Optimization of the hydrolysis of lignocellulosic residues by using radial basis functions modelling and particle swarm optimization

Pablo C. Giordano[a b], Alejandro J. Beccaria[b], Héctor C. Goicoechea[a,*] and Alejandro C. Olivieri[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] *Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina.*

[*] Corresponding authors. E-mail addresses: olivieri@iquirconicet.gov.ar (A.C. Olivieri) and hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

Abstract

The concentrations of glucose and total reducing sugars obtained by chemical hydrolysis of three different lignocellulosic feedstocks were maximized. Two response surface methodologies were applied to model the amount of sugars produced: (1) classical quadratic least-squares fit (QLS), and (2) artificial neural networks based on radial basis functions (RBF). The results obtained by applying RBF were more reliable and better statistical parameters were obtained. Depending on the type of biomass, different results were obtained. Improvements in fit between 35 % and 55 % were obtained when comparing the coefficients of determination (R^2) computed for both QLS and RBF methods. Coupling the obtained RBF models with particle swarm optimization to calculate the global desirability function, allowed to perform multiple response optimization. The predicted optimal conditions were confirmed by carrying out independent experiments.

Keywords: Glucose, Modelling, Optimization, Artificial Intelligence, Particle swarm optimization, Radial basis functions.

1. Introduction

Experimentalists have several techniques available for finding optimal process conditions. These approaches vary from the traditional one-variable-at-a-time method to more complex statistical and mathematical techniques involving experimental designs, such as full and fractional factorial, and central composite designs, followed by optimization techniques such as the response surface methodology (RSM) [1].

Experimental design and RSM have been proved to be useful for developing, improving and optimizing processes, and have been extensively used in the industrial world [2–9] and in bioprocesses [10–16], including the formulation of culture media for bacteria and fungi [17–20].

When RSM is applied, the experimental responses are usually fitted to quadratic functions by least-squares (QLS). In most of the cases which have been studied by this methodology, a second-degree polynomic relation can reasonably approximate the behavior of the systems under study.

Artificial neural networks (ANN) represent another smart tool for non-linear multivariate modeling. The power of an ANN lies in its universal structure and in its ability to learn from historical data. Among the main advantages of ANN compared to QLS, the former do not require a prior specification of a suitable fitting function and have universal approximation capability, i.e. they can approximate almost all kinds of non-linear functions, including quadratic functions. QLS, on the other hand, is only useful for quadratic approximations; it should be noticed that more complex functions require a larger number of experiments [21]. QLS and ANN have been applied in diverse areas such as in the vehiculization of therapeutic drugs [22], and in the production of recombinant proteins [23,24,25], bioinsecticides [26], biopolymer scleroglucan [21], and endonuclease derived from recombinant *Escherichia coli* [27].

Artificial neural networks based on the use of radial basis functions (RBF) have been recently introduced for nonlinear multivariate function estimation and regression tasks [28]. RBF networks have a single hidden layer of neurons incorporating gaussian transfer functions, and a linearly activated output layer. In comparison with multi-layer perceptron (MLP) networks, RBF offer some advantages such as robustness towards noisy data as well as a faster training phase [29].

In the context of regression analysis, recent RBF publications which deserve to be cited describe applications to near-infrared analysis of organic matter in soils [30], glucose in blood [31], and water content in fish products [32]. In the field of optimization, RBF was used for the prediction of optimal culture conditions for maximum hairy root biomass yield [33].

In the present report, the RBF modeling power is complemented with a stochastic procedure for finding global minima called particle swarm optimization (PSO). This latter technique has been shown to successfully optimize a wide range of continuous functions [34], based on concepts loosely related to social interaction issues. It searches a space by adjusting the trajectories of individual vectors, called “particles”, while they move in a multidimensional space. The individual particles are drawn stochastically toward the positions of their own previous best performance and the best previous performance of their neighbours [35].

The combination RBF-PSO, which has been successfully applied by Liu et al. [36] and Kitayama et al. [37], is herein applied to optimize the conditions for the chemical hydrolysis of lignocellulosic feedstocks (corn bran, wheat bran and pine sawdust). The results show that the conditions reached by RBF-PSO are much more realistic than those obtained from QLS.

2. Materials and methods

2.1. Raw materials

Corn bran, wheat bran and pine sawdust were gently provided by Marchisio-Fernandez SRL, 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 sieve, 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 Hydrolysis process

Feedstocks were chemically hydrolyzed using solutions of sulphuric acid. In each experiment, the mass of feedstock was mixed with the acid solution in 15 mL closed polypropilene tubes. Each mixture was incubated at different temperatures and during different periods of time, according to the central composite designs (CCD) employed in this study. The incubation was performed 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 minutes plus further filtration with filter paper. All liquid fractions recovered were stored at –18 ºC until sugars quantitation. A control assay was made using filter paper to take into account any contribution of this material to sugars concentration that could occur in the filtration step.

2.3 Central composite design and RBF-PSO approach

A CCD was introduced in this study to optimize the chemical hydrolysis process of three different feedstocks. According to this design, each variable was examined at five levels: $-\alpha$, -1, 0, $+1$ and $+\alpha$.

Since the application of QLS was not successful in achieving the modeling of the hydrolysis processes, an RBF-PSO approach was used to obtain the optimal factor levels that guarantee the maximization of the responses. In the present work, an RBF network combined with forward selection was used, and for PSO, the population size and the number of generations were estimated by trial and error, set as fifteen particles (wheat bran) or ten particles (corn bran and pine sawdust) and fifteen generations in both cases. The value of the global desirability function (D) was the objective function to be optimized [38].

