Bioresource Technology 102 (2011) 10602–10610

Contents lists available at SciVerse ScienceDirect

Bioresource Technology

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

Significant factors selection in the chemical and enzymatic hydrolysis of lignocellulosic residues by a genetic algorithm analysis and comparison with the standard Plackett–Burman methodology

Pablo C. Giordano [a,b], Alejandro J. Beccaria [b], Héctor C. Goicoechea [a,c,*]

[a] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA) Santa Fe, Argentina
[b] Laboratorio de Fermentaciones, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, CC 242 (S3000ZAA) Santa Fe, Argentina
[c] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Avda. Rivadavia 1917, CP C1033AAJ Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 9 June 2011
Received in revised form 8 August 2011
Accepted 1 September 2011
Available online 18 September 2011

Keywords:
Genetic algorithm
Plackett–Burman design
Hydrolysis
Carbohydrates
Significant factors

ABSTRACT

A comparison between the classic Plackett–Burman design (PB) ANOVA analysis and a genetic algorithm (GA) approach to identify significant factors have been carried out. This comparison was made by applying both analyses to data obtained from the experimental results when optimizing both chemical and enzymatic hydrolysis of three lignocellulosic feedstocks (corn and wheat bran, and pine sawdust) by a PB experimental design.

Depending on the kind of biomass and the hydrolysis being considered, different results were obtained. Interestingly, some interactions were found to be significant by the GA approach and allowed to identify significant factors, that otherwise, based only in the classic PB analysis, would have not been taken into account in a further optimization step. Improvements in the fitting of c.a. 80% were obtained when comparing the coefficient of determination (R^2) computed for both methods.

1. Introduction

To verify or reject an investigation hypothesis, the researcher has to outline and implement an investigation design, i.e.: a plan or strategy conceived to answer the investigation questionings (Miller and Miller, 1993).

There are specific methods for the application of an investigation design. These methods are applied to systems where one or more factors (independent variables) are varied, and one or more responses (dependent variables) are determined, which also, may be influenced by other variables that cannot be controlled. One of these methods is the so-called traditional method of experimentation, but since one factor is varied at a time, it is not known if the change in the response is due to changes in this factor or changes in other factors (Box et al., 1989).

Alternatively to the traditional method, the experimental design emerges, which is a methodology that allows to obtain a series of experiments in which the simultaneous changes complement each other, thus obtaining, with statistical confidence and at a low cost, the information being sought by combining the results of all experiments (Massart et al., 1997).

Once all factors that may have influence in the response of interest have been identified, it may occur that a large number of them should be considered in further experiments, becoming necessary a refinement to reduce the number of experiments to be developed, according to their influence on the response. By running a screening experimental design, it can be determined which factors have more influence on the response and which do not. A 2-level factorial design serves this purpose and requires 2^k runs to build a model, where k is the number of factors. Unfortunately, when more than three factors have to be considered, the number of runs increases rapidly (Olivieri and Magallanes, 2010). That is the reason why, as an alternative, fractional factorial designs emerge, which allow to evaluate factors efficiently using a small fraction of the experiments of the full factorial design, since it considers that interactions of order three and higher are negligible. However, the possibility of independently estimating each model term is lost, and only estimations of confounded effects can be obtained (Wu and Hamada, 2000).

Another altenative to the 2-level full factorial design is the Plackett–Burman design (PB), which is a 2-level fractional factorial design, used to study $K = N - 1$ factors in N experiments, where N is a multiple of 4 (Plackett and Burman, 1946). As an example, a 12-experiment PB allows to study up to 11 factors, while a two-level

* Corresponding author at: Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Paraje "El Pozo", Ruta Nacional No. 168 S/N, S3000ZAA Santa Fe, Argentina. Tel.: +54 342 4575205; fax: +54 342 4575 221.
E-mail address: hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

0960-8524/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2011.09.015

full factorial design requires 2048 runs. The model used for PB is first order in each factor as follows:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + e \quad (1)$$

where y stands for the response, β_0 is the independent coefficient, β_i is the coefficient associated to each factor x_i, and e collects the model error. As can be seen, the classic PB analysis can only estimate the main factors with a reduced number of experiments, but each of them has as *partial alias*, all 2-factors interactions where that main factor in question is not involved. The validity of a PB lies in the assumption that all the interaction terms are negligible. This validity is questionable since it is present the risk of misevaluating some factors (e.g.: missing an important effect, incorrectly considering an irrelevant effect) when the real effect lies on the interaction terms, which have been considered negligible (Montgomery, 1991).

As an alternative to the classic PB analysis, very recently Olivieri and Magallanes (2010) developed an approach to identify significant main factors and interactions, based on a genetic algorithm (GA) approach. This algorithm is capable to evaluate interactions according to Eq. (2):

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_i j x_i x_j + e \quad (2)$$

where $\beta_{ij} (i \neq j)$ are two-factor interaction terms and x_i, x_j are the model factors.

A GA makes a population of individuals to evolve by submitting it to random actions, which are those acting in biological evolution (e.g.: mutations, genetic recombinations) as well as to a selection process according to some criteria, according to which it is decided which are the better adapted individuals, thus surviving, and which are the worse ones so as to be discarded.

In the present case, an initial random population is generated, arranged in a group of chromosomes, which represents the possible solutions to the problem. Then, each chromosome is ranked according to a given objective function to be minimized, and only half of them having the best figures of merit is allowed to survive, mutate and recombine to generate offspring. The other half is discarded. After a number of generations in which the above steps are repeated, the final best chromosome, i.e.: the one leading to the minimum value for the objective function, is employed for model building (Olivieri and Magallanes, 2010).

Since many factors are involved in the hydrolysis of lignocellulosic materials, it becomes necessary to perform a screening in order to select only the significant ones. These lignocellulosic materials stand for the 82% of the worldwide biomass produced in a year (Almeida e Silva et al., 2003), and its industrialization generates large quantities of residues. Therefore, legislations and environmental issues are forcing industries to reduce these residuals by recovery and recycling strategies (Spigno et al., 2008). Owing to the fact of being constituted by polysaccharides (hemicelluloses and cellulose), these materials stand for a renewable source of many sugars, e.g.: glucose, xylose, arabinose (Bower et al., 2008), which can be obtained by hydrolysis processes.

There are two ways to hydrolyze polysaccharides: chemically (by the action of acids or bases) and biochemically (by the action of enzymes). Although chemical hydrolysis processes save time, the enzymatic ones require less energy and milder conditions. Thus, the overall cost of the enzymatic process is lower when compared to alkaline or acid hydrolysis (Sun and Cheng, 2002). In addition, the enzymatic hydrolysis is substrate-specific without byproduct formation, contrary to the chemical hydrolysis, which generates byproducts, such as furfural and 5-hydroxymethylfurfural, as a consequence of the degradation of pentoses and hexoses, respectively, which may be inhibitory for microorganisms and enzymes (Bower et al., 2008; Sun and Cheng, 2002).

