ISSN 1330-9862

(FTB-2256)

original scientific paper

Comparative Study of Response Surface Methodology, Artificial Neural Network and Genetic Algorithms for Optimization of Soybean Hydration

*Tushar Gulati, Mainak Chakrabarti, Anshu Singh, Muralidhar Duvuuri and Rintu Banerjee**

Microbial Biotechnology and Downstream Processing Laboratory, Agricultural and Food Engineering Department, Indian Institute of Technology, IN-721 302 Kharagpur, India

Received: March 16, 2009
Accepted: June 2, 2009

Summary

The present investigation deals with the modelling and optimization of soybean hydration for facilitating soybean processing and it focuses on maximization of mass gain, water uptake and protein retention in the bean. Process variables considered for optimization were: soybean to water ratio (1:2.48 obtained with response surface methodology, RSM, and 1:1.19 obtained with artificial neural network and genetic algorithm, ANN/GA), time (2.0 h using RSM and 8.0 h using ANN/GA) and temperature (40.0 °C using RSM and 45.1 °C using ANN/GA). The findings in this first report on optimization of soaking conditions for soybean hydration employing response surface methodology, hybrid artificial neural network and genetic algorithms reveal a substantially better alternative to the time-consuming soaking process, extensively practiced in industries, in terms of process time economy. Reasonably accurate neural network model (regression coefficient of 0.9443) was obtained based on the experimental data. The optimized set of process conditions was predicted through genetic algorithm, and the effectiveness of the ANN/GA model, validated through experiments, was indicated by significant correlations (R^2 and mean squared error (MSE) being 0.9380 and 5.9299, respectively). RSM also resulted in accurate models for predicting percentage mass gain, percentage water uptake and percentage protein retention (R^2 and MSE in the range of 0.889–0.9297 and 0.80–4.94, respectively).

Key words: response surface methodology (RSM), artificial neural network (ANN), genetic algorithms (GA), soybean soaking

Introduction

Soybean (*Glycine max*), a key member of the *Leguminoseae* family, accumulates a large amount of lipids and proteins, accounting for significant portions within the global, especially Asian, food and feed supply (*1*). Its protein content (38–44 %) is significantly higher than that of other legumes (20–30 %) and even substantially greater than that of cereals (8–15 %) (*2*). Nutritionally important soy proteins, with balanced amino acid profile, belonging to the globulin family, are termed glycinin (11S) and β-conglycinin (7S). These proteins show high aqueous solubility (>90 %) at alkaline conditions and hence can be extracted most efficiently by water or water with diluted alkali, their pH ranging from 7–9. At pH<7 the solubility decreases to a minimum at the isoelectric point (pI) of 4.5 (*3*).

Water plays a critical role in the biochemical reactions, as well as in the thermal treatment of soybeans resulting in protein denaturation and polysaccharide gelatinization (*4*). Moisture transfer in soybean takes place

*Corresponding author; Phone: ++91 3222 283 104; Fax: ++91 3222 282 244; E-mail: rb@agfe.iitkgp.ernet.in

mainly through diffusion (*5*) and imbibition (*6–8*) by soaking. Microstructural entities like the coat, micropyle and hilum predominantly impact the moisture transfer rate (*6–8*).

Soaking constitutes a vital step in soybean processing (*9*) as it reduces the requirement of processing energy as well as cooking time (*10*). Textural changes of soybeans result from water absorption during soaking and affect the subsequent grinding and soymilk extraction processes (*10*,*11*). Soaking, the way it is extensively practiced in industries, is generally an inconvenient time-consuming process, taking even up to 16 h at ambient temperatures (*12*). Minimizing soaking time is vital for reducing the proliferation of mesophilic bacteria, which affect the colour, taste and smell of the beans. The water uptake of soybeans during soaking is mainly influenced by time and temperature of soaking. The amount of absorbed water increases with the increase of the soaking time and temperature (*13–15*). Standard methods for reducing soaking time include the use of high temperatures, the addition of salts to increase alkalinity of the soaking water (*16*) and the application of high hydrostatic pressure during soaking (*17*). As an outcome of soaking, it was observed that soybeans soaked for 24 and 72 h lost 5 and 10 % of their solids, respectively (*11*). About 24 % of the loss comprised nitrogenous compounds. Any biological process, especially pertaining to food processing, which is a complex interplay of several intrinsic biochemical properties and influences process variables, warrants a strategic application of modelling and optimization tools like response surface methodology (RSM), artificial neural network (ANN), genetic algorithms (GA), *etc*. to obtain a feasible process ensuring optimum yield. Hence, determination of optimal conditions for processing is the key to ideal industrial processing. The effectiveness of RSM in optimization of processing conditions in the area of food technology ranging from raw materials to final products has been well documented (*18*,*19*).