In the present work, three or four factors were varied in order to obtain the optimal conditions for the chemical hydrolysis of pine sawdust, corn bran and wheat bran.

2.4 Analytical method

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. (1), 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.

$$C_6H_{12}O_6 + O_2 + H_2O \rightarrow C_6H_{12}O_7 + H_2O_2 \quad (1)$$

In the second step, according to Eq. (2), 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.

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

The concentration of reducing sugars was measured by using a well-known chemical method [39].

2.5. Software

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.

RBF networks were implemented using the forward selection method described by Orr in ref [40] and available at http://www.anc.ed.ac.uk/rbf/rbf.html. The complete RBF-PSO optimization algorithm was written in MATLAB R2008a (The MathWorks, Inc.).

3. Theory

3.1. Radial basis function networks

Artificial neural networks based on radial basis functions consist of three layers. The neurons of the input layer distribute the input variables (which in our case are the *F* factor values influencing a given response) to the neurons of the hidden layer. Each of the *M* neurons of the hidden layer transfers the input data through a Gaussian function to the output layer. Finally, the output neuron uses a linear transfer function, in contrast to MLP networks, which employ non-linear transfer functions. To specifically implement RBF networks, suitable parameters for the Gaussian functions of the hidden

layer are needed. They consist of the centres of the Gaussian functions (contained in the $F\times1$ vector $\mathbf{c}_m$) and the Gaussian widths σ, which are typically taken as identical for all functions. The output value from the mth. hidden neuron for a given input value $\mathbf{x}_i$, is thus given by:

$$\text{out}_m = \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{x}_i - \mathbf{c}_m\right\|^2\right) \qquad (3)$$

where $\|\mathbf{x}_i - \mathbf{c}_m\|$ is the length of the vector difference and equal to the distance between $\mathbf{x}_i$ and $\mathbf{c}_m$. The input value to the output node is the weighted sum of all the outputs of the hidden nodes. Finally, the response of the output node is linearly related to its input. Therefore, the RBF network output (out_i) for an input object $\mathbf{x}_i$ can be written as:

$$\text{out}_i = w_0 + \sum_{m=1}^{M} w_m \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{x}_i - \mathbf{c}_m\right\|^2\right) \qquad (4)$$

where w_0 is the so-called bias, and w_m is the weight ascribed to the mth. hidden output. The weights are adjusted so that the mean square error of the net output (with regard to reference values) is minimized. The parameters to be adjusted are the Gaussian centres and widths of the hidden neurons, and the weights of the output layer. The RBF networks show a guaranteed convergence in their learning procedure: from the centres of the M basis functions and a set of I training objects with known factor values ($\mathbf{x}_i$) and target response (r_i), the minimum squared error in the prediction of r can be shown to be lead to the following weights:

$$\mathbf{w} = (\mathbf{H}^{\mathrm{T}}\,\mathbf{H})^{-1}\,\mathbf{H}^{\mathrm{T}}\,\mathbf{r} \qquad (5)$$

where $\mathbf{w}$ ($M\times1$) collects the weights, $\mathbf{r}$ ($I\times1$) the target response values, and $\mathbf{H}$ ($I\times M$) is the design matrix whose elements are:

$$H(i,m) = \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{x}_i - \mathbf{c}_m\right\|^2\right) \qquad (6)$$

Several procedures exist to limit the dimensionality of the hidden layer. One alternative is to control the network complexity using a subset of possible centres, which can be found by forward selection. The latter starts with an empty model and adds new functions, centred on each data point, according to the degree in which these functions reduce the squared error. Orr [41] combined forward selection with regularization involving the additional parameter λ in Eq. (5), to penalize for large weight values:

$$\mathbf{w} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y} \qquad (7)$$

where **I** is an appropriately dimensioned unit matrix. Our specific RBF working parameters are provided below.

It may be noticed that RBF are different from MLP networks in the following aspects: 1) RBF networks have a single hidden layer, whereas MLP may have several, 2) the hidden (non-linear) RBF layer is different from output (linear) layer, while in MLP there is a common neuronal model for all layers and 3) the argument of the RBF transfer function is the Euclidean distance between the input vector and the centre, while MLP compute the inner product of the input vector and the synaptic weight vector.

3.2. Particle swarm optimization

Particle swarm optimization is a technique inspired in a natural process, in this case the collective motion of birds. In PSO, a number of particles is given initial random positions and velocities, and the positions allow to evaluate a certain objective function. In the present case, the positions are the factors, defined in a space having a number of dimensions equal to the number of factors *F*, while the objective function to be minimized is the sum of squared errors SSE (predicted vs. measured response). Both the

particle positions and velocities are subsequently tuned employing well-defined rules, with the new positions allowing one to evaluate new function values in each running cycle. Whenever a particle finds a position which is better than those previously found (because the SSE is lower), its coordinates are stored. The new position of each particle is then defined within the context of a neighbourhood which comprise the particle itself and other particles in the population. This is achieved by defining the velocity in future time steps as a linear combination of: (1) the current velocity, (2) the difference between the overall best position and the actual individual position and (3) the stochastically weighted difference between the neighbourhood best position and the individual current position:

$$v_{ia,t+1} = w(t)v_{ia,t} + c_1(p_{ia,t} - x_{ia,t}) + c_2(p_{a,t} - x_{ia,t}) \quad (8)$$

where $v_{ia,t}$ and $v_{ia,t+1}$ are the velocities for the ith. particle in the ath. dimension at times t and t+1 respectively, $x_{ia,t}$ is its current position, $p_{ia,t}$ is its best position, $p_{a,t}$ is the best position for any member of the population, $w(t)$ is a time-dependent weight, and c_1 and c_2 are adjustable parameters. The weight $w(t)$ decreases with time to ensure that position changes in the last cycles monotonically decrease:

$$w(t) = w_0 + \frac{w_\infty - w_0}{t_{\max}} t \quad (9)$$

where w_0 and w_∞ ($w_0 > w_\infty$) are adjustable parameters, and $t_{\max}$ is the maximum number of time cycles. Usually the value provided by equation (8) is compared with a certain maximum velocity $v_{\max,a}$ and the least of them is added to the particle position:

$$x_{ia,t+1} = x_{ia,t} + |v_{ia,t+1}| \times \min(|v_{ia,t+1}|, v_{\max,a}) \,/\, v_{ia,t+1} \quad (10)$$

where $|\cdot|$ implies the modulus. These rules for particle movement cause them to search between two best positions: the individually best point and the globally best one, in a manner which is related to some social activities such as bird flocking. Figure 1 shows

the flow sheet for the PSO scheme employed in this study. Specific details concerning the PSO process are provided below.

3.3. Desirability function

The use of a desirability function involves creating a function for each individual response d_i and finally obtaining a global function D that should be maximized choosing the best conditions of the designed variables. The latter function varies from 0 (value totally undesirable) to 1 (all responses are in a desirable range simultaneously), and can be defined by Eq. (17):

$$D = \left(d_1^{\,r1} \times d_2^{\,r2}\right)^{\frac{1}{r_1+r_2}} \qquad (11)$$

where d_1 and d_2 correspond to the individual desirability functions for the responses being optimized, and r_1 and r_2 measure the relative importance of each response. In the present report, both responses were assigned the same importance, i.e., $r_1 = r_2 = 1$. Individual desirabilities (d_1 and d_2) were computed with the following maximization function:

$$d_i = \begin{cases} \left(\dfrac{\hat{Y} - A}{B - A}\right)^{w_i} & , \quad A \le \hat{Y} \le B \\ 1 & , \quad \hat{Y} > B \\ 0 & , \quad \hat{Y} < A \end{cases} \qquad (12)$$

where A and B correspond to the lower and maximum limit, respectively (see values in Table 5), $\hat{Y}$ is the predicted response (by the RBF model), and w_i the weights (if a weight is 1, the d_i values will vary from 0 to 1 in a linear way while approaching to the desired value). In the present report, weights were both set to 1.

4. Results and discussion

With the aim of optimizing the chemical hydrolysis processes of three feesdstocks (corn bran, wheat bran and pine sawdust), three CCDs were built (one for each feedstock). Two of them, corresponding to corn bran and pine sawdust, consisted of twenty experiments: six center, six axial and eight factorial points. On the other hand, the one corresponding to wheat bran consisted of thirty experiences: six center, eight axial and sixteen factorial points. The independent variables taken into account to build the experimental designs were previously selected by building Plackett-Burman designs and applying a GA approach [42]. Additional variables, i.e. particle size, pretreatment and time of hydrolysis (in corn bran and pine sawdust cases), which were not found to be significant, were kept constant.

In the case of corn bran and pine sawdust, the three evaluated factors were: (1) temperature of hydrolysis (Te), (2) sulfuric acid concentration (A), and (3) acid solution/feedstock ratio (AF). In the wheat bran case, four factors were evaluated: the latter three and also the time of hydrolysis (Ti). Additionally, none of the feedstocks were chemically pretreated, and the feedstock particle sizes employed were: 1.0 mm for corn bran and 0.5 mm for both wheat bran and pine sawdust.

A literature search revealed that the sugars/raw biomass yield is usually employed as a response to be optimized, because it is assumed to be a better descriptor of the hydrolysis process. However, Vieira Canettieri et al. [43] suggested that the polysaccharide content (hemicellulose and cellulose) of the raw biomass should also be taken into account, in order to calculate an “extraction percentage”, since a good yield does not guarantee a good conversion from polysaccharides to momosaccharides. Because the aim of this study was to obtain as much monosaccharides as possible, it was decided that for the three evaluated feedstocks, the two responses to be measured

282 are the concentrations (in g L^{-1}) of glucose (G) and reducing sugars (RS). Table 1 and 2
283 summarize the twenty and thirty experiments, and the concentrations of G and RS
284 obtained for corn bran, pine sawdust and wheat bran, respectively.
285 Since the application of response surface methodology with quadratic least-squares
286 through a CCD was not successful in obtaining the optimal hydrolysis conditions for
287 each feedstock (see below), a different optimization procedure, based on RBF networks
288 coupled to PSO, was applied to achieve this objective. By employing an RBF network,
289 the multidimensional space was adequately modeled. Then, in a subsequent step, by
290 applying a PSO approach, the modelled multidimensional space was screened, and the
291 optimal hydrolysis conditions for each one of the three feedstocks were obtained, with
292 the corresponding value of desirability D.
293 Finally, a comparison of the determination coefficients (R^2) corresponding to both
294 models was carried out, in order to verify that the models obtained by RBF networks
295 were better than those yielded by the application of QLS.
296
297 *4.1 Analysis by quadratic least-squares*
298 The ANOVA tests applied to the factors and responses data demonstrated that six
299 quadratic models could fit both G and RS responses for the three feedstocks under
300 consideration. The associated probability values (p) obtained for the G response models
301 were 7×10^{-4}, 9×10^{-4} and 1×10^{-2} for wheat bran, corn bran and pine sawdust,
302 respectively, while the corresponding p values for the RS response models were 1×10^{-4}
303 for the three cases, thus indicating the significance of the models, which can be
304 mathematically expressed according to Equations (13) to (18).
305
306 For wheat bran:

307 $$Y_1 = -41.812.69X_3 - 10.49X_4 - 0.11X_1X_4 - 0.04X_2X_3 + 0.75X_4^2 \quad (13)$$

308 $$Y_2 = -383.21 + 7.39X_2 + 12.56X_3 - 1.14X_4 - 0.04X_1X_3 - 0.06X_2X_3 - 0.11X_2X_4 - 0.03X_2^2 - 0.12X_3^2 + 0.30X_4^2 \quad (14)$$

309

310 For corn bran:

311 $$Y_1 = 15.79 + 1.09X_2 + 2.86X_3 - 18.31X_4 - 0.04X_2X_3 + 0.89X_4^2 \quad (15)$$

312 $$Y_2 = -281.05 + 6.28X_2 + 4.35X_3 + 8.81X_4 - 0.16eX_2X_4 - 0.03X_2^2 - 0.09X_3^2 \quad (16)$$

313

314 For pine sawdust:

315 $$Y_1 = -0.66 + 0.06X_2 - 0.19X_3 - 2.34x10^{-3}X_2X_3 + 0.01X_3^2 \quad (17)$$

316 $$Y_2 = 1.47 + 0.27X_2 + 0.43X_3 - 3.80X_4 + 0.14X_4^2 \quad (18)$$

317

318 where Y_1 and Y_2 are G and RS responses respectively, and X_1, X_2, X_3 and X_4 are the
319 factors Ti, Te, A and AF, respectively. Only the factors that are significant for each
320 response have been included in the above equations.

321 Nevertheless, some statistical results were not satisfactory: the R^2 obtained for
322 response G were 0.648, 0.742 and 0.699 for wheat bran, corn bran and pine sawdust,
323 respectively, implying that these models could explain only about 70 % of the variabilty
324 in the responses, with the remaining 30 % explained by the residue. Moreover, the *p*
325 values corresponding to the lack of fit were all less than 1×10^{-4}, indicating that the
326 models are not suitable for prediction purposes.

327 In the case of the RS response, the R^2 obtained were 0.964, 0.852 and 0.898 for
328 wheat bran, corn bran and pine sawdust, respectively. These values indicated that the
329 models could fit satisfactorily the responses. However, in the case of pine sawdust, the *p*
330 value for the lack of fit was 0.022, once again meaning that the model could not be used

to perform predictions. In the remaining cases of wheat and corn bran, the lack of fit tests were not significant. These two models could fit the responses and could be used to perform further predictions.

Although some of the models cannot be used for prediction, an analysis of factor effects can be made. In most cases, when the individual contributions of Te, A and AF exerted positive or negative effects in a response, their interactions and/or quadratic contributions affected inversely the response, i.e.: exerted a negative effect or a positive effect, respectively. This indicates that the optimum factor values may be included in the tested ranges. With respect to the factor Ti, which was only evaluated in the case of wheat bran, two of its interactions (with AF in the G response and with A in the RS response) influence negatively the responses. According to these resullts, it is evident that these four factors exert a synergic effect on the hydrolysis processes.

It has been extensively described that these factors show a positive influence in sugar concentrations up to a certain extent, beyond which the inverse effect is observed [44–46]. Temperature is expected to have a positive effect, since it favors the rupture of heterocyclic ether bonds in the polysaccharides caused by protons, but up to a certain point, beyond which a negative effect can be observed [45,47]. Vieira Cannettieri et al. [43], working on *Eucalyptus grandis* wood, found that the time and temperature of hydrolysis have a negative effect on sugar yields due to its chemical degradation. Bower et al. [48] 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 [44]. The behaviour of responses regarding A and AF can be explained taking into account that 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 polisaccharides depolymerize, especially for hemicelluloses, producing the depletion of sugars yield [49].

4.2. Analysis by artificial neural networks

Because the models obtained by means of QLS were not satisfactory, we resorted to the application of artificial neural networks based on the use of radial basis functions. The values predicted by the RBF vs. the actual ones were employed to calculate the R^2 for both responses in the three hydrolysis process under study. The R^2 values obtained for G response were 1.000, 1.000 and 0.995, and for RS response they were 0.979, 0.859 and 0.992 for wheat bran, corn bran and pine sawdust, respectively. These values indicate that the models obtained by means of RBF show improved fitting, mainly for G response: 54.3 %, 34.77 % and 42.34 % for wheat bran, corn bran and pine sawdust, respectively. This better performance of RBF may be attributed to its ability to universally approximate non-linear systems. On the contrary, as was commented above, QLS is restricted to only second-order polynomial models [21].

The first step in the RBF modeling of the design data was the estimation of the optimal working RBF parameters, as well as the number of hidden neurons. This latter number was tuned using one of the procedures included in Orr's RBF package, i.e., forward selection combined with regularization, which were briefly commented in section 3.1. The criterion for stopping the addition of new basis functions was the obtainment of a minimum in the so-called generalized cross-validation error, as defined by Orr [ref. 40], which penalizes the mean squared error if an excessive number of parameters is employed. Once the number of hidden neurons was set: a) wheat bran: 20 for glucose and 19 for reducing sugars, b) corn bran: 8 for glucose and 9 for reducing sugars, c) pine sawdust: 8 for glucose and 15 for reducing sugars, straightforward RBF