The sugars obtained through hydrolysis, may be used for various purposes. Glucose and xylose can be converted into ethanol by microbial fermentation to produce bioenergy (Iranmahboob et al., 2002). Xylose, for example, can be used as a carbon and energy source in fermentation processes, or bioconverted to xylitol, which is a polyol with important applications as a sweetener (Aguilar et al., 2002). Moreover, glucose can be included in culture media formulations for microorganisms capable of producing the so-called *Single Cell Oil* (Ratledge, 2004). This oil, through a subsequent monoalcohol esterification process can be transformed into biodiesel, the main liquid biofuel (Ma and Hanna, 1999). Such alternative is very promising to reduce the use of fossil fuels and also to avoid the distraction of oleaginous plant food in fuels production (Duffield, 2007).

In the present report, we describe and compare the results obtained when applying classic PB analysis and a GA based strategy in order to identify the factors that are truly significant in the chemical and enzymatic hydrolysis of three lignocellulosic feedstocks: corn and wheat bran, and pine sawdust. The results to be described below show the importance of considering interactions when estimating factors in a screening experimental design, and can be very useful in future research works on recycling strategies.

2. Methods

2.1. Raw materials

Corn and wheat bran, and pine sawdust were gently provided by Marchisio-Fernandez S.R.L., Santa Fe, Argentina. Each feedstock was air-dried, milled, homogenized in a single lot and stored under dry conditions before use. The feedstocks were milled in a Wiley knife mill (Standard Model No. 3, Arthur H. Thomas, Philadelphia, USA) to pass through a 1.0 mm screen. In a further step, the milled feedstocks were passed through a 0.5 mm screen, thus obtaining 2 batches for each feedstock (one containing particles between 0.5 mm and 1.0 mm and the other one, particles with a size less than 0.5 mm).

2.2. Pretreatments

2.2.1. Pretreatment of feedstocks prior to chemical hydrolysis

Five grams of feedstock were mixed with 100 mL of ammonium hydroxide 2.9 mol L^{-1} in a 250 mL Erlenmeyer flask and incubated in an orbital shaker at 200 rpm for 24 h at 30 °C. After this process, each residue was washed with distilled water to achieve the complete removal of the ammonium hydroxide. Then, the solid fraction was separated from the liquid by centrifugation at 5000 rpm for 10 min (Spigno et al., 2008). After, the solid was dried at 100 °C to constant weight. The dried solid obtained was used for chemical hydrolysis.

2.2.2. Pretreatment of feedstocks prior to enzymatic hydrolysis

One grame of feedstock was mixed with 20 g of 70:30 glycerol:water mixture (%m/m) and 120 μL of sulfuric acid 1% m/V in a 250 mL Erlenmeyer flask. Each mixture was heated at 220 °C with an electric heater for three hours (Sun and Chen, 2008). After this process, each residue was washed with 300 mL of boiling water and filtered using a Buchner funnel (Demirbaş, 1998). Samples were not dried prior to enzyme digestibility to avoid pore collapse that can occur in the micro-structure of the biomass leading to decreased enzymatic release of glucose from the cellulose (Brown and Torget, 1996).

2.3. Hydrolysis processes

2.3.1. Chemical hydrolysis

The residues of pre-treated feedstocks were chemically hydrolyzed using solutions of sulfuric acid. In each experiment, the mass of feedstock was mixed with the acid solution in 15 mL closed polypropylene tubes. Each mixture was incubated for 30 or 60 min at 60 °C or 100 °C by dipping the tubes in a water bath. After the time of hydrolysis was complete, the liquid fraction was recovered by centrifugation at 5000 rpm for 10 min plus further filtration with filter paper. All liquid fractions recovered were stored at −18 °C until sugars quantitation. A control assay was made using filter paper to take into account any contribution of this material to the sugars concentration that could occur in the filtration step.

2.3.2. Enzymatic hydrolysis

The residues of pre-treated feedstocks were enzymatically hydrolyzed by a cellulase complex from *Trichoderma longibrachiatum* purchased from Sigma (C9748), with an activity of 74.57 FPU/mg of powder, determined according to Ghose (1987). One international filter paper unit (FPU) was defined as the amount of enzyme that releases 1 μmol of reducing sugar per minute during hydrolysis reaction (Ferreira et al., 2009). The quantities of enzyme, substrate and polyethyleneglycol (PEG) according to PB were mixed with 10 mL of buffer (acetate or citrate, 0.05 or 0.20 mol L^{-1}) in a 50 mL closed polypropylene tube. Each mixture was incubated for 24 or 72 h at 50 °C by dipping the tubes in a water bath. Stirring was performed on a rotary shaker equipped with a thermostatic chamber. After the time of hydrolysis was complete, the liquid fraction was recovered by centrifugation at 5000 rpm for 10 min plus further filtration with filter paper. All liquid fractions recovered were stored at −18 °C until sugars quantitation. A control assay was made using filter paper to take into account any contribution of this material to the sugars concentration that could occur in the filtration step.

2.4. Plackett–Burman design and genetic algorithm approach

PB were introduced in this study to identify which factors have a significant effect on chemical and enzymatic hydrolysis. According to these designs, each variable was examined at two levels: −1 (low level) and +1 (high level) (Plackett and Burman, 1946).

Since the PB does not consider the interaction effects among variables, a GA approach developed by Olivieri and Magallanes (2010) was used to identify those interactions in order to perform a further comparison of both analyses. In the present work, the population size and the number of generations were estimated by trial and error, thus they were set as 20 chromosomes (initialized with random binary digits with 20% probability for 1 and 80% for 0 values) and 100 generations, respectively. For recombination, a 50% probability was employed and a probability of 0.05 was applied to mutations after offspring were produced. The ratio of the root mean square error corresponding to the fit of the experimental responses to the current model (RMSE) and the one corresponding to the classic PB analysis (RMSE0), was the objective function to be optimized.

Before starting the GA, several data handling were made (see Olivieri and Magallanes (2010)). All GA calculations were repeated ten times for each of the analyzed cases and a histogram was built registering the average value of coefficient terms over the ten GA cycles, and finally the terms having average coefficients larger than a certain tolerance (usually 0.05) were selected and included in a final model.

In the present work, 6 and 10 factors were screened for chemical and enzymatic hydrolysis, respectively. When one factor has a positive effect on the response, it means that when the level of that factor increases from −1 to +1, the value of the response increases, while a negative effect stands for the opposite case.

2.5. Analytical

The glucose concentration was enzimatically measured by using a commercial kit (Wiener Lab., Argentina). This quantitation method consists of two steps: first, according to Eq. (3), the glucose oxidase catalyzes the oxidation reaction of glucose to gluconic acid, with the consequent consumption of oxygen and water, and the generation of hydrogen peroxide

$$C_6H_{12}O_6 + O_2 + H_2O \rightarrow C_6H_{12}O_7 + H_2O_2 \quad (3)$$

In the second step, according to Eq. (4), a peroxidase catalyzes the reaction between two molecules of hydrogen peroxide with phenol and 4-aminophenazone to generate four molecules of water and a colored compound known as 4-(p-benzoquinone monoimine)-phenazone, which has an absorption maximum at 505 nm.

$$C_6H_6O + 2H_2O_2 + C_{11}H_{13}N_3O \rightarrow C_{17}H_{15}N_3O_2 + 4H_2O \quad (4)$$

The concentration of reducing sugars was measured by using a chemical quantitation technique (Miller, 1959).

