

Modelling of high-pressure treatments of foods by an Artificial Neural Network

J.S. Torrecilla, ¹L. Otero, ¹B. Guignon, ¹P.D. Sanz, ²A. Ramos

Univ. Complutense de Madrid, Dep. Chemical Engineering. Avda. Complutense s/n, 28040 Madrid, Spain

tel. +34 913944240, fax: +34 913944243, e-mail: jstorre@quim.ucm.es ; ¹Dep. Engineering, Instituto del Frío (CSIC), c/ José Antonio Novais, 10. 28040 Madrid, Spain; ²Univ. Complutense de Madrid, Dep Applied Mathematics, Avda. Complutense s/n, 28040 Madrid, Spain.

ABSTRACT

An Artificial Neural Network (ANN) has been used to predict two parameters of interest for high-pressure food processing: the maximum or minimum temperature reached in the sample after pressurization and the time needed for thermal re-equilibration in the high-pressure system. The ANN was training using 14 different algorithms and then the best one was selected. Afterwards, an experimental design carried out in order to optimize the ANN. The Levenberg-Marquard as training function, topology of 5,12,2, learning coefficient of 0.001 and learning coefficient decrease of 1 appears to be the most suitable for such application. The prediction errors and the correlation coefficients are 0.4, 0.9994 and 0.62, 0.998 % to time and temperature, respectively. Therefore, the neural networks become now an accuracy method alternative to physical-based models in the prediction of the thermal behavior of food under pressure.

1. INTRODUCTION

The artificial neuronal networks are mathematical algorithms that learn to solve a problem with a set of known questions-answers, i.e., they do not require any phenomenological knowledge. They have already been employed successfully to predict properties of substances or to solve physically-based models where the almost total lack of appropriate thermal properties of foodstuffs (Torrecilla, Otero & Sanz, 2004; Torrecilla, Otero & Sanz, 2005). Nowadays, the interest in applying this technology is to control the microbiological and/or enzymatic activity of food products.

From a thermodynamic point of view it is very interesting the coupled pressure-temperature effect due to any high-pressure treatment. By one side, the pressure effect could be considered as instantaneous phenomena as it is regulated by the Pascal principle. By the other side, the temperature variations due to the corresponding applied pressure is regulated by the heat and mass transfer phenomena and acquire a transient behavior. Some attempts have been made by different authors to model the thermal behavior of food during high-pressure treatments, but a number of difficulties are met (Denys et al., 2000a; Denys, Van Loey and Hendrickx, 2000b; Schlüter, Heinz and Knorr, 1998; Schlüter, George, Heinz and Knorr, 1999; Otero & Sanz, 2003). The main problem of physically-based models is the almost total lack of appropriate thermal properties of foodstuffs and pressurizing fluids at elevated pressure.

The aim of this work is obtain a model that predicts with an error comparable to other commercial methods. This is reached by the selection and optimization the most adequate training function of the ANN. This modelization would be very useful to design and optimize pressure food processes because the temperature and time variables provide together with the

programmed pressure a proper notion about the thermal evolution of the sample during the pressure treatment (Ting et al, 2002).

2. MATERIALS AND METHODS

2.1. Sample

Usually the ANN uses a huge amount of data to optimize its parameters. The sample data must be characteristic of the problem to be simulated. The data have been obtained from a physically-based corresponding to a high-pressure simulation model as described in Torrecilla, Otero & Sanz (2005).

Two independent data sets were used to optimize the ANN: a training set made up of 300 sets of input and output data and a validation set made up of 50 sets, both have the same format and using the processing conditions shown in Table 1. All of these are composed of seven columns in each one is pressure applied (MPa), pressure increase rate (MPa/s), set point temperature (°C), high-pressure vessel temperature (°C) and temperature of the air surrounding the high-pressure system or ambient temperature (°C), the maximum or minimum temperature (°C) reached in the sample after compression and one representing the time (s) needed to re-equilibrate the temperature in the sample after pressurization and each run is classified in rows

Table 1.- Range of processing conditions used as input variables in the ANN model.

VARIABLES	POSSIBLE VALUES
Applied pressure (MPa)	250, 350, 450
Pressure increase rate (MPa/s)	1, 1.5, 2
Set point temperature (°C)	40, 50, 60, 70, 80
High-pressure vessel temperature (°C)	Between set point temperature and ambient temperature (10°C intervals)
Ambient temperature (°C)	10, 20, 30

2.2. ANN description

The ANN used is a Multilayer Back propagation Perceptron. It is formed by several artificial neurons arranged in two layers (an output layer and one or more hidden layers) connected to each other by weighted links, figure 1. An artificial neuron consists of two mathematical algorithms, transfer function and activation function. Each one receives inputs from all neurons of the previous layer in proportion to their connection weights and calculates a single output which will be propagated to all other neurons of the following layers (Sreekanth et al., 1999).

