ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
19 (1999) 443–452

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Use of artificial neural networks to predict quaternery phase systems from limited experimental data

R.G. Alany, S. Agatonovic-Kustrin, T. Rades, I.G. Tucker *

School of Pharmacy, University of Otago, PO Box 913, Dunedin, New Zealand

Received 3 April 1998; accepted 9 August 1998

Abstract

The aim of the present work was to develop a method for predicting the phase behaviour of four component systems consisting of oil, water and two surfactants from a limited number of screening experiments. Investigations were conducted to asses the potential of artificial neural networks (ANNs) with back-propagation training algorithm to predict the phase behaviour of four component systems. Three inputs only (percentages of oil and water and HLB of the surfactant blend) and four outputs (oil in water emulsion, water in oil emulsion, microemulsion, and liquid crystals containing regions) were used. Samples used for training represented about 15% of the sampling space within the tetrahedron body. The network was trained by performing optimization of the number and size of the weights for neuron interconnections. The lowest error was obtained with 15 hidden neurons and after 4500 training cycles. The trained ANN was tested on validation data and had an accuracy of 85.2–92.9%. In most cases the errors in the prediction were confined to points lying along the boundaries of regions and for the extrapolated predictions outside the ANNs 'experience'. This approach is shown to be highly successful and the ANNs have proven to be a useful tool for the prediction of the phase behaviour of quaternary systems with less experimental effort.

Keywords: Quaternary systems; Surfactants; Artificial neural network

1. Introduction

Systems containing oil, water and surfactant are of interest as drug delivery systems, or as bases for cosmetics. A diverse range of colloidal systems and coarse dispersions can be obtained depending on ratios of the components. Emulsions, microemulsions, creams and lyotropic liquid crystals are some examples. One of the most convenient methods to study the phase behaviour of such systems is by constructing a phase diagram using a Gibbs triangle [1]. However as the formulation may contain more than three components the complete phase behaviour cannot be represented using a triangular diagram. The phase behaviour of a four component mixture at fixed pressure and temperature can be represented using a tetrahedron. Full characterization of such systems is a tedious task requiring a large number of experiments [2].

* Corresponding author. Tel.: +64-3479-7275; fax: +64-3479-7034.

0731-7085/99/$ - see front matter
PII: S0731-7085(98)00232-5

Artificial neural networks (ANNs) are finding increasing applications in areas of prediction [3–5]. Indeed Richardson et al [6] reported on the use of ANNs to predict the pseudo-ternary phase diagrams for four component microemulsion systems using only a range of four computed physicochemical properties for the cosurfactants employed. The use of ANNs seems to be most suitable for dealing with complex multivariate nonlinear relationships. Once trained ANNs can be used to predict outputs from new sets of input data. These features make ANNs suitable for solving problems in the area of optimization of formulations in pharmaceutical product development.

The paper reports on the use of ANNs to characterize four component quaternary systems consisting of oil, water and two surfactants to minimize experimental effort.

1.1. Neural networks

ANNs are mathematical systems that mimic the way in which the human brain works [7]. They consist of fully interconnected processing units called neurons organized in layers. There is always one input and one output layer and there should be at least one hidden layer [8]. The use of a hidden layer enables ANNs to describe nonlinear systems. A problem in constructing an ANNs is to find the optimal number of hidden nodes. We have used a supervised network with back-propagation training algorithm [9] and its mode of operation is presented schematically in Fig. 1. The information of the model that is passed from one processing element to another is contained in the connections between neurons in successive layers as weights and in the bias values of a neuron (neuron activation threshold). The network is trained by optimizing of the weight for each node interconnection and bias term, until the output values at the output layer neurons are close to the actual outputs. The output of the neuron is related to the summed inputs by a nonlinear transfer function. The most commonly used function in back-propagation neural networks is a sigmoidal function. The optimization is therefore non-linear and consists of iteratively varying the weights until the output values for each sample are close to the target values. The error in prediction is then propagated through the system and the interunit connections are changed to minimize the error in the prediction. This process is continued with multiple training sets until the error is minimized across the sets. The mean squared error of the network (MSE) is defined as the squared difference between the target values t and the output y of the output neurons:

$$\mathrm{MSE} = \frac{1}{p \cdot m} \sum_{k=1}^{p} \sum_{l=1}^{m} (y_{kl} - t_{kl})^2$$

where p is the number of training sets, and m is the number of output neurons in the neural network. During training, neural techniques need to have some way of evaluating their own performance. Since they are learning to associate inputs with outputs, evaluating the performance of the network on the training data may not produce the best results. If a network is trained for too long, it will overtrain and will lose the ability to generalize. Thus, three types of data sets are used: training data (to train the network), test data (to monitor the performance of the neural network during training), and validation data (to measure the performance of a trained application).

