

Dynamic Modeling and Validation of a Biomass Hydrothermal Pretreatment Process—A Demonstration Scale Study

Remus Mihail Prunescu and Mogens Blanke

Dept. of Electrical Engineering, Automation and Control Group, Technical University of Denmark, Elektrovej, 2800, Kgs. Lyngby, Denmark

Jon Geest Jakobsen

Dept. of Process Control and Optimization, DONG Energy Thermal Power A/S, Nesa Allé, 1, 2820, Gentofte, Denmark

Gürkan Sin

Dept. of Chemical and Biochemical Engineering, CAPEC-PROCESS, Technical University of Denmark, Søtofts Plads, 2800, Kgs. Lyngby, Denmark

DOI 10.1002/aic.14954

Published online August 19, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Hydrothermal pretreatment of lignocellulosic biomass is a cost effective technology for second generation biorefineries. The process occurs in large horizontal and pressurized thermal reactors where the biomatrix is opened under the action of steam pressure and temperature to expose cellulose for the enzymatic hydrolysis process. Several by-products are also formed, which disturb and act as inhibitors downstream. The objective of this study is to formulate and validate a large scale hydrothermal pretreatment dynamic model based on mass and energy balances, together with a complex conversion mechanism and kinetics. The study includes a comprehensive sensitivity and uncertainty analysis, with parameter estimation from real-data in the 178–185°C range. To highlight the application utility of the model, a state estimator for biomass composition is developed. The predictions capture well the dynamic trends of the process, outlining the value of the model for simulation, control design, and optimization for full-scale applications. © 2015 American Institute of Chemical Engineers AIChE J, 61: 4235–4250, 2015

Keywords: computer simulations (MC and MD), mathematical modeling, statistical analysis, reaction kinetics

Introduction

Lignocellulosic biomass (wheat straw, corn stover, bagasse etc.) consists of cellulose, hemicellulose, lignin, ash, and a negligible amount of residues.¹ The cellulosic fibers contain glucose units, which are necessary for biofuel production, but layers of hemicellulose and lignin make cellulose hardly accessible. The goal of the pretreatment process is to relocate lignin, and to partially hydrolyze the hemicellulose, which opens the biomatrix for cellulose such that enzymes can easily access it in the enzymatic hydrolysis process.²

Chiaromonti et al.³ review various methods of pretreatment, for example, autohydrolysis, steam explosion, acid hydrolysis, alkaline hydrolysis, and many others. Studies show that hydrothermal pretreatment with steam excels in cost effectiveness and, therefore, has been commercialized in large scale second generation biorefineries. Integrating the biorefinery with a power plant following the Integration Biomass Utilization System (IBUS) also contributes to reducing costs.⁴ The pretreatment process partially depolymerizes hemicellulose creating several degradation and by-products, that is, organic acids, xylooligomers, xylose, and inhibitors such as furfural and 5-HMF, which

impact the downstream processes. The organic acids, that is, acetic, succinic, and lactic acid influence the pH of pretreated fibers and become an issue in the enzymatic hydrolysis process.⁵ Xylooligomers and xylose act as strong inhibitors of cellulose hydrolysis by enzymes⁶ while furfural and 5-HMF inhibit the fermentation process.⁷ Also carbohydrates react with degradation products to create spherical droplets with lignin like structure named pseudo-lignin, which can degrade the enzymatic activity.⁸ Experimental studies show that reactor temperature and retention time relate to biomass conversion.⁹

Lavarack et al.¹⁰ formulate a mechanistic acid hydrolysis model capable of predicting cellulose, xylan, and furfural concentrations but no one evaluated the model for steam pretreatment, and at a large scale. The Lavarack model has been used in later studies for model-based optimization of bioprocesses under uncertainty¹¹ and biorefinery configurations.¹² Overend et al.¹³ present an empirical modeling alternative known as the severity factor, which Petersen et al.⁹ validated in laboratory experiments for xylan recovery and furfural formation. These models are incomplete because: (1) they miss production of important by-products such as organic acids, xylooligomers, and pseudo-lignin and (2) assume a uniform thermal environment, which is not the case in a full scale reactor.¹⁴ Our study extends the existing Lavarack et al.¹⁰ model to demonstration scale using computational fluid dynamics techniques, taking into account temperature variations in a large scale thermal

Correspondence concerning this article should be addressed to G. Sin at gsin@kt.dtu.dk.

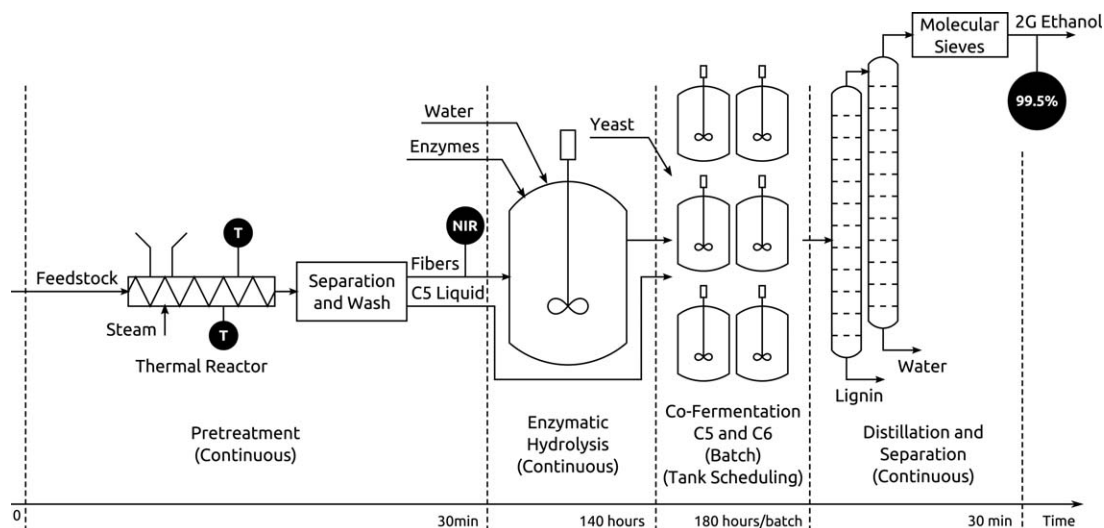


Figure 1. Process flow diagram for a second generation biorefinery: pretreatment, enzymatic hydrolysis, C5 and C6 cofermentation, and purification.

The samples for NIR analysis are extracted after the pretreatment process every 10 min. The thermal reactor is monitored with two temperature sensors: one at the top and another one at the bottom of the reactor.

reactor, and production of the above enumerated by-products. The model is then calibrated and validated against real data that were logged throughout several hours of operation at a demonstration scale facility.

This study also assesses the model reliability through a comprehensive sensitivity and uncertainty analysis. The sensitivity analysis quantifies the importance of each model parameter and creates an identifiable subset of parameters that is used for parameter estimation following the methodology from Sin et al.¹⁵ and Prunescu and Sin.⁵

Samples of fibers were collected after the pretreatment process and were analyzed with near infrared (NIR) instruments to determine their composition. The model parameters are estimated using the NIR readings. The uncertainty analysis determines a confidence interval for model predictions and is carried with respect to both model and feed parameters following the method from Sin et al.¹⁵ As a global sensitivity measure, the standardized regression coefficients (SRC) are computed in order to identify the model parameters responsible for most of the variations in model predictions.¹⁶ A residual analysis follows to identify how much of the signal is represented by the model. The study ends with a model application as a state estimator using a static Kalman filter. The state estimator not only filters the NIR measurements but also predicts by-products formation such as C5 sugars and inhibitors.

This paper has the following structure: the *Methods* section gives an overview of a state of the art demonstration scale biorefinery from where the real measurements were collected; a *Model Development* section follows, which formulates the mathematical model for the pretreatment process along with the model analysis methodology; the *Results and Discussion* section presents the model analysis and validation results, and the model application as a state estimator; the study ends with a *Conclusions* section, which summarizes all findings.

Methods

Biorefinery experimental setup

Figure 1 shows an overview of a demonstration scale second generation bioethanol plant. Biomass is first pretreated in

a continuous thermal reactor where the temperature can be set within 160–210°C, and the retention time can be adjusted from 6 to 20 min. Following the IBUS concept, a nearby power plant supplies the biorefinery with steam for pressurizing the thermal reactor, thus reaching the necessary reaction temperatures. The steam supply pipe provides saturated steam at 18×10^5 Pa. After the thermal reactor, the pretreated slurry is first washed and then the liquid part is separated from fibers in a screw press. The liquid part is rich in C5 sugars, that is, xylooligomers, xylose, and arabinose, as a result of partially hydrolyzing the hemicellulose in the thermal reactor.

The cellulosic fibers are transported on a conveyor belt to the enzymatic hydrolysis tank. Enzymes are added in the liquefaction tank where cellulose and the remaining xylan are converted to glucose and xylose. The enzymatic hydrolysis process has been detailed and analyzed in Prunescu and Sin.⁵ The C5 and C6 sugars are then cofermented for ethanol production in scheduled batch reactors with genetically modified organisms for enhancing bioethanol production.

The purification and separation phase contains two distillation columns and molecular sieves. Lignin is separated in the first distillation column, while ethanol is purified to 99.5% in the second column and in the molecular sieves. The recovered lignin is transported to an evaporation plant and solidified as biopellets, which are sent to the nearby power plant for burning.

There is a timeline indicator at the bottom of Figure 1 showing the retention time for each section of the biorefinery. The pretreatment process and distillation are the fastest processes with a duration of maximum half an hour, while the enzymatic hydrolysis and fermentation can last 5–7 days each.

The demonstration scale facility has a processing capacity of 4000 kg/h of biomass.¹⁷ Samples of pretreated fibers were extracted after the pretreatment process every 10 min for a total duration of 15 h. The samples were then analyzed with NIR instruments to determine their composition with respect to cellulose, xylan, lignin, acetic acid, and furfural. The thermal reactor was monitored with two temperature sensors, one placed at the top of the tank for measuring the temperature in the steam layer, and another one placed at the bottom of the reactor to measure the biomass temperature.