380 analysis provided the values of the optimal working parameters, i.e., the centers, radii
381 and weights which are quoted in Supplementary material.
382 Table 4 shows a comparison between the R^2 values obtained by applying QLS and
383 RBF, respectively. The improvement in model fitting for the wheat bran case can be
384 seen in Figure 2A and B, which show the correlation between actual and predicted
385 values for the responses using both models.
386 After modeling, the RBF parameters were used to find the optimal hydrolysis
387 conditions by applying a methodology based on PSO. For the optimization process, a
388 number of particles was set for each of the optimized systems, i.e., 15 particles for
389 wheat bran and pine sawdust and 10 particles for corn bran. This appeared to be enough
390 to cover the experimental factor space. Also, 15 generations were employed to find the
391 optimal points in the multidimensional space for all the cases under study. These
392 parameters (number of particles and generations) were assessed by try and error, in such
393 a way that the convergence tolerance for the optimal values of the studied factors was
394 less than 0.01%, i.e. that the difference between successive factor values after the
395 generation cycle was less than 0.01%.
396 In comparison with other potential optimizing tools, such as exhaustive grid-search
397 methods or genetic algorithms, PSO provides a reliable and fast manner of estimating
398 the values of continuous experimental factors for optimizing the desirability function.
399 Table 4 shows the criteria employed to perform the optimization. Figure 3 shows the
400 evolution of D as a function of the number of generations in the case of wheat bran.
401 For wheat bran hydrolysis, the optimal value found for *D* was 0.942, which
402 corresponds to the following combination of factors: Ti 59.6 min, Te 99.2 ºC, A 10.4%
403 m/m and AF 6.0 mLg^{-1}. The response values that correspond to this combination were:
404 54.8 gL^{-1} G (individual desirability value d_G = 0.994) and 108.2 gL^{-1} RS (d_{RS}= 0.892).

With respect to corn bran, the optimal combination was: Te 80.4 ºC, A 20.5 % m/m and AF 4.2 mLg^{-1} which corresponded to D = 1.000, 45.8 gL^{-1} G (d_G = 1.000) and 97.5 gL^{-1} RS (d_{RS} = 1.000). Finally, for pine sawdust, the optimal combination was: Te 80.2 ºC, A 36.8 % m/m and AF 9.0 mLg^{-1}, which corresponds to D = 0.900. The predicted responses values were: 3.8 gL^{-1} G (d_G = 0.996) and 19.5 gL^{-1} (d_{RS} = 0.811). All these results were validated employing multiple layer perceptrons based ANN (data not shown). Figure 4A and B show the response surface for D as a function of Ti and Te, and as a function of A and AF, respectively, for wheat bran case, both at optimal values of the other factors.

An interesting observation can be made from the results obtained: there is some agreement with the optima reached by the application of experimental design followed of ANN-PSO and the highest experimental obtained values (see trials number 7, 18 and 14, respectively, of Table 1 and 2). Nevertheless, this result is not common in the field of optimization, because most of the times in which the desirability function is applied, the optimal combination of factors do not necessarily match the best experiment. An erroneous conclusion could be extracted: the modeling is not necessary to get the optima. However, it must be strongly stated that modeling is the only way to know that there is agreement between trials maxima (corresponding to the design) and maxima reached by the modeling.

In sum, the RBF-PSO approach was capable of improving the model fitness in comparison to what was obtained by applying QLS, mainly for G responses. In addtion, the values of D, which were all near 1, are indicative that the factors and responses have simultaneously desirable values. Consequently, it can be concluded that the application of the RBF-PSO approach allows to obtain more reliable results in comparison with classical QLS analysis.

Although the three studied raw materials have the same components, the optimal combinations predicted for each of them are specific for each material. This observation may be explained taking into account the specific macromolecular structure of the studied feedstocks: the arrangement of cellulose, lignin and hemicelluloses may vary among the different raw biomass. Then, different biomasses, subjected to hydrolysis reactions, may lead to different results. Additionally, almost all the optimal values were not at the edges of the tested factor ranges, which were adequately chosen, in order to find the optimal hydrolysis conditions.

5. Conclusion

The application of QLS was not capable of fitting adequate models that could satisfactorily explain the variability, mainly in G responses. On the contrary, RBF allowed obtaining more reliable models, a fact that can be attributed to its ability to approximate non-linear systems, whereas QLS is only capable of fitting second-order polynomial models with a reasonable number of experiments.

Moreover, with the introduction of a PSO approach, the optimal combinations that guarantee the maximization of the responses in the chemical hydrolysis processes of three different feedstocks were obtained. Thus, the RBF-PSO approach performed better than QLS in this particular study.

Finally, different biomass subjected to hydrolysis may lead to very different results due to its different macromolecular structure.

Acknowledgments

The authors are grateful to Universidad Nacional del Litoral (Project CAI+D Nº 12-65 and CAI+D 2009 Tipo III R2), to CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project PIP 2988) and to ANPCyT (Agencia Nacional de Científica y la Tecnológica, Project PICT 2010-0084) for financial support, and Arturo Simonetta for sharing his milling equipment. P.C.G. thanks CONICET for his fellowship.

References

[1] Myers RH, Montgomery DC, Response Surface Methodology: Process and Product Optimization Using Designed Experiments (Wiley Series in Probability and Statistics), Wiley, New York; 2009.

[2]Shi X-Y, Jin D-W, Sun Q-Y, Li W-W. Optimization of conditions for hydrogen production from brewery wastewater by anaerobic sludge using desirability function approach, Renew. Energ. 2010; 35:1493–1498.

[3]Foudjo BUS, Kansci G, Fokou E, Lazar IM, Pontalier P-Y, Etoa F-X. Multi-response optimization of aqueous oil extraction from five varieties of cameroon-grown avocados. Environ. Eng. Manag. J. 2012; 11:2257–2263.

[4] Gadhe A, Sonawane SS, Varma MN. Optimization of conditions for hydrogen production from complex dairy wastewater by anaerobic sludge using desirability function approach. Int. J. Prod. Hydrogen 2013; 38:6607–6617.