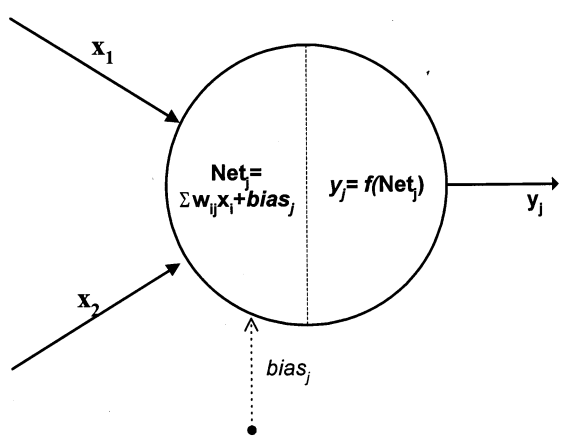

Fig. 1. General scheme of a processing unit or neuron. Y_j; w_{ij} is the weight from neuron I to neuron j; $\mathrm{bias_j}$ is the bias of neuron j; Net_j is the sum of weighted inputs of neurons j; $f(\mathrm{Net}_j)$ is the transfer function of neuron j.

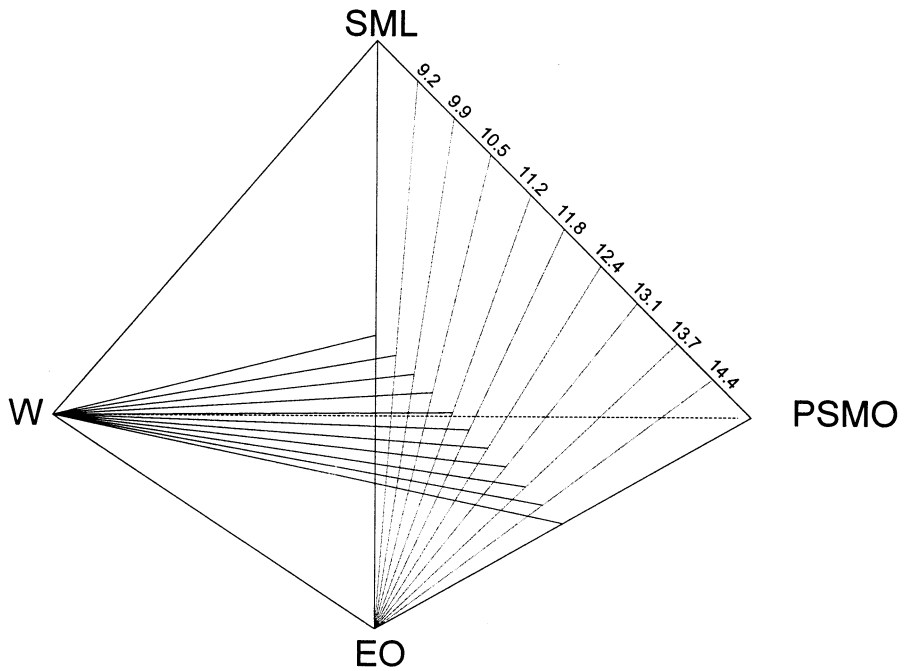

Fig. 2. Four component tetrahedron illustrating the position of the samples formulated at different HLB values. HLBs are shown on the SML-PMSO axis. W: water; EO: ethyl oleate; SML: sorbitan mono laurate; PMSO: polyoxyethylene 20 sorbitan monooleate.

2. Experimental

2.1. Materials and methods

2.1.1. Materials

Ethyl oleate (EO) was used as the oil component (Crodamol EO). Sorbitan mono laurate (SML) as surfactant component 1 (Crill 1, HLB = 8.6) and polyoxyethylene 20 sorbitan monooleate (PSMO) as surfactant component 2 (Crillet 4 super, HLB = 15) were mixed to give a range of HLB values. EO, SML, and PSMO was obtained form Croda Surfactants New Zealand as free samples. Sudan III (BDH Chemicals Poole, UK), and methylene blue (Koch-Light Laboratories, Colnbrook Bucks, England) were used as lipophilic and hydrophilic dyes respectively. Deionised water was used.

2.1.2. Methods

2.1.2.1. Apparatus. Phase contrast and polarized light microscopy were performed using a Nikon Optiphot microscope (Nikon, Tokyo, Japan). Electrical conductivity measurements were carried out using a Riac CM/100 conductivity meter, with a YSI 3418 electrode (YSI, Yellow Springs, USA).

2.1.2.2. Artificial neural networks simulator software. MS-Windows based ANNs simulator software, NNMODEL Version 1.404 (Neural Fusion) was used. Calculations were performed on a 586 personal computer.