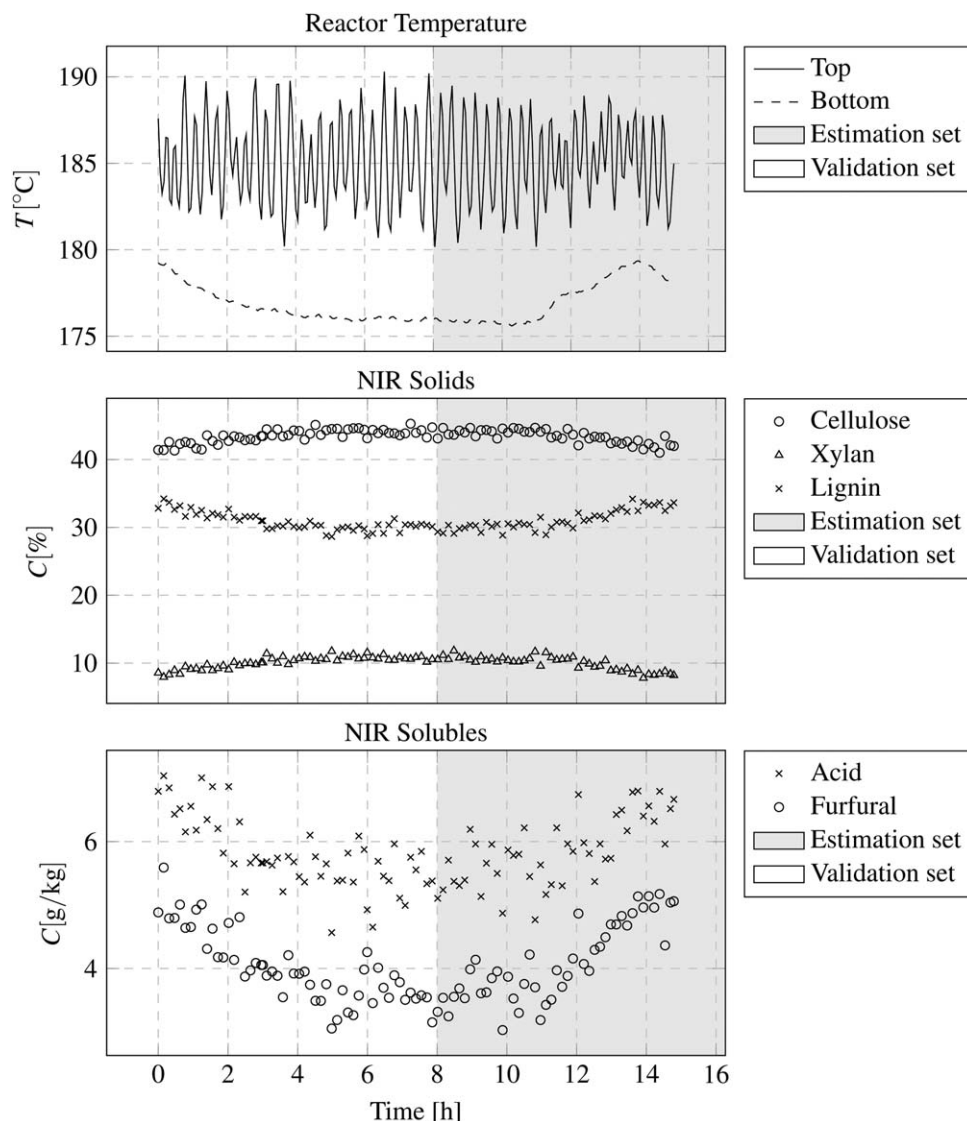


Figure 2. The raw dataset.

The top plot shows the reactor temperatures measured by the top and bottom sensors. NIR offers information on the solid and soluble content of pretreated fibers. The whole dataset is split into estimation and validation subsets.

Dataset

Figure 2 shows the raw temperature and NIR measurements collected throughout the 15 h of operation. The dataset is split in two subsets: one is used for estimation purposes, while the other for validation. The estimation dataset is identified in all figures with a grey background. The top reactor temperature oscillates within 180–190 °C due to the pressurization unit, which takes the biomass from atmospheric to reactor pressure with recycled steam from the reactor. The whole process is thoroughly explained in Prunescu et al.¹⁴ There is a temperature difference between the top and bottom part of the reactor due to the insulation properties of the biomass, and also because vertical mixing in the reactor was not possible due to a low horizontal pushing speed.

Model analysis methodology

The model analysis follows the methodology from Table 1: 1. The first step is to calibrate the model considering the entire set of parameters. Ideally, this system identification exercise should give the set of parameters that has the small-

est sum of squared errors between model predicted output and actual measurements. In the present case, this is a non-linear least squares problem and local minima can be obstacles. The parameter estimation may be solved in Matlab with the `fminsearch` function, coupled with a cost function that calculates the prediction error, if the initial parameter guess is sufficiently close to a global minimum.

2. The second step is to investigate which parameters of the model could be determined given the input and the model structure.¹⁸ This selection is achieved through assessment of sensitivity of the partial derivatives of the cost function with respect to each model parameter. After computing the sensitivities, δ^{msqr} , all parameters are ranked with respect to their value of δ^{msqr} . Parameters that have low sensitivity are more uncertain than those with high sensitivity and would not contribute to model accuracy. Therefore, a relevant subset of parameters is selected based on δ^{msqr} being higher than a threshold.

3. In the third step the reduced set of parameters is identified using the NIR measurements from the demonstration scale plant. The correlation matrix and standard deviations of the estimates are also computed.

Table 1. Methodology for Sensitivity and Uncertainty Analysis

#	Step	Description	Output
1	Model initialization	Initialization of all model parameters to obtain a good working model fit;	θ_0
2	Sensitivity analysis	List of significant parameters; Find an identifiable parameter subset.	δ^{msqr}
3	Parameter estimation	Identify parameter subset; Correlation matrix;	θ_{R_0} θ_R R_θ
4	Uncertainty analysis	Confidence interval for parameters. Calculate prediction uncertainty of the model; Sensitivity analysis with SRC.	σ 5th–95th percentile β
5	Residual analysis	Run simulation with the estimated parameters and using the entire dataset Check probability distribution of model estimation errors or residuals Compute the autocorrelation function	

The output from step $k - 1$ is the input to step k .

4. This step quantifies the prediction uncertainty. Having the covariance matrix and standard deviations from the previous step allows Latin Hypercube Sampling (LHS)¹⁹ with correlation control. The feed parameters are another source of uncertainty and are included in this analysis. N Monte Carlo simulations are then run with sampled values and the 5th–95th percentiles of the model predictions are found. A global sensitivity analysis follows by fitting a linear model from parameters to model predictions and the standardized regression coefficients (SRC) are computed to identify which parameters are the most important for explaining the output uncertainty.

5. The model estimation error or the residuals are analyzed in this step. A simulation is run with the estimated parameters using the entire set of data (not only the estimation set). The residuals distribution and autocorrelation are calculated in order to assess the quality of model predictions. A good model captures most of the signal in measurements and is characterized by residuals being Gaussian with uncorrelated increments.

Model Development

The mathematical model consists of mass and energy balances for the pretreatment process. In large scale plants, the most common continuous thermal reactor is a long tank with cylindrical shape. This study uses simplified computational fluid dynamics tools for modeling the composition and temperature profiles.

Mass balance

The thermal reactor has a continuous operation and the mass balance is established as the accumulation of mass per unit of time equals the difference between inflow and outflow rates

$$\frac{dM}{dt} = F_{\text{in}} - F_{\text{out}} \quad (1)$$

where M is the total mass of biomass inside the reactor, F_{in} is the inflow rate of pressurized biomass, and F_{out} is the outflow rate of pretreated biomass.

Composition balance

Pretreated fibers contain the following species: cellulose, xylan, arabinan, lignin, acetyl groups, ash, glucose, xylooligomers, xylose, organic acids, furfural, 5-HMF, and other components in negligible amounts. The change of species concentration with respect to time is a combination of convection and diffusion effects plus production and consumption terms, which is modeled with the generic Convection-Diffusion-Reaction equation²⁰

$$\frac{\partial C}{\partial t} = -\nabla(vC) + \nabla(D\nabla C) + R \quad (2)$$

C is the species concentration vector, v is the horizontal transportation speed, ∇ is the gradient operator, D is the diffusion coefficient, and R is the reaction rate vector. Most of the particles are in solid state and they do not diffuse. Solubles, that is, xylooligomers, xylose, arabinose, organic acids, furfural, and 5-HMF, diffuse but the snail pushing the biomass along the reactor separates the tank in chambers due to its pallets, which leads to little exchange of matter between these chambers. If the reactor is split in the same number as the chambers created by the rotational snail pallets then the diffusion coefficient can be dropped

$$\frac{\partial C}{\partial t} = -\nabla(vC) + R \quad (3)$$

The horizontal speed v is set to a constant value such that to meet a certain retention time

$$v = \frac{L_r}{t_r} \quad (4)$$

where L_r is the reactor length and t_r is the retention time. For example, a thermal reactor of 12 m with a retention time set to 15 min has a horizontal speed of 0.013 m/s, which is a low speed not capable of agitating the biomass vertically.

As the reactor tank is longer on one axis, the partial derivatives with respect to the other axis could be dropped. It is assumed that the biomass composition changes only along the reactor and remains uniform in the other directions

$$\frac{\partial C}{\partial t} = -v \frac{dC}{dz} + R \quad (5)$$

where z is the axis oriented along the reactor from inlet to outlet. Due to the low horizontal speed, the vertical uniformity assumption stands valid only if the biomass level in the reactor is low. At high biomass levels, the temperature gradient causes different pretreatment layers, and the reactor needs to be split on the vertical axis too.¹⁴ The data in this study were collected at low reactor levels and vertical uniformity can be assumed but differences between the steam and biomass layer temperatures still exist.

Figure 3 illustrates the concept of splitting a reactor tank of length L_r into N smaller reactors, or cells, of width δ_z . Equation 5 is then discretized in space using an Upwind Difference Scheme, which is a more stable technique for moving environments²¹

$$\frac{dC_k}{dt} = \frac{u_z}{\delta_z} (C_{k-1} - C_k) + R_k \quad (6)$$

C_k is the composition vector in central cell k , C_{k-1} is the composition vector from the western neighbor, and R_k is the

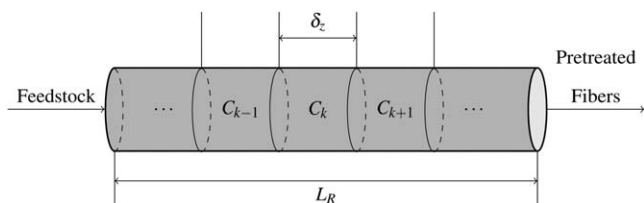


Figure 3. Computational fluid dynamics methodology for the thermal reactor.

The reactor tank is split into N cells of equal width δ_z . A cell C_k has two neighbors, that is, C_{k-1} to the west, and C_{k+1} to the east. Biomass is transported from left to right. The length of the reactor is L_r .

reaction rate vector from current cell k . Movement from west to east (left to right) is assumed. The composition vector will be detailed in the next section.

Mechanistic Modeling for Hydrothermal Mediated Pretreatment. Raw biomass consists of cellulose, hemicellulose (arabinan and xylan), lignin, acetyl groups, ash, water, and other components in negligible amounts. Table 2 shows the composition fractions of different types of lignocellulosic biomass.

