ELSEVIER

International Journal of Pharmaceutics 110 (1994) 101–108

international journal of pharmaceutics

Research Papers

Effect of neural network topology and training end point in modelling the fluidized bed granulation process

Ere Murtoniemi [a,*], Pasi Merkku [a], Pirjo Kinnunen [b], Kauko Leiviskä [b], Jouko Yliruusi [a]

[a] *Pharmaceutical Technology Division, University of Helsinki, P.O. Box 15, FIN-00014 University of Helsinki, Finland*
[b] *Control Engineering Laboratory, Department of Process Engineering, University of Oulu, Linnanmaa, FIN-90570 Oulu, Finland*

(Received 1 October 1993; Modified version received 31 January 1994; Accepted 14 February 1994)

Abstract

The effect of the topology and the training end point of artificial neural networks (ANN) in the modelling of a fluidized bed granulation process is presented. The neural network topologies were designed on the basis of an earlier study (Murtoniemi et al., *Int. J. Pharm.,* 108 (1994) 155–163). In the first part of this study, the networks contained only one hidden layer in which the number of neurons was either 10, 15, 20 or 25. The training end points with all four networks ranged from 0.15 to 0.07, with a step length of 0.01. In the second part, the training end point was fixed to be 0.12, while the number of neurons in the hidden layer varied from 10 to 25. The main purpose of this study was to find a suitable ANN in regard to the generalization ability and to compare the results to those calculated on the basis of multilinear stepwise regression analysis. The results showed that the number of hidden layer neurons did not affect the generalization ability of the networks and a proper generalization ability was achieved with rather simple networks. The training end point, however, had a significant effect on the generalization ability and it also affects the number of iteration epochs needed. In complicated systems this probably will affect remarkably the time required for the training.

Key words: Artificial neural network; Multilayer feedforward network; Neurocomputing; Training end point; Process modeling; Multilinear stepwise regression analysis; Fluidized bed granulation

1. Introduction

In our previous study (Murtoniemi et al., 1994) it was concluded that the use of artificial neural networks (ANN) is a very promising method in modelling a complex pharmaceutical agglomeration process: fluidized bed granulation. The network topologies were studied rather superficially just to obtain general information on the suitability of different network topologies in modelling the fluidized bed granulation. More accurate analysis requires more detailed studies. Our earlier study suggested that the training end point is a critical factor in ANN analysis and therefore should be studied more carefully. The term train-

* Corresponding author. Tel: +358-0-191-2765; Fax: +358-0-191-2786.

SSDI 0378-5173(94)00064-C

ing end point is defined more precisely in the earlier paper (Murtoniemi et al., 1994).

Network topology is the stucture of a neural network, which consists of parallel and serial connected artificial neurons. Several different structures are used in ANN (Lisboa, 1992). In this study different variations of the most common structure (layered feedforward network) were used. Our earlier investigation as well as the theoretical principles (Knight, 1990) suggest that the more detailed study may focus on the neural network topologies in the case where there is only one hidden layer. In the previous study there were only two response variables (mean granule size and granule friability). In normal pharmaceutical applications, however, the number of responses is generally higher. Therefore, one additional response (granule flow rate) will be included in this study.

The effect of the training end point on the generalization ability was now studied more systematically using the same method as in the earlier study (Murtoniemi et al., 1994) by calculating the average error percentage between the test data and the predicted granule properties. The numerical value of the training end point in this study varied from 0.15 to 0.07, with a step of 0.01. In the earlier study the optimal training end point, from the point of view of generalization ability, was somewhere within this range. Because the training data used in this study were exactly the same as in the earlier investigation, this training area will now be examined in more detail.

The experimental test data in this study were chosen so that each test point was within the same range in the factor space as the experimental points used to train networks (interpolation). At the end the neural network models will be compared to the models generated by multilinear stepwise regression analysis.