2.1.2.3. Sample preparation. Fig. 2 shows a first set of samples, representing the cuts made through the phase tetrahedron. The two surfactants were mixed to give the following SML: PSMO mass ratios: 1.0:0.0, 0.9:0.1, 0.8:0.2, 0.7:0.3, 0.6:0.4, 0.5:0.5, 0.4: 0.6, 0.3:0.7, 0.2:0.8, 0.1:0.9, and 0.0:1.0 with corresponding HLB values of 8.6, 9.2, 9.9, 10.5, 11.2, 11.8, 12.4, 13.1, 13.7, 14.4, and 15, respectively. The various SML: PSMO blends were heated to 50°C and stirred at high speed for 5 min on a hot plate magnetic stirrer. After cooling to room temperature,

aliquots of each blend were mixed with EO in 1:1 mass ratio. The mixtures were stirred vigorously using a magnetic stirrer for 5 min. To each system water was added to give water concentrations in the range of 0–95% (w/w) at 1–5% increments. Samples were vortexed and left overnight to equilibrate, then were characterized by visual appearance, microscopical examination, conductivity testing, staining, and dilution. Clear, isotropic, one-phase systems were characterized as microemulsion (ME); systems showing birefringence with a typical oily streak, maltese cross or fan shaped texture (Fig. 3), were characterized as systems containing a lamellar mesophase (LC). The LC systems were either one-phase systems (pure lamellar mesophase) or two-phase systems with lamellar mesophase and either ME or an aqueous continuous phase. Water continuous coarse dispersions (Fig. 4) were classified as o/w emulsions (o/w EM) and oil continuous coarse dispersions were categorized as w/o emulsions (w/o EM). The w/o EM were found to be very unstable, and cracking occurred usually within 2–5 min. The ME were stable for at least six months.

To construct a pseudoternary phase diagram, a second set of samples was prepared as above at a fixed mass ratio of the two surfactants of 4:6 (HLB = 12.4) (Fig. 5a). The results were plotted using a Gibbs triangle (Fig. 5b) and the percentage occupied by each region was determined by a cut and weigh method [10].

3. Results and discussion

3.1. Artificial neural networks structure and training

The composition of a network is related to both the number and the size of the weights. Model

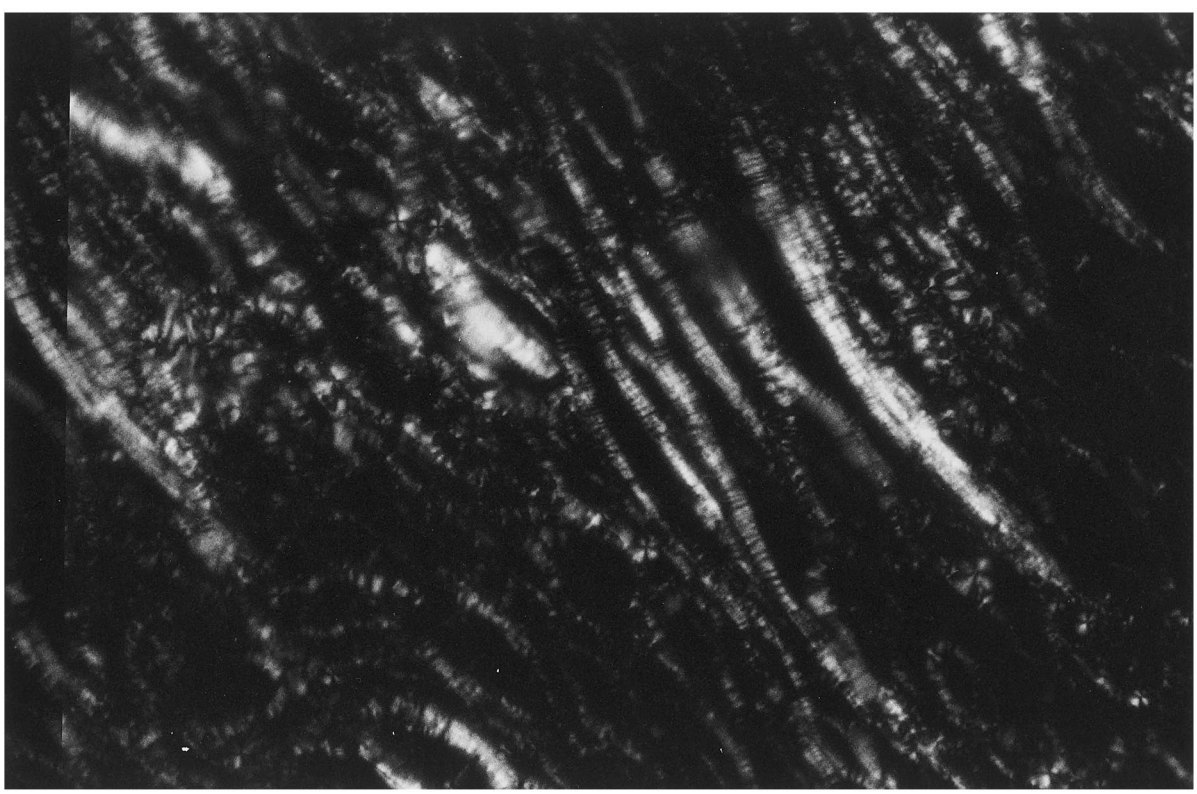

Fig. 3. Polarized light micrograph showing the lamellar mesophase with the characteristic oily streaks and maltese cross texture.