The ANN is able to interpolate the input values with a small prediction error without a prior knowledge of the relationships between the process variables (Palancar, Aragón, Torrecilla, 1998). In this, the weights are optimized by iterative calculations.

In this study a standard feed-forward back-propagation algorithm with a prediction horizon, and supervised learning function with the sigmoid transfer function was applied. From an ANN topology (neurons for each layer) point of view, the ANN was designed with one hidden layer (Torrecilla, Otero, Sanz, 2005). Therefore, the ANN is composed of three layers: input, hidden and output. The input layer consists of equal number of nodes and variables necessary to characterize the problem to be simulated. This layer had five nodes that corresponded to five input variables: pressure applied (MPa), pressure increase rate (MPa/s), set point temperature (°C), high-pressure vessel temperature (°C) and temperature of the air surrounding the high-

pressure system or ambient temperature (°C). In order to give a proper idea of thermal evolution of the sample during the pressure treatment, the maximum or minimum temperature (°C) reached in the sample after compression and one representing the time (s) needed to re-equilibrate the temperature in the sample after pressurization (Ting et al., 2002) were predicted by the ANN. Therefore, the output layer consisted of two neurons.

Matlab version 7.01.24704 (R14) Service Pack 1 and QuickBasic software, version 5.4 were used to ANN modeling and training function selection and Statgraphics Plus version 5.1 was used to carry out the statistical tests in order to select the most adequate training function and develop the experimental design.

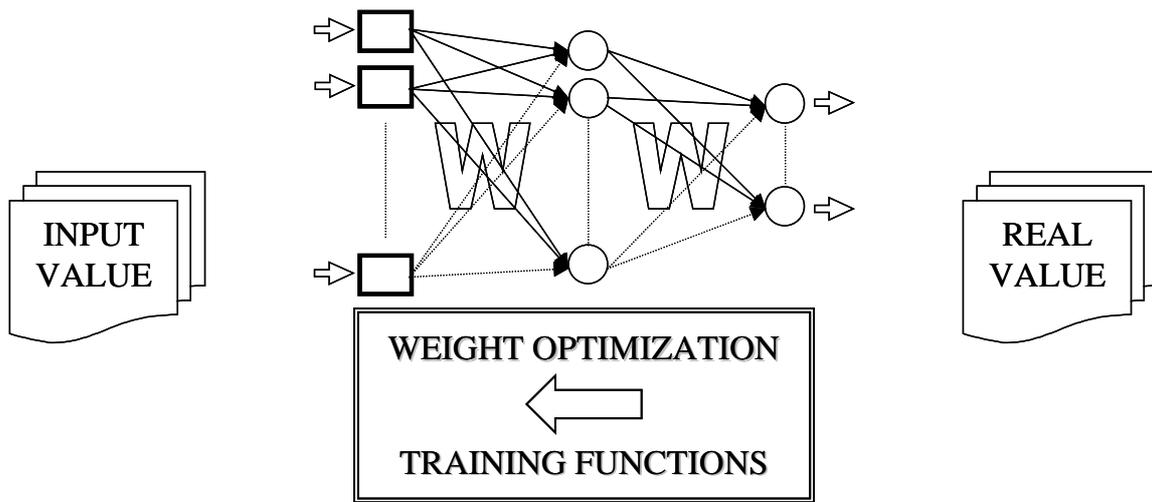


Figure 1 Training data set and structure of the neural network model schematically shown. A weight optimization with different training functions.

3. RESULTS AND DISCUSSION

The characteristics of the first ANN used in this study were optimized in Torrecilla, Otero & Sanz (2005), i.e., the topology consists of five nodes in the input layer, seven neurons in the hidden layer and two neurons in the output layer. The learning coefficient was set to 0.5. These parameters were used in this study to the selection of the better training function, and then, the new characteristics of the second ANN (using the optimized training function) were optimized.

The selection of the training function was carried out in two basic steps: Training step is applied to the ANN with the parameters mentioned above and the weights are optimized using different training functions, mentioned in table 2. In each case, the prediction error and optimized weights are saved. With these weights a verification step is applied.

3.1. Training and verification step

The training function was selected from among the 14 training functions shown in table 2. At the beginning of the training step, the weights of the ANN were initialized randomly between 0 and 1. Afterwards, the weights were adjusted to reduce the prediction error (eq 1) in each neuron from the output to the input layer using the training algorithms summarised in table 2.