The main advantage of ANN over RSM lies in the fact that it does not require a prior specification of suitable fitting function and it can also universally approximate non-linear functions of all orders, including quadratic functions (*18*). These advantages have led to the wide application of ANN in food engineering (*20*). However, in comparison with ANN, RSM with its structured nature is more useful in getting insight into sensitivity analyses and interactive effects among the various influencing components of a bioprocess (*18*).

Besides RSM, soft computing methodologies, namely ANN and GA, have recently been found to offer novel solutions to improve control and modelling in food processing. Several studies on the use of neural networks in food process modelling and optimization have been mentioned (*21*) and even GA has evolved as an ideal technique to solve diverse optimization problems in food and biochemical engineering (*22*,*23*).

The present work aims at modelling and optimization of soybean hydration conditions, taking into consideration soybean to water ratio, time and temperature of soaking as the influencing parameters in view of maximizing the mass gain, water uptake and protein retention in the soybean by employing statistical optimization techniques and soft computing tools.

Materials and Methods

Material and experimental set-up

Soybean was purchased from the local markets of Kharagpur, India. Accurately weighed clean soybeans were soaked in various ratios of soybean to water and incubated for different combinations of soaking time and temperature in a thermostatically controlled water bath based on the experimental design of RSM and ANN/GA. After soaking, the residual water was drained off and the drained volume was recorded. The hydrated beans were collected and weighed. Standard protocols were used to determine proteins in both soybeans and residual water (*24*).

Optimization using response surface methodology

RSM is an empirical statistical modelling technique employed for multiple regression analyses using quantitative data obtained from properly designed experiments to solve multivariate equations simultaneously (*25*). In the present investigation, RSM was used to determine the optimum soaking conditions for the soybeans using a Box-Behnken (BB) experimental design, generated through MINITAB 14® for three independent variables. The soaking conditions (independent variables) selected for the optimization were solid/liquid ratio, time and temperature of incubation. These three independent variables were studied at low (–1), middle (0) and high (+1) coded levels. Three different solid/liquid ratios X_1 (1:1, 1:3 and 1:5), incubation times X_2 (2, 5 and 8 h) and temperatures X_3 (45, 55 and 65 °C) were selected, and regression analysis of the data obtained from the experiments was performed. Coding of the variables was done according to the following equation:

$$x_i=\frac{X_i-X_M}{X_{max}-X_M} \qquad /1/$$

where x_i is the dimensionless coded value of an independent variable, X_i is the real value of an independent variable, X_{max} is the maximum value of the independent variable and X_M is the mean of the maximum and minimum value of the independent variable. In order to find out the significance of the developed model for relating the response Y to independent variables X, pure experimental error is required to be determined. For this reason a number of experiments were carried out at the centre points where the coded values of x were zero. This permits the checking of bias between several sets of experiments. The relationship of the three independent variables and the response was calculated by the second-order polynomial described in Eq. 2:

$$Y=b_0+b_1x_1+b_2x_2+b_3x_3+b_4x_1^2+b_5x_2^2+ +b_6x_3^2+b_7x_1x_2+b_8x_2x_3+b_9x_3x_1 \qquad /2/$$

where Y is the response variable; b_0 is a constant; b_1, b_2 and b_3 are the coefficients of the linear terms; b_4, b_5 and b_6 are the coefficients of the quadratic terms; and b_7, b_8 and b_9 are the coefficients of the interaction terms. The second-order polynomial coefficients were calculated using the software package MINITAB 14® to estimate the responses of the dependent variable, and response surface plots were generated.

Optimization is the ultimate aim of performing the RSM. Many response surface problems involve the analysis of several responses. Simultaneous consideration of multiple responses involves first building an appropriate response surface model for each dependent variable and then trying to find a set of operating conditions that in some sense optimizes all responses or at least keeps them in desired ranges (*26*). For this purpose, the use of desirability functions is one of the useful approaches for optimization of multiple responses. In this technique, the general approach is to first convert each response Y_i into an individual desirability function d_i that varies over the range of $0 \le d_i \le 1$, where if response Y_i is at its target value T, then $d_i=1$, and if it is outside an acceptable region, then $d_i=0$. The individual desirability, d_i, is calculated using the following equation:

$$d_i=\begin{bmatrix} 0 \text{ if } Y_i<1 \\ \left(\dfrac{Y_i-L}{T-L}\right)^s \\ 1 \text{ if } Y_i>T \end{bmatrix} \text{if } L \le Y_i \le T \qquad /3/$$

where T is the target value of the response, L is the lower acceptable value of the response and s is the mass. Thus, when s=1, the desirability function is linear. When s>1 is chosen, a major importance is given to the points near the target value. When s<1 is chosen, this last demand is of low importance. Here s controls the variation rate of the desirability function. By varying the value of s, one can attribute different desirability to the responses and can increase and decrease the range of acceptable values in the optimization process. The design variables were chosen to maximize the overall desirability as:

$$D=(d_1 \times d_2 \times \ldots d_i \ldots \times d_N)^{\frac{1}{N}} \qquad /4/$$

where D is the overall desirability value, d_i is the individual desirability value of the response variables and N is the number of responses. The desirability of several responses was generated using the MINITAB 14®.

Normalization of data set for ANN modelling

In order to achieve fast convergence to minimum mean square error (MSE), the input and output data were normalized within the range [–1, 1] and [0, 1], respectively. Normalization results in an even distribution of the dataset and scales it into an acceptable range for the network. As a result, all variables acquire the same significance (importance) during the learning process.

ANN model development

In the present study, there was a total of 107 datasets each with 6 components (x_1, x_2, x_3, y_1, y_2 and y_3), which were used for training and testing the neural networks. The first three components, namely solid to liquid ratio, time and temperature of soaking, were the input variables, whereas the remaining three, namely percentage mass gain, percentage water uptake and percentage protein retention, were the output variables. The dataset was randomly divided into 90 and 17 datasets for training and testing the neural networks, respectively.

The training of neural networks was continued until a minimum error was reached during validation. The estimation of the performance of the trained network was based on the accuracy of the network using this test data set. A feed-forward (multilayer perception) artificial neural network trained by back propagation (BP) algorithm was selected to predict the outputs (*27,28*). Different neural networks were formed and the optimum values of network parameters were obtained by trial and error. Adjustment of ANN parameters included the number of hidden layers and neurons, the type of transfer function and learning rate. The performance of the ANN was statistically measured by mean squared error (MSE) and regression coefficient (R^2) obtained as follows:

$$MSE=\frac{1}{N}\sum_{i=1}^{n}(y_o-y_e)^2 \qquad /5/$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_o-y_e)^2}{\sum_{i=1}^{n}(y_o-y_m)^2} \qquad /6/$$

where n is the number of experiments used for training the ANN, y_o is the predicted value of the ANN model, y_e is the actual or experimental value and y_m is the average of actual values. The final network was selected on the basis of the lowest error when training and testing datasets. Preliminary trials indicated that the learning and prediction ability of one hidden layer network were better than those with two hidden layer networks.

Optimization with genetic algorithms

Genetic algorithms (GA) are stochastic optimization techniques that simulate a natural evolution process based on the mechanism of genetics and Darwin's principle of natural selection (*23*). Briefly, GA involve an initial population generated at random, a fitness function and the development of new generations *via* the application of genetic operators, namely selection, crossover and mutation. The procedure is repeated several times (*29*). An optimal set of soaking conditions was searched for by the hybrid ANN/GA models. For the present study, a code was written in MATLAB v. 7.1 environment to search for optimal conditions for soybean hydration.

ANN/GA optimization of soybean soaking conditions

An initial or starting population was randomly generated for the optimization of soybean hydration conditions through GA. A pre-specified number of generations was used as the stopping criterion. In order to evaluate the fitness of an individual in GA population, the ANN model was used to predict the percentage water uptake, percentage mass gain and percentage protein retention in the soybean. The fitness of an individual is defined as:

$$\text{Fitness}=y_{o\ \text{mass gain}}+y_{o\ \text{water uptake}}+y_{o\ \text{protein retention}} \qquad /7/$$

where $y_{o\ mass\ gain}$, $y_{o\ water\ uptake}$ and $y_{o\ protein\ retention}$ are predicted values of percentage mass gain, percentage water uptake and percentage protein retention in the hydrated soybeans, respectively, based on the ANN model in Eq. 7. GA subsequently searches the soaking conditions that maximize the fitness.