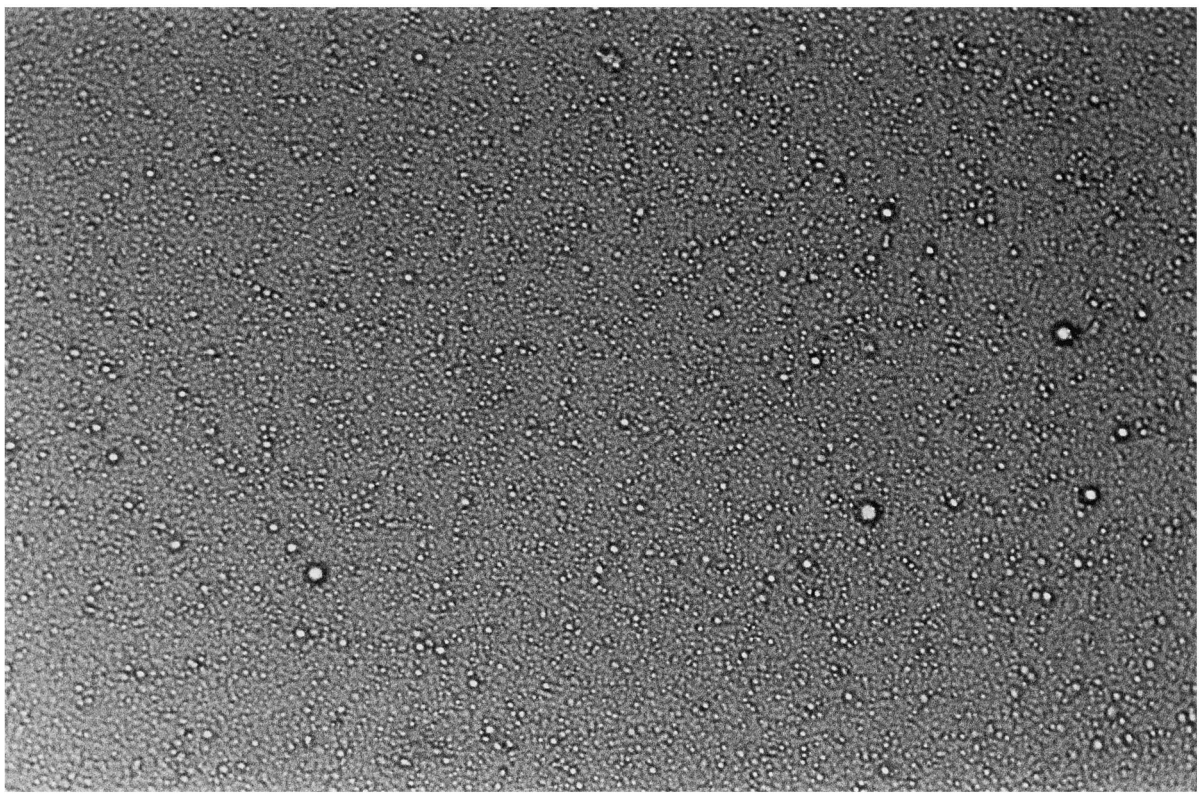

Fig. 4. Phase contrast micrograph showing an emulsion (o/w EM).

selection is concerned with determining the optimal number of hidden units and the weights for these neurons. The number of connections in the network depends upon the number of neurons in the hidden layer. In 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 network performance. Since there are no theoretical principles for choosing the proper network topology different structures were tested. The ANN was trained using a different number of hidden neurons (5–25) and training cycles (0–6000).

At the start of each training run, both weights and biases were initialized with random values. During training, modifications of the network weights and biases were made by back-propagation of error. The magnitudes of the changes for the weights and biases in the hidden and output layer were controlled by the error signal in each layer.

To compare the predictive power of the different ANN models, mean squared errors were calculated. Since the test set error is usually a better measure of performance than the training error, while the network was being optimized, $\approx 10\%$ of the experimental data, called a test set, was fed forward through the network to evaluate the trained network. The performance of the network on this test set gives a reasonable estimate of the ANNs prediction ability [7]. The lowest testing error was obtained with 15 hidden neurons and after 4500 training cycles (Table 1).

