Use of Artificial Neural Networks to Predict Drug Dissolution Profiles and Evaluation of Network Performance Using Similarity Factor

Kok Khiang Peh,[1,3] Chee Peng Lim,[2] Siow San Quek,[2] and Kean Hock Khoh[2]

Received May 10, 2000; accepted July 21, 2000

Purpose. To use artificial neural networks for predicting dissolution profiles of matrix-controlled release theophylline pellet preparation, and to evaluate the network performance by comparing the predicted dissolution profiles with those obtained from physical experiments using similarity factor.

Methods. The Multi-Layered Perceptron (MLP) neural network was used to predict the dissolution profiles of theophylline pellets containing different ratios of microcrystalline cellulose (MCC) and glyceryl monostearate (GMS). The concepts of leave-one-out as well as a time-point by time-point estimation basis were used to predict the rate of drug release for each matrix ratio. All the data were used for training, except for one set which was selected to compare with the predicted output. The closeness between the predicted and the reference dissolution profiles was investigated using similarity factor (f_2).

Results. The f_2 values were all above 60, indicating that the predicted dissolution profiles were closely similar to the dissolution profiles obtained from physical experiments.

Conclusion. The MLP network could be used as a model for predicting the dissolution profiles of matrix-controlled release theophylline pellet preparation in product development.

KEY WORDS: multilayered perceptron; artificial neural networks; similarity factor; drug dissolution profiles.

INTRODUCTION

Advances in the area of soft computing have resulted in the development of a variety of intelligent systems. One of the more popular soft computing techniques, artificial neural networks (ANNs), has been developed and used as a problem-solving tool in various fields, among them pattern recognition and classification, signal and image processing, robot control, weather prediction, financial forecasting, and medical diagnosis (1). In general, ANNs are generalizations of mathematical models of biological nervous systems in our brain. One key benefit of ANNs is their ability to build a model of the problem using data from experimental measurements of the problem domain. Rather than being programmed by a user in a traditional sense, ANNs gather their knowledge by learning relationships of variables in data and building a model, implicitly, to relate the input and output variables of the problem.

Lately, ANNs have been applied to solve problems in the pharmaceutical fields, such as pharmaceutical process optimization, product development, prediction and estimation of pharmaceutical process coefficients, and pharmacokinetic parameters (2–6). Takahara *et al.* (2,3) demonstrated that the multi-objective simultaneous optimization technique incorporating ANNs was useful in optimizing formulae for pharmaceutical responses that are nonlinearly related to the process variables. Hussian *et al.* (4) showed that ANNs were able to predict the response variables that characterized the drug release profile of a hydrophilic matrix capsule system, more precisely than the response surface methodology. In another study (5), ANNs were used to predict pharmacokinetic parameters in human based on animal data. In addition, ANNs were reported to possess added advantages over some theoretical approaches in pharmaceutical data analysis (6).

For instance, the theoretical approach of nonlinear mixed effect modeling employed mathematical formulae to build a model that described the behavior of some pharmacokinetic processes (6). However, this approach may fail if the underlying principles governing the processes are not sufficiently understood, because the results are dependent on the accuracy of the model. On the other hand, ANNs are basically a data-based learning approach that does not require specific pharmacokinetic model in prediction.

Among the many possible ANN architectures, the multilayer perceptron (MLP) network (7) is one of the most widely used. This network has been proven to be a universal approximator (8). When given sufficient processing elements, it can approximate any nonlinear function with arbitrary accuracy. Since predicting drug release profiles can also be viewed as a function approximation problem, the MLP network has been selected in the present study.

The objective of the present study was to utilize the MLP network to predict the *in vitro* dissolution profiles of a matrix-controlled release theophylline pellet preparation. Empirically, mathematical models are used to represent behaviors and dynamics between various interacting components in many pharmaceutical processes. Hitherto, other researchers (2–6) used ANNs to predict the formulation and/or process parameters based on certain mathematical models that characterize the dissolution profiles. Instead of estimating parameters that fit models of the profiles as conducted by other researchers, we employed in the present study a different approach in prediction, by treating the entire dissolution profile as a time-series curve and estimating the whole profile based on a time-point by time-point estimation basis. Each time point was used as a dependent feature in which information contained in one time point affected further predictions subject to subsequent inputs. In addition, the network performance was evaluated by comparing the predicted dissolution profiles with those obtained from physical experiments using similarity factor (9–12).