The ANN/GA optimization scheme is initiated considering randomly generated population for the first generation. The trained ANN model is then used to predict the respective outputs for each individual. The fitness of each individual is subsequently calculated using Eq. 7. Three operators including selection, crossover, and mutation, having their specified parameters, are applied to produce a new generation. The new generation thus obtained is considered as the current generation for producing the next generation. The above set of operations continues iterating until the termination criterion is met with. The performance of GA is affected by the size of initial population, the number of generations and GA parameters, namely mutation rate and crossover rate. Hence the best fitness of the end population was evaluated under different sizes of initial population and generations as well as different mutation rates and crossover rates. A 'tournament' selection method was used for selecting elite populations for crossover. For the present study, initial population of 100, number of generations of 5, mutation rate of 0.15, and crossover rate of 0.60 were obtained by trial and error to obtain the best fitness.

With the above GA parameters, the ANN/GA model was run several times to obtain optimal conditions for soybean hydration. Different runs for ANN/GA gave different optimal conditions even for the same set output values and the results were selected from the end populations of different runs.

To evaluate the efficacy of the ANN/GA model, the soybeans were soaked according to the optimum conditions predicted by the model. The percentage mass gain, percentage water uptake and percentage protein retention in the hydrated soybeans were experimentally determined and statistically compared with the predicted results. The regression line, regression coefficient and MSE between the experimental values and the ANN/GA responses were obtained to assess the performance of the hybrid ANN/GA-based optimization approach for soybean hydration.

Results and Discussion

To examine the combined effect of solid/liquid ratio, time and temperature of soaking (independent variables) on percentage mass gain, percentage water uptake and percentage protein retention (response variables) in the soaked beans, a Box-Behnken design comprising 17 experiments was performed. Second-order polynomial equations were used to correlate the independent process variables, X_i, with the responses. The second-order polynomial coefficient for each term of the equation was determined through multiple regression analysis using MINITAB 14®. Table 1 shows the BB design along with the actual and RSM-predicted results. The results were analyzed by using ANOVA, *i.e.* analysis of variance suitable for the experimental design. ANOVA tables for mass gain, water uptake and protein retention were obtained by using MINITAB 14® software to evaluate the significance of different parameters. The results are shown in Tables 2a–c. For percentage mass gain, percentage water uptake and percentage protein retention, the model F-values of 27.51, 59.65 and 38.45, respectively, imply that the models are significant. Model F-value is calculated as ratio of mean square regression and mean square

Table 1. Box-Behnken design of independent variables and their corresponding experimental and predicted results of percentage mass gain, percentage water uptake and percentage protein retention

Soaking conditions			Experimental			RSM-predicted		
Solid/liquid	Time/h	Temperature/°C	Mass gain/%	Water uptake/%	Protein retention/%	Mass gain/%	Water uptake/%	Protein retention/%
1:3	8	65	116.4	53.3	88.4	113.50	52.14	89.68
1:3	2	65	109.3	55.6	88.0	110.93	55.81	88.33
1:3	8	45	122.7	51.1	90.5	121.08	50.89	90.18
1:5	8	55	122.5	33.3	90.9	128.06	31.65	91.08
1:5	2	55	119.1	46.7	92.0	120.14	43.66	93.13
1:5	5	45	123.2	32.0	95.8	119.26	33.86	95.95
1:1	5	45	85.0	100.0	100.0	87.66	97.19	101.45
1:5	5	65	126.5	29.3	92.3	123.84	32.11	90.85
1:1	2	55	87.3	100.0	100.0	81.74	101.65	99.83
1:1	8	55	85.9	100.0	100.0	84.86	103.03	98.88
1:1	5	65	69.9	100.0	100.0	73.74	98.14	99.85
1:3	2	45	109.7	56.7	95.8	112.60	57.86	94.43
1:3	5	55	124.5	55.6	85.3	126.14	50.22	84.94
1:3	5	55	127.3	48.9	84.6	126.14	50.22	84.94
1:3	5	55	127.1	44.4	84.6	126.14	50.22	84.94
1:3	5	55	123.3	53.3	84.6	126.14	50.22	84.94
1:3	5	55	128.5	48.9	84.6	126.14	50.22	84.94

residue. Model p-value has been found to be very low ($p<0.0001$) for all the three response variables. This resignifies the efficacy of the model. The p-values were used as a tool to check the significance of each of the coefficients, which, in turn, is necessary to understand the pattern of the mutual interactions between the independent variables. The smaller the magnitude of p, the more significant the corresponding coefficient. Values of $p<0.05$ indicate that the model terms are significant. The coefficient estimates and the corresponding p-values suggest that, among the independent variables used in the study, the linear terms X_1 (solid/liquid), X_2 (time), X_3 (temperature) and the quadratic terms X_1^2, X_2^2 and X_3^2 ($p<0.0001$) have the greatest effect on each of the response variables. The mutual interaction between independent variables has been found to be of least importance as p-values are well above 0.05, which is evident from Tables 2a-c. The fit of the models was also expressed with the coefficient of determination R^2, which was found to be 0.973, 0.987 and 0.980 respectively for percentage mass gain, percentage water uptake and percentage protein retention. The corresponding second-order response models for water uptake, mass gain and protein retention that were obtained upon regression analyses were:

$$\begin{aligned}\text{Water uptake}=&50.22-32.34X_1-2.66X_2-0.20X_3+\\&+15.46X_1^2+4.13X_2^3-0.36X_3^2-\\&-3.35X_1X_2-0.67X_1X_3+0.83X_2X_3\end{aligned} \tag{8}$$

$$\begin{aligned}\text{Mass gain}=&126.14+20.40X_1+2.76X_2-2.31X_3-\\&-17.91X_1^2-4.53X_2^2-7.08X_2^3+\\&+1.20X_1X_2+4.60X_1X_3-1.47X_2X_3\end{aligned} \tag{9}$$

$$\begin{aligned}\text{Protein retention}=&84.94-3.625X_1-0.75X_2-\\&-1.675X_3+8.568X_1^2+2.218X_2^2+\\&+3.518X_3^2-0.275X_1X_2-\\&-0.875X_1X_3+1.425X_2X_3\end{aligned} \tag{10}$$

Table 2a. Analysis of variance for the experimental results of percentage mass gain

Sources	d.f.	Sum of squares	Mean square value	F-value	p-value
Regression	9	5302.83	589.20	27.51	<0.0001
Linear terms	3	3433.11	1144.37	53.43	<0.0001
Quadratic terms	3	1770.61	590.20	27.56	<0.0001
Interaction terms	3	99.10	33.03	1.54	0.2860
Residual error	7	149.91	21.42	–	–
Lack-of-fit	3	131.32	43.77	9.42	0.0280
Pure error	4	18.59	4.65	–	–

d.f. – degrees of freedom

Table 2b. Analysis of variance for the experimental results of percentage water uptake

Sources	d.f.	Sum of squares	Mean square value	F-value	p-value
Regression	9	9592.58	1065.84	59.65	<0.0001
Linear terms	3	8422.74	2807.58	157.12	<0.0001
Quadratic terms	3	1120.40	373.47	20.90	0.0010
Interaction terms	3	49.30	16.48	0.92	0.4780
Residual error	7	125.09	17.87	–	–
Lack-of-fit	3	49.30	16.43	0.87	0.5280
Pure error	4	75.79	18.95	–	–

d.f. – degrees of freedom

Table 2c. Analysis of variance for the experimental results of percentage protein retention

Sources	d.f.	Sum of squares	Mean square value	F-value	p-value
Regression	9	554.46	61.61	38.45	<0.0001
Linear terms	3	132.07	44.02	27.47	<0.0001
Quadratic terms	3	410.90	136.97	85.47	<0.0001
Interaction terms	3	11.49	3.83	2.39	0.1540
Residual error	7	11.22	1.60	–	–
Lack-of-fit	3	10.31	3.44	15.07	0.0120
Pure error	4	0.91	0.23	–	–

d.f. – degrees of freedom

Optimization of soaking process using RSM

For optimization of soaking conditions, desirability functions of RSM were used. When generating the optimum conditions, percentage water uptake and percentage protein retention in the soybeans were allotted higher mass values than percentage mass gain. Percentage water uptake and percentage protein retention significantly impact the soybean hydration process in terms of processing ease during subsequent grinding operations and also securing nutritional benefits by retaining maximum protein. The mass values (s) for percentage mass gain and percentage protein retention were taken as one, and for percentage water uptake the s value was considered to be 0.1, yielding an overall maximized desirability D of 0.7947, as seen in Eq. 4. The optimal values for the test variables of soybean hydration, based on the above desirability value, were obtained as a soybean to water ratio of 1:2.48, soaking temperature of 40 °C and soaking time of 2 h.

Validation of the RSM model

To evaluate the effectiveness of the developed regression equations, randomly selected experimental results were statistically compared with response surface predictions. Table 3 shows the selected soybean soaking conditions, regression equation responses and experimental data resulting from the selected conditions. Comparison of the predicted results with experimental data shows a very small error (R^2=0.889–0.9297 and MSE=0.80–4.90), which shows good agreement between RSM results and experimental data.