Figure 4 illustrates the reaction paths that occur in the thermal reactor. There is little change to cellulose, which is recovered almost entirely in the fiber fraction after the pretreatment process under optimal conditions.⁹ However, a small fraction of cellulose could be hydrolyzed to glucose, which is further degraded to 5-HMF. Arabinan is completely hydrolyzed to arabinose.² The xylan to xylose path has the xylooligomers intermediate product, which is important to predict because it acts as a very strong inhibitor in the enzymatic hydrolysis process along with xylose.⁶ Pentoses, that is, arabinose and xylose, are further decomposed into furfural and other degradation products. Carbohydrates, that is, xylooligomers, xylose, arabinose, and glucose, react with furfural and 5-HMF to create spherical droplets with a lignin like structure, also known as pseudo-lignin.⁸ Furfural and 5-HMF production is important to monitor due to its inhibitory effects on fermentation,⁷ while organic acids influence the pH of the enzymatic hydrolysis process.²⁵ Pseudo-lignin can degrade the enzymatic activity.⁸

Reaction rates are modeled as first order Arrhenius temperature dependent equations. Glucose production rate is defined as

$$r_G = k_G \exp\left(-\frac{E_G}{R_g \cdot T_K}\right) C_{C_s} \quad (7)$$

where r_G is the glucose production rate, k_G is the reaction constant, E_G is the activation energy, R_g is the universal gas constant, T_K is the environment temperature in Kelvin degrees, and C_{C_s} is the concentration of cellulose (solid), or the substrate.

Glucose degradation rate to 5-HMF or r_H is modeled similarly with concentration of glucose C_G as substrate

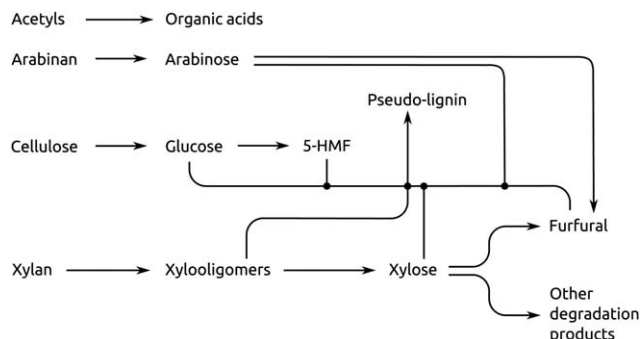


Figure 4. Hydrothermal pretreatment conversion mechanism.

$$r_H = k_H \exp\left(-\frac{E_H}{R_g \cdot T_K}\right) C_G \quad (8)$$

Arabinan hydrolysis forms arabinose following the below rate expression

$$r_A = k_A \exp\left(-\frac{E_A}{R_g \cdot T_K}\right) C_{A_s} \quad (9)$$

where C_{A_s} is the concentration of arabinan (solid).

Xylan degradation produces xylooligomers with rate r_{X_o} , which further decompose to xylose with rate r_X defined as follows

$$r_{X_o} = k_{X_o} \exp\left(-\frac{E_{X_o}}{R_g \cdot T_K}\right) C_{X_s} \quad (10)$$

$$r_X = k_X \exp\left(-\frac{E_X}{R_g \cdot T_K}\right) C_{X_o} \quad (11)$$

where C_{X_s} and C_{X_o} are concentrations of xylan (solid), and xylooligomers (liquid).

Furfural is produced from pentoses, that is, from both arabinose and xylose as follows

$$r_F = k_F \exp\left(-\frac{E_F}{R_g \cdot T_K}\right) (C_X + C_A) \quad (12)$$

where C_X and C_A are concentrations of xylose (liquid) and arabinose (liquid). The amount of furfural produced from xylose is denoted as r_{F_X} , and the one from arabinose as r_{F_A} .

Furfural and 5-HMF bond to carbohydrates to create pseudo-lignin,⁸ which is modeled in this study in the following way

$$r_L = k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) (C_{X_o} + C_X + C_A + C_G) (C_F + C_H) \quad (13)$$

where C_F is the concentration of furfural (liquid) and C_H is the concentration of 5-HMF (liquid). Equation 13 can be split into

Table 2. Composition of Different Raw Biomass Types in Percentage of Dry Matter

Biomass	Cellulose	Xylan	Arabinan	Lignin	Ash	Acetyls	Other	Refs.
Bagasse	39.0	21.8	1.8	24.8	3.9	3.3	5.4	Templeton et al. ²²
Wheat straw	39.8	24.5	2.8	22.6	4.2	n/m	6.1	Kristensen et al. ²
Wheat straw	42.1	23.4	1.9	21.6	2.1	n/m	8.9	Hansen and Kristensen ²³
Miscanthus	38.2	20.9	1.5	26.4	2.6	4.1	6.3	Chiaramonti et al. ³
Corn stover	35.0	18.5	n/m	13.9	3.8	3.2	25.6	Weiss et al. ²⁴
Corn stover	34.0	19.2	2.5	12.3	4.7	2.9	24.4	Templeton et al. ²²

Some components were not measured (n/m).

$$\begin{aligned}
r_{L_{Xo}} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_{Xo} (C_F + C_H) \\
r_{L_X} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_X (C_F + C_H) \\
r_{L_A} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_A (C_F + C_H) \\
r_{L_G} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) C_G (C_F + C_H)
\end{aligned} \quad (14)$$

which denote pseudo-lignin produced from xylooligomers, xylose, arabinose, and glucose when they bind to both furfural and 5-HMF. Then Equation 13 becomes

$$r_L = r_{L_{Xo}} + r_{L_X} + r_{L_A} + r_{L_G} \quad (15)$$

Equation 13 can also be split into

$$\begin{aligned}
r_{L_F} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) (C_{Xo} + C_X + C_A + C_G) C_F \\
r_{L_H} &= k_L \exp\left(-\frac{E_L}{R_g \cdot T_K}\right) (C_{Xo} + C_X + C_A + C_G) C_H
\end{aligned} \quad (16)$$

where r_{L_F} is the production rate of pseudo-lignin with furfural participation while in r_{L_H} 5-HMF participates.

Acetyls are released during hemicellulose hydrolysis and lead to organic acids formation with rate r_{Ac}

$$r_{Ac} = k_{Ac} \exp\left(-\frac{E_{Ac}}{R_g \cdot T_K}\right) C_{AcS} \quad (17)$$

where C_{AcS} is the concentration of acetyl groups in the hemicellulose (solid).

Xylose, glucose and arabinose degrade further into negligible amounts of other components with rates r_{O_X} , r_{O_G} , and r_{O_A} , which are set to 0 in this study.

The composition vector C_k from Equation 6 contains all components from the mechanistic scheme. The reaction rates from this section are put into a reaction rates vector R_k . C_k , and R_k are shown next in vector form

$$C_k = \begin{bmatrix} C_{Cs} \\ C_{Xs} \\ C_{As} \\ C_{Ls} \\ C_{AcS} \\ C_G \\ C_{Xo} \\ C_X \\ C_A \\ C_{Ac} \\ C_F \\ C_H \\ C_W \\ C_O \end{bmatrix} \quad R_k = \begin{bmatrix} -r_G \\ -r_{Xo} \\ -r_A \\ r_L \\ -r_{Ac} \\ r_G - r_{O_G} - (1-\alpha)r_{L_G} \\ r_{Xo} - r_X - (1-\alpha)r_{L_{Xo}} \\ r_X - r_{F_X} - r_{O_X} - (1-\alpha)r_{L_X} \\ r_A - r_{O_A} - r_{F_A} - (1-\alpha)r_{L_A} \\ r_{Ac} \\ r_F - \alpha r_{L_F} \\ r_H - \alpha r_{L_H} \\ 0 \\ r_{O_X} + r_{O_G} + r_{O_A} \end{bmatrix} \quad (18)$$

where C_W is the water content, and α is a stoichiometric parameter for furfural and 5-HMF participation in pseudo-

lignin formation. To close the mass balance, the sum of all elements in R_k has to be 0, and the sum of all elements in vector C_k is 1000 g/kg at any time t

$$\sum R_k = 0 \text{ g/(kgs)} \quad \sum C_k = 1000 \text{ g/kg} \quad (19)$$

Energy Balance. The steam layer energy balance together with a temperature controller for the thermal reactor have been formulated in Prunescu et al.²⁶ The energy balance for the biomass layer has been studied in Prunescu et al.¹⁴ and is simplified in this article by a distributed parameters model on one axis, which is discretized along the reactor, or the z axis

$$\frac{dh}{dt} = -v \frac{\partial h}{\partial z} + Q_k \Rightarrow \frac{dh_k}{dt} = \frac{v}{\delta z} (h_{k-1} - h_k) + Q_k \quad (20)$$

v is the horizontal speed, h_k is the biomass enthalpy in cell k and h_{k-1} is the enthalpy in the western neighbor. Q_k represents the transfer of energy from steam to biomass in cell k . As part of this coupled partial differential equation (PDE) system, the bottom and top temperature measurements (sensors) are used to construct the boundary conditions. The PDE model is then solved for obtaining the reactor temperature gradient, which is then utilized as the temperature of the reaction in calculating the reaction rates vector R_k .

Steam is injected through the bottom of the reactor and gets in direct contact with the biomass. The steam injection heat transfer is lumped into the boundary conditions of Equation 20. The biomass is assumed to heat till the steam temperature near the inlet of the reactor, and used as a western boundary condition. The heat-transfer rate is computed as

$$Q_0 = F_{in} \frac{h_f - h_0}{h_{s_0} - h_0} \quad (21)$$

where Q_0 is the heat-transfer rate from the boundary conditions, F_{in} is the flow rate of biomass, h_f is the final enthalpy of heated biomass, h_0 is the initial biomass enthalpy, and h_{s_0} is the fresh steam enthalpy. h_0 is obtained by measuring the temperature of the biomass entering the reactor T_0

$$h_0 = c_b (T_0 - T_r) \quad (22)$$

where c_b is the specific heat of biomass, and $T_r = 0$ is the reference temperature. Biomass is assumed to have a constant specific heat of approximately $c_b = 3.8$ kJ/kg, a value slightly lower than water (4.18 kJ/kg) since the pretreatment slurry is a mix of condensed water and biomass.

h_{s_0} is derived from saturated steam tables following the IAPWS-IF97 standard²⁷ and from a temperature sensor T_{s_0} mounted in the steam supply pipe

$$h_{s_0} = f(T_{s_0}) \quad (23)$$

h_f is computed using the temperature of the steam layer T_s measured by the top temperature sensor, and assuming that the biomass heats till the steam temperature near the reactor inlet

$$h_f = c_b (T_s - T_r) \quad (24)$$

It is natural to use the same grid in Equation 20 as the one from the composition balance section. The model tracks the biomass enthalpy throughout each cell of the grid. The conductive heat from the steam to the biomass layer is neglected due to the fact that biomass acts as an insulator. Therefore, only convective effects remain in the biomass layer and $Q_k = 0, 0 < k \leq N$.

Table 3. Fixed Model Parameters

Parameter	Description	Value	Unit
L_r	Thermal reactor length	12	m
t_r	Pretreatment retention time	15	min
N	Grid cell resolution	10	—
δz	Grid cell width	1.2	m
v	Reactor horizontal speed	0.013	m/s
F_{in}	Feedstock flow rate	6	kg/s
h_0	Feedstock enthalpy	117	kJ/kg
h_{s0}	Fresh steam enthalpy (saturated)	2795	kJ/kg
c_b	Specific heat of pretreated biomass (constant)	3.8	kJ/(kgK)
R_g	Universal gas constant	8.3145	J/(molK)
C_0	Feedstock composition:		
	Cellulose	160	g/kg
	Xylan	95	g/kg
	Arabinan	8	g/kg
	Lignin	80	g/kg
	Acetyls	16	g/kg
	Glucose	0	g/kg
	Xylooligomers	0	g/kg
	Xylose	0	g/kg
	Arabinose	0	g/kg
	Acetic acid	0	g/kg
	Furfural	0	g/kg
	5-HMF	0	g/kg
	Water	600	g/kg
	Other	41	g/kg
		1000	g/kg

Feedstock is soaked before entering the thermal reactor till approximately 40% dry matter.