Table 1
Factor levels of the input variables (inlet air temperature (T), atomizing air pressure (p) and binder solution amount (m)) and flow rate of granules

T (°C)	p (bar)	m (g)	Flow rate (s) ($x \pm \epsilon$) [a]
40	1.0	150	12.0 ± 0.1
40	1.0	300	13.0 ± 0.5
40	1.0	450	13.2 ± 0.3
40	1.5	150	12.1 ± 0.1
40	1.5	300	12.1 ± 0.2
40	1.5	450	12.6 ± 0.1
40	2.0	150	12.0 ± 0.2
40	2.0	300	11.8 ± 0.1
40	2.0	450	12.0 ± 0.1
50	1.0	150	12.1 ± 0.3
50	1.0	300	12.3 ± 0.3
50	1.0	450	12.7 ± 0.1
50	1.5	150	11.8 ± 0.3
50	1.5	300	11.7 ± 0.2
50	1.5	450	11.7 ± 0.1
50	2.0	150	11.8 ± 0.2
50	2.0	300	12.0 ± 0.2
50	2.0	450	11.7 ± 0.1
60	1.0	150	11.8 ± 0.1
60	1.0	300	12.9 ± 0.5
60	1.0	450	12.7 ± 0.3
60	1.5	150	11.6 ± 0.1
60	1.5	300	11.8 ± 0.1
60	1.5	450	12.0 ± 0.1
60	2.0	150	11.7 ± 0.1
60	2.0	300	12.0 ± 0.1
60	2.0	450	12.4 ± 0.1

[a] x, mean; ϵ, maximum error estimate defined as $\frac{1}{2}\cdot$(max − min) ($n = 3$).

2. Materials and methods

2.1. Study design

The study design is based on our earlier study (Murtoniemi et al., 1994). The inlet air temperature (T), atomizing air pressure (p) and binder solution amount (m) were used as the input variables. The factor levels (a 3^3 factorial design) used in the production of granules at different experimental points are shown in Table 1. All 2^3 factorial points were made in duplicate and the center point in quadruplicate.

2.2. Materials and preparation of granules

The materials and preparation of the granules were the same as in the earlier studies, introducing the advantages of multilinear stepwise regression analysis (Merkku and Yliruusi, 1993; Merkku et al., 1993, 1994) and artificial neural networks (Murtoniemi et al., 1994).

2.3. Granule properties

The mean granule size and the granule friability were measured as previously described by Murtoniemi et al. (1994). The granule flowability was measured by a flow-time and cone angle testing instrument (PharmaTest PTG, Pharma-Test, Germany) using three parallel measurements. The flowability was expressed as the flow time (s) for a 100 ml sample to flow through an 8 mm orifice.

2.4. Multilinear stepwise regression analysis

In earlier studies (Merkku and Yliruusi, 1993; Merkku et al., 1993, 1994), regression models were developed for the mean granule size, granule friability and granule flow rate using stepwise multilinear regression. The regression models for the mean granule size and granule friability were the same as in the previous paper (Murtoniemi et al., 1994). The regression model for the granule flow rate had the following form:

$$Y = 0.204T^2 + 0.27p^2 - 0.17pm - 0.285p + 0.246m + 11.77 \quad (1)$$

where Y is the granule flow rate (s), T denotes the inlet air temperature (°C), p is the atomizing air pressure (bar) and m represents the binder solution amount (g).

2.5. Training data

The factor levels and the numerical values for the mean granule size and granule friability were the same as in the previous study (Murtoniemi et al., 1994). Table 1 lists the factor levels used in the production of granules and the flow rates of the granules now used also as training data. Before training, average values for the flow rates were calculated in replicate experimental points and were used in the training of the networks. Both input and output variables were converted as in the earlier study (Murtoniemi et al., 1994) to values between 0 and 1 with 10% headroom before the training because the output of a neuron is restricted to values between 0 and 1 by the sigmoidal transform function (Dayhoff, 1990; Davalo, 1991).

