in the Supplementary Material accompanying this article. The substrate for enzymatic hydrolysis corresponds to corn stover pretreated by dilute sulfuric acid in NREL's continuous pilot-scale vertical reactor as reported elsewhere (Schell et al., 2003). Once diluted to the target solid fraction, the resulting pretreated corn stover solids and slurry compositions are shown in Table 1. Spezyme CP from DuPont-Genencor (lot#301-04075-054) without β-glucosidase supplementation was used as the cellulase preparation throughout these studies with a protein content of 106 mg/ml, measured using the enhanced BCA protocol (Pierce Biotechnology, Rockford, IL) and 59 FPU/ml.

In order to address the proposed objectives, the set of experimental data contains data on hydrolysis of washed solids, full slurry (pH conditioned hydrolyzate and solids); washed solids mixed with a concentrated stock solutions of hydrolyzate sugars, and, finally washed solids with supplemental acetic acid. The initial insoluble solids content in the hydrolysis experiments range from 10% to 25% (weight to total weight). All enzymatic hydrolysis experiments were performed at 45 °C and initial pH of 5.0 in shake flasks.

2.2. High solids cellulase binding to dilute acid pretreated corn stover

Dilute acid pretreated corn stover (PCS) was graciously provided by Dan Schell (National Renewable Energy Laboratory). To remove solubles for the binding studies, the PCS insoluble solids were subjected to multiple cycles of centrifugation, decanting, and resuspension in distilled water until a pH of greater than 4 was reached. The moisture content of washed and decanted PCS insoluble solids was determined and biomass was stored sealed at 4 °C to prevent moisture loss. Next, approximately 1–4 g of wet biomass was placed into 15 mL centrifuge tubes. Volumes of enzyme (Accellerase 1500, DuPont Danisco Cellulosic Ethanol LLC, Itasca, IL, USA), Na-citrate buffer (pH 5.0, 1 M), and water were calculated and added to yield final solids loadings of 5%, 10%, 15% and 20% (weight to total weight) and protein loadings (according to the Bradford Assay) at 2.5, 5, 7.5 and 10 mg/g dry biomass for each solids level. Duplicate sets of samples were subsequently vortexed and subjected to mixing at 4 °C by end-to-end rotation (Revolver Rotator, Labnet International, Inc., Edison, NJ, USA) for 22 h in a walk-in refrigerator. Following incubation, free protein was determined colorimetrically using the Bradford Assay and fraction bound protein was determined as reported in our previous work (Williams and Hodge, 2014). Following sampling for binding, samples were hydrolyzed by rotary incubation at 50 °C with periodic sampling to determine hydrolysis yields.

2.3. Parameter estimation and fitting evaluation

The predicted values of the state variables $\boldsymbol{X}_i = (S, G_2, G)_i^T$, with G_2 (cellobiose), G (glucose) and S (cellulose), for a particular experimental condition i in Table 1 are calculated by solving a set of algebraic and ordinary differential equations, Eqs. (5) and (6). Where $\boldsymbol{f}$ is a vector of reaction rates $(r_1, r_2, r_3)^T$ defined in Eqs. (9)–(11) and $\boldsymbol{g}$ stands for Eqs. (12)–(15).

Table 1
Summary of experimental conditions used for the estimation of models parameters.

Tag	Comments[a]	SF[b]	Glucan(%)[c]	E_T[d]	G_{20}	G_0	X_0	Ar_0	Ac_0
EXP1	F	0.15	63.2	25.0	1.26	24.07	63.87	10.57	12.20
EXP2	F	0.15	63.2	8.2	1.26	24.17	64.20	10.57	12.20
EXP3	W	0.10	60.1	9.6	0.00	0.00	0.00	0.00	0.00
EXP4	F	0.10	60.1	9.6	1.57	16.47	44.14	7.01	8.16
EXP5	F	0.15	63.2	9.6	1.26	24.05	64.44	10.57	12.20
EXP6	W + Ac. A	0.15	63.2	9.6	0.00	0.00	0.00	0.00	12.20
EXP7	W	0.15	63.2	9.6	0.00	0.00	0.00	0.00	0.00
EXP8	W + Sug.	0.15	63.2	9.6	1.26	24.05	64.44	10.57	0.00
EXP9	W	0.13	60.1	9.6	0.00	0.00	0.00	0.00	0.00
EXP10	W + G	0.10	53.2	5.4	0.00	153.05	0.00	0.00	0.00
EXP11	W	0.25	53.2	16.2	0.00	0.00	0.00	0.00	0.00
EXP12	W	0.25	53.2	5.4	0.00	0.00	0.00	0.00	0.00

Initial soluble solids content [g/kg]: G_{20}, cellobiose; G_0, glucose; X_0, xylose; Ar_0, arabinose; Ac_0, acetic acid.
[a] F: full slurry; W: washed solids; Ac. A: acetic acid; Sug: sugar stock solution: G: glucose.
[b] Solid fraction as weight percent of insoluble solids.
[c] Glucan content in solids.
[d] Enzyme dosage in FPU per gram of glucan.

$$\frac{d\boldsymbol{X}_i}{dt} = \boldsymbol{f}(\boldsymbol{X}_i, \boldsymbol{\theta}, t) \tag{1}$$

$$\boldsymbol{g}(\boldsymbol{X}_i, \boldsymbol{\theta}) = 0 \tag{2}$$

Estimation of $\boldsymbol{\theta}$, the m-size vector of model parameters, is achieved by weighted least squares, being the least square estimator $\boldsymbol{\theta}^*$ defined as the value minimizing the objective function $J(\boldsymbol{\theta}) = \boldsymbol{\epsilon}^T \boldsymbol{W} \boldsymbol{\epsilon}$. Where $\boldsymbol{\epsilon}$ is the n-size vector formed by the difference between model outputs and experimental values. Entries in the $n \times n$ diagonal matrix of weights $\boldsymbol{W} = diag(w_1, w_2, w_n)$, were chosen to provide an adequate scaling of the measured values in different experimental conditions. A thorough description is given in Supplementary Material S1. The $1-\alpha$ marginal confidence interval ($CI^{1-\alpha}$) for an individual parameter θ_j is calculated as (Seber and Wild, 1989):

$$CI^{1-\alpha} = \theta_j \pm t_{n-p}^{\alpha/2} \sqrt{COV^{j,j}(\theta^*)} \tag{3}$$

where $t_{n-p}^{\alpha/2}$ is the upper $\alpha/2$ quantile for Students t distribution with $n-p$ degrees of freedom and $COV^{j,j}$ are the diagonal elements of the covariance matrix:

$$\boldsymbol{COV}(\theta^*) = \frac{J(\boldsymbol{\theta}^*)}{n-m} \left(\boldsymbol{V}(\theta^*)^T \boldsymbol{W} \boldsymbol{V}(\theta^*) \right)^{-1} \tag{4}$$

with, $\boldsymbol{V}$ the $n \times m$ derivative matrix of model outcomes respect to parameters evaluated at θ^*. Throughout this work the relative (half) confidence interval, defined as $rCI^{1-\alpha} = (t_{n-p}^{\alpha/2} \sqrt{COV^{j,j}(\theta^*)})/\theta_j^*$, is used. Due to the curvature of the non-linear model, the above defined confidence intervals might be misleading (Seber and Wild, 1989). Hence, a method based on the F distribution and in the contours of $J(\boldsymbol{\theta}^*)$ will be used as comparison, the confidence region for joint parameters is described by the following equation (Seber and Wild, 1989):

$$\left\{ \boldsymbol{\theta} : \frac{J(\boldsymbol{\theta}) - J(\boldsymbol{\theta}^*)}{J(\boldsymbol{\theta}^*)} \leqslant \frac{p}{n-p} F_{p,n-p}^{\alpha} \right\} \tag{5}$$

where $F_{p,n-p}^{\alpha}$ is the upper α critical value of the $F_{p,n-p}$ distribution. This region is commonly considered exact since it does not rely on any approximation of the covariance matrix; however, determining the region for more than three parameters is computationally time consuming. Hence only the axis aligned extreme values of the region will be calculated by solving the following optimization problem:

$$\begin{aligned} &\min_{\theta} \quad v\theta_j \\ &\text{s.t.} \quad \tfrac{J(\boldsymbol{\theta}) - J(\boldsymbol{\theta}^*)}{J(\boldsymbol{\theta}^*)} \leqslant \tfrac{p}{n-p} F_{p,n-p}^{\alpha} \end{aligned} \tag{6}$$