Table 2.- Summary of training function used (Demuth; Beale; Hagan, 2005)

Training Function	Description
Gradient descendent BP (traingd)	Slow response, can be used in incremental mode training.
Gradient descendent with momentum BP (traingdm)	Generally faster than traingd. Can be used in incremental mode training.
Gradient descendent with momentum & adaptive linear BP (traingdx)	Adaptive learning rate. Faster training than traingd, but can only be used in batch mode training.
Gradient descendent with adaptive learning learning rate BP (traingda)	
Random order incremental update (trainr)	Simple batch mode training algorithm with fast convergence and minimal storage requirements
Resilient PB (Rprop) (trainrp)	
Fletcher Powell conjugate gradient BP (traingcf)	The algorithm has the smallest storage requirement of the conjugate gradient algorithms.
BFGS Quasi-Newton Backpropagation (trainbfg)	BFGS quasi-Newton method Requires storage of approximate Hessian matrix and has more computation in each iteration than conjugate gradient algorithms, but usually converges in less iteration.
One step secant BP (trainoss)	Compromise between conjugate gradient methods and quasi-Newton methods.
Levenberg-Marquardt BP (trainlm)	The fastest training algorithm for networks of moderate size. It has memory reduction feature for use when the training set is large.
Bayesian Regularization (trainbr)	Modification of the Levenberg-Marquardt training algorithm to produce networks that generalizes well. It reduces the difficulty of determining the optimum network architecture.
Polak-Ribiere conjugate gradient BP (traingcp)	Slightly larger storage requirements than traingcf. Faster convergence on some problems.
Powell-Beale conjugate gradient Backpropagation (traingcb)	Slightly larger storage requirements than traingcp. Generally faster convergence.
Scaled conjugate gradient BP (traingscg)	The only conjugate gradient algorithm that requires no line search. A very good general purpose training algorithm.

The prediction error (eq 1) was calculated again and back distributed across the network for the next modification. It was used by each training function to optimize the weights using the training data.

$$E_k = \frac{1}{2} \sum_k (r_k - y_k)^2 \quad (1)$$

For the given training data of inputs to the network, the response of each neuron in the output layer (y_k) was then calculated by the ANN and compared with the corresponding real

output response (r_k). Then the error associated with the output response (E_k) was computed and back distributed to the previous layers across the network. The weights were modified once for each row of the training set, that is, from the beginning to the end of the training set the weights were optimized 300 times, and then, the input data set was fed to the ANN once more, and a new estimation was made. This process was repeated while the prediction error decreased, and at this moment the weights are assumed as optimized.

In the verification step the competence of the trained network was evaluated. This second step consists of the input to the ANN of a new data set never shown before, and predicts time and temperature values without updated of the weights. The most optimal training function was selected from among algorithms mentioned in table 2. The criterion used to select the training function adequate is based on final error (error at the end of the learning process) the number of iterations needed to end the learning process and several statistical tests applied to the predicted and real values.

3.1.1. Training function selection

Training and verification steps were carried out for each fourteen different training functions shown in table 2. The data predicted in the verification step using all training functions were compared with the real values data, i.e., the fourteen predicted data set were compared with the same real data one by one. It was carried out by the application of several statistical tests, these could be classified as follow: non-parametric methods were applied based either on measures of central tendency (Kolmogorow-Smirnov test, Mann-Whitney-Wilcoxon test and Kruscal-Wallis test), on the variance (Kruscal-Wallis test, Cochran's test, Barlett's test, and Levene test) and inferential parametric test for significance (F-Test and t-test).

All of these statistical analyses (at 95% confidence level) were performed to determine if there were significant statistical differences between temperature and time data provided by the physically-based model (real data) and those predicted by the ANN. The null hypothesis assumes that statistical parameters of both series are equal. Otherwise, an alternative hypothesis is defined. P-value was used to check each hypothesis. Its threshold value was 0.05. If p value is greater than this value the null hypothesis is fulfilled. If P-value is closer to 1, the null hypothesis is fulfilled with more confidence. Figure 2 shows every calculated statistical tests versus p-values for every training function to time and temperature prediction and from a p-value point of view, all statistical tests confirm that real and predicted temperature and time data have a similar distribution. Taking into account that a training function has been considered as better as its p-value is closer to 1. In figures 3 is shown the two training function closer to 1 for the time and temperature. Given that the ANN is used to predict the time and temperature values, was necessary select only one training function. As can be seen in figure 3 (time and temperature predictions) the best training function in both cases is the Levenberg-Marquard Back propagation because the mean p-value is higher than the others functions in both cases. Therefore, the training function selected was Levenberg-Marquard Back propagation.