Table 2
Experimental test data used in studying the generalization ability of the networks (T, inlet air temperature; p, atomizing air pressure; m, binder solution amount)

Batch	T (°C)	p (bar)	m (g)	Granule size (μm)	Friability (%)	Flow rate (s)
1_a	45	1.8	225	382	44.5	12.4
1_b	45	1.8	225	435	40.0	12.6
2_a	55	1.8	225	367	53.9	12.2
2_b	55	1.8	225	425	32.6	12.3
3_a	45	1.3	375	529	14.0	13.1
3_b	45	1.3	375	532	11.5	12.8
4_a	55	1.3	375	437	33.3	12.7
4_b	55	1.3	375	458	16.5	12.5
5_a	55	1.3	225	373	34.5	12.3
5_b	55	1.3	225	386	26.3	12.5

$(1-5)_a$ and $(1-5)_b$ are replicated experiments.

2.6. Validation data

In order to study the generalization ability of neural networks, five additional granulations were performed (Table 2). The factor levels were selected so that they were within the range of the original experimental data. The generalization abilities were studied by supporting the test data to a network and examining the output values, which were the predicted granule properties.

2.7. ANN simulator software

A commercially available MS Windows based artificial neural network simulator program package, NeuDesk V.2.10 (Neural Computer Sciences, U.K.), was used throughout the study in a 486 Personal Computer with an accelerator card, NeuSprint (Neural Computer Sciences, U.K.).

3. Results and discussion

3.1. Generalization ability of neural networks with different training end points

It is commonly known that the training influences the performance of the network (Hush and

Table 3
Average error percentages of four different networks using different training end points in the case of mean granule size

n	Training end point								
	0.15	0.14	0.13	0.12	0.11	0.10	0.09	0.08	0.07
10	19.68	16.77	14.03	12.46	12.69	13.08	14.45	14.75	14.62
15	21.65	16.13	13.91	12.21	12.36	12.16	12.38	12.98	13.01
20	18.11	12.43	13.48	12.39	12.36	12.11	12.44	12.61	12.67
25	21.26	15.52	15.09	12.25	12.29	12.82	13.31	12.74	12.92
Sum	80.70	60.85	56.51	49.31	49.70	50.17	52.58	53.08	56.22

n, number of neurons in the hidden layer.

Table 5
Average error percentages of four different networks using different training end points in the case of granule flow rate

n	Training end point								
	0.15	0.14	0.13	0.12	0.11	0.10	0.09	0.08	0.07
10	4.25	3.86	3.66	4.02	4.13	4.32	4.81	5.09	5.92
15	3.43	3.22	3.80	4.07	4.17	4.44	4.71	4.99	4.94
20	2.84	2.80	3.66	3.98	4.16	4.47	4.75	4.97	4.96
25	4.93	3.96	3.71	4.14	4.15	4.43	4.80	4.93	5.07
Sum	15.45	13.84	14.83	16.21	16.61	17.66	19.07	19.98	20.89

n, number of neurons in the hidden layer.

Horne, 1993; Lodewyck and Deng, 1993). The training affects the generalization ability in two phases. Soon after starting the training, the generalization ability of the network is improved. When the training is continued a certain point will be reached after which the generalization ability becomes worse. The training end point is the criterion used for the termination of the training.

Tables 3–5 show the significance of the training end point to the average error percentage of the neural network. The average error percentage is in this study used to characterize the generalization ability of the neural network. First, the diminution of the training end point improved the generalization ability (decreased the average error percentage) of the network but quite soon, if the training end point was too small, the average error percentages started to increase. After the training end point 0.12, irrespective of the different responses (mean granule size, granule friability and granule flow rate), every network (10, 15, 20 or 25 hidden neurons) became overtrained. The term overtrained has been explained in the preceding paper (Murtoniemi et al., 1994).

Table 4
Average error percentages of four different networks using different training end points in the case of granule friability

n	Training end point								
	0.15	0.14	0.13	0.12	0.11	0.10	0.09	0.08	0.07
10	18.43	19.06	16.00	17.16	16.57	16.78	17.86	18.56	18.39
15	14.69	17.14	14.63	16.40	16.78	17.28	18.20	19.03	18.39
20	16.16	16.48	15.66	16.19	16.68	17.04	17.91	18.88	18.58
25	16.64	15.66	14.88	16.68	16.76	17.38	18.13	18.49	18.56
Sum	65.92	68.34	61.17	66.43	66.79	68.48	72.10	74.96	73.92

n, number of neurons in the hidden layer.