If $v = 1$, then the smallest value of θ_j satisfying Eq. (5) is found. On the other hand, if $v = -1$ then the solution of the optimization problem gives the largest value of θ_j that satisfies Eq. (5). The confidence intervals calculated by using Eq. (6) will be denoted as $CI_F^{1-\alpha}$.

2.4. Collinearity analysis and identifiable subsets detection

An algorithmic adaptation of the ideas published in the work of Brun et al., 2001 is used in this work to find an identifiable subset $K \subset \boldsymbol{K}$ of size k of the m parameters in a model, being $\boldsymbol{K}$ the combinations of m parameters taken k at a time. The procedure is thoroughly described in Supplementary Material S1. The Identifiable Subsets Search Algorithm (ISSA) is designed to identify a subset of identifiable parameters among the set of model parameters and proceeds as shown in Fig. 1. The algorithm requires the calculation of a sensitivity index (δ_j^{msqr}) and a collinearity index (γ_K). Sensitivity is defined as follows:

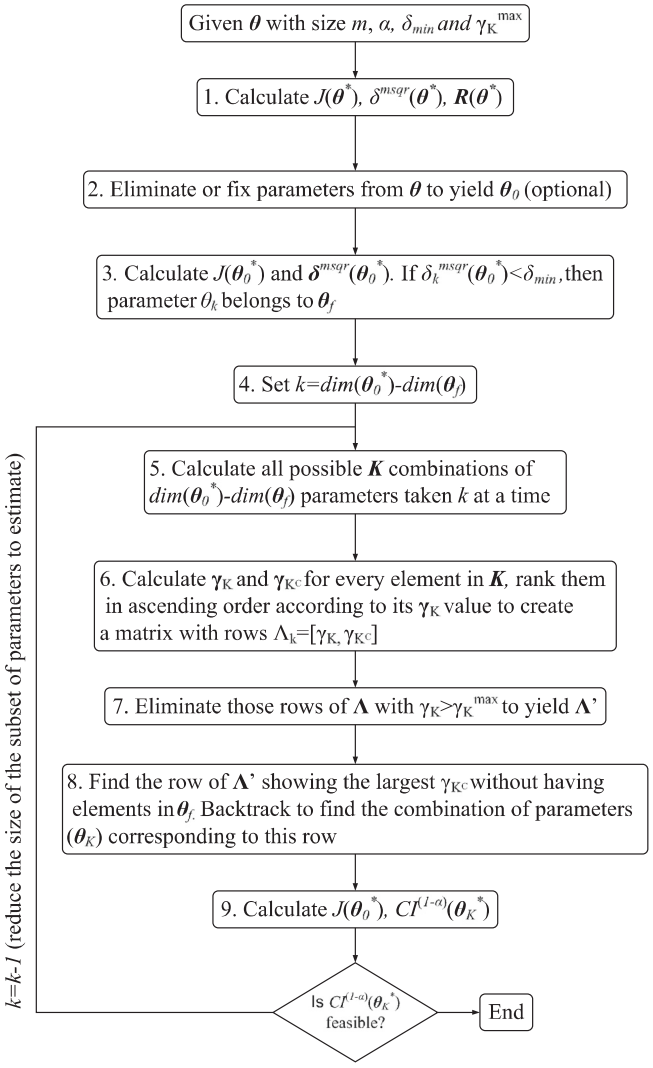

Fig. 1. Identifiable Subsets Search Algorithm (ISSA).

$$\delta_j^{msqr} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} s_{ij}^2} \tag{7}$$

with $s_{i,j} = v_{i,j} \frac{\theta_j}{w_i}, i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. $v_{i,j}$ is an element of $\boldsymbol{V}$, the $n \times m$ derivative matrix, and w_i an element of the scaling matrix $\boldsymbol{W}$. While collinearity index is defined as $\gamma_K = 1/\sqrt{\lambda_k}$, where λ_k is the smallest eigenvalue of $\tilde{\boldsymbol{S}}_K^T \tilde{\boldsymbol{S}}_K$ with $\tilde{\boldsymbol{S}}_K$ being the subset of the K columns of the matrix $\tilde{\boldsymbol{S}}$ with entries $\tilde{\boldsymbol{s}}_j = s_j / \|s_j\|$ and $j = 1, 2, \ldots, m$. According to experiences reported in literature, for a set to be deemed as identifiable, γ_K must be lower than a threshold value of 10–15 (Brun et al., 2002; Sin et al., 2010).

2.5. Comparison of model using the F ratio test

Models with a large number of parameters will always be able to fit the experimental data at least as well (and generally better) as a model with just a few parameters. Hence, a procedure to determine whether it is worthwhile to use more parameters to decrease $J(\boldsymbol{\theta}^*)$ is required to discriminate among models with different number of parameters. In this work, the F ratio test is used for this purpose. Consider two competing models 1 and 2 having p_1 and p_2 parameters respectively, with $p_2 > p_1$, and with weighted

residual sum of squared $J(\theta_1^*)$ and $J(\theta_2^*)$ with $J(\theta_1^*) \geqslant J(\theta_2^*)$. The F ratio can be calculated as:

$$F = \left(\frac{J(\theta_1^*) - J(\theta_2^*)}{p_2 - p_1}\right)\left(\frac{J(\theta_2^*)}{n - p_2}\right)^{-1} \quad (8)$$

Assuming that the F ratio is approximately F-distributed since the sample size is large, and under the null hypothesis that model 2 does not provide a significantly better fit than model 1, the null hypothesis is rejected if the calculated F ratio is greater than the critical value of the F-distribution with $(p_2 - p_1, n - p_2)$ degrees of freedom for some desired false rejection probability (in this work, 0.05 was considered).

2.6. Implementation, software and computational tools

The equality equations in the original DAE (Eqs. (1) and (2)) were removed via algebraic manipulations to yield a system of differential equations (ODE). MATLAB™ was used to perform all calculations in this work. To solve systems of ODEs, the MATLAB™ built-in `ODE45` algorithm based on explicit Runge–Kutta formula was used. The sensitivity matrix $\boldsymbol{V}$ was calculated via forward sensitivity analysis; i.e., by simultaneous solution of the ODEs shown in Eq. (1), and the forward sensitivity equations obtained by applying the chain rule of differentiation over θ to the original ODEs described in Eq. (1).

Parameters minimizing $J(\theta)$ were estimated using OPTI (Currie and Wilson, 2012), a MATLAB™ toolbox and interface for building and solving optimization problems using open source and academic solvers. The problem was formulated as a Dynamic System Parameter Estimation problem in OPTI, and is automatically converted into a standard Nonlinear Least Squares problem and solved using the Intel MKL Trust Region solver. The optimization problem posed in Eq. (6) was solved using the interior point method built-in in MATLAB™ `fmincon` routine.