Neural network modelling

In order to use GA to search for the optimal soaking conditions, objective fitness function was used to guide the search direction. The neural network concept was used to develop various models required by GA for optimization. Each ANN model was trained by a set of data and the performance of the trained ANN model was then tested by another set of independent data not used for training. Fig. 1 shows the test results of the ANN model using the log sigmoid function which gave the best results compared to other transfer functions.

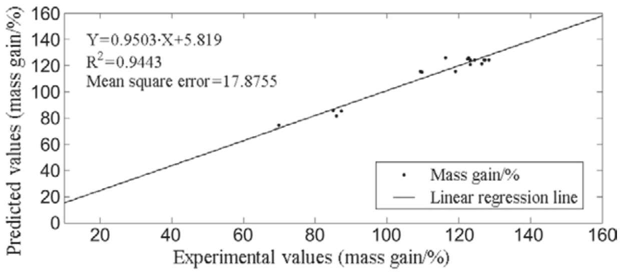

Fig. 1. Correlation between the values predicted by ANN and experimental data of percentage mass gain under different experimental conditions

The neural network architecture comprised an input layer consisting of 3 neurons, a hidden layer consisting of 5 neurons, and an output layer consisting of 3 neurons with a learning rate of 0.75. The figure shows that the values predicted by ANN closely matched the results of experiments. The high R^2 (0.9443) and low mean square error (17.8755) confirmed that the neural network models can adequately describe the relationship between soaking function parameters and each response variable. Consequently, it was reliable for the trained ANN models to be employed as fitness functions in GA.

Genetic algorithms for optimization

Fig. 2 shows the ANN/GA optimization profile for soaking conditions. This figure illustrates the variation

Table 3. RSM-predicted and experimental values of water uptake, mass gain and protein retention resulting from randomly selected soybean soaking conditions for RSM model validation

RSM response variables			Experimental results (y_e)			RSM-predicted results (y_o)		
Solid/ liquid ratio	Temperature/ °C	Time/h	Water uptake/%	Mass gain/%	Protein retention/%	Water uptake/%	Mass gain/%	Protein retention/%
1:1.5	50	3	53.8	120.4	88.9	54.12	121.18	88.62
1:1.5	60	4	54.6	121.8	87.2	51.71	123.86	87.39
1:1.5	50	6	53.8	126.4	86.2	49.66	126.19	86.41
1:2	60	3	71.2	103.4	91.3	72.45	104.42	90.07
1:2	50	4	75.0	109.6	90.8	71.02	110.53	91.08
1:2	60	6	69.8	109.7	88.5	70.50	107.36	89.43
1:2.5	50	3	63.2	113.5	90.3	62.53	115.74	89.90
1:2.5	60	4	60.1	113.9	85.5	60.09	115.34	86.77
1:2.5	50	6	56.7	118.6	88.3	58.91	120.44	87.77
1:3.5	60	3	50.8	128.9	85.1	46.72	124.21	85.56
1:3.5	50	4	43.4	126.3	88.3	44.96	127.16	87.15
1:3.5	60	6	44.7	125.0	83.9	42.26	128.04	84.71
1:4	50	3	43.6	123.2	89.6	43.09	125.35	89.26
1:4	60	4	41.0	127.4	86.2	39.32	128.71	85.40
1:4	50	6	36.8	128.5	85.8	36.97	130.96	86.92

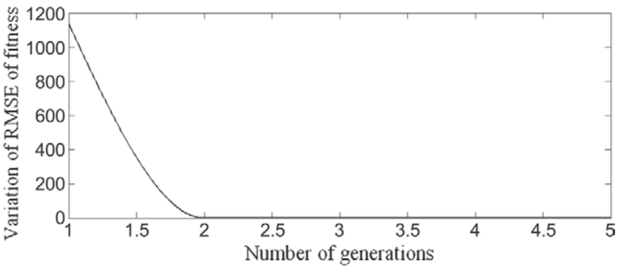

Fig. 2. The ANN/GA optimization process for maximizing percentage mass gain, percentage water uptake and percentage protein retention

of root mean square error (RMSE) of fitness value down the number of generations. As shown, the value of RMSE decreases across generations until it becomes zero at around 2 generations. For this reason, the termination criterion for the ANN/GA model was set to 5 in all the experiments.