MATERIALS AND METHODS

Development of Pellets and *In vitro* Drug Release Studies

The development and the *in vitro* dissolution studies of the matrix-controlled release theophylline pellet preparations

[1] School of Pharmaceutical Sciences, University of Science Malaysia, 11800 Penang, Malaysia.

[2] School of Industrial Technology, University of Science Malaysia, 11800 Penang, Malaysia.

[3] To whom correspondence should be addressed (e-mail: kkpeh@usm.my)

0724-8741/00/1100-1384$18.00/0

have been described elsewhere (13). A series of formulations containing a constant proportion of theophylline, but different proportions of microcrystalline cellulose (MCC) and glyceryl monostearate (GMS) at ratios of 10:10:0, 10:8:2, 10:7:3, 10:6:4, 10:5:5, and 10:4:6 were prepared. The GMS was first dispersed in a sufficient quantity of distilled water heated at about 80°C, followed by the addition of theophylline with constant stirring until a slurry was formed. The hot slurry was immediately mixed and blended with MCC in the Kenwood planetary mixer for 10 min. The wet powder mass was then extruded at a displacement rate of 30 cm/min using a Ram Extruder (SDX, Penang, Malaysia) fitted with a single-holed die of hole 1 mm in diameter and 4 mm in length. The extrudates were spheronized using a 22.5 cm Spheroniser (GB Caleva, Ascot, Berks, TJK) at 1000 rpm for 10 min. The spherical pellets obtained were dried in a fluidized bed drier at 40°C for 30 min.

The *in vitro* drug release was determined using the paddle method of the USP 23 dissolution test apparatus. The test was conducted in 900 ml of distilled water maintained at 37 ± 0.5°C with a paddle rotation speed of 100 rpm. Samples of the 3 ml volume were collected at predetermined time intervals of 0, 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0, and 10.0 hours. The drug concentrations were determined at 273 nm using a spectrophotometer (Hitachi, Tokyo, Japan) after appropriate dilution. Each test was run in sets of six and the average percentage of drug release versus time was calculated and plotted.

Profile Prediction Using Artificial Neural Networks

The MLP network generally can be exemplified as a feedforward type of ANN comprising a number of layers, namely, the input layer, one or more hidden layers, and the output layer. Each layer consists of a number of artificial neurons known as processing elements or nodes. An example of a three-layer MLP network is depicted in Fig. 1. The nodes in neighboring layers are fully connected with links that store the modifiable strength connections called weights. Nodes in the hidden layer receive input signals from the nodes in the input layer. These inputs are weighted by the connection strengths and then linearly summed to produce the net neuron internal activity level as follows:

$$c_j = \sum_{i=1}^{p} w_{ij} x_i$$

where w_{ij} is the weight from node i in the input layer to node j in the hidden layer, x_i is the i-th input element, and p is the number of nodes in the input layer. Generally, a sigmoid/logistic function is used to regulate the linearly combined output of a node, as follows:

$$o_j = \varphi(c_j) = \frac{1}{1 + \exp(-c_j)} \qquad -\infty < c_j < \infty$$

where o_j is the output of the j-th node in the hidden layer. Subsequently, output from the hidden layer is used as input to the output node. Finally, the overall response from the network is obtained via the output node in the output layer.

In the present study, the NEURAL program (14) was used to implement the MLP network. The conjugate gradient and simulated annealing algorithm were used to train the MLP network. The conjugate gradient is a useful method to minimize the mean-squared error function, and is generally faster and more robust compared with the traditional backpropagation algorithm. While the conjugate gradient algorithm is effective at finding the nearest minimum point from the starting weights, this might be a local minimum of the error function. It is therefore imperative for MLP to escape from local minima of the error function during training. To achieve this, simulated annealing (14) can be used to get away from the current point to a lower point, and eventually to arrive at the global minimum of the error function. Hence, in the training of MLP it is useful to combine the local power of the conjugate gradient with the global power of simulated annealing to attain a global minimum among local minima (14). In addition, genetic optimization is used for weight initialization in the MLP network. Like simulated annealing, the genetic algorithm is a stochastic search technique that can elude local minima. It works with a population of strings encoding parameters of the function to be optimized, and searches many solution points in parallel. Genetic operations such as reproduction, crossover, and mutation are used to create strong traits that would optimize the function, and this is suitable for finding starting weights.