3. Results and discussion

3.1. Suitability of an existing model for fitting the experimental data

The 2004 NREL model for the enzymatic hydrolysis of dilute acid pretreated corn stover (Kadam et al., 2004) was used as a starting point to fit the experimental data to a suitable kinetic model. A total of 12 parameters were adjusted using the values reported by Sin et al. (2010) as a starting point (θ_0) and bounds for the estimated parameters were set as $0.001\theta_0$ and $1000\theta_0$. The solution of the parameters estimation problem is shown in Table 2 (subset K-S0), as it can be seen, the size of the relative 95% confidence interval for some parameters is a thousand times the parameter value for subset K-S0, large enough to include negative values for the parameters, which is non-realizable in the model. Moreover, some parameters are non identifiable as evidenced by the correlation coefficients in the correlation matrix (see Supplementary Material S2, Table S13). This means that the effect of a change in the value of a parameter over the sum of squared errors can be compensated by changing the value of another parameter. Hence, it is not possible to find a unique estimate for these correlated parameters, but rather one can calculate sets of parameter values showing the same fit to the experimental values. It is possible to reduce the correlation between parameters and, as a consequence, the parameter uncertainty to levels acceptable for engineering purpose by: (i) altering the model structure by eliminating parameters; (ii) increasing the information content of experimental data by using an adequate design of experiments, and (iii) by setting the values of a subset of the parameters to an *a priori* value, thus obtaining a subset of parameters showing less correlation which can be reliably estimated from the available experimental data. Table 3, shows the minimum value of collinearity index for parameters subsets of different number of parameters. These subsets are calculated by taking k out of m parameters at a time to create $\binom{m}{k}$ combinations. Thus, the dimensions of a particular subset $\boldsymbol{K}$ of combinations of parameters are k columns and $\binom{m}{k}$ rows. As it can be seen, for subsets with more than 8 parameters out of 12, the minimum collinearity index is larger than the threshold value of 15 where the subsets can be considered as identifiable. Hence, in order to reduce the confidence interval of the estimated parameters by reducing the collinearity of the parameters subset, only 7 parameters will be estimated while the remaining 5 will be left fixed. However, to do so it is necessary not only to find a subset of parameters with reduced collinearity index (γ_K) but also the collinearity index of the complement of this subset (the fixed parameters), γ_K^C, needs to have a high collinearity index to avoid a conditional or biased estimate (Brun et al., 2001). In order to reduce this potential dependency between both sets, the collinearity index for the set of parameters must be low while the collinearity index for its complement needs to be as high as possible. Table 3 shows the minimum collinearity index (γ_K) for subsets of different number of parameters and the value of the collinearity index for the complement of the

Table 2
Solution of the parameter estimation problem for Kadam et al., 2004 model. Values for γ_K and γ_K^C are shown for subsets of size 6, this is the subsets of size 7 in Table 3 after removing K_{2IG2}. Cells marked with *f* correspond to parameters whose values were fixed to the ones reported by Sin et al., 2010.

Subset	K-S0			K-S1		K-S2		K-S3	
Parameter	θ^*	$rCI^{95\%}$	δ^{msqr}	θ^*	$rCI^{95\%}$	θ^*	$rCI^{95\%}$	θ^*	$rCI^{95\%}$
K_{1IG2}	0.002	513.7	26.75	0.003	0.493	0.007	0.386	*f*	–
K_{2IG2}	91.07	5881	0.020	*f*	–	*f*	–	*f*	–
K_{1IX}	6.118	826.5	0.020	6.040	155.5	5.939	57.16	8.580	9.230
K_{2IX}	0.001	1027	0.020	0.132	0.817	0.0009	108.65	0.001	31.26
K_{3IX}	320.1	43.70	0.360	*f*	–	*f*	–	*f*	–
K_{2IX}	84.86	513.7	26.79	*f*	–	*f*	–	2.112	0.385
k_{1r}	8.229	658.9	0.020	7.511	113.6	7.522	42.05	11.03	7.066
K_{1IG}	0.586	98.90	8.680	5.682	0.266	1.592	0.833	*f*	–
k_{2r}	0.202	94.40	8.460	*f*	–	*f*	–	0.019	0.842
K_{2IG}	93.29	114.7	28.32	*f*	–	*f*	–	*f*	–
k_{3r}	855.5	113.8	28.27	560.4	0.494	*f*	–	*f*	–
K_{3M}	5.195	4.904	24.43	*f*	–	0.545	0.274	0.543	0.272
$J(\theta^*)$	139.7	–	–	215.9	–	140.9	–	142.4	–
γ_K	–	–	–	6.0	–	5.6	–	5.6	–
γ_K^C	–	–	–	63.1	–	6682.6	–	6620.3	–

Table 3
Minimum collinearity index of all possible combinations of parameters subsets of different sizes when using the NREL 2004 model. Complement's collinearity index ($\gamma_{\boldsymbol{K}^c}$) is also shown for the parameters combination having the minimum collinearity index. Additionally, the minimum and maximum values of $\gamma_{\boldsymbol{\Lambda}^c}$ are shown, this is the complement's collinearity index of combinations of parameters having a collinearity index lower than a threshold value of 15.

dim ($\boldsymbol{K}$)[a]	min ($\gamma_{\boldsymbol{K}}$)	$\gamma_{\boldsymbol{K}^c}$	min ($\gamma_{\boldsymbol{\Lambda}^c}$)	max ($\gamma_{\boldsymbol{\Lambda}^c}$)	dim ($\boldsymbol{\Lambda}'$)[a]
(220, 3)	1.559	9947	99.63	60,502	(172, 3)
(495, 4)	2.561	8171	94.45	56,837	(295, 4)
(792, 5)	3.897	8233	71.84	52,264	(316, 5)
(924, 6)	4.802	7631	66.02	7631	(207, 6)
(792, 7)	13.03	6274	59.33	6845	(76, 7)
(495, 8)	14.62	2932	54.01	2932	(12, 8)
(220, 9)	99.63	2.826	–	–	(0, 9)
(66, 10)	1145	2.697	–	–	(0, 10)
(12, 11)	10,055	1.000	–	–	(0, 11)
(1, 12)	63,463	–	–	–	–

[a] Dimension of set (rows: number of parameters combinations in the set, columns: number of parameters to estimate)

particular combination showing the minimum γ_K. Additionally, Table 3 shows the minimum and maximum values for the collinearity index of the complement for combinations of parameters showing γ_K lower than the threshold value of 15, this is γ_K^C for elements in $\boldsymbol{\Lambda}'$. If 7 out of 12 parameters need to be estimated, fixing the value of the remaining 5 to pre-established values, then there are 792 possible combination for the parameters. Among them, there is one combination showing a minimum γ_K of 13.03 and simultaneously a value of 6274 for γ_K^C. In the set of combinations of 7 parameters showing γ_K less than 15 (set $\boldsymbol{\Lambda}'$) there are 76 elements whose values for γ_K^C range from 59.33 to 6845. As it can be seen, the value of γ_K^C for the combination of 7 parameters showing the minimum collinearity index is close to the maximum value of γ_K^C for the elements in $\boldsymbol{\Lambda}'$, however there is no guarantee that this will always be the case. Hence in order to maintain the independence between the set of parameters that needs to be estimated and its complement, the Identifiable Subsets Search Algorithm (ISSA) looks for the row of $\boldsymbol{\Lambda}'$ showing the largest value of γ_K^C. In order to illustrate the importance of selecting a subset of parameters showing a value of γ_K below the identifiability threshold and with high γ_K^C value, three subsets of 7 parameters with different combinations of γ_K and γ_K^C were used as shown in Table 2.