In order to examine the generalization ability of the best ANNs model, 180 additional experiments were performed to interrogate the model. Forty five sets of data were systematically selected from

each of the four pseudoternary phase diagrams at HLB values 9, 11.5, 13, and 14.7.

3.2. Artificial neural networks training and testing data

Samples representing the phase tetrahedron cuts (8 samples at each cut) at fixed 1:1 oil surfactant blend mass ratio (Fig. 2), as well as of the phase diagram (40 samples) at HLB 12.4 (Fig. 5), were used as training data providing 128 input-output pairs. A further 15 samples were randomly selected from these cuts as testing data. Sampling of the tetrahedron for full characterization would require about 1000 samples, thus about 15% of the space was sampled. The inputs were percentage of oil, percentage of water, and HLB of the surfactant blend. The outputs were the four different outcomes: ME, LC, w/o EM, and o/w EM. Samples were coded as $+1$ to signify the presence of a particular system, and -1 to indicate its absence. Thus, a region consisting of ME would have values of $+1, -1, -1, -1$; pure LC was $-1, +1, -1, -1$; w/o EM was $-1, -1, +1, -1$; and o/w EM was categorized as $-1, -1, -1, +1$. ME and LC would have values of $+1, +1, -1, -1$, whereas LC and o/w EM were categorized as $-1, +1, -1, +1$. The coexistence of w/o EM and o/w EM were classified as $-1, -1, +1, +1$. A perfectly trained network should recover such values for perfect phase classification. Any deviation from $+1$ and -1 would reflect error in the classification process.

3.3. Prediction of the phase nature based on overall trained network

Fig. 6 shows the predicted pseudoternary phase diagrams at HLB 9, 11.5, 13, and 14.7. The ANN was able to classify regions containing multiple phases. For example the LC region shown in Fig. 6 consist of monophasic LC region $(-1, +1, -1, -1)$, an aqueous LC dispersion $(-1, +1, -1, +1)$ and LC dispersion in ME $(+1, +1, -1, -1)$. At HLB 9, there were no LC or ME regions. Rather the whole triangular area was occupied by coarse dispersions of either o/w EM (36% of the total area of the phase triangle), or w/o EM (64%). At HLB 11.5, the predicted pseudoternary phase diagram revealed the presence of an appreciable ME region covering about 19% of the total triangle area. The maximum water solu-

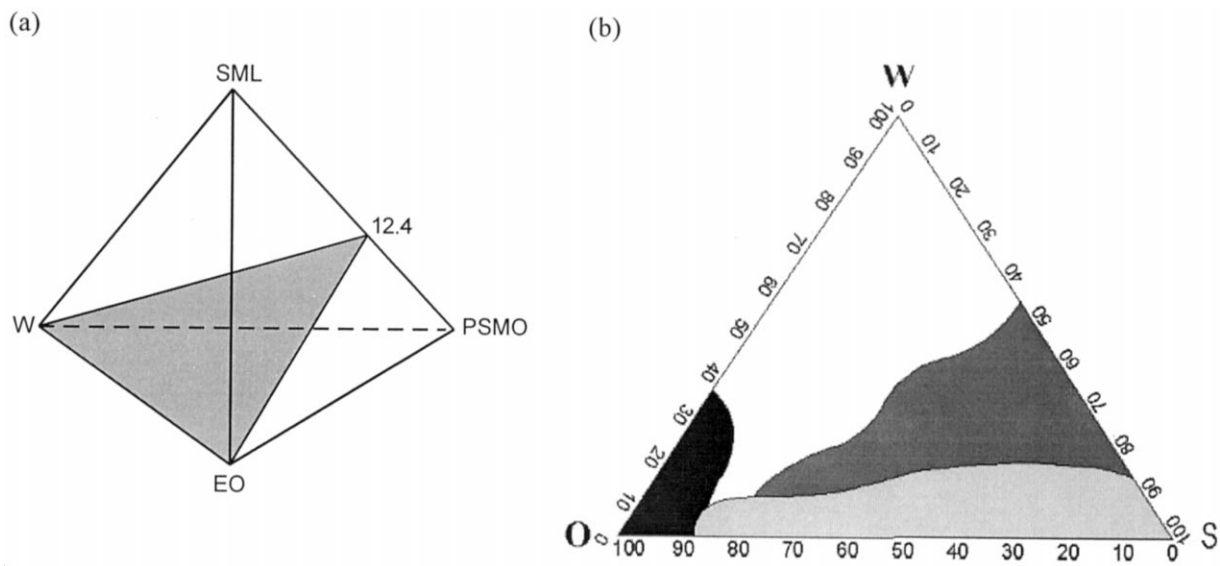

Fig. 5. (a) Constructed pseudoternary phase diagram locus within the tetrahedron body; (b) the pseudoternary phase diagram at HLB 12.4 W = 100% water; O = 100% EO; S = 100% SML: PSMO.