This study showed that the average error percentage varied remarkably when comparing the response variables (Tables 3–5). The smallest error (2.80) was found with granule flow rate when the network was trained to the end point of 0.14. The estimations of granule size and friability were clearly more inaccurate. This might be due to original measurement accuracy of different granule properties. It was assumed that the best training end point corresponds to the minimum sum of average error percentage. For all responses the best training end point was between 0.12 and 0.14. However, when observing different responses, the same network loses the generalization ability at different training end points. In the case of mean granule size the minimum sum of errors (61.17) was achieved using the training end point 0.13 (Table 3). On the other hand, granule friability had the minimum sum of errors (49.31) (Table 4) at the training end point 0.12 and the granule flow rate at the training end point 0.14 (13.84) (Table 5).

3.2. *Comparison of four different neural network topologies to describe the experimental data*

The number of hidden layers is difficult to decide, but typically no more than one hidden layer is used in a network (Hush and Horne, 1993). Therefore, this study started with one hidden layer. More hidden layers were not used, since all networks converged well. The number of connections in the network is directly dependent on the number of neurons in the hidden layer. In

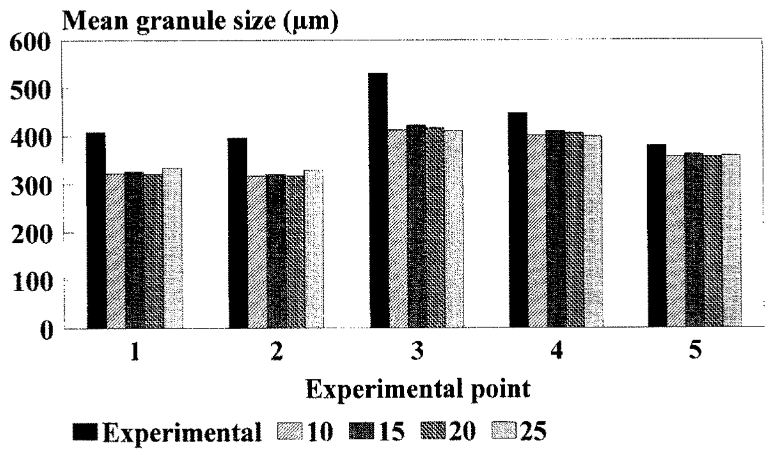

Fig. 1. Mean granule size predictions of four different networks containing 10, 15, 20 and 25 hidden layer neurons and the corresponding experimental value in five experimental point.

the training phase the information of the training data is transformed to weight values of the connections. Therefore, the number of connections might have a significant effect on the performance of the network. Because there are no theoretical principles for choosing the proper network topology several structures were tested in the study.

Four neural network topologies with different numbers of hidden layer neurons (10, 15, 20 and 25) were tested in order to predict granule properties. The training end points used were 0.13 (mean granule size), 0.12 (granule friability) and 0.14 (granule flow rate) as presented in the previous section. Neural model predictions and experimental values are compared in Fig. 1–3. The experimental values are average values of the replicate experiments. According to this study, all four neural networks tested underestimate the values of different responses to some extent in most cases when compared to the experimental data. Only granule friability at experimental point 3 was clearly overestimated (Fig. 2). The underestimations were significant, especially with regard

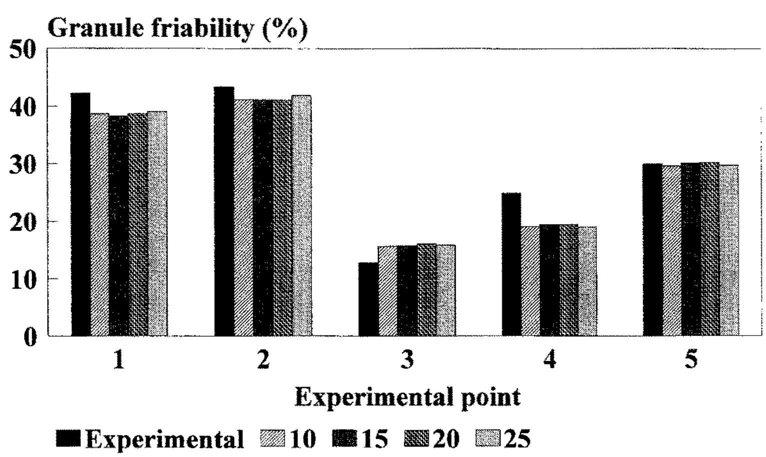

Fig. 2. Granule friability predictions of four different networks containing 10, 15, 20 and 25 hidden layer neurons and the corresponding experimental values in five experimental point.