The concept of leave-one-out was used to predict the rate

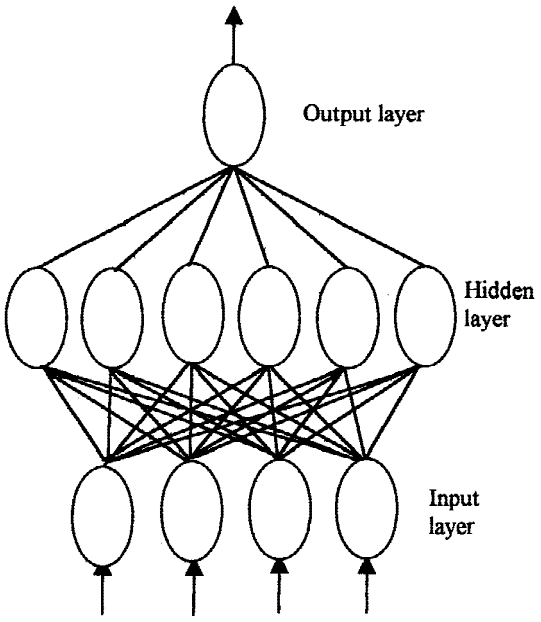

Fig. 1. A three-layer MLP network.

Table 1. The Mean f_2 Values Obtained from the MLP Network with Varying Numbers of Hidden Nodes

	Number of hidden nodes			
MCC:GMS	4	6	8	10
4:6	68.5	68.9	69.6	66.9
5:5	82.2	82.5	82.0	78.0
6:4	86.9	86.8	86.4	85.7
7:3	82.8	84.9	84.2	83.6
8:2	80.2	79.6	74.3	76.7
10:0	59.7	63.2	54.9	61.3

Table 2. The f_2 Values of Varying Matrix Ratios of MCC and GMS Obtained Using Six Hidden Nodes

MCC:GMS	P1	P2	P3	P4	P5	P6	Mean	SD[a]	CV%[b]
4:6	70.4	64.2	65.5	66.2	74.5	72.4	68.9	4.2	6.1
5:5	80.1	77.1	83.3	82.1	84.0	88.1	82.5	3.7	4.5
6:4	87.0	83.8	87.9	89.1	86.4	86.4	86.8	1.8	2.1
7:3	84.7	85.1	80.5	83.2	88.5	87.2	84.9	2.8	3.4
8:2	80.0	82.8	77.8	75.0	81.5	80.4	79.6	2.8	3.5
10:0	60.2	61.1	67.6	63.3	63.4	63.3	63.2	2.6	4.1

[a] Standard deviation.
[b] Coefficient of variation.

of drug release. In the experiment, all, the dissolution data were used for training except one set, which was used to compare with the predicted output. For example, data sets of MCC and GMS at ratios of 10:0, 8:2, 7:3, 6:4, and 5:5 were used to train the network and to predict the release profile of ratio 4:6. Since dissolution studies were run in sets of six for each matrix ratio, there were a total of 30 training samples. Each training sample comprised four inputs and one output. The first two inputs represented the matrix ratio, the third input represented the time point of the measurement of percent dissolved, and the fourth one encoded the difference between the release rate of the preceding two time points of the predicted profile. This fourth input was essentially a second derivative (rate of change of the slope) of the percent dissolved against time curve for the previous two time points of the network predicted profile. Prediction of the dissolution profile was conducted based on a time-point by time-point basis, with a total of eleven time points (0, 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0, and 10.0 hours) to obtain a complete dissolution profile. Since information on the rate of change of the slope was not available for the first and second time points, the fourth input was set to zero and to the predicted percent dissolved from the first time point respectively.

Six dissolution profiles were generated from the network for each matrix ratio and each profile was used to compare with one of the six dissolution profiles obtained from the physical experiment. This approach provides an alternative prediction of the dissolution profile as a whole time-series curve using ANNs, compared with predicting the parameters of certain models that fit the dissolution profile, a technique normally employed by other researchers. As a result, errors in parameter estimation as well as in mathematical modeling of the dissolution profile can be reduced.