All the collected data were transferred to a PC Intel Celeron D for their further interpretation. Design Expert™ version 8.05.0 (Stat-Ease, Inc., Minneapolis, USA, 2010) was used to perform experimental design and Matlab R2008a (The MathWorks, Inc.) to perform data analysis by the GA approach. The Matlab routine was kindly provided by Prof. Olivieri.

3. Results and discussion

In order to select significant factors in both chemical and enzymatic hydrolysis of wheat and corn bran, and pine sawdust, three PB were built (one for each feedstock), each one consisting in 12 experiments, for each hydrolysis treatment. Both the classic PB and the GA analyses were applied to achieve significant factors selection.

In the case of chemical hydrolysis, the 6 factors evaluated were: time of hydrolysis (Ti), temperature of hydrolysis (Te), pretreatment (P), feedstock particle size (FS), sulfuric acid concentration (A), and acid solution/feedstock ratio (AF). On the other hand, the 10 factors evaluated for enzymatic hydrolysis were: pH, formal buffer concentration (FBC), buffer type (BT), stirring (S), time of hydrolysis (Ti), feedstock concentration (F), enzyme loading (E), pretreatment (P), feedstock particle size (FS), and PEG loading (PC). In both cases, two responses were measured: concentrations (in gL^{-1}) of glucose (G) and reducing sugars (RS). Tables 1 and 2 summarize the 12 experiments and the concentrations of G and RS obtained for each of the three lignocellulosic residues, for chemical and enzymatic hydrolysis, respectively.

Since the number of PB runs is not enough to work out the complete system of equations for more than four factors, the classic analysis can only estimate the parameters corresponding to the main factors, which are confused with the interacting terms thus leading to incorrect results in reference to the significance of coefficients. Once the terms were obtained through the application of the least squares method (Montgomery, 1991), a root mean square error (RMSE0) was calculated for each response, by considering actual and predicted responses values. This RMSE0 corresponded to the fit of the experimental responses to the current model.

In contrast to the classic analysis, GA analysis includes not only the main factors but also their two-factor interactions, and therefore leads to more reliable results. An initial population was produced in

Table 1
Plackett–Burman design built to find the significant factors in the chemical hydrolysis of corn and wheat bran, and pine sawdust.

Experiment	Factors[a]						Responses[b]					
	Ti (min)	Te (°C)	P	FS (mm)	A (% m/m)	AF (g acid/g residue)	G (g L^{-1})			RS (g L^{-1})		
							CB	WB	PS	CB	WB	PS
1	30	60	No	<0.5	10	5	0.69	1.89	0.36	24.66	24.09	2.34
2	30	60	Yes	<0.5	30	11	0.46	0.43	0.24	37.57	60.35	16.14
3	60	100	No	<0.5	10	11	27.56	33.19	3.12	52.71	53.33	18.59
4	60	60	Yes	0.5–1	30	5	0.35	0.37	0.25	111.36	96.63	28.37
5	30	100	No	0.5–1	30	5	0.35	0.30	0.16	85.48	55.95	28.86
6	60	60	No	<0.5	30	5	0.28	1.12	0.20	55.29	111.74	20.93
7	30	100	Yes	<0.5	30	11	0.28	0.29	0.18	40.59	94.67	21.02
8	60	100	Yes	<0.5	10	5	47.67	60.86	3.79	86.16	122.23	26.93
9	60	100	No	0.5–1	30	11	0.07	0.32	0.25	43.16	34.51	21.67
10	30	100	Yes	0.5–1	10	5	46.69	47.99	2.19	116.09	98.51	23.45
11	60	60	Yes	0.5–1	10	11	0.56	0.34	0.17	17.27	17.09	1.55
12	30	60	No	0.5–1	10	11	0.46	1.12	0.30	10.91	14.19	1.04

[a] Ti: time of hydrolysis, Te: temperature of hydrolysis, P: pretreatment, FS: feedstock particle size, A: sulfuric acid concentration, AF: acid solution/feedstock ratio.
[b] G: concentration of glucose, RS: concentration of reducing sugars, CB: corn bran, WB: wheat bran, PS: pine sawdust.

Table 2
Plackett–Burman design built to find the significant factors in the enzymatic hydrolysis of corn and wheat bran, and pine sawdust.

Experiment	Factors[a]										Responses[b]					
	pH	FBC (mol L^{-1})	BT[c]	S	Ti (h)	F (% m/V)	E (FPU/g feedstock)	P	FS (mm)	PC (mg PEG/g feedstock)	G (gL^{-1})			RS (gL^{-1})		
											CB	WB	PS	CB	WB	PS
1	6.0	0.20	Cit	No	24	10.0	5.0	Yes	0.5–1	300.0	2.64	5.20	1.38	6.81	24.74	7.26
2	6.0	0.20	Cit	Yes	72	10.0	5.0	No	<0.5	0.0	3.57	2.48	1.62	20.50	17.57	11.31
3	4.0	0.20	Ace	No	72	10.0	40.0	No	0.5–1	0.0	25.36	11.87	16.16	44.41	41.06	34.59
4	4.0	0.05	Ace	No	72	10.0	5.0	Yes	<0.5	300.0	1.67	1.38	1.29	7.31	6.47	5.00
5	4.0	0.05	Cit	Yes	24	10.0	40.0	No	<0.5	300.0	14.68	12.13	15.08	40.44	35.06	34.63
6	6.0	0.20	Ace	No	24	2.0	40.0	No	<0.5	300.0	0.75	2.76	2.39	9.39	15.03	7.48
7	4.0	0.20	Cit	Yes	72	2.0	40.0	Yes	0.5–1	300.0	4.11	2.57	1.87	10.35	10.40	7.04
8	6.0	0.05	Ace	Yes	72	2.0	5.0	No	0.5–1	300.0	0.59	N/D	0.19	4.49	0.83	2.14
9	4.0	0.05	Cit	No	24	2.0	5.0	No	0.5–1	0.0	0.24	0.63	0.36	2.04	3.48	1.61
10	6.0	0.05	Ace	Yes	24	10.0	40.0	Yes	0.5–1	0.0	10.19	10.76	9.39	39.47	42.54	33.85
11	4.0	0.20	Ace	Yes	24	2.0	5.0	Yes	<0.5	0.0	0.17	0.24	0.24	1.21	0.97	0.86
12	6.0	0.05	Cit	No	72	2.0	40.0	Yes	<0.5	0.0	0.50	0.75	1.48	6.50	5.96	6.22

[a] FBC: formal buffer concentration, BT: buffer type, S: stirring, Ti: time of hydrolysis, F: concentration of feedstock, E: enzyme loading, P: pretreatment, FS: feedstock particle size, PC: polyethyleneglycol loading.
[b] G: concentration of glucose, RS: concentration of reducing sugars, CB: corn bran, WB: wheat bran, PS: pine sawdust.
[c] Cit: citrate buffer, Ace: acetate buffer.

the form of an $(X \times Y)$ random binary matrix, where X is the total number of terms in Eq. (2) (main and two-factor associations), and Y is a predetermined number of chromosomes (binary strings containing genes which encode the experimental variables) (Olivieri and Magallanes, 2010). Then, for each response and for each one of the different Y initial models, a RMSE was calculated, which corresponded to the fit of the experimental responses to the current model. Afterward, each model was ranked according to the RMSE/RMSE0 ratio, which was the objective function to be minimized. The final best chromosome was the one leading to the minimum value for the objective function, in other words, the model that better fitted the responses in comparison to the classic analysis.

