Electronic Supplementary Information

Yield optimization and rational function modeling of enzymatic hydrolysis of wheat straw pretreated by NaOH-delignification, autohydrolysis and their combination

V. Pihlajaniemi,* M.H. Sipponen, O. Pastinen, I. Lehtomäki, S. Laakso

Aalto University, School of Chemical Technology, PL161, 00076 Espoo, Finland.

E-mail: ville.pihlajaniemi@aalto.fi

Construction of the rational function model

The current model is based on the basic features of a rational function

$$A * \frac{f(x)}{f(x)+1} + f(0) \qquad (S1)$$

where A is the asymptote $(x \to \infty)$ and $f(x)$ determines how it is being approached.

For a two-variable function, the asymptote is not necessary constant, and may instead be a function of the variables. For enzymatic hydrolysis yield $Y_E = f(E,t)$, it is assumed that the maximum yield at infinite t is a function of E, and at infinite E a function of t. It is also assumed that both behave asymptotically, which can be expressed as a pair of simple rational functions:

$$\lim_{t \to \infty} Y_E = Y_{E,max} \frac{aE}{aE+1} \quad \text{AND} \quad \lim_{E \to \infty} Y_E = Y_{E,max} \frac{bt}{bt+1} \qquad (S2)$$

where $Y_{E,max}$ is the absolute maximum at infinite E and t. Combining these gives:

$$Y_{E,Asymptote} = Y_{E,max} \left(\frac{aE}{aE+1}\right)\left(\frac{bt}{bt+1}\right) \qquad S(3)$$

This alone may already give a decent fit for hydrolysis. However, to affect the way the asymptote is being approached, a rational function can be added to provide flexibility and further improve fit, comprising a linear combination of different E and t terms:

$$Y_E = Y_{E,Asymptote} * \frac{cEt + dE + et}{cEt + dE + et + 1} \qquad (S4)$$

Adding the pretreatment yield Y_{AH} leads to the complete model:

$$Y_{total} = Y_{E,max} \left(\frac{aE}{aE+1}\right)\left(\frac{bt}{bt+1}\right)\left(\frac{cEt + dE + et}{cEt + dE + et + 1}\right) + Y_{AH} \qquad (S5)$$

$Y_{E,max}$ and Y_{AH} are separately modeled as polynomial correlations to pretreatment severity S.

Since the yield is strictly increasing as a function of E and t, all the parameters of the rational terms must be non-negative. Otherwise a root exists for the denominator, leading to a singularity somewhere in the response surface.

Table S1. 95% confidence intervals for the parameter values of the rational function model.

	Direct delingification			**Autohydrolysis**			**Double treatment**		
	Lower confidence bound	Parameter value	Upper confidence bound	Lower confidence bound	Parameter value	Upper confidence bound	Lower confidence bound	Parameter value	Upper confidence bound
a	0.213	0.317	0.420	0.155	0.167	0.179	0.187	0.206	0.226
b	-0.634	1.910	4.454	0.319	0.352	0.384	0.196	0.219	0.243
c	$-4.5\cdot10^{-5}$	0.00195	0.00394						
d	0.103	0.138	0.173						
e	-0.00073	0.00591	0.01255						
α_1	-0.944	-0.817	-0.691				0.104	0.121	0.138
α_2	14.8	17.2	19.5	16.5	20.0	23.5			
α_3	12.8	17.4	21.9	-35.3	-22.4	-9.5	50.4	53.1	55.8
β_1				-52.0	-28.3	-4.6			
β_2				59.2	239.3	419.5			
β_3				-827	-484	-142	16.2	16.7	17.3

Demonstration of the invalidity of a quadratic model for asymptotic behaviour

The poor suitability of a second degree polynomial model to the total sugar yield data is displayed in Figure S1. When the conventional quadratic model (Eq. S6) is only fitted to the experiment data points (Fig. S1 A), it leads to good fit (Table S2) within data range, but fails to show zero hydrolysis at zero enzyme or hydrolysis time and extrapolation above data range shows downward curvature, displaying the inherently parabolic behavior of the second degree polynomial. Adding the zero response points to the model shows severe lack of fit (Fig. S1 B) due to wrong shape of the curvature. The quadratic model can be forced through the zero response points by including Et in every term (Eq. S7). However, this leads to poor fit and the curvature turns downwards already within data range (Fig. S1 C). The fitting parameters for the conventional quadratic model are given in Table S1, for each three pretreatments.

$$Y = \beta_1 S^2 + \beta_2 E^2 + \beta_3 t^2 + \beta_4 S + \beta_5 E + \beta_6 t + \beta_7 SE + \beta_8 St + \beta_9 Et + \beta_{10} \qquad (S6)$$

$$Y = (\beta_1 (Et)^2 + \beta_2 E^2 t + \beta_3 Et^2 + \beta_4 Et)\left(1 + \beta_5 S + \beta_6 S^2\right) \qquad (S7)$$

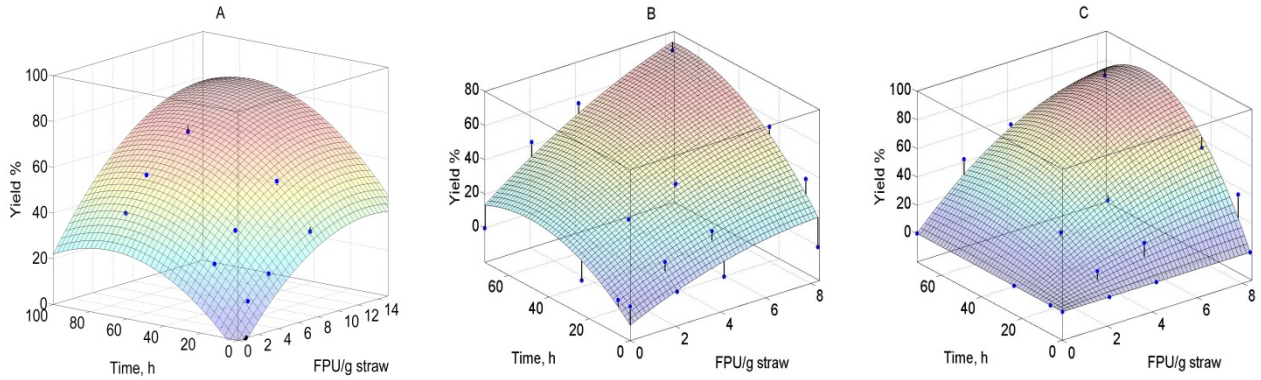

Figure S1. Fitting a quadratic model to the total yield data from direct delignification (shown: 12% NaOH-loading). A) Only experiment data points included. B) Experiment data points and zero response points are included. C) Model forced through zero response points.

Table S2. Quadratic model (Eq. S6) fitting parameters (zero response points excluded).

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	R^2
Direct delignification	-0.3214	-0.2530	-0.0072	5.1119	4.9277	0.7196	0.146896	0.023382	0.008742	-21.4807	98.12%
Autohydrolysis	-20.4161	-0.1311	-0.0062	189.5249	2.9493	0.7228	0.041096	-0.01	0.006031	-408.202	98.37%
Double treatment	0.0751	-0.1414	-0.0044	-1.2067	2.8428	0.4392	0.12848	0.003852	0.010685	20.22643	98.22%