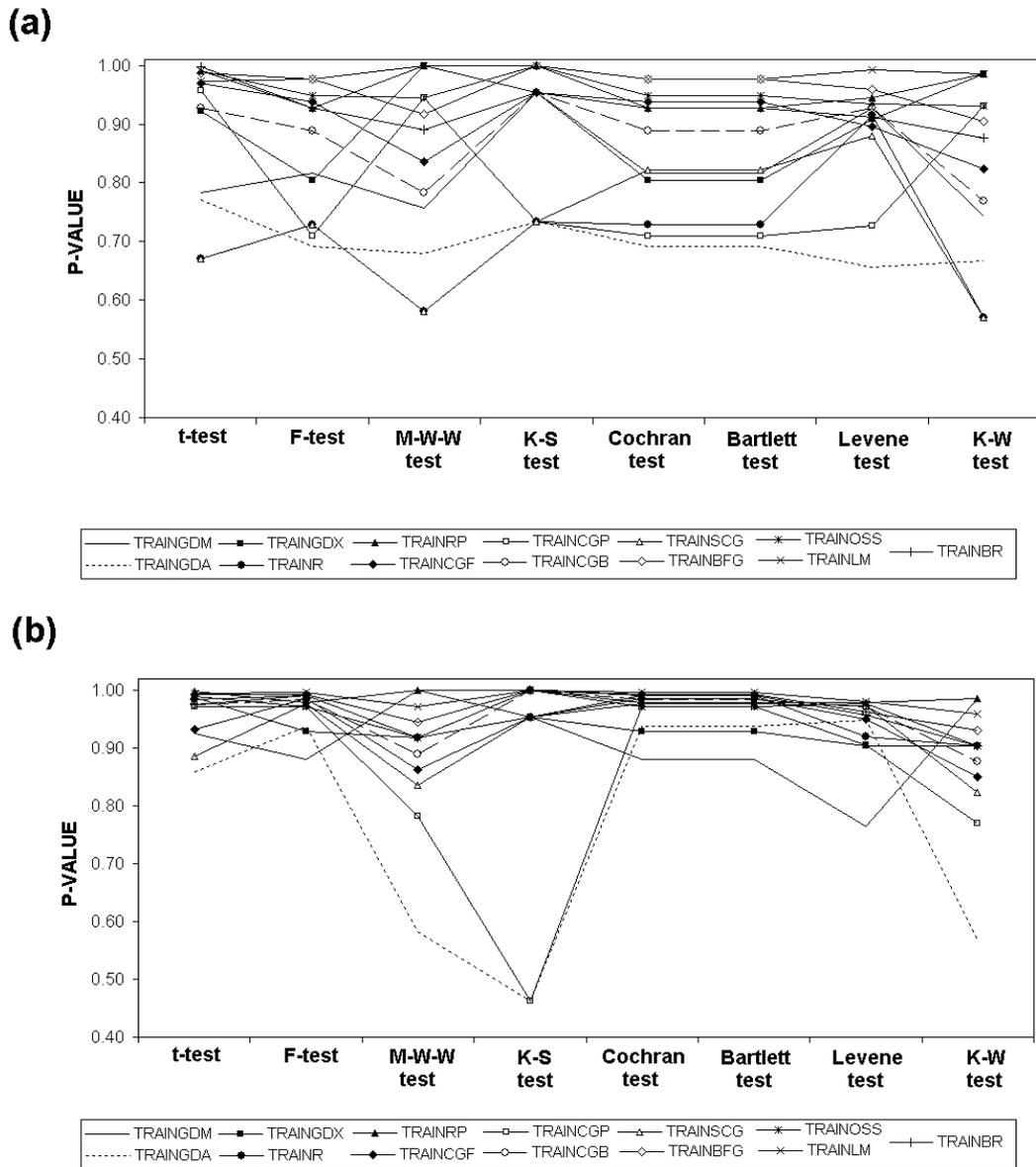


Figure 2 Statistical analyses versus p-value for all training functions studied; (a) time; (b) Temperature.