Subsets K-S2 and K-S3 correspond to subsets with near minimal collinearity index and with a high value of collinearity index for its complement while subset K-S1 shows low collinearity index for both the set itself and its complement. Parameter K_{2IG2} presents the lowest sensitivity of all 12 parameters to be estimated. Hence, although all of the subsets with acceptable collinearity index contained K_{2IG2}, this parameter is still non-identifiable, since large changes in their values will only slightly change the models output. For this reason, it is convenient to exclude them from the identified subsets leaving subsets with size 6 (see ISSA in Fig. 1).

Parameters values for the low collinearity index subsets shown in Table 2 were estimated by fixing the remaining 6 parameters to reported values (Sin et al., 2010). Since the optimal values shown in Table 2 (Subset K-S0) do not match the reported ones, one might expect an important lack of fit when 6 of the parameters are fixed to those values. However, only a minor increase in the objective function value $J(\boldsymbol{\theta}^*)$ was found for subsets having a large collinearity index for the complement of the subset (K-S2 and K-S3). On the other hand, when 6 parameters where set to the reported values (Sin et al., 2010) in subset K-S1, the objective function augments 1.5 times indicating an important worsening in fit. As shown in Table 2, the most important difference between subsets K-S1 to K-S3 lies in their (γ_K^C) values. For those sets whose (γ_K^C) value is high, only a small worsening in the value of $J(\boldsymbol{\theta}^*)$ was calculated after fixing 6 parameters, indicating the relevance and convenience of maintaining the independence of both sets as high as possible.

As it can be seen in Table 2 the relative confidence intervals for the estimated parameters diminished substantially, especially for subset K-S3 where the larger relative confidence interval, corresponding to parameter K_{2IX}, was calculated to be 31.26, an important decrease when compared to values in Table 2 (Subset K-S0), however these values still represent an unacceptably large confidence interval for assessing the effect of parameters uncertainty in model output.

3.2. A modified semi-mechanistic model to fit the experimental data

As shown in Fig. 4, NREL's 2004 model (Kadam et al., 2004) does not completely fit the experimental data used in this study, especially for experiences made at 10% w/w solid loading where a lower glucose production is observed when compared to the experimental data. Unlike previous work, where the solid loading is fixed or low values are used (Kadam et al., 2004; Ruiz et al., 2012), in this study a model is presented that represents the enzymatic hydrolysis of lignocellulosic materials at solid loadings ranging between 10% and 25% w/w. Additionally, experiments including inhibitors from the hydrolysate such as glucose, xylose, xylo-oligomers (unquantified) and acetic acid where included by using unwashed solids or washed solids supplemented with inhibitors (Table 1).

Several attempts have been made in previous years to establish the causes of the diminishing conversion observed at high solid loadings even at constant enzyme-to-substrate ratios, see Fig. 2(C). Hodge et al., 2008, found that conversion did not decrease until a threshold solid loading is achieved. Clearly, this threshold value depends on the reaction's time; being close to 25% w/w for typical hydrolysis times of 128–168 h. Authors hypothesize that this decrease in conversion might be caused by mass transfer limitations. However, Kristensen et al. (2009) performed a series of experiment to test whether the observed decrease in conversion was caused by product inhibition, the decrease in water activity as solid loading increases, lack of mixing or a decrease in the fraction of adsorbed enzyme. In this study, it was found that neither of the above mentioned factors could explain the decrease in conversion observed at high solid loadings except the decrease in enzyme adsorption. Specifically, it was found that the ratio between enzymes adsorbed to the solid after 22 h of reaction and the total enzyme content also decreases as solid loading increases (results supported by our experiments, Fig. 2(B)). In the 2004 NREL model, cellulases are assumed to adsorb to cellulose following a Langmuir type equation. Fig. 2(A) shows the protein bound to the substrate using Accellerase 1500 at enzyme loadings between 2.5 and 10 mg per gram of solid substrate for initial solid contents of 5%, 10%, 15% and 20%. Since, the analysis is restricted to the fraction of CBH/EG fraction of Accellerase 1500, it was assumed that this fraction represents 55% of the protein content (Nagendran et al., 2009). Using the adsorption parameters previously reported (Kadam et al., 2004), the calculated bound enzyme is shown in Fig. 2(A) and (B). It can be seen that the Langmuir model used in the 2004 NREL model, not only does not fit the experimentally measured adsorbed enzyme, but fails to predict the observed decrease in the fraction of bound enzyme as the insoluble solid fraction increases.

These observations suggest that the original 2004 NREL model needs to be modified to produce a decrease in the fraction of enzymes adsorbed when solid loading increases (or a decrease in enzyme activity bonded to the solid fraction). In order to do so, two possibilities were investigated: (i) modifying the activity of the adsorbed enzymes by introducing the linear decreasing terms

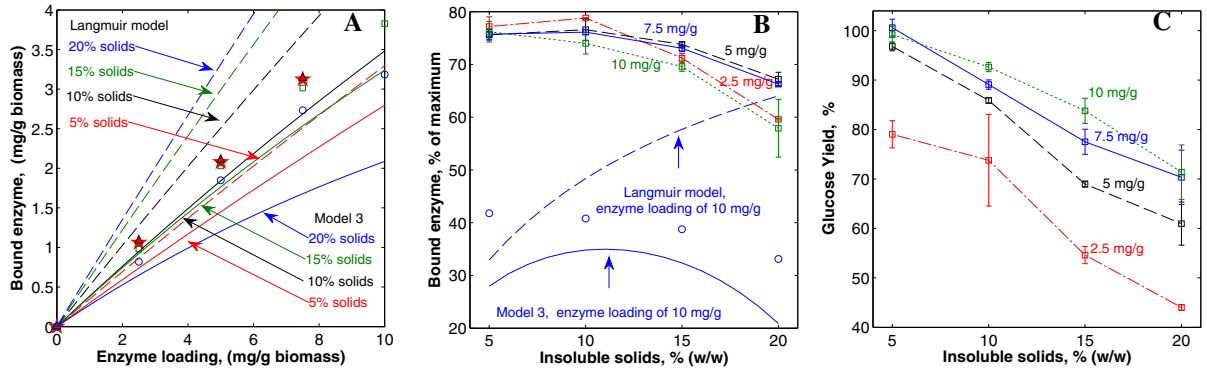

Fig. 2. Bound enzyme as a function of the initial solid using Accellerase 1500 at different enzyme loadings. (A) Bound enzyme versus enzyme loading assuming that CBH/EG correspond to 55% of the protein content in Accellerase 1500 at 5%, 10%, 15% and 20% insoluble solids. Blue circles, green squares, black triangles and red stars correspond to calculated CBH/EG bound protein (55% of total measured bound protein) at 20%, 15%, 10% and 5% solids respectively at different enzyme loadings. Dashed line, prediction of adsorbed CBH/EG enzymes percentage using a Langmuir type equation as in Kadam et al., 2004. Solid line, prediction of adsorbed CBH/EG enzymes percentage using Model 3. (B) Percentage of bound enzyme at several enzyme loadings and different solid fractions. Black squares correspond to measures of bounded enzyme at 10 mg/g, blue circles are obtained assuming that CBH/EG fraction represents 55% of the protein content in Accellerase 1500. (C) Glucose yield at different solid fractions and enzyme loadings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Solution of the parameter estimation problem for Model 2. Estimated parameters and confidence intervals for subsets of size 16, 14, 11 and 8 along with confidence intervals for the calculated parameter values.