The temperature profile is obtained by dividing the enthalpy from each cell with the specific heat for biomass constant c_b

$$T_k = \frac{h_k}{c_b} \quad (25)$$

where T_k is the biomass temperature from cell k , and c_b is the specific heat of pretreated biomass.

Model summary. The thermal reactor model tracks $n_C = 14$ species concentrations shown in vector C_k from Eq. 18: cellulose, xylan, arabinan, lignin, acetyls, glucose, xylooligomers, xylose, arabinose, acetic acid, furfural, 5-HMF, water, and other components.

The total number of states n_x is variable depending on the initial value of N , the amount of cells in the reactor grid. n_x can be calculated as follows

Table 4. Calibrated Parameters Using the Estimation Dataset from Figure 2, Which Provides System Dynamics in the 178–185°C Range

Parameter	Value	Unit
k_{Xo}	2.78×10^{31}	1/s
E_{Xo}	298,011	J/mol
k_X	1.31×10^{34}	1/s
E_X	304,680	J/mol
k_G	1.11×10^{35}	1/s
E_G	335,614	J/mol
k_{PL}	1.03×10^{33}	1/s
E_{PL}	325,629	J/mol
k_F	5.09×10^{33}	1/s
E_F	327,253	J/mol
k_{Ac}	4.88×10^{24}	1/s
E_{Ac}	242,687	J/mol
k_H	1×10^{31}	1/s
E_H	300,000	J/mol
a_{PL}	0.1019	—

$$n_x = N \cdot (n_C + n_h) + (n_s + n_m) \quad (26)$$

where n_C is the number of species, $n_h = 1$ meaning one state for each grid cell enthalpy, $n_s = 2$ is the number of states from the steam layer (mass and enthalpy), and $n_m = 1$ is the total mass of biomass in the reactor. In this study, N is set to 10 leading to 153 states in total.

The model has 2 bus inputs: one for feedstock, and another one for steam. The feedstock input has 16 components: flow rate (1), feedstock concentrations (14), and enthalpy (1). The steam input has 2 components: flow rate (1) and enthalpy (1). In total there are 18 inputs.

There are 2 bus outputs: pretreated fibers and the liquid rich in C5 sugars. Each bus has 16 components: flow rate (1), species concentrations (14), and enthalpy (1).

Table 3 shows the fixed model parameters. The kinetics parameters are determined in the model calibration section of this study.

Results and Discussion

This section is split into a model analysis and validation part, and the model application as a state estimator. The model analysis and validation section contains the sensitivity and uncertainty analysis, parameter estimation, and residual analysis.

Model analysis and validation

Model Initialization. Model parameters are calibrated with respect to the following NIR measurements: cellulose, xylan, lignin, acetic acid, and furfural. The data were obtained from a demonstration scale thermal reactor throughout 15 h of operation. Only a subset of 7 h is used for calibration and parameter estimation, while the entire set of measurements is used for validation and residual analysis. The measurements of cellulose, xylan, and lignin are reported as percentage of dry matter, while acetic acid and furfural concentrations as g/kg of slurry. Table 4 illustrates the estimated values while Figure 5 shows the fitting result.

Sensitivity Analysis. The sensitivity analysis contributes to a good process understanding by quantifying the relation between outputs and model parameters. The analysis is carried with respect to all model parameters from Table 4 and aims at ranking these parameters by their significance. In this way a reduced set of relevant parameters can be identified and used for a proper parameter estimation procedure. Also, by fitting less parameters, the model calibration procedure simplifies and can be run more often in a real industrial application.

A measure of sensitivity with respect to model parameters, and suitable for time varying signals, is the delta mean square δ_{ik}^{msqr} defined by Brun et al.²⁸

$$\delta_{ik}^{msqr} = \sqrt{\frac{1}{N} s_{nd,ik}^\top s_{nd,ik}} \quad (27)$$

where k is the parameter index, i is the model output index, N is the number of samples, and $s_{nd,ik}$ is a vector with the nondimensional sensitivity calculated in each sample

$$s_{nd,ik} = \frac{\partial y_i}{\partial \theta_k} \frac{\theta_k}{s_{ci}} \quad (28)$$

$\partial y_i / \partial \theta_k$ represents the output variation with respect to parameter θ_k , and s_{ci} is a scaling factor with the same physical dimension as the corresponding observation in order to make

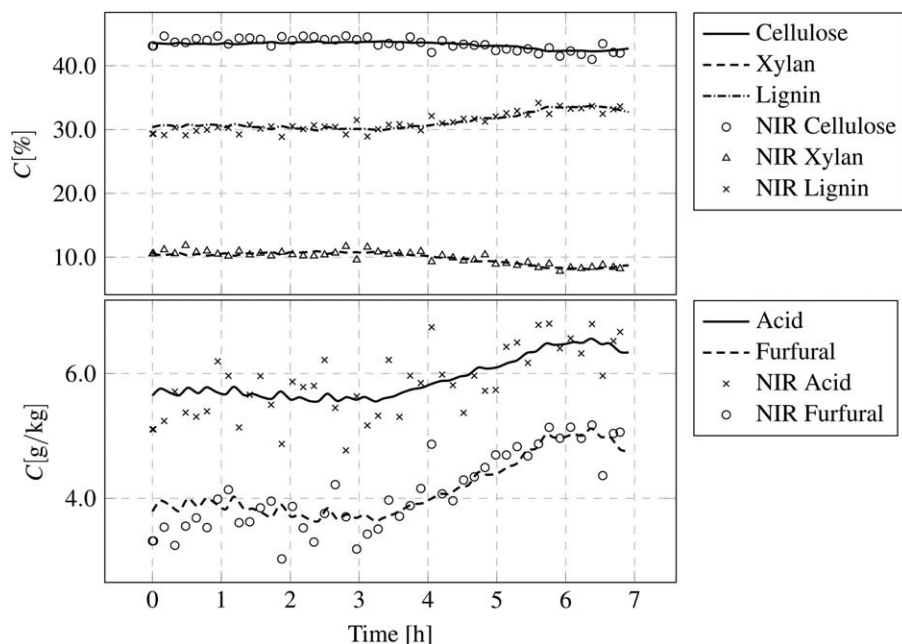


Figure 5. Solid and liquid content of pretreated biomass: cellulose, xylan, lignin, acetic acid, and furfural.

this measure nondimensional. In this study, the scaling factor is chosen as the mean value of output i

$$sc_i = \frac{1}{N} \sum_{k=1}^N y_i(k) \quad (29)$$

All parameters are ranked according to δ_{ik}^{msqr} for each output i . As the sensitivity measure is nondimensional, a cumulative variable is also defined as the sum of sensitivities for a given parameter in all outputs. Because the model has to predict all defined outputs, the subset of significant parameters contains all parameters with a cumulative sensitivity above a threshold,

which is set to 2% of the maximum sensitivity. The cumulative delta mean square is defined as

$$\delta_k^{msqr} = \sum_{i=1}^{n_y} \delta_{ik}^{msqr} \quad (30)$$

where n_y is the total number of outputs, that is, $n_y = 5$ in this study: concentrations of cellulose, xylan, lignin, acetic acid, and furfural.

The same sensitivity analysis methodology has been applied on a cellulosic hydrolysis model in previous studies by Sin et al.,¹⁵ and Prunescu and Sin.⁵

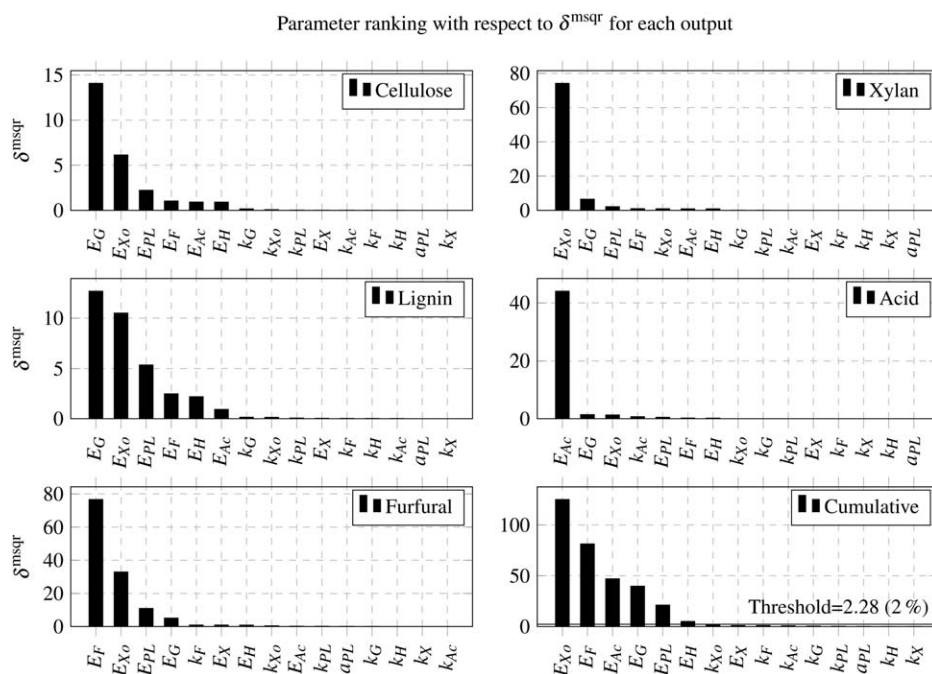


Figure 6. Sensitivity analysis by calculating the δ^{msqr} for each output with respect to each model parameter.

Although the model has 15 parameters in total, only 6 parameters are of interest as indicated in the bottom right figure.

Table 5. Parameter Estimation Results: Parameter Estimates in J/mol with Standard Deviation σ , and 95% Confidence Interval

Parameter	Value (J/mol)	Standard Deviation, σ	Lower Bound	Upper Bound
E_{X_0}	298,010	98	297,918	298,102
E_F	327,255	285	326,989	327,520
E_{Ac}	242,693	174	242,531	242,855
E_G	335,616	249	335,383	335,848
E_{PL}	325,632	1573	324,165	327,099
E_H	299,999	2639	297,537	302,461

The results are valid in the 178–185°C range.

The sensitivity analysis results for the pretreatment process can be observed in Figure 6:

- Cellulose is mostly sensitive to E_G , which is expected since cellulose is directly hydrolyzed into glucose and this reaction is sensitive to the reactor temperature.
- Xylan is sensitive to the activation energy of xylooligomers production E_{X_0} , which is not surprising since xylooligomers are direct products of xylan hydrolysis.
- Lignin as percentage of dry matter follows the changes from xylan and cellulose content, which is the solid content of biomass. This means that if more xylan is hydrolyzed then the percentage of lignin in the remaining slurry after separation will increase. There is also lignin production as pseudo-lignin and E_{PL} is ranked second. E_F , E_G , and E_X influence the amount of carbohydrates that participate in pseudo-lignin formation.
- Organic acids, mostly represented by acetic acid, are sensitive to the activation energy for reaction rate r_{Ac} , that is, E_{Ac} .
- The last output, furfural, is mostly sensitive to the activation energy E_F of reaction rate r_F . E_{X_0} appears second because it directly affects the amount of xylose, which degrades to furfural.