[5] Dopar M, Kusic H, Koprivanac N. Treatment of simulated industrial wastewater by photo-Fenton process. Part I: The optimization of process parameters using design of experiments (DOE). Chem. Engin. J. 2011; 173: 267–279.

[6] Kılıç M, Uzun BB, Pütün E, Pütün AE. Optimization of biodiesel production from castor oil using factorial design. Fuel Process. Technol., 2013; 111: 105–110.

[7] Severini C, Baiano A, De Pilli T, Romaniello R, Derossi A, Lebensm A, Prevention of enzymatic browning in sliced potatoes by blanching in boiling saline solutions. Wiss Technol 2003;36: 657-665.

[8] Nardi JV, Acchar W, Hotza D. Enhancing the properties of ceramic products through mixture design and response surface analysis. J Eur Ceramic Soc 2004;24: 375-379.

[9] Abnisa F, Wan Daud WMA, Sahu JN. Optimization and characterization studies on bio-oil production from palm shell by pyrolysis using response surface methodology. Biomass Bioenergy 2011;35: 3604-3616.

[10] Lee KM, Gilmore DF. Formulation and process modelling of biopolymer (polyhydroxyalkanoates: PHAs) production from industrial wastes by novel crossed experimental design. Process Biochem 2005;40: 229-246.

[11] Giordano PC, Martínez HD, Iglesias AA, Beccaria AJ, Goicoechea HC. Application of response surface methodology and artificial neural networks for optimization of recombinant Oryza sativa non-symbiotic hemoglobin 1 production by Escherichia coli in medium containing byproduct glycerol. Bioresour. Technol. 2010; 101:7537–7544.

[12] Demain A, Davies J. Manual of industrial microbiology and biotechnology. Second Edition Washington: Am Soc Microbiol; 1999.

[13] Zhi W, Song J, Ouyang F. Application of response surface methodology to the modeling of α-amylase purification by aqueous two-phase systems. J Biotechnol 2005;118: 157-165.

[14] Liu R-S, Tang Y-J. Tuber melanosporum fermentation medium optimization by Plackett–Burman design coupled with Draper–Lin small composite design and desirability function. Biores. Technol. 2010; 101:3139–3146.

[15] Lim S, Le K-T. Optimization of supercritical methanol reactive extraction by Response Surface Methodology and product characterization from Jatropha curcas L. seeds. Biores Technol, 2013; 142:121–130.

[16] Contesini FJ, Ibarguren C, Ferreira Grosso CR, de Oliveira Carvalho P, Harumi Sato H. Immobilization of glucosyltransferase from Erwinia sp. using two different techniques. J Biotechnol, 2012; 158:137–143.

[17] Sella SRBR, Masetti C, Figueiredo LFM, Vandenberghe LPS, Minozzo JC, Soccol CR. Soybean molasses-based bioindicator system for monitoring sterilization process: Designing and performance evaluation. Biotechnol Bioproc Eng 2013; 18: 75–87.

[18]Braga ARC, Gomes PA, Kalil SJ. Formulation of Culture Medium with Agroindustrial Waste for β-Galactosidase Production from Kluyveromyces marxianus ATCC 16045. Food Bioproc Technol 2012; 5: 1653–1663.

[19] Larentis AL, Quintal Nicolau JFM, Argondizzo APC, Galler R, Rodrigues MI, Medeiros MA. Optimization of medium formulation and seed conditions for expression of mature PsaA (pneumococcal surface adhesin A) in Escherichia coli using a sequential experimental design strategy and response surface methodology. J Ind Microbiol Biotechnol 2012; 39: 897–908.

[20] Liu YT, Long CN, Xuan SX, Lin BK, Long MN, Hu Z. Evaluation of culture conditions for cellulase production by two *Penicillium decumbens* under liquid fermentation conditions. J Biotechnol 2008;136: S328.

[21] Desai KM, Survase SA, Saudagar PS, Lele SS, Singhal RS. Comparison of artificial neural network (ANN) and response surface methodology (RSM) in

fermentation media optimization: Case study of fermentative production of scleroglucan. Biochem Eng J 2008;41: 266–273.

[22] Leonardi D, Lamas MC, Salomón CJ, Olivieri AC. Development of novel formulations for Chagas' disease. Optimization of benznidazol chitosan microparticles based on artificial neural networks. Int J Pharm 2009;367: 140–147.

[23] Cheng S, Song Q, Wei D, Gao B. High-level production penicillin G acylase from *Alcaligenes faecalis* in recombinant *Escherichia coli* with optimization of carbon sources. Enzyme Microb Technol 2007;41: 326–330.

[24] Didier C, Forno G, Etcheverrigaray M, Kratjie R, Goicoechea, HC. Novel chemometric strategy based on the application of artificial neural networks to crossed-mixture design for the improvement of recombinant protein production in continuous culture. Anal Chim Acta 2009;650: 167-174.

[25] Giordano PC, Martínez HD, Iglesias AA, Beccaria AJ, Goicoechea HC. Application of response surface methodology and artificial neural networks for optimization of recombinant *Oriza sativa* non-symbiotic hemoglobin 1 production by *Escherichia coli* in medium containing byproduct glycerol. Bioresour Technol 2010;101: 7537-7544.

[26] Moreira GA, Micheloud GA, Beccaria AJ, Goicoechea HC. Optimization of the *Bacillus thuringiensis* var. kurstaki HD-1 δ-endotoxins production by using experimental mixture design and artificial neural networks. Biochem Eng J 2007;35: 48–55.

[27] Günay ME, Nikerel IE, Oner ET, Kirdar B, Yildirim R. Simultaneous modeling of enzyme production and biomass growth in recombinant *Escherichia coli* using artificial neural networks. Biochem Eng J 2007;42: 329–335.