Subset	S0 (size 16)			S1 (size 14)		S2 (size 11)		S3 (size 8)				
Parameter	θ^*	$rCI^{95\%}$	δ^{msqr}	θ^*	$rCI^{95\%}$	θ^*	$rCI^{95\%}$	θ^*	$CI^{95\%}$		$CI_F^{95\%}$	
K_{1IG2}	0.055	122.1	6.523	0.043	110.6	0.041	111.1	0.041(f)	–	–	–	–
K_{2IG2}	1.912	7.233	17.19	3.332	4.002	4.264	3.403	4.264(f)	–	–	–	–
K_{1IX}	0.503	121.4	4.823	0.418	110.0	0.397	110.5	0.395	0.184	0.607	0.173	3.472
K_{2IX}	2.147	6.503	3.556	4.858	3.002	5.758	2.468	5.877	3.152	8.601	1.108	15.02
k_{1r}	26.89	121.4	11.70	27.14	110.1	28.638	110.6	28.65	22.42	34.88	18.07	43.07
K_{1IG}	21.18	126.5	0.035	21.26	114.8	22.658	113.0	22.658(f)	–	–	–	–
k_{2r}	0.630	6.187	27.33	0.525	2.969	0.422	2.400	0.422	0.322	0.522	0.239	0.652
K_{2IG}	246.0	13.78	0.794	241.4	6.729	∞	–	∞	–	–	–	–
k_{3r}	310.7	4.605	28.33	128.4	–	128.4(f)	–	128.4(f)	–	–	–	–
K_{3M}	0.024	1620	23.55	0.301	–	0.301(f)	–	0.301(f)	–	–	–	–
K_{3IG}	0.014	1617	23.54	0.612	–	0.612(f)	–	0.612(f)	–	–	–	–
K_{1IA}	1.055	123.8	0.388	2.145	110.6	∞	–	∞	–	–	–	–
K_{2IA}	0.607	6.613	2.136	2.452	2.814	2.892	2.372	2.970	0.938	5.002	0.542	20.68
K_{3IA}	19.52	3067.4	0.003	56.73	201.4	∞	–	∞	–	–	–	–
b_1	6.869	0.043	6.650	6.821	0.040	6.817	0.038	6.822	6.567	7.078	6.097	7.000
b_2	6.183	0.130	65.57	6.377	0.063	6.376	0.059	6.376	6.014	6.738	5.301	6.872
$f_{\beta G}$	–	–	–	0.756	0.248	0.770	0.201	0.766	0.639	0.892	0.528	1.107
$J(\theta^*)$	62.9	–	–	66.53	–	67.50	–	67.49	–	–	–	–

$(1 - b_1 S_o)$ and $(1 - b_2 S_o)$ (Model 2) or, (ii) modifying the amount of enzyme that can be adsorbed to the substrate (E_{iB}) by introducing the linear decreasing terms $(1 - a_{iad} S_o)$ in the original adsorption equations (Model 3). Both modifications create a parabolic shape in the fraction of adsorbed enzyme (activity for Model 2 and percentage of protein in Model 3) as a function of the initial solid loading. This occurs because, when a constant enzyme dosage per gram of substrate is adopted, the total concentration of protein increases with increasing solid loading. As shown in Fig. 2, the proposed modifications yield a closer representation of the trends in the experimental data. Additional modifications to the 2004 NREL model include the incorporation of parameters K_{iIA} accounting for the inhibitory effect of acetic acid found in Hodge et al. (2008) and the elimination of the inhibitory effect of xylose on β-glucosidase activity as this is not supported by published data (Xiao et al., 2004).

$$r_1 = (1 - b_1 S_o) \frac{k_{1r} E_{1B} R_s S}{1 + G_2/K_{1IG2} + G/K_{1IG} + C_5/K_{1IX} + A/K_{1IA}} \quad (9)$$

$$r_2 = (1 - b_2 S_o) \frac{k_{2r}(E_{1B} + \omega E_{2B}) R_s S}{1 + G_2/K_{2IG2} + G/K_{2IG} + C_5/K_{2IX} + A/K_{2IA}} \quad (10)$$

$$r_3 = \frac{f_{\beta G} k_{3r} E_{2F} G_2}{K_{3M}(1 + G/K_{3IG} + A/K_{3IA}) + G_2} \quad (11)$$

$$E_{iB} = (1 - a_{iad} S_o) \frac{E_{imax} K_{iad} E_{iF} S}{1 + K_{iad} E_{iF}} \quad (12)$$

With $R_s = S/S_o$. Moreover, the way the different enzyme fractions are accounted for is different from the 2004 NREL model. In this work, the total protein in the enzymatic preparation (E_T) is distributed in a cellulolitic fraction (E_1: CBH/EG) and a cellobiose hydrolyzing fraction (E_2 : β-glucosidase) according to Eqs. (14) and (15), being f_2 the fraction of β-glucosidase protein in the preparation. This fraction is calculated considering an activity of 16.277 UI/mg for pure β-glucosidase at 50 °C (Chauve et al., 2010) and 16.5 UI/ml in 123 mg/ml of protein in Spezyme CP. The specific β-glucosidase activity was calculated from a reported value (Pryor and Nahar, 2010) considering a constant FPU to CBU ratio and a measured filter paper activity of 59 FPU/ml. Using these values f_2 was calculated as 0.0083.

$$E_{iT} = E_{iF} + E_{iB} \quad (13)$$

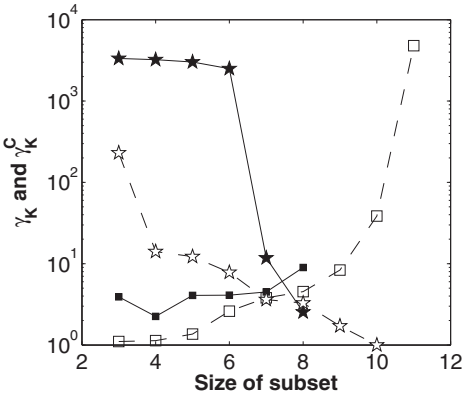

Fig. 3. Collinearity indices for subsets of parameters of different size (and for their complements) for Model 2. Black filled squares, γ_K of sets identified using ISSA; non-filled squares, minimum value of γ_K for a subset of given size. Black filled stars, γ_K^c of sets identified using ISSA; and non-filled stars, γ_K^c values of the sets that show the minimum value of γ_K for the subset of a given size.

$$E_{1T} = (1 - f_2)E_T \tag{14}$$

$$E_{2T} = f_2 E_t + E_{2S} \tag{15}$$

Summarizing, 2004 NREL model is represented by Eqs. (9)–(12) with $b_1 = b_2 = 0, K_{ilA} = \infty, j = 1,2,3, \omega = 1$ and $a_{iad} = 0$ with $i = 1,2$. Models 2 and 3 involve Eqs. (9)–(15) with $\omega = 0$. For Model 2, $a_{iad} = 0, i = 1,2$ while for model 3 $b_i = 0, i = 1,2$. Models 2 and 3 have 16 parameters that need to be estimated. Model 3, including parameters a_{1ad} and a_{2ad}, result in a minimum weighted sum of squared error $J(\theta^*)$ of 83.27 while this fitting indicator for Model 2, including a linear decrease in activity with increasing solid loadings, was calculated as 62.9. Although both models show an improved fit to the experimental data when compared with the NREL 2004 model (with $J(\theta^*)$ equal to 139.7), Model 2 was selected for further analysis due to its improved fit.