Table 1
Testing MSE using different numbers of hidden neurons and training cycles

Number of hidden neurons	Number of training cycles					
	1000	2000	3000	4000	5000	6000
5	0.162	0.162	0.173	0.190	0.202	0.206
10	0.128	0.130	0.135	0.146	0.152	0.153
15	0.114	0.085	0.080	0.078	0.078	0.079
20	0.142	0.150	0.149	0.153	0.162	0.162
25	0.137	0.150	0.151	0.148	0.139	0.127

bilized by ME was predicted to be $\approx 18\%$. An LC region (28%) was detectable at this HLB. Regarding the w/o EM region (10%), theory would predict that its area at HLB 11.5 would be less than that at HLB 9, which is in agreement with the ANN prediction. At HLB 13, the percentages occupied by the four different regions were ME 17.5, LC 45.5, o/w EM 32, and w/o EM 5%. There was a pronounced increase in the LC area compared to HLB 11.5. This can be explained by the increased mass fraction of PSMO in the surfactant blend. PSMO is a polyethyleneoxide containing surfactant, and it is well established that these surfactants have a tendency to form mesophases [11]. Predictions at HLB 14.7 yielded the following ratios: ME 0.7, LC 88.3, o/w EM 9, and w/o EM 2%. The pure ME region was predicted to be very small, the dominant region being LC. This predicted behaviour might be explained with the fact that the surfactant blend at this HLB mainly consists of PSMO.

Knowledge of the physicochemical properties of the four components comprising these systems would allow only qualitative guessing in terms of the possible phase behaviour. On the contrary ANNs utilisation enables quantitative predictions.

For the systems included in the training set data, the different regions were reproduced with a coefficient of correlation ranging from 0.886 for pure LC to 0.978 for o/w EM with a mean value of 0.9453. The monophasic LC region was very limited and this is probably the reason for the relatively low coefficient of correlation of 0.886.

For the validation data set, various critical values (0, ± 0.25, ± 0.50, and ± 0.75) were used to classify these data. That is, if an output value was >0 the corresponding phase was assumed to be present and if <0 then absent. In the most stringent case, a phase was assumed present if the output was >0.75 and absent if <-0.75. Values in the range between -0.75 and 0.75 were considered unclassifiable. In this way it was hoped to modify the reliability of the predictions. ANNs generalized the phase behavior with accuracies of 85.2–92.9%, depending on the output critical values used for the classification (Table 2). Predictions were only counted correct when all phases present in a region were correctly predicted. If a critical value of zero was used, 86.9–95.6% of data were classified correctly. For critical values ± 0.25, 86.7–95.6% of data were classified correctly and 0–2.2% of data were unclassified. With a more rigorous critical values of ± 0.5, 82.3–98.0% of data were classified correctly and 2.2–6.6% data were unclassified. Increasing the critical values to ± 0.75 increased the unclassified percentage, as expected to 8.1 overall and to 24.0% at HLB 9.0, but this increase was mainly at the expense of correct predictions, not wrong predictions (Table 2). Consequently, increasing the critical values does not improve the reliability of the predictions. Reasons for this are unclear and require further study.

There are two different types of generalization: interpolation and extrapolation. Interpolation applies to cases that are more or less surrounded by nearby training data sets; everything else is extrapolation. Validation samples that were close to those in the training set were mainly correctly classified. At HLB 11.5, 13.0, and 14.7, almost all the mistakes were made in the small LC region close to the ME region. The trained ANN pre-