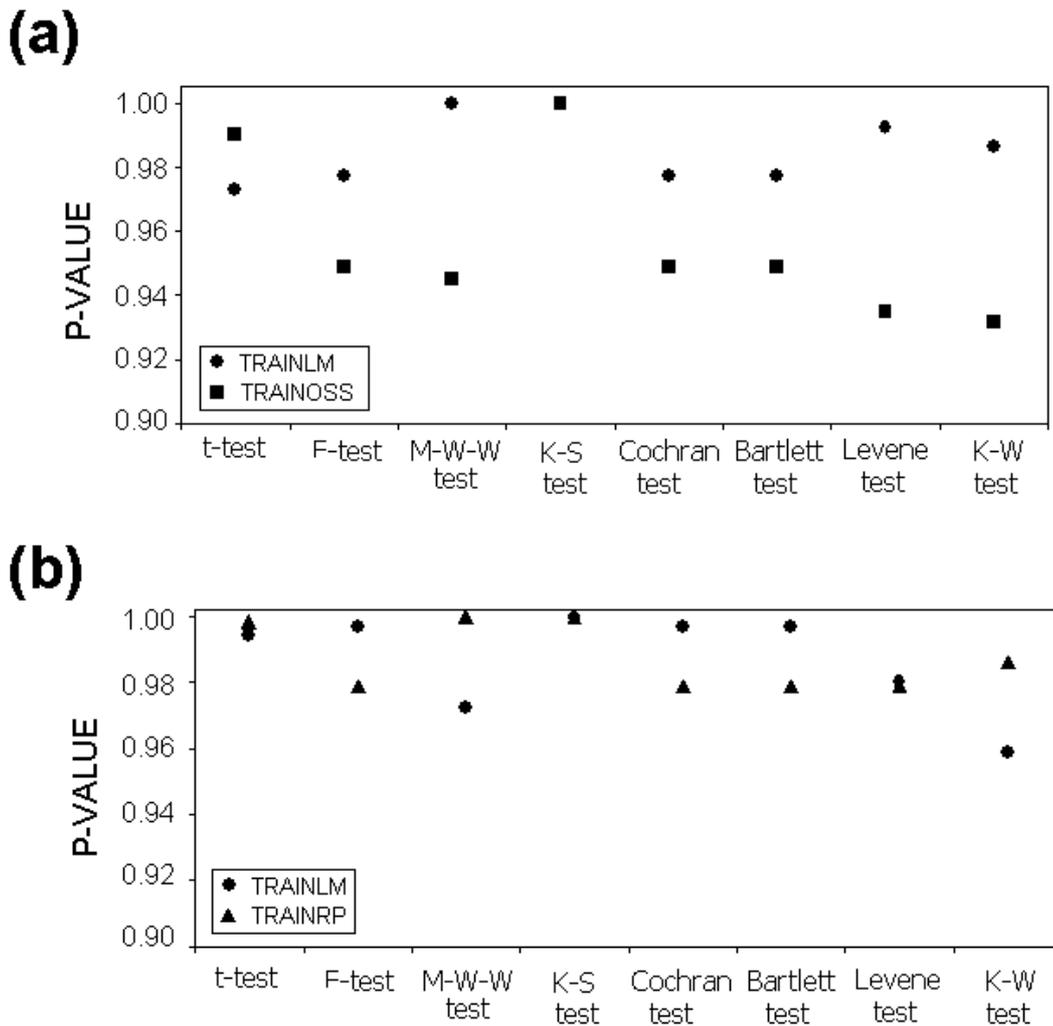


Figure 3 Selected statistical analyses versus p-value for all training functions studied; (a) time; (b) Temperature.

3.2. ANN topology and learning coefficients optimization

The training function of the ANN has changed from characteristics mentioned above to Levenberg-Marquard Back propagation. Therefore, its parameters (topology, Learning Coefficient (μ), Learning coefficient decrease (μd)) must be optimized. This optimization was carried out by an experimental design. It was a Screening Factorial, the factors were Topology (from 1 to 12 neurons in the hidden layer), Learning coefficient (from 0.001 to 1) and learning coefficient decrease (from 0.1 to 1). These ranges were taken from Vacic, 2005. The responses were the number of iteration, final prediction error, correlation coefficient to time and temperature (real versus predicted one), mean prediction error of the time and the temperature, and the number of weights of the ANN. For every conditions of factor (topology, learning coefficient and learning coefficient decrease) training and verification steps were carried out.

The aim of this experimental design was to find out the adequate parameters values of the ANN in order to predict the temperature and time with the lesser error as possible, and the correlation coefficients (real vs. prediction) must be as close to unity as possible. As can be seen in figure 4, the responses mainly depend on the topology. Bear in mind the prediction error and the correlation coefficient values; the number of hidden neurons was fixed 12. On the other hand, the lowest value of learning coefficient produce less prediction error and correlation coefficient closer to 1, figure 4. Given that the ANN was set to 12 neurons and the learning coefficient is 0.001, the ANN prediction with the least error was developed to learning coefficient decrease equal to 1, therefore the learning coefficient decrease was fixed to 1

As can be seen in figure 4, the topology has the biggest influence in the prediction error and correlation coefficient. Therefore, it was optimized carefully. Different training and verification steps were carried out to eight different topologies tested (7 to 13 neurons in the hidden layer). From an experimental design point of view, this range of topologies were selected because an ANN with a higher number of hidden neurons predict with lesser error, but more than 13 hidden neurons involves the optimization of more than 93 weights values and the amount of memory used by the algorithm is too many and the compute time of the PC used is too long.