The cumulative sensitivity measure is illustrated in the right bottom plot of Figure 6. The most sensitive parameters are picked to be the first six

$$\theta_R = [E_{X_0} \ E_F \ E_{Ac} \ E_G \ E_{PL} \ E_H] \quad (31)$$

which are all activation energies directly involved in the reaction temperature dependency. This is natural as it has been observed in experimental studies that small changes in reactor temperature impact significantly the composition of pretreated fibers, and is in agreement with process expert knowledge.⁹ This analysis also ranks the activation energies among themselves. E_{X_0} (related to xylooligomers production) is ranked first as the most sensitive parameter, which tells that hemicellulose hydrolysis is the main phenomenon taking place in the reactor. E_F is ranked second showing that furfural is the main by-product followed by acetic acid (E_{Ac}). E_G is ranked 4th, which means that cellulose hydrolysis also occurs in the reactor but at a much lower rate than hemicellulose hydrolysis. The other two by-products, that is, pseudo-lignin and 5-HMF, have a lower significance.

Parameter Estimation. The reduced set of parameters θ_R is identified based on the real NIR measurements from the demonstration scale facility. A nonlinear least square method is run to obtain the parameter estimates $\hat{\theta}_R$ along with their standard deviation σ and correlation matrix \mathbf{R}_θ . Tables 5 and 6 show the results. The estimated parameter values are deemed

Table 6. Parameter Estimation Results: Correlation matrix \mathbf{R}_θ

Parameter	E_{X_0}	E_F	E_{Ac}	E_G	E_{PL}	E_H
E_{X_0}	1					
E_F	−0.51	1				
E_{Ac}	0.17	−0.12	1			
E_G	0.74	−0.54	0.26	1		
E_{PL}	−0.51	0.74	−0.15	−0.63	1	
E_H	0.14	−0.61	0.01	0.16	−0.85	1

The results are valid in the 178–185°C range.

credible as the parameter estimation error indicated by the standard deviation is rather low, that is, less than 1%. However some of the parameter estimates are found to be significantly correlated, for example, correlation between E_G and E_{X_0} is 0.74, which implies poor identifiability. The reason for this is the dataset used for parameter estimation. The data are obtained from an industrial scale facility during normal operational conditions under small temperature disturbances. Such data with limited dynamics cannot be expected to provide rich information for complete identification of all the parameters²⁹ and design of experiments for identification should be pursued on lower scale facilities. Other measurements should also be included in the parameter estimation analysis, such as xylooligomers, xylose, and glucose content of the liquid part, which are missing in this study.

Uncertainty Analysis. To assess the uncertainty of the model, a large number of Monte Carlo simulations are run in 3 different scenarios that cover most sources of uncertainty: model parameters, feed parameters, and combined.

1. Model parameters uncertainty: the standard deviation and the correlation matrix for model parameters were obtained in the previous step. A number of 200 samples are generated using LHS with correlation control. Figure 7a shows the biomass composition predictions with the 5th–95th percentile interval considering only model parameters uncertainty. The model parameters uncertainty cannot explain the entire output variations, especially in the solid content of the mixture. This is expected since there are other sources of uncertainty, that is, initial biomass composition or feed parameters.

2. Feed parameters uncertainty: Latin hypercube uniform sampling is performed to cover 7% variation of the initial composition and another 200 samples are generated considering fixed model parameters. The simulation results with NIR measurements and percentile intervals are displayed in Figure 7b. Cellulose and lignin concentrations are well within the bounds while xylan exceeds the interval in some samples. Acetic acid and furfural are not entirely captured by the percentile interval, but it might be due to model parameters uncertainty.

3. Combined uncertainty: the results are shown in Figure 7c. The uncertainty for acetic acid and furfural is wider but still not entirely within bounds. NIR instruments have a lower accuracy for low concentrations and it is expected that the remaining prediction inaccuracy is due to non Gaussian measurement error. Acetic acid and furfural have a much lower concentration compared to the solids, that is, cellulose, xylan, and lignin. Therefore, it is expected to have significant measurement errors at these low concentrations.

Reactor Profiles. The thermal reactor was split in 10 smaller cells and reactor profiles can be drawn in order to

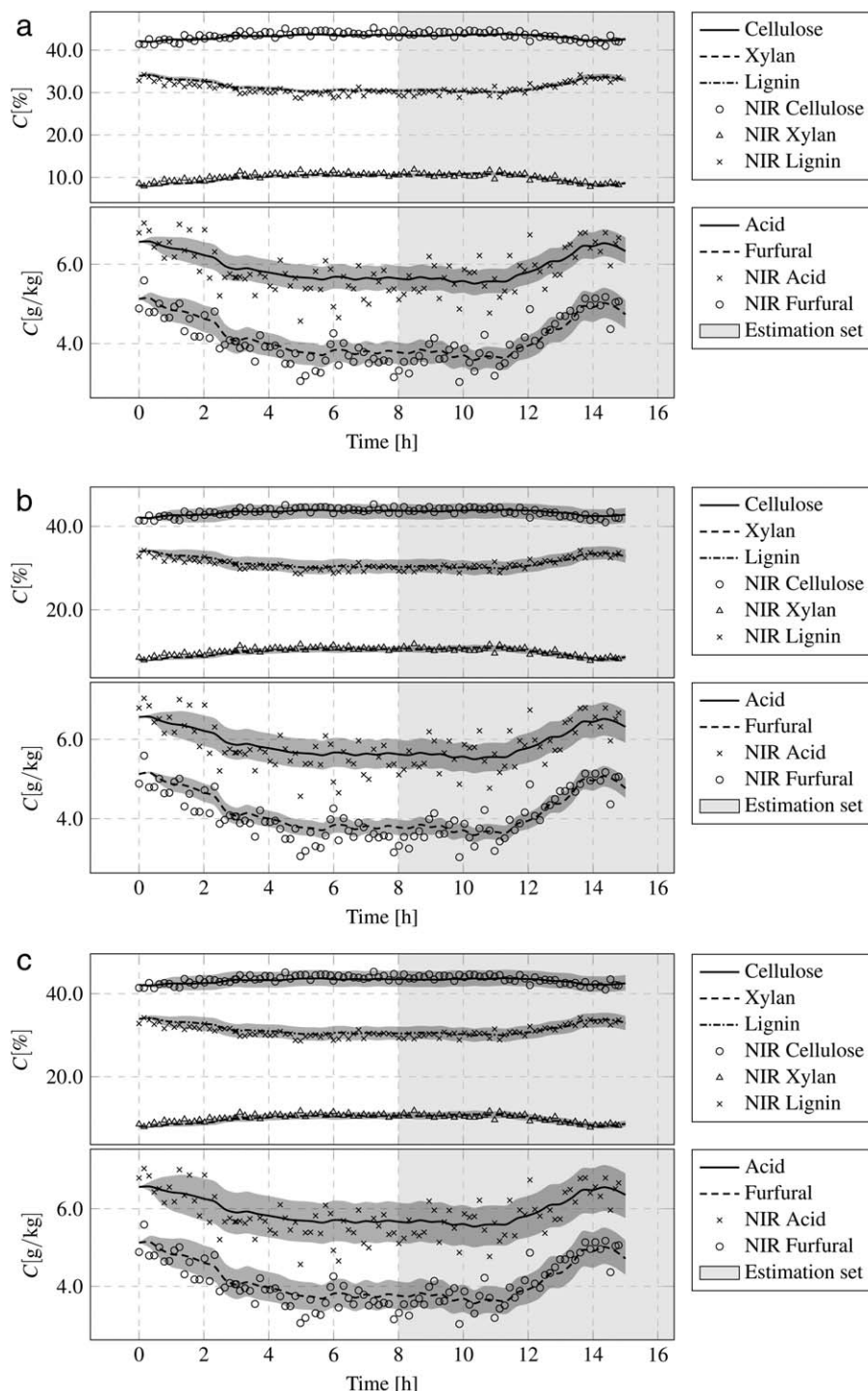


Figure 7. Model predictions uncertainty with respect to model parameters, feed parameters, and combined.

observe how biomass changes composition along the reactor. The top plot from Figure 8 indicates the reaction temperature, which is not constant but rather a gradient built with the help of the top and bottom temperature sensors. The biomass is subject to a range of pretreatment temperatures. The remaining 3 plots from Figure 8 show how biomass composition changes along the reactor. The figure x axis is the length of the reactor, while the y axis contains the species concentration in g/kg. The second plot indicates that arabinan is fully hydrolyzed, while cellulose and xylan are partially hydrolyzed with more xylan conversion than cellulose. The following plot shows sugar production, that is, glucose, xylooligomers, xylose, and

arabinose. By-products such as acetic acid and furfural are displayed in the bottom plot of Figure 8.

Standardized Regression Coefficients. A linear model is first fitted from model parameters to each model output from the Monte Carlo simulations^{16,30}

$$y_{\text{reg}} = a + \sum_k b_k \theta_k \quad (32)$$

where y_{reg} is the i th output, and a and b_k are the linear model parameters. The SRC β are a global sensitivity measure and are defined as

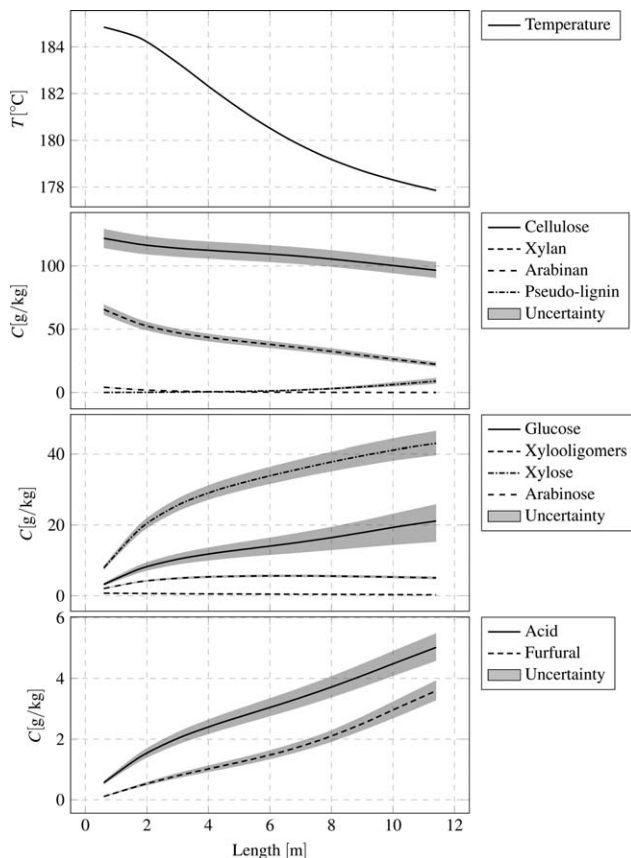


Figure 8. The top plot shows the reactor horizontal temperature gradient, which is used for calculating the reaction rates.