3.3. Model 2, enzymatic activity decreases linearly with initial solid loading

Table 4 shows the solution of the parameter estimation problem for Model 2 with 16 estimated parameters (Subset S0). Although Model 2 and the original 2004 NREL model share a subset of parameters, their values differ as shown in Table 2 and 4. This can be explained by the modifications made to the model and because of the high correlation between parameters that prevents finding a unique estimate of the parameters values. At this point it is convenient to evaluate whether the increase in the number of parameters from 12 in NREL model to 16 in model 2 is worthwhile. Using the number of parameters and the calculated $J(\theta)$ values for NREL model and Model 2, the F ratio statistic was calculated as 42.9 while the critical value of the F-distribution with (16 – 12, 156 – 16) degrees of freedom ($F^{0.05}_{4,140}$) is equal to 2.43. Hence, the improvement in fit by Model 2 is statistically significant at a false rejection probability of 0.05. However, from the analysis of the relative 95% confidence intervals, most parameters are far from being deemed as identifiable. In order to reduce the uncertainty in parameters values, a subset of sensitive parameters with low collinearity index needs to be found. Furthermore, parameters k_{3r}, K_{3M} and K_{3IG} are highly correlated as evidenced by their entries

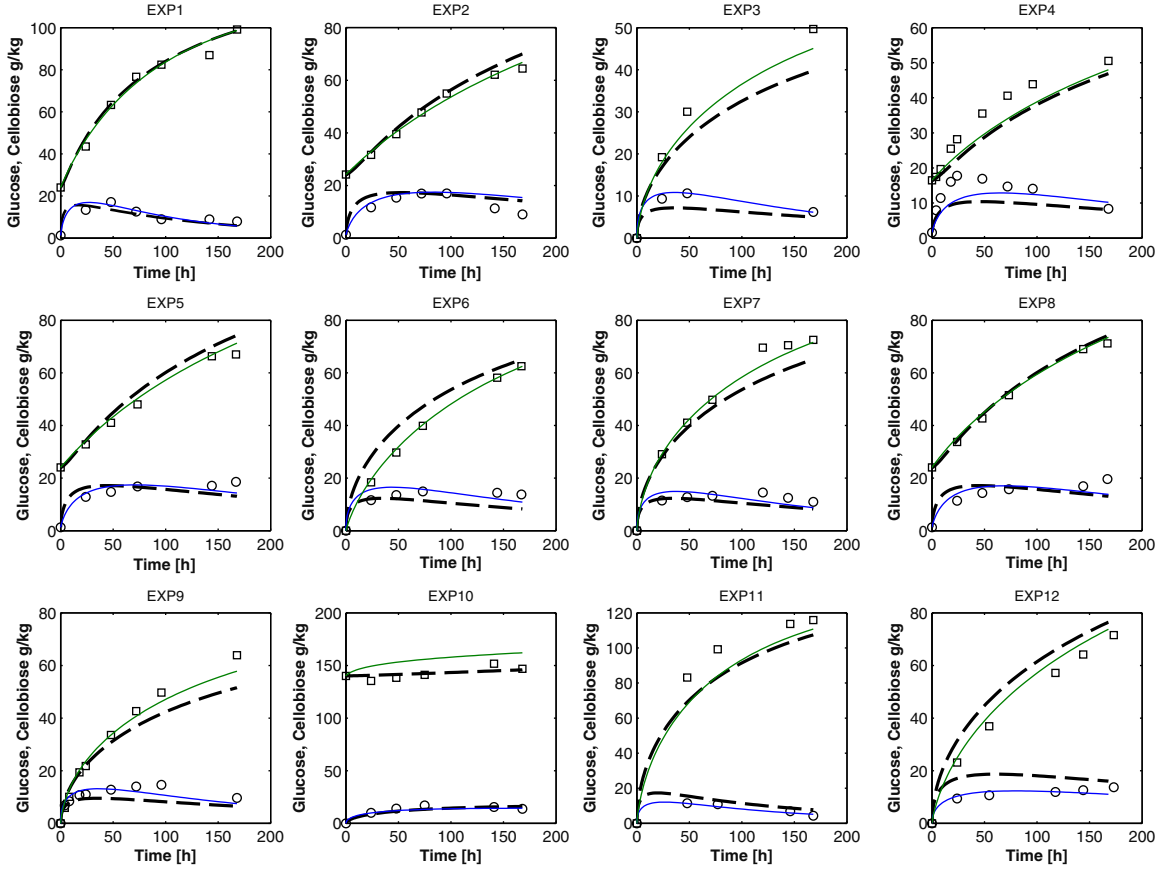

Fig. 4. Models fit to experimental data. Comparison of NREL's 2004 model with subset K-S3 of reestimated parameters (dashed lines) and Model 2 with subset S3 (solid lines). Circles and squares correspond to cellobiose and glucose experimental measurements respectively.

in the correlation matrix (see Supplementary Material S2, Table S14) and hence needs to be fixed to the optimal values shown in Table 4 or to parameters calculated from dedicated experiments. The second alternative was adopted and the values reported by Chauve et al. (2010) were used. The value of k_{3r} used represents the maximum catalytic activity of an *Aspergillus niger* β-glucosidase enzyme (11.9 UI/mg at 45 °C). However, considering the extended reaction times, the maximum catalytic activity might be reduced due to inactivation. In order to account for this phenomena, a new parameter needs to be introduced ($f_{\beta G}$) representing the average fraction of the maximum catalytic activity attained during the reaction course. So far, three parameters are known to be included in the subset of fixed values and their values do not match with the ones presented in Table 4 (Subset S0); hence the value of the remaining parameters needs to be re-estimated. After solving the parameter estimation problem, a value of $J(\boldsymbol{\theta}^*) = 66.5$ was obtained (Table 4, Subset S1). Results suggest that eliminating K_{1IA}, K_{3IA} and K_{2IG} would not importantly worsen the objective function, since their values are already high (when compared to the measured concentrations of acetic acid and glucose) and the objective function sensitivity toward these parameters is low as evidenced by their δ^{msqr} values.

As it can be seen in Table 4, $J(\boldsymbol{\theta}^*)$ value of subset S2 is only 7.3% higher than the best value found when all parameters were used, even after eliminating 3 parameters and fixing the values of k_{3r}, K_{3M} and K_{3IG} to previously reported values. After these changes, a reduction in the relative confidence interval of some parameters can be appreciated (compare subsets S0 and S1 in Table 4). However, some parameters still have relative confidence intervals of hundreds of times its estimated value. In order to further reduce these values, the second part of the algorithm outlined in Fig. 1 was applied, consisting in the identification of a subset of parameters with a collinearity index lower than a pre-established value and with a large collinearity index for the subset of fixed parameters. As shown in Fig. 3, if a threshold value for the collinearity index of 10 is used, then the larger subset of parameters that can be estimated contains 9 elements. However, as evidence by its δ^{msqr} value, the objective function shows a very low sensitivity towards K_{1IG} at $\boldsymbol{\theta}^*$. This means that, even if K_{1IG} is not collinear with other parameters, it is not possible to obtain a narrow uncertainty interval because large variation in the value of K_{1IG} only produces small changes in $J(\boldsymbol{\theta})$. Attending to the preceding argument, K_{1IG} must belong to the set of fixed parameter $\boldsymbol{\theta}_f$. As it can be seen in Fig. 3, the combinations of parameters identified are (generally) different from the combination that shows the minimum γ_K value for each subset. This is caused by the requirement that the parameter combination set must not only have a γ_K value lower than a certain threshold, but also must have the maximum possible γ_K^C value in order to reduce the potential dependency between both sets (K and its complement K^C). In particular, although certain