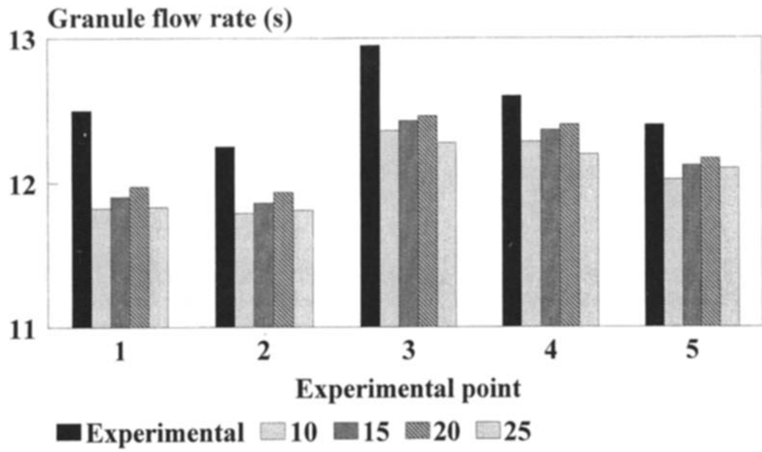

Fig. 3. Granule flow rate predictions of four different networks containing 10, 15, 20 and 25 hidden layer neurons and the corresponding experimental value in five experimental point.

to the mean granule size and flow rate of granules (Fig. 1 and 3). This effect might be attributable to some unknown and varying process parameter, since the original experiment series and that for collecting test data were made at different times.

Fig. 1 and 2 show clearly that the number of hidden layer neurons only slightly affected the estimated values. The predicted values given by all four network are quite similar at different experimental points. In the case of granule flow rate the predictions varied more depending on the topology of the network, and the best predictions were obtained when the number of hidden layer neurons was 20 (Fig. 3). However, this effect was not significant. The networks tended to become easily overtrained, expecially in the case of granule flow rate (Table 5). The average error percentages start to increase significantly when the network is trained above the training end point 0.14. It is obvious that the number of neurons in all tested networks was large enough for

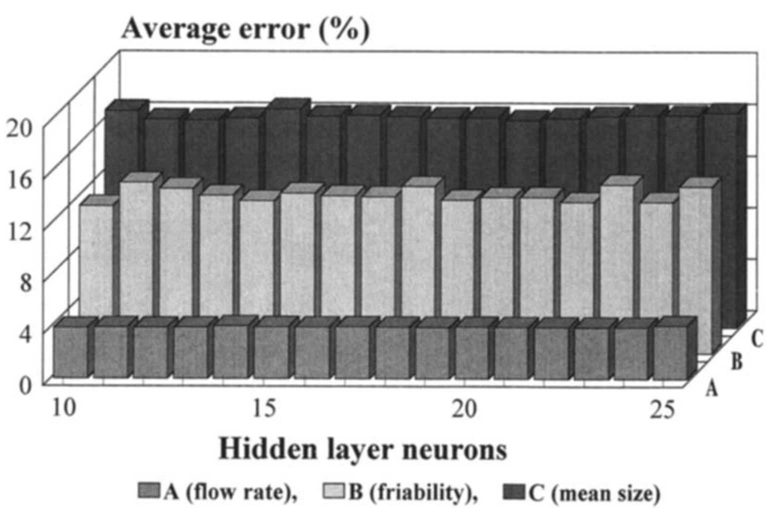

Fig. 4. Effect of number of hidden layer neurons on the average error percentage of granule flow rate (A), granule friability (B) and mean granule size (C).

modelling the particular process, but the number of cases was too limited for more precise summations of the effect of the number of hidden layer neurons.