Finally, the comparison of the R^2 values corresponding to both models was done in order to verify that the model obtained by GA was better than that yielded by the application of the classic PB analysis.

3.1. Chemical hydrolysis

3.1.1. Pine sawdust

The classic PB analysis gave Te (positive effect) and A (negative effect) as significant factors for G response, with an associated probability value (p) of 0.027 and 0.021, respectively. The value of $R^2 = 0.623$ indicated that the model explained 62% of the variability in the responses, the remaining 38% being explained by the residue.

In the case of RS response, significant factors were Te ($p = 0.004$) (positive effect), A ($p = 0.008$) (positive effect) and AF ($p = 0.022$) (negative effect). This result gave a model that had a $R^2 = 0.818$.

On the other hand, applying the GA method, Te ($p = 2 \times 10^{-4}$) (positive effect) and A ($p = 1 \times 10^{-4}$) (negative effect) were found as significant factors for G response, with the addition of a significant interaction between them (Te/A) ($p = 4 \times 10^{-4}$) (negative effect), yielding a $R^2 = 0.946$. On the other hand, for RS response, Te ($p = 2 \times 10^{-4}$) (positive effect), A ($p = 1 \times 10^{-4}$) (positive effect) and AF ($p = 0.037$) (negative effect) resulted to be the significant factors, plus the interaction Te/A ($p = 0.004$) (negative effect), with a determination coefficient of $R^2 = 0.966$. Fig. 1A and B shows the histograms built registering the average value of the coefficient terms over the 10 GA cycles, for G and RS responses, respectively.

Consequently, the R^2 values for both G and RS responses were improved in a 51.8% and 18.1%, respectively, applying the GA approach. This considerable enhancing of the modeling can be attributed to the fact that Te/A interaction has been taken into account, for being significant its effect, by the GA approach.

3.1.2. Corn Bran

The application of the classic PB analysis to the data corresponding to corn bran in Table 1 showed that Te ($p = 0.024$) (positive

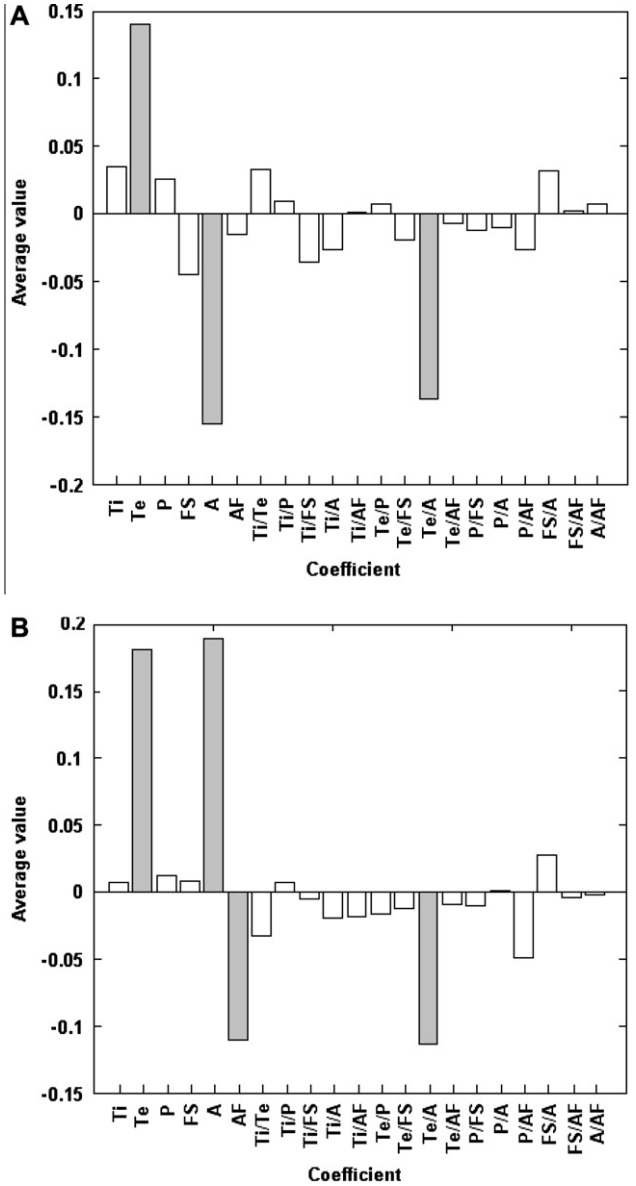

Fig. 1. Histograms built registering the average value of the coefficient terms over the 10 GA cycles performed, for (A) G response, and (B) RS response, for the chemical hydrolysis of pine sawdust. Ti: time of hydrolysis, Te: temperature of hydrolysis, P: pretreatment, FS: feedstock particle size, A: sulfuric acid concentration, AF: acid solution/feedstock ratio. Combinations factor/factor stand for the interaction between 2 main factors.