Performance Measurement Using Similarity Factor

Dissolution profiles predicted from the MLP network was compared with those generated from physical experiment using similarity factor (f_2). The similarity factor is a function of the reciprocal of mean square-root transform of the sum of square distances at all points, and is a measure of the similarity in the percent rate of drug release between two dissolution profiles. The value of f_2 ranges between 0 to 100 with a higher f_2 value indicating more similarity between the two profiles. The equation of f_2 is expressed as follows:

$$\text{Similarity factor, } f_2 = 50 \cdot \log\{[1 + (1/n)\sum_{t=1}^{n} (\mu_{ti} - \mu_{ri})^2]^{-1/2} \cdot 100\}$$

where μ_{ri} and μ_{ti} represent the percentage of drug dissolved measured at the i-th time point of the experimental and predicted curves, and n is number of time points tested.

RESULTS AND DISCUSSION

In ANN applications, a problem that often encountered is the determination of the "optimal" number of hidden nodes. Normally, one has to resort to empirical methods to obtain a good network structure that can produce satisfactory performance. Table 1 summarizes the mean f_2 values from experiments with different matrix ratios of MCC and GMS, using four, six, eight, and ten hidden nodes. We can see that six hidden nodes yielded, on average, the best performance. As such, the MLP network with six hidden nodes was selected throughout all the experiments.

Table 2 shows the complete results comprising f_2 values from individual predictions as well as the mean f_2 values from six trials using different matrix ratios of MCC and GMS. The performance of the MLP network appeared to be satisfactory as the f_2 values for all the matrix ratios were above 60. The small coefficient of variation values of less than 6.5% suggested that the dissolution profiles generated from the net-

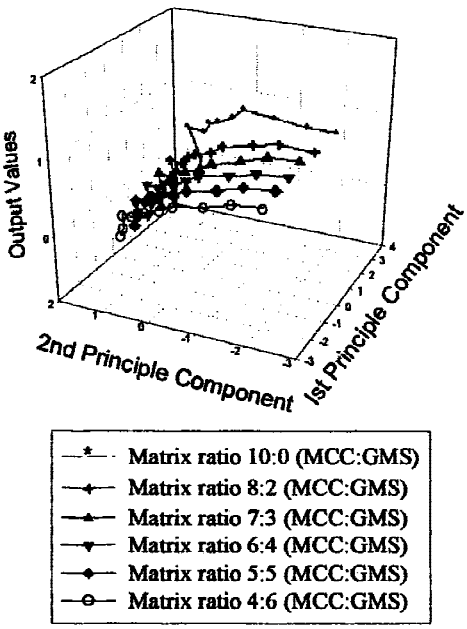

Fig. 2. The percentage of drug dissolved (output) against the two principal components of the input features extracted using principal component analysis.