[28] Haykin S. Neural networks. A comprehensive foundation. Second edition Upper Saddle River, NJ: Prentice-Hall; 1999.

[29]Derks EPPA, Sanchez Pastor MS, Buydens LMC. Robustness analysis of radial base function and multilayered feedforward neural network models. Chemom Intell Lab Syst 1995;28: 49-60.

[30] Fidêncio PH, Poppi RJ, de Andrade JC. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. Anal Chim Acta 2002;453: 125-134.

[31] Fischbacher C, Jagemann KU, Danzer K, Muller UA, Papenkordt L, Schuler J. Enhancing calibration models for non-invasive near-infrared spectroscopical blood glucose determination. Fresenius J Anal Chem 1997;359: 78-82.

[32] Carlin M, Kavli T, Lillekjendlie B. A comparison of four methods for non-linear data modelling. Chemom Intell Lab Syst 1994;23: 163-177.

[33] Prakash O, Mehrotra S, Krishna A, Mishra BN. A neural network approach for the prediction of in vitro culture parameters for maximum biomass yields in hairy root cultures. J Theor Biol 2010;265: 579-585.

[34] Kennedy J, Eberhart RC. Particle swarm optimization. In: Proc IEEE Int Conf Neural Networks, Perth, Australia, 1995, pp. 1942-1948.

[35] Clerc M, Kennedy J. The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space. IEEE Trans Evol Comput 2002;6: 58-73.

[36] Liu L, Sun J, Zhang D, Du G, Chen J, Xu W. Culture conditions optimization of hyaluronic acid production by *Streptococcus zooepidemicus* based on radial basis function neural network and quantum-behaved particle swarm optimization algorithm. Enz Mic Tec 2009;44: 24-32.

[37] Kitayama S, Yasuda K, Yamazaki K. Integrative optimization by RBF network and particle swarm optimization. Elec Com Japan 2009;92: 31-42.

[38] Derringer G, Suich R. Simultaneous optimization of several response variables, J Qual Technol 1980;12: 214-219.

[39] Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 1959;31: 426-428.

[40] Orr MJL. Matlab functions for radial basis function networks. Technical report. Institute for Adaptive and Neural Computation, Division of Informatics, Edinburgh University; 1999.

[41] Orr MJL. Regularisation in the selection of radial basis function centres. Neural Comp 1995;7: 606-623.

[42] Giordano PC, Beccaría AJ, Goicoechea HC. 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. Bioresour Technol 2011;102: 10602-10610.

[43] Vieira Canettieri E, Jackson de Moraes Rocha G, Andrade de Carvalho Jr. J, Batista de Almeida e Silva J. Optimization of acid hydrolysis from the hemicellulosic fraction of *Eucalyptus grandis* residue using response surface methodology. Bioresour Technol 2007;98: 422-428.

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

[45] Aguilar R, Ramírez JA,. Garrote J, Vázquez M. Kinetic study of the acid hydrolysis of sugar cane bagasse. J Food Eng 2002;55: 309–318.

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

[47] Yoo CG, Lee CW, Kim TH. Optimization of two-stage fractionation process for lignocellulosic biomass using response surface methodology (RSM). Biomass Bioenergy 2011;35: 4901-4909.

[48] Bower S, Wickramasinghe R, Nagle NJ, Schell DJ. Modeling sucrose hydrolysis in dilute sulfuric acid solutions at pretreatment conditions for lignocellulosic biomass. Bioresour Technol 2008;99: 7354-7362.

[49] Sanchez G, Pilcher L, Roslander C, Modig T, Galbe M, Liden G. Dilute-acid hydrolysis for fermentation of the Bolivian straw material Paja Brava. Bioresour Technol 2004;93: 249–256.

462 **Figure captions**

463

464 **Fig. 1.** Optimization flowchart by using particle swarm optimization.

465

466 **Fig. 2.** Correlation between actual and predicted values for responses glucose (A) and
467 reducing sugars (B), fitted applying quadratic least-squares fit methodology and
468 artificial neural networks based in radial basis functions, for wheat bran.

469

470 **Fig. 3.** Evolution of the global desirability function (D) as a function of the number of
471 generations when applying radial basis functions and particle swarm optimization in
472 the case of wheat bran.

473

474 **Fig. 4.** (A) Response surface for the desirability as a function of time of hydrolysis
475 (minutes), temperature of hydrolysis (ºC). (B) Response surface for the desirability as a
476 function of sulphuric acid concentration (% m/m) and acid solution/feedstock ratio (g
477 acid sol/g residue). Both figures at optimal values of the other factors and for wheat
478 bran case.

Table 1 Central composite design built to find the optimal conditions of the chemical

Experiment	Factors[a]			Responses[b]			
	Te	A	AF	G		RS	
				CB	PS	CB	PS
1	100.0	10.0	12.0	26.9	1.6	56.7	8.9
2	80.0	20.0	9.0	0.0	0.3	70.3	8.3
3	113.6	20.0	9.0	0.0	0.1	54.2	16.4
4	80.0	20.0	9.0	0.1	0.2	74.0	8.4
5	46.4	20.0	9.0	0.0	0.2	10.7	2.6
6	80.0	20.0	9.0	0.1	0.0	65.9	8.8
7	60.0	30.0	12.0	0.0	0.1	49.8	3.2
8	100.0	10.0	6.0	41.1	3.0	95.5	19.5
9	80.0	20.0	9.0	0.0	0.1	73.3	8.3
10	100.0	30.0	6.0	0.0	0.0	91.6	23.7
11	80.0	3.2	9.0	1.3	2.4	19.0	2.6
12	60.0	30.0	6.0	0.2	0.1	52.0	13.4
13	80.0	20.0	14.1	0.0	0.7	48.8	3.8
14	80.0	40.0	9.0	0.3	3.6	50.9	19.4
15	100.0	30.0	12.0	0.0	0.2	53.0	18.5
16	80.0	20.0	9.0	0.0	0.5	69.1	5.8
17	60.0	10.0	12.0	0.0	0.5	34.2	1.25
18	80.0	20.0	3.9	45.4	0.2	97.2	18.8
19	80.0	20.0	9.0	0.2	0.2	55.1	6.9
20	60.0	10.0	6.0	0.6	0.4	33.2	1.8