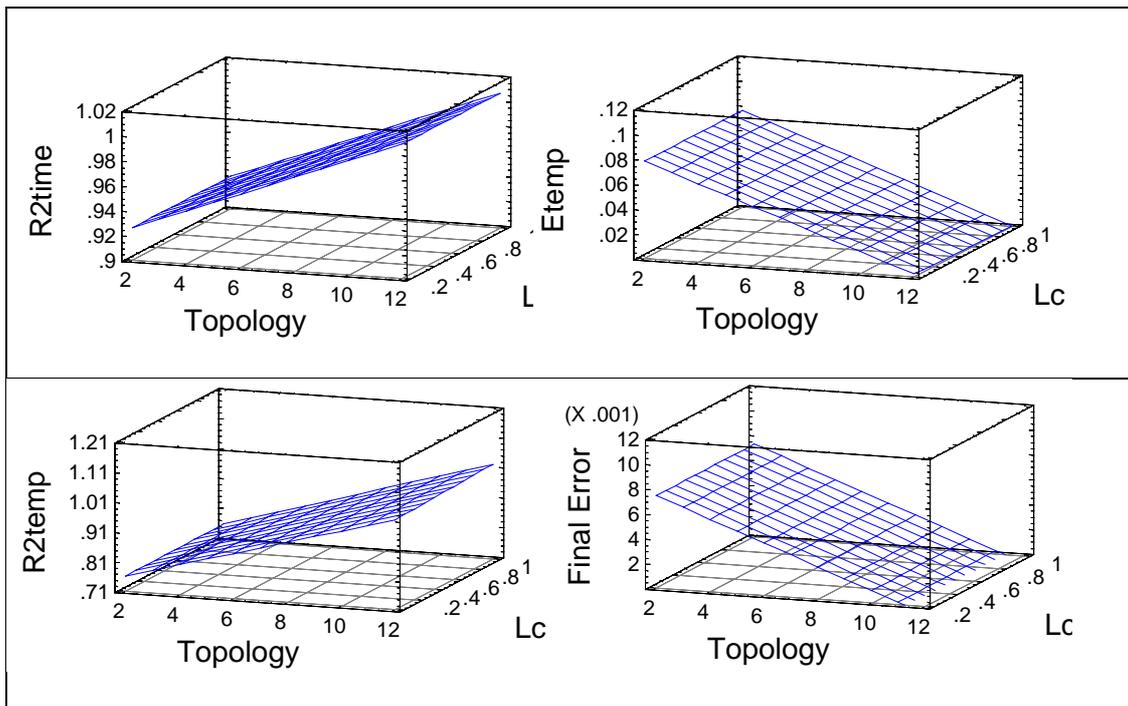


Figure 4 Dependence study by experimental design

The minimum prediction error and number of learning runs required to finish the learning process corresponded to 12 neurons. The correlation coefficients were low for topologies lesser than 12 neurons in the hidden layer. For 13 neurons in the hidden layer, the correlation coefficient decreased, Figure 5. Thus, the topology 5, 12, 2 (5 nodes in the input layer, 12 neurons in the hidden layer and 2 neurons in the output layer) was selected as the optimal topology.

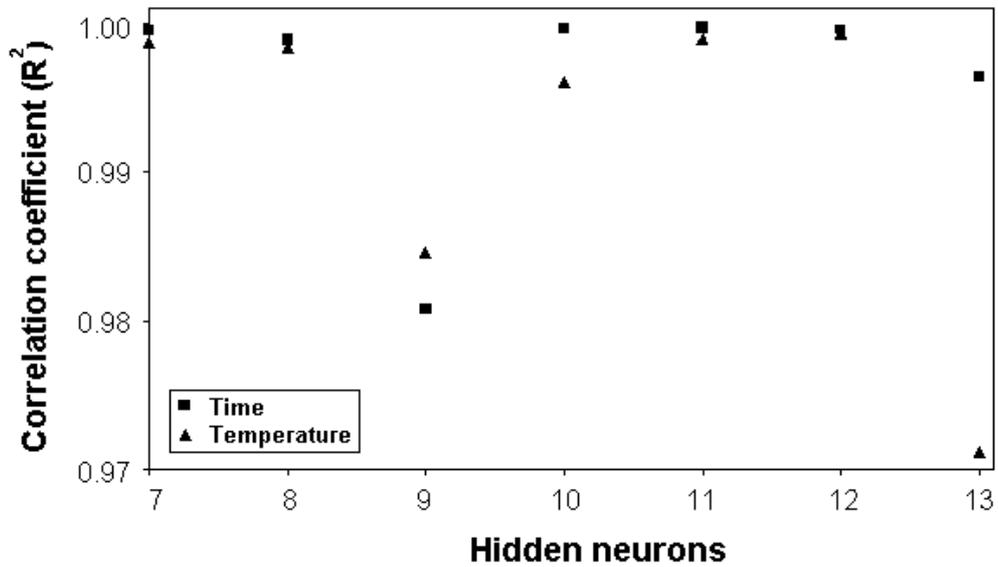


Figure 5 Optimization of hidden neurons using the Levenberg-Marquard.