3.3. Effect of number of neurons in hidden layer on the average error percentage

In the second test series the training end point was fixed to be 0.12 in each case. The number of neurons in the hidden layer varied from 10 to 25. The generalization ability was studied by calculating the average error percentages (Murtoniemi et al., 1994). Fig. 4 shows clearly that the effect of the number of hidden layer neurons on the average error is minimal. The accuracy of the predictions varied depending on the response examined. The accuracy of the predictions was the best in the case of granule flow rate (Fig. 4) with all tested networks, whereas the differences between the networks were less than 1%. It seems impossible to find any logical relationship between the number of neurons in the hidden layer and the average error percentage. The effects of hidden layer size on network training and performance have been studied recently by Lodewyck and Deng (1993). They found that the hidden layer size did not affect the performance of networks used for the information system planning.

3.4. Comparison of the best network and the regression model

Regression models (Merkku and Yliruusi, 1993; Merkku et al., 1993, 1994) were used to predict granule properties (mean granule size, granule friability and granule flow rate) from test input data. The average error percentage was 19.0 for granule mean size, 13.7 for friability and 5.2 for flow rate. The corresponding percentages using the neural model were 14.6, 12.1 and 2.8. These results show that it is possible to predict certain granule properties more accurately using neural models than with regression models. The neural network analysis is very flexible as regards the amount of experimental data (training data) used to generate a model, which is not always the case in regression analysis. The accuracy of neural network models is suggested to improve on increasing the amount of the training data and by adding additional process input variables to the training data.

It can be concluded that the training end point is a very critical factor in training a neural network, and in future studies it must be selected carefully. On the other hand, the number of hidden layer neurons does not affect significantly the generalization ability and obviously it is not possible to improve the performance of the ANN model merely by increasing the size of hidden layer. The fundamental reason for the discovered underestimation of the predictions could not be explained sufficiently. There are several different training algorithms available, and their suitability should be studied in the future. Network topologies other than the feedforward networks used in this study might be suitable for process modelling.

Acknowledgements

This study was supported by the Technology Development Centre in Finland (TEKES) and the Finnish pharmaceutical industry (Leiras Oy and Orion Corporation, Orion-Farmos Pharmaceuticals). The authors wish to thank Jaana-Liisa Aro, B.Sc. (Pharm.), for technical help and assistance.

References

Davalo, E. and Naïm, P., *Neural Networks*, Macmillan, London, 1991, pp. 19–24.

Dayhoff, J.E., *Neural Network Architectures: An Introduction*, Van Nostrand Reinhold, New York, 1990, pp. 63–66.

Hush, D. and Horne, B.G., Progress in supervised neural networks. *IEEE Signal Processing Mag.*, 10 (1993) 8–39.

Knight, K., Connectionist ideas and algorithms, *Communications ACM*, 33 (1990) 59–74.

Lisbon, B.G.J., *Neural Networks: Current Applications*, Chapman & Hall, London, 1992, pp. 9–34.

Lodewyck, R.W. and Deng, P.-S., Experimentation with a back-propagation neural network. *Information Management*, 24 (1993) 1–8.

Merkku, P. and Yliruusi, J., Use of 3^3 factorial design and multilinear stepwise regression analysis in studying the fluidized bed granulation process: I. *Eur. J. Pharm. Biopharm.*, 39 (1993) 75–81.

Merkku, P., Antikainen, O. and Yliruusi, J., Use of 3^3 factorial design and multilinear stepwise regression analysis in studying the fluidized bed granulation process: II. *Eur. J. Pharm. Biopharm.*, 39 (1993) 112–116.

Merkku, P., Lindqvist, A.-S., Leiviskä, K. and Yliruusi, J., Influence of granulation and compression process variables on flow rate of granules and on tablet properties, with special reference to weight variation. *Int. J. Pharm.*, 102 (1994) 117–125.

Murtoniemi, E., Kinnunen, P., Merkku, P. and Leiviskä, K., The advantages by the use of neural networks in modelling the fluidized bed granulation process. *Int. J. Pharm.*, 108 (1994) 155–163.