[a]Te (ºC): temperature of hydrolysis, A (% m/m): sulphuric acid concentration, AF (g acid sol/g residue): acid solution/feedstock ratio.

[b]G (g L^{-1}): concentration of glucose, RS (g L^{-1}): concentration of reducing sugars, CB: corn bran, PS: pine sawdust.

hydrolysis of corn bran and pine sawdust.

Table 2 Central composite design built to find the optimal conditions of the the chemical hydrolysis of wheat bran.

Experiment	Factors[a]				Responses[b]		Experiment	Factors[a]				Responses[b]	
	Ti	Te	A	AF	G	RS		Ti	Te	A	AF	G	RS
1	45.0	80.0	20.0	9.0	0.0	80.8	16	30.0	100.0	30.0	12.0	0.2	52.3
2	60.0	60.0	10.0	6.0	2.7	30.3	17	60.0	60.0	10.0	12.0	1.3	14.5
3	45.0	120.0	20.0	9.0	0.1	51.3	18	45.0	80.0	20.0	9.0	0.2	69.4
4	75.0	80.0	20.0	9.0	0.0	80.4	19	30.0	100.0	30.0	6.0	0.2	91.1
5	45.0	80.0	20.0	9.0	0.1	76.1	20	60.0	60.0	30.0	12.0	0.0	52.0
6	45.0	80.0	40.0	9.0	0.4	52.3	21	30.0	60.0	10.0	12.0	1.4	6.3
7	60.0	100.0	10.0	6.0	55.1	106.8	22	30.0	100.0	10.0	12.0	26.9	48.6
8	60.0	100.0	30.0	12.0	0.1	46.4	23	45.0	80.0	20.0	3.0	52.6	117.4
9	45.0	80.0	20.0	15.0	0.0	54.4	24	60.0	100.0	30.0	6.0	0.2	77.9
10	45.0	80.0	20.0	9.0	0.0	69.0	25	60.0	60.0	30.0	6.0	21.0	50.1
11	30.0	60.0	30.0	12.0	0.2	51.0	26	45.0	40.0	20.0	9.0	1.2	10.3
12	45.0	80.0	20.0	9.0	0.0	76.5	27	30.0	100.0	10.0	6.0	0.2	84.1
13	45.0	80.0	20.0	9.0	0.1	78.1	28	30.0	60.0	30.0	6.0	0.3	79.8
14	15.0	80.0	20.0	9.0	0.1	67.5	29	30.0	60.0	10.0	6.0	1.8	13.9
15	45.0	80.0	0.0	9.0	0.7	3.7	30	60.0	100.0	10.0	12.0	27.8	56.6

[a]Ti (minutes): time of hydrolysis, Te (ºC): temperature of hydrolysis, A (% m/m): sulphuric acid concentration, AF (g acid sol/g residue): acid solution/feedstock ratio.

[b]G (g L^{-1}): concentration of glucose, RS (g L^{-1}): concentration of reducing sugars

Table 3 Statistics obtained by means of QLS and RBF

Feedstock		Wheat bran		Corn bran		Pine Sawdust	
Response[a]		G	RS	G	RS	G	RS
QLS[b]	Model	Quadratic (p=0.0007)	Quadratic (p<0.0001)	Quadratic (p=0.0009)	Quadratic (p<0.0001)	Quadratic (p=0.0114)	Quadratic (p<0.0001)
	Lack of fit	Significant (p<0.0001)	Not significant (p=0.1833)	Significant (p<0.0001)	Not significant (p=0.1063)	Significant (p=0.0021)	Significant (p=0.0219)
	R^2	0.648	0.964	0.742	0.852	0.699	0.898
RBF[c]	R^2	1.000	0.979	1.000	0.859	0.995	0.992

[a] G: concentration of glucose; RS: concentration of reducing sugars.

[b] QLS: quadratic least-squares fit methodology

[c] RBF: artificial neural networks based in radial basis functions.

Table 4 Criteria used for the optimization of multiple responses.

Factors[a] and responses[b]	Optimization criteria	Lower limit[c]			Upper limit[c]		
		WB	CB	PS	WB[c]	CB	PS
Ti (min)	In range	15.0	–	–	75.0	–	–
Te (ºC)	In range	40.0	46.4	46.4	120.0	113.6	113.6
A (% m/m)	In range	0.0	3.2	3.2	40.0	36.8	36.8
AF (mLg^{-1})	In range	3.0	3.9	3.9	15.0	14.1	14.1
G (gL^{-1})	Maximize	0.0	0.0	0.0	55.1	45.4	3.6
RS (gL^{-1})	Maximize	3.7	10.7	2.6	117.4	97.2	23.7

[a] Ti: time of hydrolysis. Te: temperature of hydrolysis. A: concentration of sulphuric acid. AF: acid solution/feedstock ratio.

[b] G: concentration of glucose. RS: concentration of reducing sugars.

[c] WB: wheat bran, CB: corn bran, PS: pine sawdust.