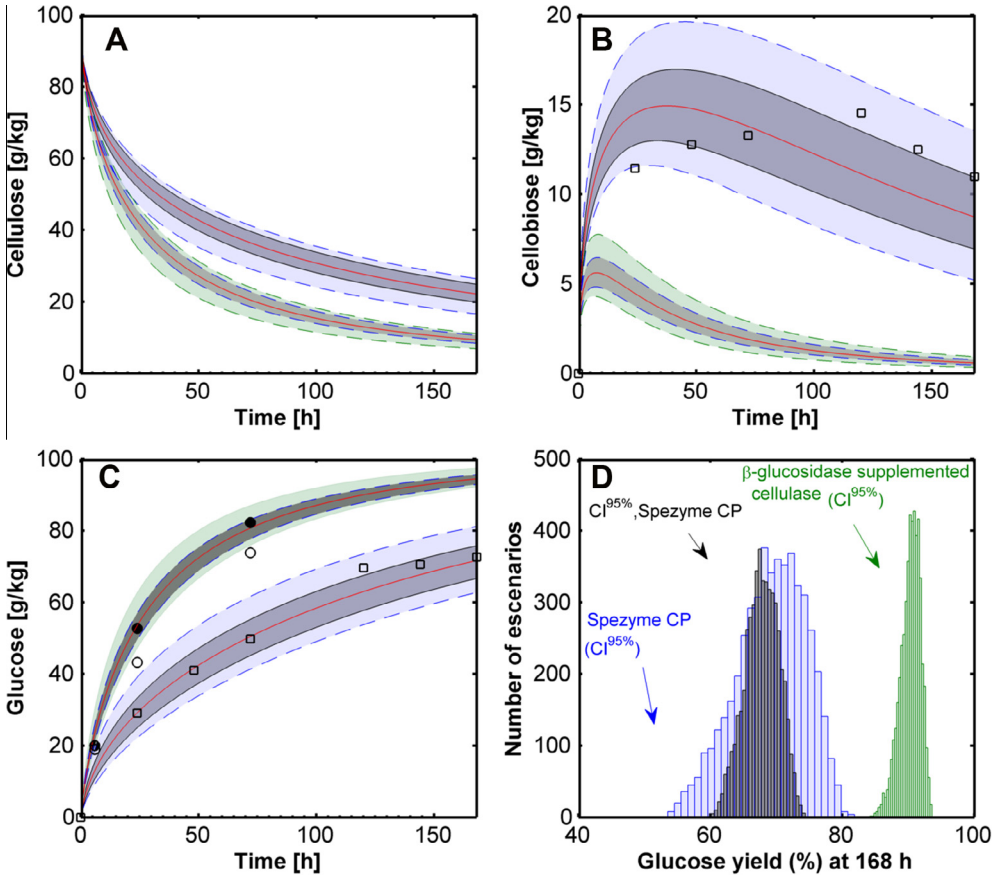

Fig. 5. Uncertainty analysis of the outputs (predictions) of Model 2 using the experimental conditions of EXP7 (washed pulp, 9.6 FPU/g of glucan, 15% solids). Red lines corresponds to outputs calculated using the optimized parameters values ($\boldsymbol{\theta}^*$) in Table 4 for subset S3. Calculated bounds for predictions correspond to 5% and 95% percentiles obtained by performing 5000 Monte Carlo Simulations. Grey lines and shaded regions correspond to the outputs of the base enzyme cocktail (Spezyme CP) and parameters confidence intervals calculated using conventional method, Eq. (3). Dashed lines (blue lines and shaded regions) correspond to scenarios calculated using the confidence intervals obtained using Eq. (6). Green lines and green shaded regions correspond to model outputs calculated using a β-glucosidase supplemented cellulase at a 2.5 CBU/FPU ratio. Filled and non-filled circles correspond to experimental measurements (15% solid) using Accellerase 1500 at 7.5 and 5 mg/g solid respectively (A) cellulose predictions, (B) cellobiose predictions, (C) glucose predictions and (D) histogram of glucose conversion. Squares show the experimental data for EXP7. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

elements of the set of elements of size 9 show values of γ_K below the threshold value, all of these combinations include a low sensitivity parameter (K_{1IG}), hence no combination satisfies all the requirements imposed by ISSA. For subsets of size 8 and lower, ISSA successfully identifies candidate combinations of parameters that satisfy the imposed constraints.

The solution of the parameter estimation problem for a subset with 8 parameters out of 17 (Subset S3) is shown in Table 4 and a comparison of the model fit to experimental data, for both the 2004 NREL model and Model 2, is presented in Fig. 4. Finally, after eliminating the low sensitivity parameter K_{1IG} and those parameters that were correlated in Subset S2, the relative 95% confidence intervals were drastically reduced. In fact, the maximum $rCI^{95\%}$ is equal to 0.68 for K_{2IA}. Considering these promising results, the more reliable method for calculation of confidence intervals of nonlinear models (Eq. (6)) was applied to Subset S3 in order to investigate whether significant differences can be found. As shown in Table 4, none of the estimated parameters have infeasible confidence intervals (i.e. there are no negative values within the confidence interval). Moreover, the results obtained by applying Eq. (6) show that the approximate method for confidence intervals calculation (Eq. (3)) underestimate its values and produces symmetric intervals, a result not supported by the values of $CI_F^{95\%}$ in Table 4.

3.4. Propagation of parameters uncertainty to model outputs

A common situation in parameter estimation of enzymatic processes is to calculate large confidence intervals. Since is not possible to calculate the model output for infeasible parameters (e.g. negative reaction rates), commonly expert judgment is used to derive reasonable bounds (Helton and Davis, 2003) or intervals are truncated to make them feasible. In this work this was unnecessary since the calculated confidence intervals are tight and feasible. Fig. 5 shows the results of propagating the calculated confidence intervals to model outputs using 5000 Monte Carlo scenarios and Latin hypercube sampling, using both calculated confidence intervals for subset S3 in Model 2. It can be seen that, the confidence intervals calculated with Eq. (6) produced a less symmetrical output compared with propagation of conventional confidence intervals. As shown in Fig. 5(C) and (D), the distributions are skewed to highest glucose concentrations and yields.

Fig. 5 also shows the predicted result of increasing β-glucosidase activity as to achieve a ratio of cellulase activity (FPU) to β-glucosidase (CBU) equal to 2.5 (labeled as β-glucosidase supplemented cellulase with green panels and lines in Fig. 5) compared with the base enzymatic preparation used in this work (Genencor Spezyme CP). Increasing β-glucosidase activity greatly increases total glucose conversion by alleviating inhibition by cellobiose; this is in agreement with experimental results utilizing enzyme preparations with high CBU to FPU ratios (McMillan et al., 2011) and with the experimental conversions found in this work when Accellerase 1500 was used to saccharify washed solids at 15% loading. Another interesting observation is that, albeit a uniform distribution for the random values of parameters was used, when the β-glucosidase supplemented cellulase preparation was used in the model, not only the maximum glucose yield increases but the also the distribution of the conversion yields for the random scenarios changes. In fact it is narrow and skewed towards higher yields (Fig. 5(C) and (D)). Hence augmenting the CBU to FPU ratio increases the final glucose yield and the probability of achieving high glucose yield despite the uncertainty in the value of parameters. This result is an example of the importance of model-based process design and analysis (in this case the analysis of the composition of the enzyme preparation). However, it remains unclear if the model will be able to predict the effect of β-glucosidase supplementation at higher FPU to CBU ratios. Since the model already slightly over predicts glucose conversion at a 2.5 CBU/FPU ratio and no provision was taken in model formulation to account for a nonlinear effect in β-glucosidase supplementation, the model must be used only for small increases in the CBU to FPU ratio. Despite this limitation, the model presented in this work can be used to assess new process configurations and to guide the search for improved operating conditions like initial solid loading or enzyme dosage in an economic optimization framework.