ANN/GA model for the optimization of soaking conditions for soybeans indicated that the problem of the optimization lies in many-to-one mapping. In other words, for different combinations of solid/liquid ratio, time and temperature of incubation can result in the same optimum conditions for maximizing percentage mass gain, percentage water uptake and percentage protein retention in the soaked beans. A few optimal conditions were selected from a pool of optimal ANN/GA responses, as shown in Table 4. These results imply that in the case of soybean soaking, maximized as well as identical percentage mass gain, percentage water uptake and percentage protein retention in hydrated soybeans can be reached under different optimal soaking conditions. A feasible set of conditions can be selected among the different proposed ANN/GA optimal soaking conditions. Predicted optimal conditions using ANN/GA were applied in actual settings of soybean soaking. Percentage mass gain, percentage water uptake and percentage protein retention in the hydrated beans were determined experimentally.

Table 4. Soaking conditions obtained from ANN/GA response pool leading to the same percentage water uptake, percentage mass gain and percentage protein retention in the soaked beans

ANN/GA optimal operating conditions		
Solid/liquid ratio	Temperature/°C	Time/min
1:1.19	45.1	8.0
1:1.14	46.2	8.0
1:1.09	46.3	7.2
1:1.27	49.2	7.7
1:1.03	45.9	7.4
ANN/GA predicted maximum mass gain, water uptake and protein retention		
Mass gain/%	Water uptake/%	Protein retention/%
103.39	100.00	99.99
103.39	100.00	99.99
103.39	100.00	99.99
103.39	100.00	99.99
103.39	100.00	99.99

Validation of the ANN/GA model

To evaluate the effectiveness of the scheme, randomly selected experimental results were statistically compared with ANN/GA predictions. Fig. 3 shows the selected soybean soaking conditions as well as ANN/GA responses and experimental data resulting from the selected conditions. Comparing the ANN/GA results with the experimental data, a very small error (R^2 and MSE being 0.9380 and 5.9299, respectively) is seen, which shows a very good agreement between the ANN/GA results and the experimental data.

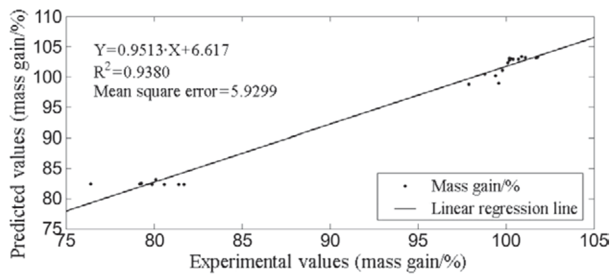

Fig. 3. Correlation between ANN/GA-predicted results and the experimental data of percentage mass gain under different experimental conditions

Conclusion

An ANN model for soybean hydration was proposed using a set of experimental data. The evaluation of the ANN model with test data indicated considerable robustness of the model. The application of this model lies in obtaining optimum reaction conditions through GA, which results in maximization of percentage mass gain, percentage water uptake and percentage protein retention in the hydrated soybean. Validation of the GA model indicated its considerable effectiveness. The optimization of hydration conditions was also performed by RSM. The optimum process conditions obtained from hybrid ANN/GA as well as RSM indicated a considerable reduction in the soybean soaking time compared to the time-consuming industrial process. Moreover, the optimum time of soaking obtained by RSM is just 2 h, compared to the optimum soaking time of 8 h predicted by ANN/GA. The RSM optimum temperature of soaking (40 °C) is also lower than that predicted by ANN/GA (45 °C). The optimum RSM process conditions with reduction in time are better than the existing soaking process practiced in the industry today.

References

1. C.H. Lee: *Fermentation Technology in Korea*, Korea University Press, Seoul, Korea (2001).
2. H.E. Snyder, T.W. Kwon: *Soybean Utilization*, Van Nostrand Reinhold Company, New York, USA (1987).
3. A.M. Pearson: Soy Proteins. In: *Developments in Food Proteins*, B.J.F. Hudson (Ed.), Applied Science Publishers, Essex, UK (1983) pp. 67–108.

4. S. Bandyopadhyay, N.C. Roy, A semi-empirical correlation for prediction of hydration characteristics of paddy during parboiling, *Int. J. Food Sci. Technol. 13* (1978) 91–98.

5. C. Engels, M. Hendrickx, S. De Samblanx, I. De Grijze, P. Tobback, Modeling water diffusion during long-grain rice soaking, *J. Food Eng. 5* (1986) 55–73.

6. G.N. Agbo, G.L. Hosfield, M.A. Uebersax, K. Klomparens, Seed microstructure and its relationship to water uptake in isogenic lines and a cultivar of dry beans (*Phaseolus vulgaris* L.), *Food Microstruct. 6* (1987) 91–102.

7. S.S. Deshpande, M. Cheryan, Microstructure and water uptake of *Phaseolus* and winged beans, *J. Food Sci. 51* (1986) 1218–1223.