The other plots illustrate the reactor conversion profiles with confidence bounds due to both model and feed parameters uncertainty.

$$\beta_k = \frac{\sigma_{\hat{\theta}_{Rk}}}{\sigma_{y_i}} b_k \quad (33)$$

where β_k is the β coefficient, $\sigma_{\hat{\theta}_{Rk}}$ is the standard deviation of the parameter estimate, σ_{y_i} is the standard deviation of output i , and b_k is the linear model parameter. β_k is an indicator for how much the parameter uncertainty contributes to the prediction uncertainty.

The β coefficients are displayed in Table 7 and explained below:

- Concentration of solid cellulose C_{Cs} : is sensitive to the initial concentration of cellulose and lignin from feedstock,

to the activation energy of glucose reaction r_G , and pseudo-lignin reaction r_L . Lignin appears due to the fact that the function of lignin in nature is to protect cellulose from degradation, thus a larger amount of lignin necessarily affects the kinetics of cellulose degradation by reducing it (the corresponding β coefficient has a negative value). There is also a percentage redistribution of biomass content when lignin or pseudo-lignin increase. Thus, production of pseudo-lignin causes a decrease in cellulose concentration in pretreated fibers and E_{PL} appears among the top parameters.

- Concentration of solid xylan C_{Xs} : is sensitive to the initial concentration of xylan in feedstock, and to xylan hydrolysis or the activation energy of xylooligomers reaction rate E_{Xo} . Higher concentrations of cellulose and lignin in feedstock can cause a decrease in xylan content due to percentage redistribution.

- Concentration of solid lignin C_{Ls} : high sensitivity is detected in initial lignin concentration from feedstock and in pseudo-lignin production. Initial concentration of cellulose also has an impact due to percentage redistribution of fiber content.

- Concentration of acid C_{Ac} : as expected, acetyls content in feedstock and activation energy of the acid production reaction rate E_{Ac} have the greatest impact on acid content in pretreated fibers.

- Concentration of furfural C_F : activation energy of the furfural production reaction E_F and pseudo-lignin E_{PL} are the most sensitive parameters. Furfural participates in pseudo-lignin formation and is expected to find E_{PL} among the top parameters.

Parameters related to feed composition have a higher sensitivity than the kinetic parameters, even though only a deviation of 7% was introduced in the initial biomass composition. This indicates the importance of measuring the initial composition of feedstock for more accurate model predictions. The SRC based sensitivity results are credible as the degree of linearization indicated by Pearson correlation coefficient R^2 is high for all the outputs.¹⁶

Residual Analysis. Statistical analysis of residuals is important for model validation especially as it provides information on any model bias in the predictions.³¹ The prediction estimation error or residual is calculated as

$$e = y - \hat{y} \quad (34)$$

where e is the residual vector, y is the real measurement, and \hat{y} is the estimated output. There are 5 measurements regarding biomass composition, that is, cellulose (%), xylan (%), lignin (%), includes pseudo-lignin), acetic acid (g/kg), and furfural

Table 7. SRC Coefficients for Model and Feed (Combined) Parameters

θ	Cellulose	θ	Xylan	θ	Lignin	θ	Acetic acid	θ	Furfural
C_{Cs}	0.82	C_{Xs}	0.70	C_{Ls}	0.77	C_{AcS}	0.72	E_F	-1.00
C_{Ls}	-0.43	E_{Xo}	0.54	E_{PL}	-0.72	E_{Ac}	-0.60	E_{PL}	0.75
E_G	0.38	C_{Cs}	-0.38	C_{Cs}	-0.47	C_{Cs}	0.06	C_{Xs}	0.56
E_{PL}	0.36	C_{Ls}	-0.23	E_H	-0.47	C_{Ls}	0.04	E_{Xo}	-0.16
E_H	0.23	E_{PL}	0.20	E_G	-0.24	E_{PL}	-0.04	C_{As}	0.08
C_{Xs}	-0.23	E_H	0.14	E_{Xo}	-0.11	E_H	-0.02	E_H	0.07
E_{Xo}	-0.08	E_G	-0.12	E_F	-0.06	E_G	0.02	C_{Ls}	0.04
C_{AcS}	-0.05	C_{AcS}	-0.03	C_{Xs}	-0.04	C_{Xs}	0.01	C_{Cs}	0.03
E_F	0.02	E_F	0.02	C_{AcS}	-0.04	E_{Xo}	0.01	E_G	0.02
E_{Ac}	-0.02	E_{Ac}	-0.02	E_{Ac}	-0.02	E_F	-0.00	C_{AcS}	0.01
C_{As}	-0.01	C_{As}	-0.01	C_{As}	0.02	C_{As}	-0.00	E_{Ac}	0.00
R^2	1.00	R^2	1.00	R^2	0.99	R^2	1.00	R^2	0.93

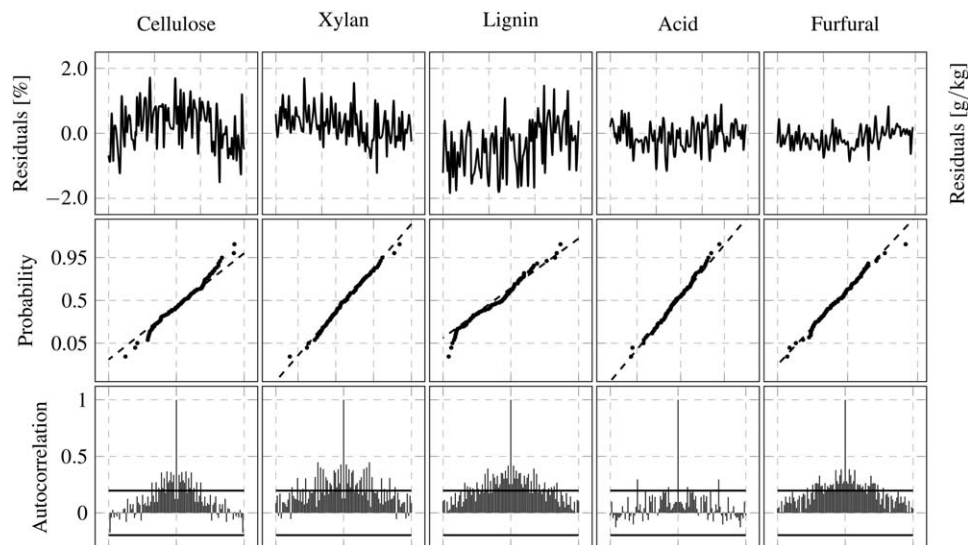


Figure 9. Residual analysis for the validation set.

The top plot series show the residuals; the middle series compare in a Gaussian probability plot the distribution of residuals to a Normal distribution; the bottom series display the autocorrelation function and the 95% confidence interval.

content (g/kg). The residuals are calculated based on the validation set and they are displayed in the top plot series from Figure 9. The estimation errors for cellulose, xylan, and lignin stay within -2 to 2% , while acetic acid and furfural are slightly lower, -1 to 1 g/kg.

If the residuals are white noise (i.e., Gaussian distribution with mean zero and a certain standard deviation) and uncorrelated then there is no systematic error or bias in the model. To check the hypothesis of residual Gaussian distribution, a 5% t -test is performed. The test is passed and the Gaussian probability plots from Figure 9 (the middle series) show how close the residuals are to a Normal distribution. The autocorrelation determines if there still is any information in the residuals that

is not captured by the model. The bottom series of plots from Figure 9 display the autocorrelation function at 50 lags with the 95% confidence interval. There are spikes in almost all plots except acetic acid production that exceed the confidence interval but they are not significant. The feedstock composition and the assumption that it remains constant might be a

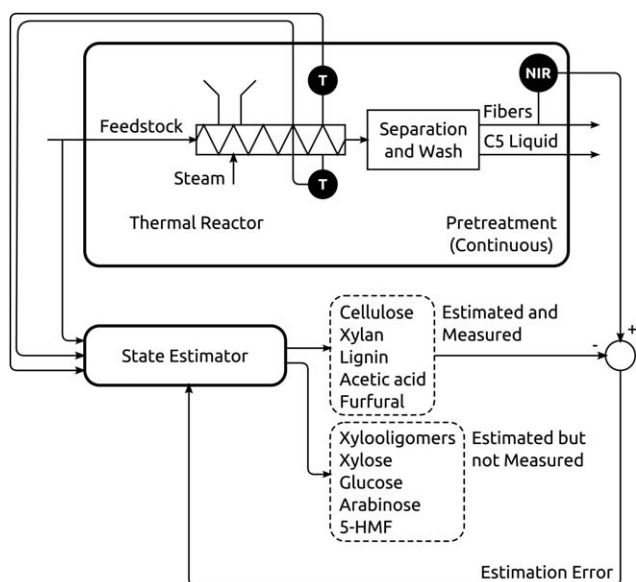


Figure 10. State estimator block diagram.

The state estimator uses 2 temperature sensors and the NIR measurements to infer pretreated biomass composition. The state estimator acts both as measurement filter and soft sensor.

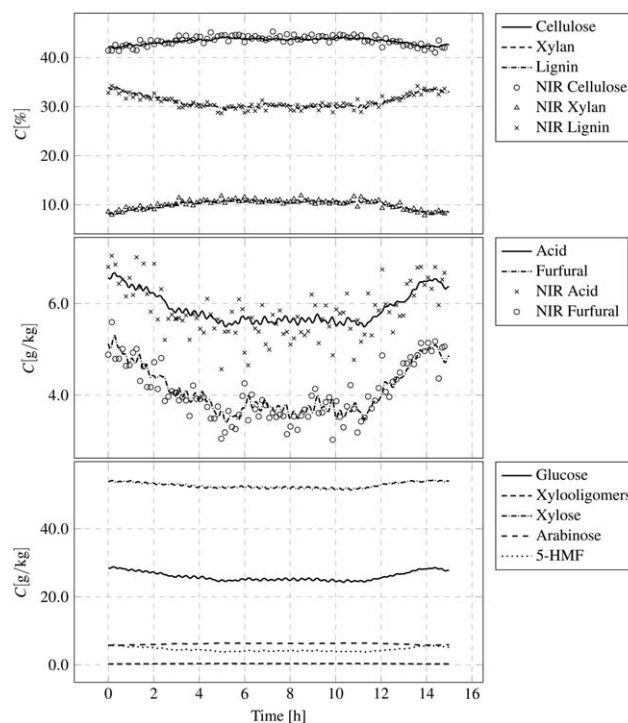


Figure 11. Pretreatment state estimator throughout 15 h of operation.

The plots show the solid and liquid content of pretreated biomass: cellulose, xylan, and lignin in the top plot, acetic acid and furfural in the second plot, and estimations of by-products production in the bottom plot.