In order to evaluate the competence of the optimised ANN. Fifty independent test data sets were employed. In this step, the selected topology (5, 12, 2), the optimised learning coefficient ($\mu=0.001$) and learning coefficient decrease (1) were used with the adjusted weights previously. Now, no corrections of these weights were made and the ANN was only used to predict the maximum or minimum temperature reached in the sample after pressure loading and the time needed to reach the thermal equilibrium. The real time and temperature values versus the predicted ones and the line fitting are shown in figure 6. The correlation coefficient (R^2) values were higher than 0.999 and 0.998 for time and temperature, respectively. On the other hand, the mean error values are about 0.62% for temperature predictions and 0.40% for time ones. The improving in the modelation process respect the last study is shown in table 3. These prediction errors and correlation coefficient values can be considered as accurate enough to predict and model.

Table 3.- Improvement of the ANN predictions

	Torrecilla, Otero & Sanz, 2005		Actual Study	
	Mean Prediction error (%)	Correlation Coefficient	Mean Prediction error (%)	Correlation Coefficient
Time	0.95	0.996	0.4	0.9994
Temperature	1.6	0.98	0.62	0.998

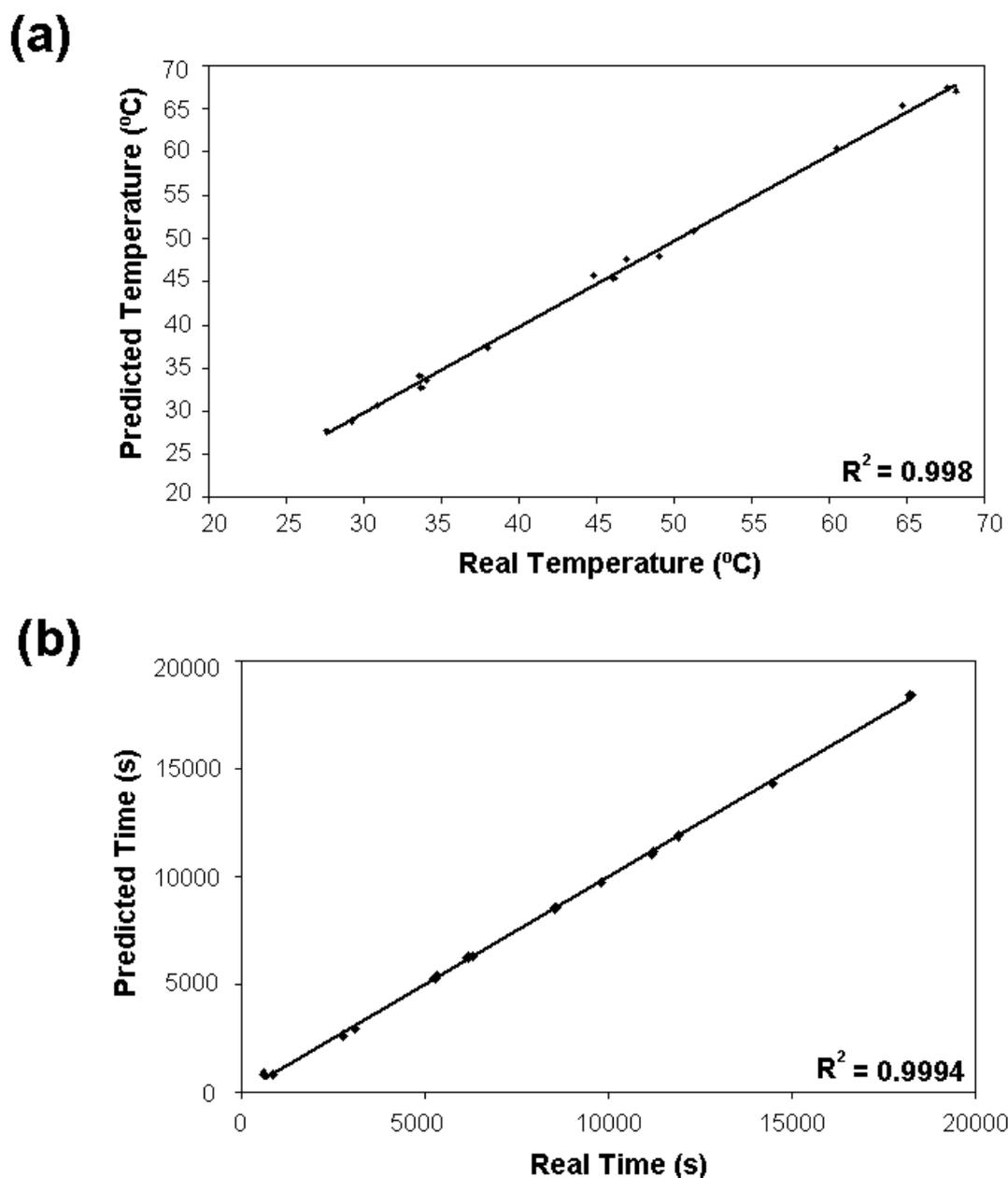


Figure 6 Competence evaluation of the optimization of the ANN; (a) Time; (b) Temperature.