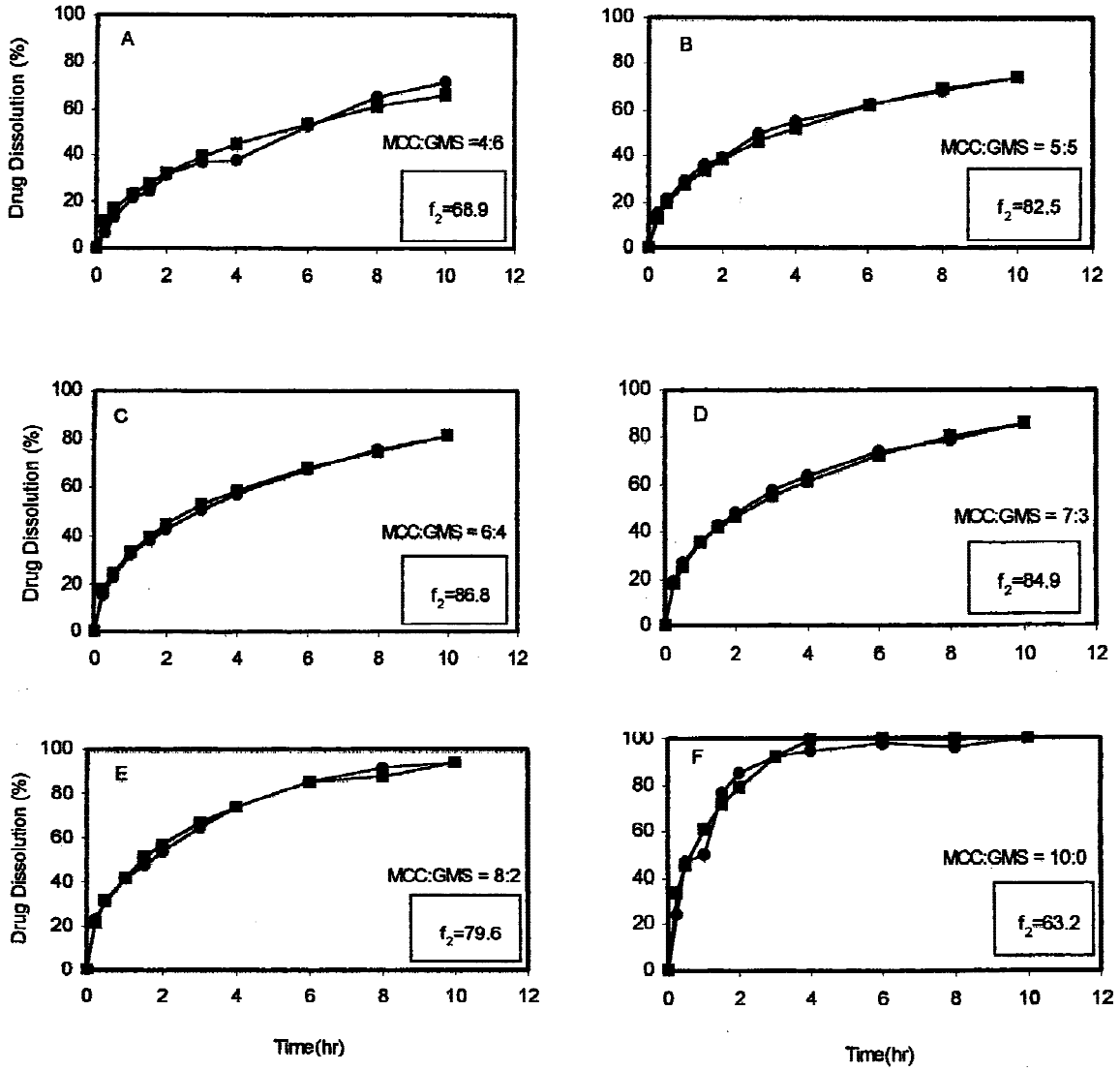

Fig. 3. Mean drug dissolution profiles obtained from network (●) and physical experiment (■) for theophylline pellets consisting of various compositions of microcrystalline cellulose (MCC) and glyceryl monostearate (GMS). (A) 4:6, (B) 5:5, (C) 6:4, (D) 7:3, (E) 8:2, and (F) 10:0.

work had little variation and the performance of network was stable.

We noticed that the results of f_2 with matrix ratios of 4:6 and 10:0 were inferior compared with those from other matrix ratios. This might be due to the phenomenon of interpolation and extrapolation in network training. As we know, efficacy of an ANN, as well as other statistical approaches such as logistic and multiple regression methods, is very much affected by the training data. Typically, the network would perform better in interpolation compared with extrapolation of the training data. Here, we use a graphical approach to visualize the distribution of the training data in a three-dimensional space. To accomplish this, the principal component analysis (15)—which was an effective procedure frequently used to reduce the dimension of the input features in ANN applications—was used. It was found that the original four-dimensional training data could be reduced to two-dimensional data that were a linear combination of the original data. From the analysis, the first and second principal components represented more than 75% of the original training data. Hence, to give a visualization of the relationship between the training data and outputs, the first two principal components against the percentage of drug dissolved was plotted as shown in Figure 2. It can be seen that the data samples from MCC and GMS at ratios of 4:6 and 10:0 were scattered at the outer regions of other (5:5, 6:4, 7:3, 8:2) data samples. Therefore, the profile prediction of matrix ratios of 4:6 and 10:0 could be considered as extrapolating the underlying function because these two data samples were distributed outside the training data range. This observation might explain inferiority of the predicted results for matrix ratios of 4:6 and 10:0 compared with that of other ratios.