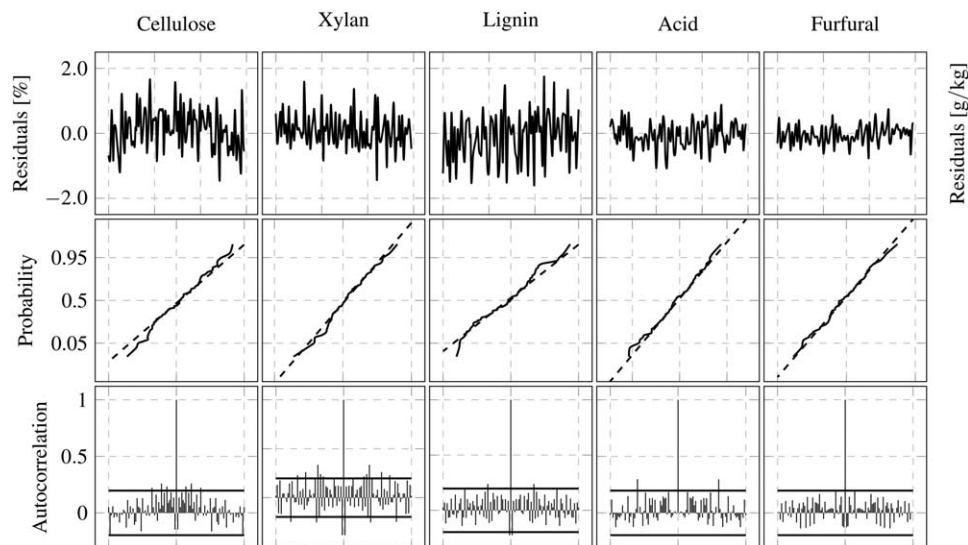


Figure 12. Residual analysis for the validation set with estimator.

The top plot series show the residuals; the middle series compare in a Gaussian probability plot the distribution of residuals to a Normal distribution; the bottom series display the autocorrelation function and the 95% confidence region.

good source for these results. In reality feedstock has composition variations.

Model application: State estimator

The utility of the model for industrial application is highlighted through the development of a state estimator using a Kalman filter. The state estimator is useful for performance monitoring by estimating in real-time the entire biomass composition, that is, both sugars and inhibitors production. The estimator infers all species concentrations from pretreated fibers using a reduced number of measurements provided by the NIR instrument.

A state estimator is naturally based on the nonlinear process model

$$\begin{cases} \dot{\hat{x}} = f(\hat{x}, u) \\ \hat{y} = g(\hat{x}) \end{cases} \quad (35)$$

where \hat{x} is the state estimates vector, \hat{y} is the output estimates vector, $f(\hat{x}, u)$ is the nonlinear state derivatives as a function of states and inputs, and $g(\hat{x})$ is the nonlinear output function.

Several methods exist for use of nonlinear process models in state estimation. Classical approaches include an extended Kalman filter for combined state and parameter estimation³² for a linearized system at a particular point of operation, or direct inclusion of the nonlinear process model in the filter.³³ Later developments have included the unscented Kalman and particle filters to better explore and approximate non-Gaussian nature of the process noise in a nonlinear system.³⁴

To compensate for model-mismatches between estimates and real measurements, an extra correction term is added to the state derivatives equation from (35)

$$\begin{cases} \dot{\hat{x}} = f(\hat{x}, u) + \mathbf{L}e \\ \hat{y} = g(\hat{x}) \end{cases} \quad (36)$$

where e is the estimation error defined in Equation (34), and \mathbf{L} is a gain matrix that needs to be designed. The expression in (36) is known as a high-gain observer. The block diagram of the state estimator for the pretreatment process is shown in Figure 10. In this application, feedback from the NIR instru-

ment is used to create the estimation error term. A constant feedstock composition is assumed and it can be determined in reality by analyzing stock samples offline in the laboratory.

There are various methods to calculate the gain matrix \mathbf{L} among which the most common ones include pole-placement and extended Kalman filter (EKF). Pole-placement and extended Kalman filters rely on a linearized version of the process model. The extended Kalman filter uses the linearized model to calculate the error covariance matrix and, from this, the observer gain. If the linearized model and process noise covariance matrices are fixed then the error covariance matrix and the observer gain can be calculated offline. Kalman filters are known to be optimal in the sense of minimizing the estimation error covariance matrix in the presence of process and measurement noise,³⁵ and are preferred to pole-placement techniques. Kalman filters require known process noise covariance matrices to work optimally, which is rather difficult in reality to approximate but alternatives exist, which estimate the noise structure online,³² at the expense of more complexity of the estimation algorithm.

The operational point in this study does not change significantly and this is the reason why a static extended Kalman filter is chosen for the state estimator. The EKF design process follows these steps:

1. The first step is to obtain a stochastic linear model by linearizing the nonlinear process model from (35) around the nominal operational point seen in the datasets

$$\begin{cases} \dot{\hat{x}} = \mathbf{A}\hat{x} + \mathbf{B}u + \mathbf{G}w \\ \hat{y} = \mathbf{C}\hat{x} + v \end{cases} \quad (37)$$

where \mathbf{A} is the dynamic matrix of the linearized model, \mathbf{B} is the input matrix, \mathbf{G} is the state noise propagation matrix, and \mathbf{C} is the output matrix. State noise w and measurement noise v are assumed to be 0 mean uncorrelated white noise sequences with variances \mathbf{Q} and \mathbf{R}

$$w \sim (0, \mathbf{Q}) \quad v \sim (0, \mathbf{R}) \quad (38)$$

The linear model matrices are calculated by differentiating the nonlinear process model around the nominal operational point (x_c, u_c)

$$\begin{aligned} \mathbf{A} &= \left. \frac{\partial f(\hat{x}, u)}{\partial \hat{x}} \right|_{\hat{x}=x_e, u=u_e} & \mathbf{B} &= \left. \frac{\partial f(\hat{x}, u)}{\partial u} \right|_{\hat{x}=x_e, u=u_e} \\ \mathbf{G} &= \left. \frac{\partial f(\hat{x}, u)}{\partial w} \right|_{\hat{x}=x_e, u=u_e} & \mathbf{C} &= \left. \frac{\partial g(\hat{x})}{\partial \hat{x}} \right|_{\hat{x}=x_e} \end{aligned} \quad (39)$$

where x_e and u_e form the nominal operational point in terms of states and inputs. It is not known how the process noise propagates inside the system dynamics, and \mathbf{G} is set to \mathbf{I}_{n_x} (identity matrix of size n_x or total number of states).

2. The second step is to approximate the state and measurement noise covariance matrices, that is, \mathbf{Q} and \mathbf{R} , which are set to

$$\mathbf{Q} = 10^{-4} \cdot \mathbf{I}_{n_x} \quad \mathbf{R} = \begin{bmatrix} 50 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 50 & 0 & 0 \\ 0 & 0 & 0 & 600 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{bmatrix} \quad (40)$$

where \mathbf{I}_{n_x} is an identity matrix, and n_x is the number of states. The concentrations of solids are higher and more reliable, therefore lower variances are used in the first 3 diagonal terms from \mathbf{R} , which correspond to cellulose, xylan, and lignin (measured in % of dry matter). The other 2 diagonal numbers are the variances for acetic acid and furfural (measured in g/kg), which are in low concentrations and have larger measurement errors.

3. In the last step of the design process, the static Kalman gain is calculated³⁵

$$\mathbf{L} = \mathbf{P}\mathbf{C}^T\mathbf{R}^{-1} \quad (41)$$

where \mathbf{P} is the error covariance matrix found from solving the Riccati equation³⁵

$$\dot{\mathbf{P}} = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T - \mathbf{P}\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\mathbf{P} \quad (42)$$

when $\dot{\mathbf{P}} = 0$.

The Kalman state estimator is tested throughout the whole dataset of 15 h. Figure 11 shows the model outputs overlapped with the NIR measurements. The model and the Kalman filter succeed in following the dynamic trends of the process. The state estimator filters the NIR measurements and also acts as a soft sensor for by-products production: glucose, xylooligomers, xylose, arabinose, and 5-HMF shown in the bottom plot of Figure 11.

The residuals are displayed in Figure 12. The variance of the raw residuals is similar to the one from Figure 9. However, the Kalman filter is able to capture more information from the process causing the autocorrelation function to stay within the confidence interval (the bottom plot series from Figure 12). The residuals distribution slightly changes as indicated by the middle series of plots from Figure 12. This happens because the noise covariance matrices, and the propagation of noise through the system are unknown and set to arbitrary values.

The state estimator successfully embeds the real measurements to compensate on model predictions mismatches under uncertainties such variation in feedstock composition or in model parameters. Also, the state estimator acts as a soft sensor for several unmeasured variables, some of which act as inhibitors in downstream processing especially in enzymatic

hydrolysis and fermentation. Such information includes: xylose, xylooligomers, arabinose, pseudo-lignin, glucose, and 5-HMF production.

Conclusions

This study presented a dynamic model for a large scale biomass hydrothermal pretreatment process. The model was capable of predicting the composition of pretreated fibers, and has been properly analyzed assessing its sensitivity and uncertainty taking into account both model and feed parameters. The sensitivity analysis showed that the set of 15 model parameters can be reduced to a subset of 6 significant parameters that deal with the activation energy of most reactions. This finding was expected as the activation energy appears in the temperature dependency term of the reaction equation, and small reactor temperatures deviations can lead to significant changes in pretreated biomass. Real data were extracted from a demonstration scale biorefinery and used for parameter estimation and model validation. The data exhibit limited system dynamics and provide a range of validity for the parameter estimates between 178 and 185°C. To extend the validity range of the parameter estimates to different temperature ranges, additional experimental data should be collected either by experimenting at lower scale facilities or changing the operational points of the biorefinery, which is not easily accessible.

The uncertainty analysis was conducted with LHS of both model and feed parameters, and Monte Carlo simulations. The 5th–95th percentile includes most of the real data variations, except the very low concentrations of acetic acid and furfural, which is expected as the NIR instrument is affected by high measurement errors at low concentrations. The global sensitivity analysis showed that feedstock composition parameters have a higher significance than model parameters, suggesting that feedstock composition should be measured upstream for more accurate predictions.

In the last section of the study, a Kalman state estimator was designed as an application of the process model. The estimator was then tested on the whole dataset, that is, throughout 15 h of operation. The model predictions were found to follow the dynamic trends of the process, making it valuable for control design, diagnosis, real-time optimization, and inferential sensors development at full scale applications. The state estimator acts both as a measurement filter and soft sensor for estimating unmeasured variables, that is, C5 sugars production and inhibitors, which are important in downstream processes.