4. CONCLUSIONS

A back propagation perceptron with Levenberg-Marquard as training function, topology of 5 input nodes, 12 hidden neurons and 2 neurons in the output layer, learning coefficient of 0.001 and learning coefficient decrease of 1 provide a very accurate notion about the thermal evolution of a sample during a pressure treatment. The prediction error in this work (0.4% and 0.62% to time and temperature, respectively) was lesser than the previous (0.95% and 1.6% to time and temperature, respectively), the correlation coefficient was also improved in this work (0.9994 and 0.998 to time and temperature, respectively) was closer to 1 than the last one (0.996

and 0.98 to time and temperature, respectively), these improvements were reached by the selection of training function. By using this simulation, the influence of the pressure level, the pressurization rate, the ambient and target temperatures or the initial high-pressure vessel temperature can be easily determined and the values of the process variables therefore adjusted to obtain the desired thermal evolution.

ACKNOWLEDGEMENTS

This paper was carried out with financial support from the Spanish ‘‘Plan Nacional de I + D + I (2004– 2006) MCYT’’ through the AGL2003-06862-C02-01 project and the Commission of the European Communities, specific RTD programme ‘‘Quality of Life and Management of Living Resources’’, QLK1-CT-2002-D2230 project, ‘‘Low temperature-pressure processing of foods: Safety and quality aspects, process parameters and consumer acceptance’’. L. Otero and J. S. Torrecilla were supported by Ramón y Cajal research contracts from the Ministerio de Educación y Ciencia in Spain.

REFERENCES

- Demuth H.; Beale, M.; Hagan, M. (2005), ‘‘Neural Network Toolbox For Use with MATLAB’’; Neural Network Toolbox User’s Guide (2005) by The MathWorks, Inc.
- Denys, S., Van Loey, A. M., & Hendrickx, M.E. (2000a). A modelling approach for evaluating process uniformity during batch high hydrostatic pressure processing: combination of a numerical heat transfer model and enzyme inactivation kinetics. *Innovative Food Science & Emerging Technologies*, 1, 5-19.
- Denys, S., Van Loey, A. M., & Hendrickx, M.E. (2000b). Modeling conductive heat transfer during high-pressure thawing processes: Determination of latent heat as a function of pressure. *Biotechnology Progress*, 16, 447-455.
- Otero, L., & Sanz, P. D. (2003). Modelling heat transfer in high pressure food processing: a review. *Innovative Food Science and Emerging Technologies*, 4, 121-134.
- Palancar, M. C., Aragón, J. M., & Torrecilla, J. S. (1998). pH-control system based on artificial neural networks. *Industrial and Engineering Chemistry Research*, 37(7), 2729-2740.
- Schlüter, O., George, S., Heinz, V., & Knorr, D. (1999). Pressure assisted thawing of potato cylinders. In H. Ludwig, *Advances in High Pressure Bioscience and Biotechnology* (pp. 475-480). Springer-Verlag.
- Schlüter, O., Heinz, V., & Knorr, D. (1998). Freezing of potato cylinders during high-pressure treatment. In N. S. Isaacs, *High Pressure Food Science, Bioscience and Chemistry* (pp. 317-324). Cambridge, UK: The Royal Society of Chemistry.
- Sreekanth, S., Ramaswamy, H. S., Sablani, S. S., & Prasher, S. O. (1999). A neural network approach for evaluation of surface heat transfer coefficient. *Journal of Food Processing Preservation*, 23, 329-348.
- Ting, E., Balasubramaniam, V. M., & Raghubeer, E. (2002). Determining thermal effects in high-pressure processing. *Food Technology*, 56 (2), 31-35.
- Torrecilla, J. S., Otero, L., & Sanz, P. D. (2004). A neural network approach for thermal/pressure food processing. *Journal of Food Engineering*, 62, 89-95.

Torrecilla, J. S., Otero, L., & Sanz, P. D. (2005). Artificial neural networks: a promising tool to design and optimize high-pressure food processes. *Journal of Food Engineering*, 69, 299-306.

Vladimir Vacic (2005); "Summary of the training functions in Matlab's NN toolbox"; http://www.cs.ucr.edu/~vladimir/cs171/nn_summary.pdf.