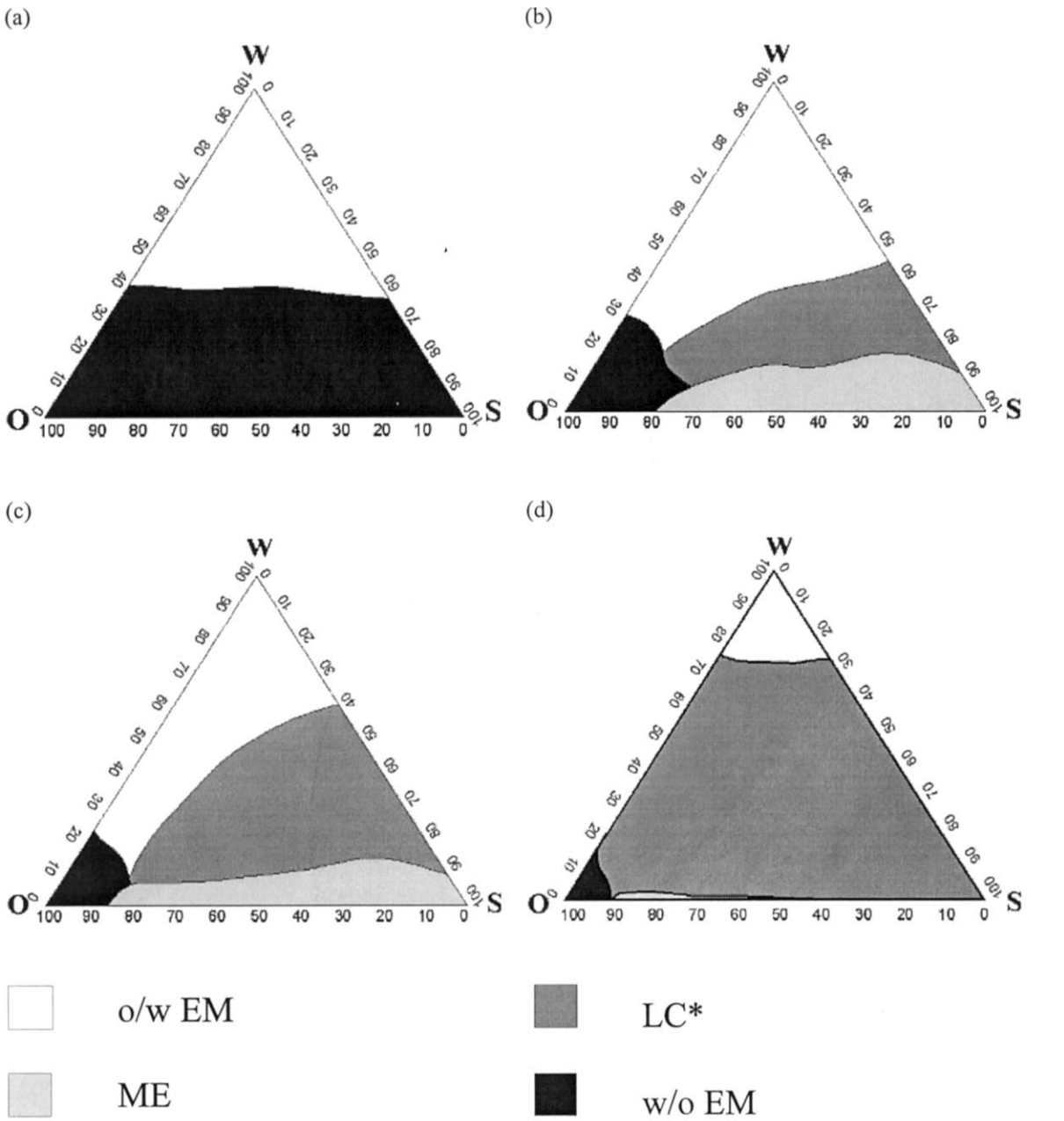

Fig. 6. Predicted phase triangles at HLB (a) 9.0; (b) 11.5; (c) 13.0; (d) 14.7.

Table 2
Accuracy in ANNs prediction[a]

HLB	Predictions[b]			Predictions[c]			Predictions[d]			Predictions[e]		
	W (%)	U (%)	C (%)	W (%)	U (%)	C (%)	W (%)	U (%)	C (%)	W (%)	U (%)	C (%)
9.0	11.1	2.0	86.9	11.1	2.2	86.7	11.1	6.6	82.3	8.8	24.0	67.2
11.5	4.4	0.0	95.6	4.4	0.0	95.6	0.0	2.0	98.0	0.0	2.0	98.0
13.0	6.6	0.0	93.4	2.2	2.2	95.6	4.4	2.2	93.4	6.6	2.2	91.2
14.7	4.4	0.0	95.6	4.4	2.2	93.4	6.6	2.2	91.2	11.1	4.4	84.5
	6.6	0.5	92.9	5.5	1.6	92.8	5.5	3.3	91.2	6.6	8.1	85.2

[a] C: correct; U: unclassified; W: wrong.
[b] Phases predicted as present if output >0; absent if output <0.
[c] Phases predicted as present if output >0.25; absent if <-0.25, and unclassified if $-0.25<$ output <0.25.
[d] Phases predicted as present if output >0.50; absent if <-0.50; unclassified if $-0.50<$output<0.50.
[e] Phases predicted as present if output >0.75; absent if <-0.75; unclassified if $-0.75<$ output <0.75.