Notation

- α = furfural and 5-HMF participation in pseudo-lignin production, -
- β = standardized regression coefficients, -
- β_k = standardized regression coefficients, -
- δ^{msqr} = delta mean square, -
- $\delta_{ik}^{\text{msqr}}$ = delta mean square of output i with respect to model parameter k , -
- δ_z = reactor cell width, m
- θ_R = estimated parameters subset,
- \hat{x} = estimated model states,
- \hat{y} = estimated model outputs, % or g/kg
- \mathbf{L} = Kalman static gain,
- \mathbf{Q} = state covariance matrix,
- \mathbf{R}_θ = model parameters correlation matrix,
- \mathbf{R} = output covariance matrix,
- ∇ = gradient operator, -
- σ = model parameters standard deviation,
- $\sigma_{\hat{\theta}_{R_k}}$ = standard deviation of parameter estimate k ,

σ_{y_i} = standard deviation of output i ,
 θ_0 = initial model parameters,
 θ_0 = parameter k in vecto,
 θ_{R_0} = identifiable parameter subset,
 a = linear model parameter used in SRC computations,
 b_k = linear model parameter used in SRC computations,
 C = biomass composition vector, g/kg
 C_k = composition vector in grid cell k , g/kg
 C_{As} = concentration of arabinan (solid), g/kg
 C_{AcS} = concentration of acetyls (solid), g/kg
 C_A = concentration of arabinose (liquid), g/kg
 c_b = biomass specific heat constant, kJ/(kgK)
 C_{CS} = Concentration of cellulose (solid), g/kg
 C_F = Concentration of furfural (liquid), g/kg
 C_G = Concentration of glucose (liquid), g/kg
 C_H = Concentration of 5-HMF (liquid), g/kg
 C_{k-1} = Composition vector in cell $k - 1$, g/kg
 C_O = Concentration of other components (liquid), g/kg
 C_W = Concentration of water, g/kg
 C_{Xs} = Concentration of xylan (solid), g/kg
 C_{Xo} = Concentration of xylooligomers (liquid), g/kg
 C_X = Concentration of xylose (liquid), g/kg
 D = Diffusion coefficient, m²/s
 e = Residual vector, % or g/kg
 E_G = Glucose activation energy constant, J/mol
 E_{Ac} = Organic acid activation energy constant, J/mol
 E_A = Arabinose activation energy constant, J/mol
 E_F = Furfural activation energy constant, J/mol
 E_H = 5-HMF activation energy constant, J/mol
 E_L = Pseudo-lignin activation energy constant, J/mol
 E_{Xo} = Xylooligomers activation energy constant, J/mol
 $f(\hat{x}, u)$ = Nonlinear state derivatives function,
 F_{in} = Feedstock inflow rate, kg/s
 F_{out} = Pretreated biomass outflow rate, kg/s
 $g(\hat{x})$ = Nonlinear model outputs function,
 h_k = Biomass enthalpy in cell k , kJ/kg
 h_f = Biomass final enthalpy after steam heating, kJ/kg
 h_{k-1} = Biomass enthalpy in cell $k - 1$, kJ/kg
 h_s = Fresh steam enthalpy (saturated steam), kJ/kg
 k_G = Glucose reaction rate constant, 1/s
 k_{Ac} = Organic acid reaction rate constant, 1/s
 k_A = Arabinose reaction rate constant, 1/s
 k_F = Furfural reaction rate constant, 1/s
 k_H = 5-HMF reaction rate constant, 1/s
 k_L = Pseudo-lignin reaction rate constant, 1/s
 k_{Xo} = Xylooligomers reaction rate constant, 1/s
 L_r = Thermal reactor length, m
 M = Mass of biomass in the thermal reactor, kg
 N = The thermal reactor is split into N cells, -
 Q_0 = Lumped heat transfer from steam to biomass, kJ/(kgs)
 Q_k = Energy transfer from steam to biomass in cell k , kJ/(kgs)
 R_g = Universal gas constant, J/(Kmol)
 r_G = Glucose production rate, g/(kgs)
 R_k = Reaction rate vector in cell k , g/(kgs)
 r_{Ac} = Organic acid production rate, g/(kgs)
 r_A = Arabinose production rate, g/(kgs)
 r_{FA} = Furfural production rate from arabinose, g/(kgs)
 r_{FX} = Furfural production rate from xylose, g/(kgs)
 r_F = Furfural production rate, g/(kgs)
 r_H = 5-HMF production rate, g/(kgs)
 r_{LA} = Pseudo-lignin production rate from arabinose, g/(kgs)
 r_{LG} = Pseudo-lignin production rate from glucose, g/(kgs)
 r_{LXo} = Pseudo-lignin production rate from xylooligomers, g/(kgs)
 r_{LX} = Pseudo-lignin production rate from xylose, g/(kgs)
 r_L = Pseudo-lignin production rate, g/(kgs)
 r_{OA} = Arabinose degradation rate to other unidentified components, g/(kgs)
 r_{OG} = Glucose degradation rate to other unidentified components, g/(kgs)
 r_{OX} = Xylose degradation rate to other unidentified components, g/(kgs)
 r_{Xo} = Xylooligomers production rate, g/(kgs)
 $s_{nd,ik}$ = Nondimensional sensitivity measure of output i with respect to parameter k , -
 SC_i = $s_{nd,ik}$,
 T_0 = Feedstock temperature, K
 T_K = Environment temperature in Kelvin degrees, K

T_k = Biomass temperature in cell k , K
 t_r = Retention time, s
 T_{s0} = Steam supply temperature, K
 v = Horizontal transportation speed, m/s
 y = Model outputs: cellulose, xylan, lignin, acetic acid and furfural content, % or g/kg
 y_{reg} = Linear model fit used in SRC computations,

Literature Cited

- Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. Determination of structural carbohydrates and lignin in biomass. *Technical Report NREL/TP-510-42618*, National Renewable Energy Laboratory, 2008.
- Kristensen JB, Thygesen LG, Felby C, Jørgensen H, Elder T. Cell-wall structural changes in wheat straw pretreated for bioethanol production. *Biotechnol Biofuels*. 2008;1:1–9.
- Chiaramonti D, Prussi M, Ferrero S, Oriani L, Ottonello P, Torre P, Cherchi F. Review of pretreatment processes for lignocellulosic ethanol production, and development of an innovative method. *Biomass Bioenergy*. 2012;46:25–35.
- Larsen J, Petersen MØ, Thirup L, Li HW, Iversen FK. The IBUS process—lignocellulosic bioethanol close to a commercial reality. *Chem Eng Technol*. 2008;31:765–772.
- Prunescu RM, Sin G. Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process—a demonstration scale study. *Bioresour Technol*. 2013;150:393–403.
- Qing Q, Yang B, Wyman CE. Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes. *Bioresour Technol*. 2010;101:9624–9630.
- Cantarella M, Cantarella L, Gallifuoco A, Spera A, Alfani F. Effect of inhibitors released during steam-explosion treatment of poplar wood on subsequent enzymatic hydrolysis and SSF. *Biotechnol Prog*. 2004;20:200–206.
- Sannigrahi P, Kim DH, Jung S, Ragauskas A. Pseudo-lignin and pretreatment chemistry. *Energy Environ Sci*. 2011;4:1306–1310.
- Petersen MØ, Larsen J, Thomsen MH. Optimization of hydrothermal pretreatment of wheat straw for production of bioethanol at low water consumption without addition of chemicals. *Biomass Bioenergy*. 2009;33:834–840.
- Lavarack B, Griffin G, Rodman D. The acid hydrolysis of sugarcane bagasse hemicellulose to produce xylose, arabinose, glucose and other products. *Biomass Bioenergy*. 2002;23:367–380.
- Morales-Rodriguez R, Meyer AS, Gernaey KV, Sin G. A framework for model-based optimization of bioprocesses under uncertainty: lignocellulosic ethanol production case. *Comput Chem Eng*. 2012;42:115–129.
- Morales-Rodriguez R, Meyer AS, Gernaey KV, Sin G. Dynamic model-based evaluation of process configurations for integrated operation of hydrolysis and co-fermentation for bioethanol production from lignocellulose. *Bioresour Technol*. 2011;102:1174–1184.
- Overend RP, Chornet E, Gascoigne JA. Fractionation of lignocellulose by steam-aqueous pretreatments. *Philos Trans R Soc Lond Ser A, Math Phys Sci*. 1987;321:523–536.
- Prunescu RM, Blanke M, Jensen JM, Sin G. Temperature modelling of the biomass pretreatment process. *Proceedings of the 17th Nordic Process Control Workshop*, 2012;8–17.
- Sin G, Meyer AS, Gernaey KV. Assessing reliability of cellulose hydrolysis models to support biofuel process design— identifiability and uncertainty analysis. *Comput Chem Eng*. 2010;34:1385–1392.
- Sin G, Gernaey KV, Neumann MB, van Loosdrecht MCM, Gujer W. Global sensitivity analysis in wastewater treatment plant model applications: prioritizing sources of uncertainty. *Water Res*. 2011;45:639–651.
- Larsen J, Haven MØ, Thirup L. Inbicon makes lignocellulosic ethanol a commercial reality. *Biomass Bioenergy*. 2012;46:36–45.
- Walter E, Pronzato L. *Identification of Parametric Models from Experimental Data*. Springer, 1997.
- Helton J, Davis F. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab Eng Syst Safety*. 2003;81:23–69.
- Bird RB, Stewart WE, Lightfoot EN. *Transport Phenomena*. Wiley, 2007.
- Egeland O, Gravdahl JT. *Modeling and Simulation for Automatic Control*. Marine Cybernetics, 2002.

22. Templeton DW, Scarlata CJ, Sluiter JB, Wolfrum EJ. Compositional analysis of lignocellulosic feedstocks. 2. Method uncertainties. *J Agric Food Chem.* 2010;58:9054–9062.
23. Hansen MAT, Kristensen JB. Pretreatment and enzymatic hydrolysis of wheat straw (*Triticum aestivum* L.)—the impact of lignin relocation and plant tissues on enzymatic accessibility. *Bioresour Technol.* 2011;102:2804–2811.
24. Weiss ND, Farmer JD, Schell DJ. Impact of corn stover composition on hemicellulose conversion during dilute acid pretreatment and enzymatic cellulose digestibility of the pretreated solids. *Bioresour Technol.* 2010;101:674–678.
25. Prunescu RM, Blanke M, Sin G. Modelling and L1 adaptive control of pH in bioethanol enzymatic process. *Proceedings of the 2013 American Control Conference*, Institute of Electrical and Electronics Engineers, 2013;1888–1895.
26. Prunescu RM, Blanke M, Sin G. Modelling and L1 adaptive control of temperature in biomass pretreatment. *Annual Conference on Decision and Control*, Institute of Electrical and Electronics Engineers, 2013;3152–3159.
27. Cooper JR, Dooley RB. *Revised Release on the IAPWS Industrial Formulation 1997 for the Thermodynamic Properties of Water and Steam*. The International Association for the Properties of Water and Steam, 2007.
28. Brun R, Reichert P, Künsch HHR. Practical identifiability analysis of large environmental simulation models. *Water Resour Res.* 2001;37:1015–1030.
29. Holmberg A. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Math Biosci.* 1982;62:23–43.
30. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S. *Global Sensitivity Analysis. The Primer*. Wiley, 2007.
31. Power M. The predictive validation of ecological and environmental models. *Ecol Model.* 1993;68:33–50.
32. Ljung L. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans Automat Control.* 1979;24:36–50.
33. Zhou WW, Blanke M. Identification of a class of non-linear state space models using RPE techniques. *IEEE Trans Automat Control.* 1989;34:312–316.
34. Ristic B, Arulampalam S, Gordon N. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
35. Brown RG, Hwang PYC. *Introduction to Random Signals and Applied Kalman Filtering*, 3rd ed. Wiley, 1996.

Manuscript received Apr. 17, 2015, and revision received July 6, 2015.