8. I. Marbach, A.M. Mayer, Permeability of seed coats to water as related to drying conditions and metabolism of phenolics, *Plant Physiol. 54* (1974) 817–820.

9. T. Toda, A. Sakamoto, T. Takayanagi, K. Yokotsuka, Changes in isoflavone composition of soybean during soaking in water, *Food Sci. Technol. Res. 7* (2001) 171–175.

10. K. Liu: Nonfermented Oriental Soyfoods. In: *Soybeans Chemistry, Technology, and Utilization*, K. Liu (Ed.), International Thomson Publishing, New York, USA (1997) pp. 152–153.

11. W.Y.L. Lo, K.H. Steinkraus, D.B. Hand, L.R. Hackler, W.F Wilkens, Soaking soybeans before extraction as it affects chemical composition and yield of soymilk, *Food Technol. 22* (1968) 1188.

12. A. Gowen, N. Abu-Ghannam, J. Frias, J. Oliveira, Influence of pre-blanching on the water absorption kinetics of soybeans, *J. Food Eng. 78* (2007) 965–971.

13. H.L. Wang, E.W. Swain, C.W. Hesseltine, H.D. Heath, Hydration of whole soybeans affects solids losses and cooking quality, *J. Food Sci. 44* (1979) 1510–1513.

14. P.A. Sopade, J.A. Obekpa, Modelling water absorption in soybean, cowpea and peanuts at three temperatures using Peleg's equation, *J. Food Sci. 55* (1990) 1084–1087.

15. R. Chopra, D.N. Prasad, Standardization of soaking conditions for soybean seeds/cotyledons for improved quality of soymilk, *Indian J. Anim. Sci. 64* (1994) 405–410.

16. N. Haladjian, R. Fayad, I. Toufeili, S. Shadarevian, M. Sidahmed, E. Baydoun, M. Karve, pH, temperature and hydration kinetics of faba beans (*Vicia faba* L.), *J. Food Process. Preserv. 27* (2003) 9–20.

17. E. Sangronis, A. Ibarz, G.V. Barbosa-Cánovas, B.G. Swanson, Effect of high hydrostatic pressure on water imbibition, cooking times and microstructure of *Phaseolus vulgaris*, *Arch. Latinoam. Nutr. 52* (2002) 301–306.

18. H.E. Gan, R. Karim, S.K.S. Muhammad, J.A. Bakar, D.M. Hashim, R.A. Rahman, Optimization of the basic formulation of a traditional baked cassava cake using response surface methodology, *Lebensm. Wiss. Technol. 40* (2007) 611–618.

19. P.S. Madamba, The response surface methodology: An application to optimize dehydration operations of selected agricultural crops, *Lebensm. Wiss. Technol. 35* (2002) 584–592.

20. S.S. Sablani, H.S. Ramaswamy, S.O. Prasher, A neural network approach for thermal processing applications, *J. Food Process. Preserv. 19* (1995) 283–301.

21. X. Liu, X. Chen, W. Wu, G. Peng, A neural network for predicting moisture content of grain drying process using genetic algorithm, *Food Control, 18* (2007) 928–933.

22. S. Nandi, P. Mukharjee, S.S. Tambe, R. Kumar, B.D. Kulkarni, Reaction modeling and optimization using neural networks and genetic algorithms: Case study involving TS-1 catalyzed hydroxylation of benzene, *Ind. Eng. Chem. Res. 41* (2002) 2159–2169.

23. D. Sarkar, J.M. Modak, Optimisation of fed-batch bioreactors using genetic algorithms, *Chem. Eng. Sci. 58* (2003) 2283–2296.

24. O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent, *J. Biol. Chem. 193* (1951) 265–275.

25. K.J. Rao, C.H. Kim, S.K. Rhee, Statistical optimization of medium for the production of recombinant hirudin from *Saccharomyces cerevisiae* using response surface methodology, *Process Biochem. 35* (2000) 639–647.

26. D.C. Montgomery: *Design and Analysis of Experiments*, Wiley, New York, USA (2001).

27. C. Gan, V. Limsombunchai, M. Clemes, A. Weng, Consumer choice prediction: Artificial neural networks versus logistic models, *J. Soc. Sci. 1* (2005) 211–219.

28. B. Yegnanarayana: *Artificial Neural Networks*, Prentice-Hall, New Delhi, India (2003).

29. E.G. Shopova, N.G. Vaklieva-Bancheva, BASIC – A genetic algorithm for engineering problems solution, *Comput. Chem. Eng. 30* (2006) 1293–1309.