4. Conclusion

This work shows that the Langmuir model, commonly used to account for enzyme adsorption onto lignocellulosic substrates, was unable to predict the observed decrease in the fraction of bound enzyme as the insoluble fraction increases. Modifications made to the original model provide a statistically significant improved fit towards a set of experimental data that extends the range of conditions where the model can be used for engineering purposes. Despite a subset of parameters with tight confidence intervals was identified, care should be taken when attaching a physical meaning to the model parameters due to the identifiability issues detected.

Acknowledgements

Financial support granted to F. Scott by CONICYT's scholarship program (Comisión Nacional de Investigación Científica y Tecnológica, Chile) is gratefully acknowledged. Muyang Li was supported in part by a grant from the National Science Foundation – United States (NSF CBET 1336622). This work was funded by Innova Chile Project 208-7320 Technological Consortium Bioenercel S.A.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biortech.2014.11.062.

References

Bansal, P., Hall, M., Realff, M.J., Lee, J.H., Bommarius, A.S., 2009. Modeling cellulase kinetics on lignocellulosic substrates. Biotechnol. Adv. 27, 833–848.
Brun, R., Kühni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter subsets. Water Res. 36, 4113–4127.
Brun, R., Reichert, P., Künsch, H.R., 2001. Practical identifiability analysis of large environmental simulation models. Water Resour. Res. 37, 1015–1030.
Chauve, M., Mathis, H., Huc, D., Casanave, D., Monot, F., Lopes Ferreira, N., 2010. Comparative kinetic analysis of two fungal beta-glucosidases. Biotechnol. Biofuels 3, 3.
Chundawat, S.P.S., Beckham, G.T., Himmel, M.E., Dale, B.E., 2011. Deconstruction of lignocellulosic biomass to fuels and chemicals. Annu. Rev. Chem. Biomol. 2, 121–145.
Currie, J., Wilson, D.I., 2012. OPTI: Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User, in: Foundations of Computer-Aided Process Operations, Savannah, Georgia, USA.
Decker, S.R., Siika-Aho, M., Viikari, L., 2008. Enzymatic depolymerization of plant cell wall hemicelluloses. In: Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy. Blackwell, London, pp. 353–373.
Galbe, M., Zacchi, G., 2012. Pretreatment: the key to efficient utilization of lignocellulosic materials. Biomass Bioenerg. 46, 70–78.
Harris, P.V., Xu, F., Kreel, N.E., Kang, C., Fukuyama, S., 2014. New enzyme insights drive advances in commercial ethanol production. Curr. Opin. Chem. Biol. 19, 162–170.
Helton, J., Davis, F., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Safe. 81, 23–69.
Hodge, D., 2005. Optimization of High Solids Lignocellulosic Biomass Conversion for Ethanol Production (Ph.D. thesis). Colorado State University.
Hodge, D.B., Karim, M.N., Schell, D.J., McMillan, J.D., 2008. Soluble and insoluble solids contributions to high-solids enzymatic hydrolysis of lignocellulose. Bioresour. Technol. 99, 8940–8948.

Hodge, D.B., Karim, M.N., Schell, D.J., McMillan, J.D., 2009. Model-based fed-batch for high-solids enzymatic cellulose hydrolysis. Appl. Biochem. Biotechnol. 152, 88–107.
Kadam, K.L., Rydholm, E.C., McMillan, J.D., 2004. Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass. Biotechnol. Prog. 20, 698–705.
Kreutz, C., Raue, A., Timmer, J., 2012. Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Syst. Biol. 6, 120–129.
Kristensen, J.B., Felby, C., Jrgensen, H., 2009. Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose. Biotechnol. Biofuels 2, 11.
Li, P., Vu, Q.D., 2013. Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst. Biol. 7, 91.
Lynd, L.R., Weimer, P.J., van Zyl, W.H., Pretorius, I.S., 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. 66, 506–577.
McMillan, J.D., Jennings, E.W., Mohagheghi, A., Zuccarello, M., 2011. Comparative performance of precommercial cellulases hydrolyzing pretreated corn stover. Biotechnol. Biofuels 4, 29.
Morales-Rodriguez, R., Meyer, A.S., Gernaey, K.V., Sin, G., 2011. Dynamic model-based evaluation of process configurations for integrated operation of hydrolysis and co-fermentation for bioethanol production from lignocellulose. Bioresour. Technol. 102, 1174–1184.
Nagendran, S., Hallen-Adams, H.E., Paper, J.M., Aslam, N., Walton, J.D., 2009. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus Amanitabisporigera, based on the secretome of *Trichoderma reesei*. Fungal Genet. Biol. 46, 427–435.
Pryor, S.W., Nahar, N., 2010. Deficiency of cellulase activity measurements for enzyme evaluation. Appl. Biochem. Biotechnol. 162, 1737–1750.
Qing, Q., Yang, B., Wyman, C.E., 2010. Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes. Bioresour. Technol. 101, 9624–9630.
Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929.
Rooney, W.C., Biegler, L.T., 2001. Design for model parameter uncertainty using nonlinear confidence regions. AIChE J. 47, 1794–1804.
Ruiz, H., Vicente, A., Teixeira, J., 2012. Kinetic modeling of enzymatic saccharification using wheat straw pretreated under autohydrolysis and organosolv process. Ind. Crop. Prod. 36, 100–107.
Schell, D.J., Farmer, J., Newman, M., Mcmillan, J.D., 2003. Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor: investigation of yields, kinetics, and enzymatic digestibilities of solids. Appl. Biochem. Biotechnol. 105, 69–86.
Scott, F., Conejeros, R., Aroca, G., 2013. Attainable region analysis for continuous production of second generation bioethanol. Biotechnol. Biofuels 6, 171.
Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. Wiley Series in Probability and Statistics. John Wiley & Sons Inc, Hoboken, NJ, USA.
Sin, G., Meyer, A.S., Gernaey, K.V., 2010. Assessing reliability of cellulose hydrolysis models to support biofuel process design – identifiability and uncertainty analysis. Comput. Chem. Eng. 34, 1385–1392.
Tejirian, A., Xu, F., 2011. Inhibition of enzymatic cellulolysis by phenolic compounds. Enzyme Microb. Technol. 48, 239–247.
Teugjas, H., Väljamäe, P., 2013. Product inhibition of cellulases studied with 14C-labeled cellulose substrates. Biotechnol. Biofuels 6, 104.
Wang, W., Kang, L., Wei, H., Arora, R., Lee, Y.Y., 2011. Study on the decreased sugar yield in enzymatic hydrolysis of cellulosic substrate at high solid loading. Appl. Biochem. Biotechnol. 164, 1139–1149.
Williams, D.L., Hodge, D.B., 2014. Impacts of delignification and hot water pretreatment on the water induced cell wall swelling behavior of grasses and its relation to cellulolytic enzyme hydrolysis and binding. Cellulose 21, 221–235.
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. Appl. Biochem. Biotechnol. 115, 1115–1126.