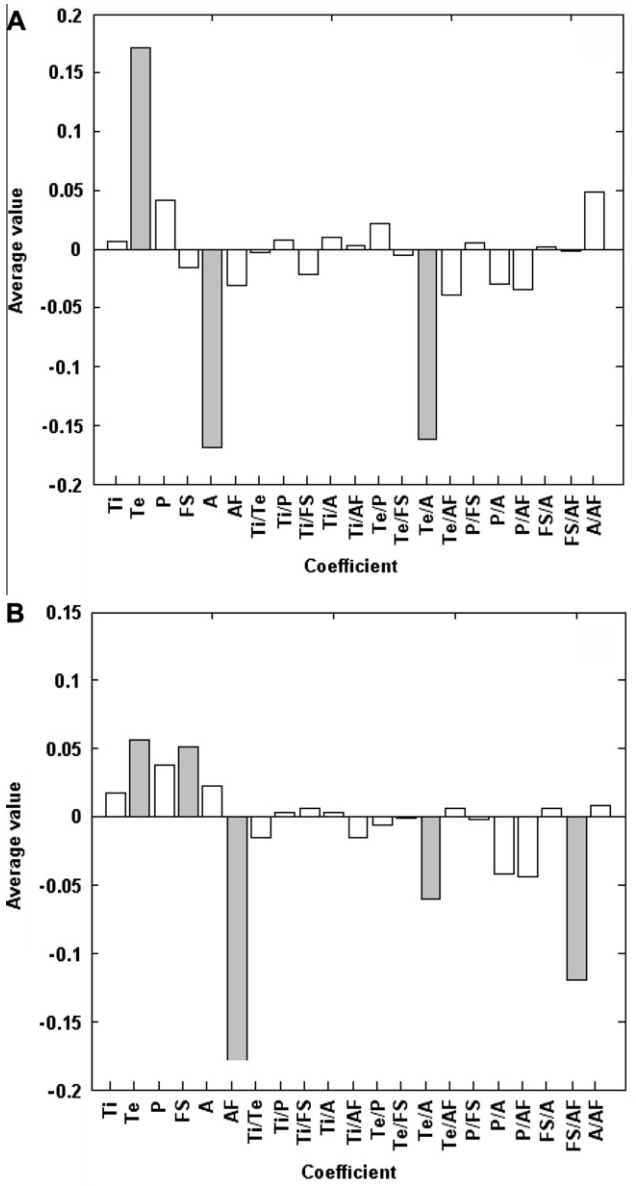

Fig. 2. Histograms built registering the average value of the coefficient terms over the ten GA cycles performed, for (A) G response, and (B) RS response, for the chemical hydrolysis of corn bran. Ti: time of hydrolysis, Te: temperature of hydrolysis, P: pretreatment, FS: feedstock particle size, A: sulfuric acid concentration, AF: acid solution/feedstock ratio. Combinations factor/factor stand for the interaction between 2 main factors.

effect) and A ($p = 0.022$) (negative effect) were the significant factors for G response, leading to a $R^2 = 0.624$. As can be seen, like in the previous analysis, the explained variance by the model was indicative of a rather poor fit.

For RS response, when applying the classic PB analysis, Te ($p = 0.049$) (positive effect) and AF ($p = 0.005$) (negative effect) were the significant factors with a $R^2 = 0.749$.

When the GA approach was applied to this experimental system, for G response, it was found that Te ($p = 4 \times 10^{-4}$) (positive effect) and A ($p = 2 \times 10^{-4}$) (negative effect) were the significant factors, but also the interaction between them (Te/A) (negative effect) with a $p = 4 \times 10^{-4}$. This led to a $R^2 = 0.958$, showing an improvement of 53.5% on the fitting of the model, in comparison to the classic PB analysis.

When the GA approach was used to analyze RS response, interesting results were obtained concerning main factors and their associations. Similarly to the classic PB analysis, Te ($p = 0.020$) (positive effect) and AF ($p = 8 \times 10^{-4}$) (negative effect) were found as significant factors, but also FS ($p = 0.019$) (positive effect) resulted to be another significant factor. In addition, the interactions Te/A ($p = 0.018$) (negative effect) and FS/AF ($p = 0.038$) (negative effect) were also significant, yielding a model $R^2 = 0.966$, which implied an improvement of 28.9% in the fitting, compared to the classic PB analysis. Fig. 2A and B shows the histograms built registering the average value of the coefficient terms over the ten GA cycles, for G and RS responses, respectively.

3.1.3. Wheat bran

Considering G response, the classic PB analysis yielded Te ($p = 0.041$) (positive effect) and A ($p = 0.036$) (negative effect) as the significant factors implying a $R^2 = 0.584$, indicating that this was a poor fit. While, for RS response, the only factor that seemed

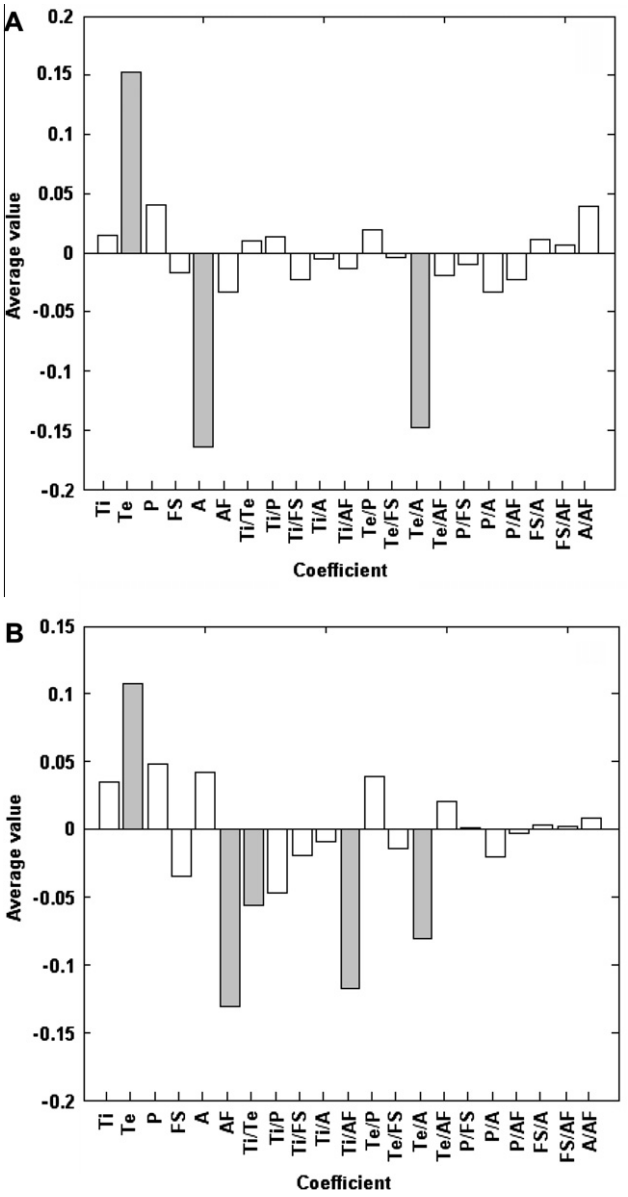

Fig. 3. Histograms built registering the average value of the coefficient terms over the ten GA cycles performed, for (A) G response, and (B) RS response, for the chemical hydrolysis of wheat bran. Ti: time of hydrolysis, Te: temperature of hydrolysis, P: pretreatment, FS: feedstock particle size, A: sulfuric acid concentration, AF: acid solution/feedstock ratio. Combinations factor/factor stand for the interaction between 2 main factors.

to be significant was AF ($p = 0.036$) (negative effect), which gave a $R^2 = 0.509$, representing the worst fit among all the cases analyzed.

The analysis by GA showed that Te ($p = 2 \times 10^{-4}$) (positive effect) and A ($p = 2 \times 10^{-4}$) (negative effect) were the only two significant factors in the case of G response, confirming the outcome of the classic PB analysis, though the interaction between them (Te/A) (negative effect) was also significant ($p = 5 \times 10^{-4}$), giving a $R^2 = 0.939$, implying a 60.8% improvement in comparison to the classic PB analysis.