Figure 3 shows the mean dissolution profiles obtained from network and physical experiment for various matrix ratios. It can be seen that an increase in f_2 value was associated with more similarity between the predicted and the experimental dissolution profiles. In view of the f_2 values, the mean dissolution profile generated from the network could be con-

sidered similar to that of the physical experiment for all the matrix ratios evaluated.

From findings in the present study, the benefit of using ANNs in product development, especially in the prediction of dissolution profiles, is evident. Instead of determining the dissolution profile of each and every matrix ratio by conducting actual physical experiments, a suitable ANN system can be used to predict the trend of the drug dissolution profile, which is associated with the composition of the matrix materials. By using this approach, a lengthy and time-consuming experimentation to determine the appropriate matrix ratio for the preparation can be shortened. When a satisfactory release profile is obtained, a confirmation test can then be carried out experimentally to verify the predicted profiles.

CONCLUSIONS

In summary, ANNs—specifically the MLP network—could be used as a model for the prediction of *in vitro* dissolution profiles of matrix-controlled release theophylline pellet preparation during product development. The f_2 results indicated that the predicted dissolution profiles were closely similar to those obtained from the physical experiments for different matrix ratios. The present study has demonstrated the potential of ANNs as a useful tool in predicting drug dissolution profile as a time-series curve. Nevertheless, it would be beneficial to conduct further investigations into the stability and reliability aspects of using ANNs in pharmaceutical product development.

ACKNOWLEDGMENTS

The authors would like to thank University of Science Malaysia as well as the Ministry of Science, Technology, and Environment Malaysia (Project no.: 06-02-05-8002) for providing grants that have contributed to this article.

REFERENCES

1. J. G. Taylor. *The Promise of Neural Networks,* Springer-Verlag, London (1993).
2. J. Takahara, K. Takayama, and T. Nagai. Multi-objective simultaneous optimization based on artificial neural network in sustained released formulations. *J. Contr. Rel.* **49**:11–20 (1997).
3. J. Takahara, K. Takayama, K. Isowa, and T. Nagai. Multi-objective simultaneous optimization based on artificial neural network in a ketoprofen hydrogel formula containing *O*-ethylmenthol as a percutaneous absorption enhancer. *Int. J. Pharm.* **15**:203–210 (1997).
4. A. S. Hussian, X. Yu, and R. D. Johnson. Application of neural computing in pharmaceutical product formulation. *Pharm. Res.* **8**:1248–1252 (1991).
5. A. S. Hussian, R. D. Johnson, N. Vachhrajani, and W. A. Ritshel. Feasibility of developing a neural network for prediction of human pharmacokinetic parameters from animal data. *Pharm. Res.* **10**:466–469 (1993).
6. E. Brier, J. M. Zurada, and G. R. Aronoff. Neural network predicted peak and trough of gentamicin concentrations. *Pharm. Res.* **12**:406–412 (1995).
7. D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representation by error propagation. In D. E. Rumelhart and J. L. McLelland (eds.), *Parallel Distributed Processing,* I. MIT Press, Cambridge, MA, 1986. pp. 318–362.
8. G. Cybenko. Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals and Systems.* **2**:303–314 (1989).
9. FDA Guidance for Industry: Immediate release solid dosage forms: Scale-up and Post Approval Changes (SUPAC-IR): Chemistry, Manufacturing and Controls. In vitro dissolution testing and in vivo bioequivalence documentation, November, 1995.
10. FDA Guidance for Industry: Dissolution testing of immediate release solid oral dosage forms, August, 1997.
11. FDA Guidance for Industry: Extended release solid oral dosage forms: Development, evaluation and application of in vitro/in vitro correlations, September, 1997.
12. FDA Guidance for Industry: Modified release solid dosage forms: Scale-up and Post Approval Changes (SUPAC-MR): Chemistry, Manufacturing and Controls. In vitro dissolution testing and in vivo bioequivalence documentation, September, 1997.
13. K. K. Peh and K. H. Yuen. Development and in vitro evaluation of a novel multi-particulate matrix controlled release formulation of theophylline. *Drug Dev. Ind. Pharmacy* **21**:1545–1555 (1995).
14. T. Masters. *Practical Neural Network Recipes in C++*: Academic Press, San Diego, CA, 1993.
15. B. F. J. Manly. *Multivariate statistical methods: A primer.* Chapman & Hall, New York, 1986.