For RS response, once again remarkable results were obtained: apart from AF ($p = 0.018$) (negative effect), also Te ($p = 0.001$) (positive effect) was found as a significant factor. Moreover, three interactions were also significant: Ti/Te ($p = 0.021$) (negative effect), Ti/AF ($p = 0.001$) (negative effect) and Te/A ($p = 0.004$) (negative effect). Then, the R^2 for this particular model was 0.947, which represents an 86.1% increase in the fit. Fig. 3A and B shows the histograms built registering the average value of the coefficient terms over the ten GA cycles, for G and RS responses, respectively. It should be noted the fact that, as in the previous case, new factors should be considered after the GA analysis. This indicated the robustness of this approach, which was able to detect factors which had not been taken into account by the classic PB analysis.

3.1.4. Further analysis

In the pine sawdust case, both PB and GA methods led to the same conclusions due to the fact that the interactions found to be significant by GA occured between significant factors, which also were found by applying the classic PB analysis.

Interesting results were obtained in the corn bran case. The classic PB analysis would have led to a wrong factors selection, which should be taken into account in a further optimization step. This was because GA analysis clearly showed that FS (found as not significant in classic PB analysis for RS response) was a significant factor, not only its linear contribution but also its interaction with AF, thus it must be considered when optimizing the chemical hydrolysis of corn bran.

The case of wheat bran was similar to that of corn bran. Again, the classic PB analysis led to the loss of one significant factor: Ti, which would not be considered in a subsequent optimization phase. The GA analysis demonstrated that, even if Ti was not significant as a main factor, it was involved in two interactions, with Te and AF, respectively.

Moreover, in all the three analyzed cases, the GA approach allowed to obtain important improvements in both responses, in regards to the fit, reflected by the R^2 values.

Now, taking into account the significance of factors, pretreatment resulted to be non-significant for glucose concentration nor for reducing sugars concentration, for the three feedstocks. This might be explained by the fact that the substrate was previously subjected to another previous pretreatment: milling. Then, for these substrates and hydrolysis conditions, this physical pretreatment was enough to break the physical barrier that lignin stands for, in order to obtain acceptable sugar yields, with no need of putting the milled substrates through a subsequent chemical pretreatment, contrary to what was obtained by Spigno et al. (2008), working on grape stalks.

Ti was non-significant in almost all cases, except for wheat bran in RS response. In this case, two interactions envolving Ti (with Te and AF) were significant carrying negative effects. Iranmahboob et al. (2002), working with a mixture of hardwood and softwood in similar conditions to this study, found that the time of hydrolysis was non-significant on sugar yields, if it was varied between 0.5 and 1.0 h. Vieira Canettieri et al. (2007), working on *Eucalyptus grandis* wood, found that the interaction between time of hydrolysis and temperature has a negative effect on sugar yields, due to its chemical degradation. However, contrary to what was found in this study, no effect was found for acid solution/feedstock ratio.

With regard to Te, this factor and its interaction with A resulted to be significant in all the analyzed cases. Temperature is expected to have a positive effect since it favors the rupture of heterocyclic ether bonds in the polysaccharides caused by protons (Aguilar et al., 2002). Bower et al. (2008) also found that an interaction between temperature and acid concentration exerted a negative effect on sugar yields, what could be explained, again, by sugars degradation to furfural and 5-hydroxymethylfurfural, mainly (Chotěborská et al., 2004).

FS was only significant for RS response only in corn bran case, being non-significant in all other cases. Generally, reducing the feedstock particle size ought to improve mass transfer and increase the available area to the acid reagent (Guo et al., 2008). However, this effect was not obvious, and in corn bran case, the opposite effect was observed: reducing the feedstock particle size decreased

sugar yields. Moreover, an interaction between FS and AF was evidenced with a negative effect on sugar yield. It could be possible that, for the particular case of corn bran, a larger feedstock particle size favors a faster liberation of sugars, but the degradation is also faster when the AF factor increases.

Since a higher AF value implies a higher acid concentration, A and AF significances will be analyzed together. The sulfuric acid concentration was found significant in all the cases for G response with a negative effect; while AF was found significant only for RS responses, but in all cases, it has a negative effect. This behavior can be explained taking into account the following: at high acid concentrations, the speed at which sugars degrade to furanes increases to the extent that it can be 10-times the speed at which polysaccharides depolymerize, especially for hemicelluloses, producing the depletion of sugars yield (Sanchez et al., 2004). For RS response, only in the pine sawdust case, contrarily, A exhibited a positive effect, but Rahman et al. (2007) found the same effect on oil palm empty fruit bunch acid hydrolysis. Thus, the results may vary depending on the characteristics of the biomass being studied.

Finally, differences in effects on the responses G and RS were observed when factors A and AF were analyzed for responses corn bran and pine sawdust. This behavior can be attributed to that both responses were measured using analytical methods which present differences in their specificity.

3.2. Enzimatic hydrolysis

3.2.1. Pine sawdust

According to the classic PB analysis, only 2 factors among the 10 evaluated were significant for both responses: F (positive effect) and E (positive effect). The corresponding p values, in the case of G response, were 0.011 and 0.008 for F and E, respectively; while for the case of RS response, the p value for F (positive effect) was 0.001, and 0.002 for E (positive effect). The value of the determination coefficients for both responses were $R^2 = 0.708$ and $R^2 = 0.819$ for G and RS responses, respectively.

When data were analyzed applying the GA approach, no differences were found to the previous analysis. Again, E and F resulted to be the significant factors for both G and RS responses. In the case of G response, the p values obtained were 3×10^{-4} and 8×10^{-5} for F (positive effect) and E (positive effect), respectively; while, for RS response, they were 2×10^{-5} both for F (positive effect) and E (positive effect). The main difference to classic PB analysis was that, for both responses, the interaction between E and F, E/F (positive effect), resulted to be significant ($p = 5 \times 10^{-4}$ and $p = 2 \times 10^{-5}$ for G and RS, respectively). Thus, both fits were improved: 33.5% in the case of G response ($R^2 = 0.945$) and 20.8% in RS response case ($R^2 = 0.989$).

3.2.2. Corn bran

In analogy to the case of pine sawdust, only 2 factors resulted to be significant for both responses, according to the classic PB analysis: F ($p = 0.020$ and $p = 0.001$ for G and RS responses, respectively) (positive effect) and E ($p = 0.031$ for G, and $p = 0.004$ for RS) (positive effect). Thus leading to determination values of $R^2 = 0.617$ and $R^2 = 0.799$ for G and RS responses, respectively.

When applying the GA analysis, again a similar result was obtained. For G response, F ($p = 0.003$) (positive effect), E ($p = 0.006$) (positive effect) and their interaction E/F ($p = 0.020$) (positive effect) were the significant factors affecting the response, and the corresponding $R^2 = 0.800$ implied an improvement of 29.7% with respect to the classic PB analysis. In the case of RS response, the analysis was quite similar due to the fact that again F ($p = 2 \times 10^{-5}$) (positive effect), E ($p = 4 \times 10^{-5}$) (positive effect) and E/F ($p = 7 \times 10^{-4}$) (positive effect) resulted to be the significant factors implying a $R^2 = 0.948$ which corresponds to an 18.6% in fit improvement.

As in the pine sawdust case, both classic PB and GA analysis identified the same significant factors for both responses.

3.2.3. Wheat bran

In reference to G response, the classic PB analysis outcome was that F ($p = 0.001$) and E ($p = 0.004$) were the only two significant factors giving a $R^2 = 0.806$. In the case of RS response, again the two significant factors were F ($p = 3 \times 10^{-4}$) (positive effect) and E ($p = 0.003$) (positive effect) yielding a $R^2 = 0.844$.

Analyzing the data with the GA approach, not much different results were obtained: F ($p = 2 \times 10^{-5}$) (positive effect) and E ($p = 4 \times 10^{-5}$) (positive effect) plus the interaction between them, E/F ($p = 5 \times 10^{-4}$) (positive effect) resulted to be the significant factors in the case of G response. The resulting $R^2 = 0.952$ showed an 18.1% of fit improvement with respect to the classic PB analysis.

If we now turn the attention to RS response, it has to be said that something particular occurred: the GA approach obtained the same conclusions of the classic PB analysis, i.e.: only F ($p = 2 \times 10^{-4}$) (positive effect) and E ($p = 0.001$) (positive effect) were the significant factors, with no interactions involved among factors. Thus, obviously, GA yielded the same value for the determination coefficient, i.e.: $R^2 = 0.844$.

3.2.4. Further analysis

Both classic PB and GA analyses identified the same significant factors for both responses in the three cases being evaluated. Then, if only the classic PB analysis would be applied, it would have led to a right selection of the factors to be taken into account in a further optimization step. This is because the interactions found to be significant by GA analysis, did not occur between factors whose linear contribution was not significant. Nevertheless, in all cases, except for RS in wheat bran, all the determination coefficients (R^2) were improved, between 18% and 33%, by the application of the GA approach, in comparison to the classic PB analysis. This could be attributed to the fact that in all of these cases, an interaction was missed by the classic PB analysis.

Only two factors were found as significants: feedstock concentration and enzyme loading, plus their interaction (except in one case). The concentrations of saccharides were increased as the feedstock concentration was augmented, and no reversal on this trend was evidenced, contrary to what was found by Jeya et al. (2009), who tested feedstock (rice straw) concentrations between 0.75% and 3.75% and encountered that the saccharification ratio was enhanced until a certain point beyond which the trend was reversed, and Ferreira et al. (2009) who also found a similar behavior working on rock-rose. Contrary to the study of Ferreira et al. (2009), who took as response the ratio of glucose mass released to dry biomass, in this study the responses were defined as concentrations. The reason for this is that there is a risk of underestimate that ratio since the feedstock may contain moisture, perhaps in a low proportion, but it would be introducing an error in the calculations. Furthermore, the greater the mass of feedstock, the greater the amount of water present. Then, from our point of view, concentrations would be more reliable and representative responses for the evaluation of the process.

The same behavior for feedstock concentration was found for the enzyme loading. Many studies have evaluated the influence of the cellulase loading in enzymatic hydrolysis. On the one hand, Chen et al. (2007) demonstrated that in order to obtain high saccharides yields from corncob, cellulase loadings up to 150 FPU/g feedstock were necessary. On the other hand, Lu et al. (2007) reached the optimum at a cellulase dosage of 22 FPU/g feedstock.

These studies demonstrate that different biomass subjected to different pretreatment and enzymatic hydrolysis reaction conditions, may lead to very different results.

The interaction between enzyme loading and feedstock concentration also has a positive effect on the responses, as might be expected since the enzyme adsorption on the substrate is the first step of the cellulose hydrolysis process (Bansal et al., 2009).

The buffer type, pH, formal buffer concentration, stirring, time of hydrolysis, pretreatment, feedstock particle size and PEG loading were non-significant factors for neither the glucose concentration nor the reducing sugars concentration.

Buffer type and buffer concentration resulted non-significant variables, which can be related with an optimum environment for enzyme activity since many possible combinations of this two factors (in the evaluated ranges) are described extensively in the bibliography for a large series of substrates. An exception is the study performed by Ferreira et al. (2009), who reported a negative effect on saccharides yield from broom hydrolysis when increasing the buffer concentration from 0.01 mol L^{-1} to 0.25 mol L^{-1}.

Time of hydrolysis and stirring being non-significant may be attributed to the enzyme inactivation or lack of stability after some time (Ferreira et al., 2009; Yang et al., 2009), or to the fact that the saccharides are completely released when 24 h of hydrolysis are reached.

In respect of PEG loading, its significance may depend on substrate lignin content: when the lignin content is high, the PEG loading may be more significant than in the opposite case (Ferreira et al., 2009). Lignin had been described to cause, by hydrophobic interactions, unspecific and non-productive binding with cellulases. Several studies reported improvements in the enzymatic hydrolysis of lignocellulosic biomass with the supplementation with surfactants (Kaar and Holtzapple, 1998) and polymers (Borjesson et al., 2007), which were described to fill hydrophobic regions in the lignin surface, preventing the adsorption of cellulases, thus producing a higher availability of enzymes for cellulose degradation (Borjesson et al., 2007). In this study, it can be said that this effect was not obvious since PEG loading resulted to be non-significant.

The non-significance of pH was an expected result since many studies have demonstrated that sugar yields are very similar when performing the enzymatic hydrolysis at pH 4 or pH 6, and that the maximum value can be reached at pH between 4.6 and 5.3. Jeya et al. (2009) applied response surface methodology in order to optimize the hydrolysis conditions of rice straw and found that the sugar yield has a parabolic dependence on pH, at least between 4 and 6, with the optimum ca. 5.3. Ferreira et al. (2009) obtained similar results working on rock-rose and broom, for which the optimum pH resulted to be 4.8 and 4.5, respectively. Tangarone et al. (1989) developed stability and activity studies on a cellulase isolated from *Trichoderma longibrachiatum*. They showed that the optimum pH for the enzyme activity was 4.8 and that beyond 4 and 6 decreases rapidly, while it was stable at pH between 4.5 and 9.3.

In reference to pretreatment, it has to be taken into account that the substrate was previously subjected to another previous pretreatment: milling. Thus, in this particular case, a physical pretreatment is enough to break the cellulose cristallinity in order to obtain acceptable sugar yields (Sun and Cheng, 2002), with no need of putting the milled substrates through a subsequent chemical pretreatment.

Generally, reducing the feedstock particle size should improve mass transfer and increase the accessible area available to the enzyme (Guo et al., 2008). Nevertheless, the effect of feedstock particle size on the yield of glucose and reducing sugars was not significant in this study and the results were similar with corn and wheat bran, and pine sawdust. These results suggested that feedstock particle size does not affect the performance of the enzymatic hydrolysis at an input size below 1 mm.

4. Conclusion

It can be concluded that the classic PB analysis, in some cases, leads to wrong significant factors identification, since through the application of a GA approach, it was demonstrated that even if a factor results to be non-significant, it may be involved in a significant interaction. Consequently, this factor should be included in a further optimization step.

Moreover, the GA approach, by identifying significant interactions between factors, generally improves the determination coefficients, i.e.: the experimental data could be better fitted, in comparison to the classic PB analysis.

Acknowledgements

The authors are grateful to Universidad Nacional del Litoral (Project CAI+D NO. 12-65 and CAI+D 2009 Tipo III R2) and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 2988) for financial support. P.C.G. thanks CONICET for his fellowship. The authors want to thank to Prof. Alejandro Olivieri for providing the Matlab routine.

References

Aguilar, R., Ramírez, J.A., Garrote, G., Vázquez, M., 2002. Kinetic study of the acid hydrolysis of sugar cane bagasse. J. Food Eng. 55, 309–318.

Almeida e Silva, J.B., Lima, U.A., Taqueda, M.E.S., Guaragna, F.G., 2003. Use of response surface methodology for selection of nutrient levels for culturing *Paecilomyces variotii* in eucalyptus hemicellulosic hydrolyzate. Bioresour. Technol. 87, 45–50.

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.

Borjesson, J., Engqvist, M., Sipos, B., Tjerneld, F., 2007. Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. Enzyme Microb. Technol. 41, 186–195.

Bower, S., Wickramasinghe, R., Nagle, N.J., Schell, D.J., 2008. Modeling sucrose hydrolysis in dilute sulfuric acid solutions at pretreatment conditions for lignocellulosic biomass. Bioresour. Technol. 99, 7354–7362.

Box, G.E.P., Hunter, W.G., Hunter, J.S., 1989. Estadística para experimentadores, Ed. Reverté S.A., Barcelona.

Brown, L., Torget, R., 1996. Enzymatic saccharification of lignocellulosic biomass. National Renewable Energy Laboratory (NREL), Golden, CO.

Chen, M., Xia, L.M., Xue, P.J., 2007. Enzymatic hydrolysis of corncob and ethanol production from cellulosic hydrolysate. Int. Biodeterior. Biodegrad. 59, 85–89.

Chotěborská, P., Palmarola-Adrados, B., Galbe, M., Zacchi, G., Melzoch, K., Rychtera, M., 2004. Processing of wheat bran to sugar solution. J. Food Eng. 61, 561–565.

Demirbaş, A., 1998. Aqueous glycerol delignification of wood chips and ground wood. Bioresour. Technol. 63, 179–185.

Duffield, J., 2007. Biodiesel: production and economic issues. Inhalation Toxicol. 19, 1029–1031.

Ferreira, S., Duarte, A.P., Ribeiro, M.H.L., Queiroz, J.A., Domingues, F.C., 2009. Response surface optimization of enzymatic hydrolysis of *Cistus ladanifer* and *Cytisus striatus* for bioethanol production. Biochem. Eng. J. 45, 192–200.

Ghose, T.K., 1987. Measurement of cellulase activities. Pure Appl. Chem. 59, 257–268.

Guo, G.L., Chen, W.H., Chen, W.H., Men, L.C., Hwang, W.S., 2008. Characterization of dilute acid pretreatment of silvergrass for ethanol production. Bioresour. Technol. 99, 6046–6053.

Iranmahboob, J., Nadim, F., Monemi, S., 2002. Optimizing acid-hydrolysis: a critical step for production of ethanol from mixed wood chips. Biomass Bioenergy 22, 401–404.

Jeya, M., Zhang, Y.W., Kim, I.W., Lee, J.K., 2009. Enhanced saccharification of alkali-treated rice straw by cellulase from *Trametes hirsuta* and statistical optimization of hydrolysis conditions by RSM. Bioresour. Technol. 100, 5155–5161.

Kaar, W.E., Holtzapple, M.T., 1998. Benefits from Tween during enzymic hydrolysis of corn stover. Biotechnol. Bioeng. 59, 419–427.

Lu, X.B., Zhang, Y.M., Yang, J., Liang, Y., 2007. Enzymatic hydrolysis of corn stover after pretreatment with dilute sulfuric acid. Chem. Eng. Technol. 30, 938–944.

Ma, F., Hanna, M.A., 1999. Biodiesel Production: a review. Bioresour. Technol. 70, 1–15.

Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J., 1997. Handbook of Chemometrics and Qualimetrics: Part A. Elsevier, Amsterdam.

Miller, G.L., 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. 31, 426–428.
Miller, J.N., Miller, J.C., 1993. Estadística y Quimiometría Para Química Analítica. Prentice Hall, Madrid.
Montgomery, D.C., 1991. Diseño y Análisis de Experimentos, Grupo Editorial S. A. de C. V., México.
Olivieri, A.C., Magallanes, J.F., 2010. The effect of factor interactions in Plackett–Burman experimental designs. Comparison of Bayesian–Gibbs analysis and genetic algorithms. Chemometr. Intell. Lab. Syst. 102, 8–14.
Plackett, R.L., Burman, J.P., 1946. The design of optimum multifactorial experiments. Biometrika 33, 305–325.
Rahman, S.H.A., Choudhury, J.P., Ahmad, A.L., Kamaruddin, A.H., 2007. Optimization studies on acid hydrolysis of oil palm empty fruit bunch fiber for production of xylose. Bioresour. Technol. 98, 554–559.
Ratledge, C., 2004. Fatty acid biosynthesis in microorganisms being used for single cell oil production. Biochimie 86, 807–815.
Sanchez, G., Pilcher, L., Roslander, C., Modig, T., Galbe, M., Liden, G., 2004. Dilute-acid hydrolysis for fermentation of the Bolivian straw material Paja Brava. Bioresour. Technol. 93, 249–256.
Spigno, G., Pizzorno, T., De Faveri, D.M., 2008. Cellulose and hemicelluloses recovery from grape stalks. Bioresour. Technol. 99, 4329–4337.
Sun, Y., Cheng, J., 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour. Technol. 83, 1–11.
Sun, F., Chen, H., 2008. Enhanced enzymatic hydrolysis of wheat straw by aqueous glycerol pretreatment. Bioresour. Technol. 99, 6156–6161.
Tangarone, B., Royer, J.C., Nakas, J.P., 1989. Purification and characterization of an Endo-(1, 3)-3-D-glucanase from *Trichoderma longibrachiatum*. Appl. Environ. Microbiol. 55, 177–184.
Vieira Canettieri, E., Moraes Rocha, G.J., Carvalho, J.A., Almeida e Silva, J.B., 2007. Optimization of acid hydrolysis from the hemicellulosic fraction of *Eucalyptus grandis* residue using response surface methodology. Bioresour. Technol. 98, 422–428.
Wu, J.C.F., Hamada, M., 2000. Experiments, Planning, Analysis, and Parameter Designs Optimization. Wiley-Interscience Publications, New York.
Yang, J., Zhang, X., Yong, Q., Yu, S., 2009. Three-stage hydrolysis to enhance enzymatic saccharification of steam-exploded corn stover. Bioresour. Technol. 101, 4930–4935.