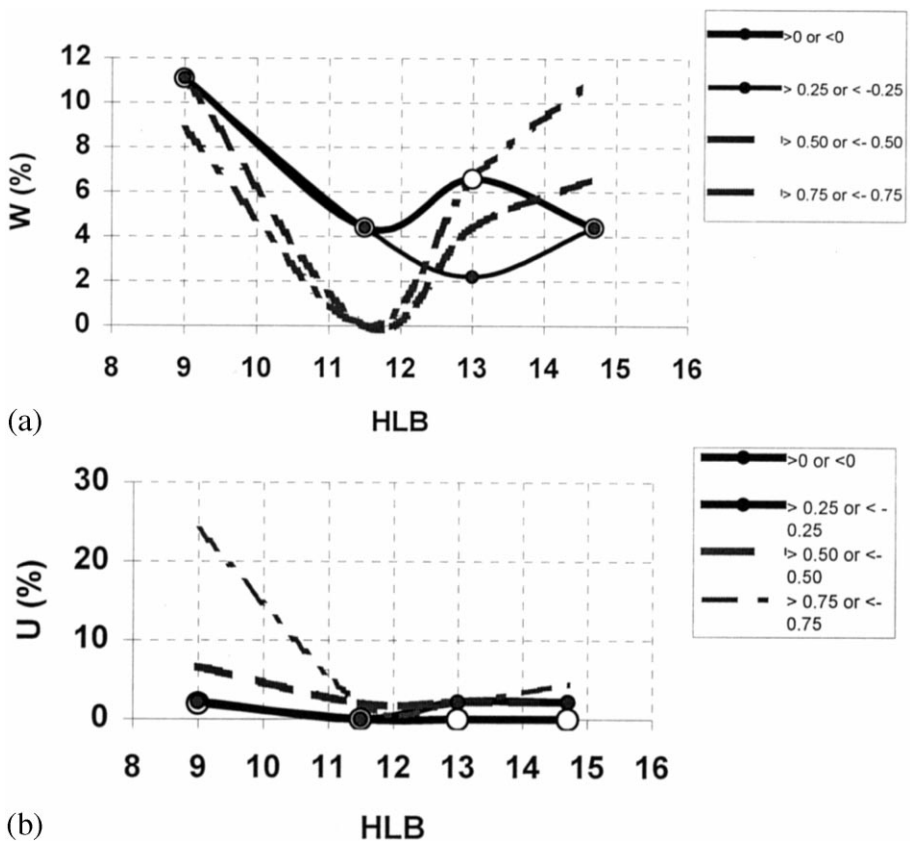

Fig. 7. Accuracy of ANNs predictions for various critical values, (a) percentage wrong predictions at different HLB values; (b) percentage unclassified predictions at different HLB values.

dicted both liquid crystal and microemulsion presence, although only LC was seen under microscope. In addition to these wrong predictions near the LC region errors also occurred in regions well

removed from the training data. The sampling strategy provided no data near the SML and PSMO apexes (Fig. 2) resulting in errors in these regions. Additionally, the phase diagram at HLB 9.0 does not contain either LC or ME regions, which may then explain the higher percentage of wrong and unclassified data at this HLB. At HLB 9.0 the trained ANN predicted the presence of a small LC region and in some cases the presence of w/o EM in addition to o/w EM when only o/w EM was detected experimentally.

Narrowing the criterion of classification had little influence on the number of the wrongly classified data but increased the percentage of unclassified data. Nevertheless the mean percentage success calculated over the validation data is very encouraging (Fig. 7), given that only about 15% of the tetrahedron was sampled.

4. Conclusion

The results of the present study show that successful prediction of phase behaviour of four component systems is possible with an optimized ANNs model structure using a range of three inputs only. In addition, the training process required a relatively small number ($\approx 15\%$ of the sample space) of experimental data sets for training and testing. Detailed experimental data were gathered from several 'slices' within a tetrahedron region and after training. The ANNs was reasonably successful in predicting other regions of that tetrahedron. The reliability of the predictions was not improved by changing the critical values for classification. Further work is required to investigate how the reliability of prediction is influenced by the sampling strategy (i.e. fraction of the tetrahedron space sampled and the distribution of samples) and the critical values used for classification.

Analyzing 180 validation points yielded an average of 90.5% correct answers, 3.4% unclassified and only 6.1% incorrect predictions. This low error rate suggests that a trained ANNs would be helpful in predicting the phase behavior of quaternary systems with less experimental effort.

Acknowledgements

We wish to thank Croda Surfactants New Zealand for free samples of surfactants and oils.

References

[1] H.N. Bhargava, A. Naruker, L.M. Lieb, Pharm. Tech. 3 (1987) 45–46.
[2] M. Bourell, R.S. Schecter, Microemulsions and Related Systems, Marcel Dekker, New York, 1988, pp. 141–142.
[3] L.S. Anker, P.C. Jurs, Anal. Chem. 64 (1992) 157–1164.
[4] J.W. Ball, P.C. Jurs, Anal. Chem. 65 (1993) 3615–3621.
[5] S. Agatonovic-Kustrin, M. Vasiljevic, Lj. Zivanovic, I. G. Tucker, J. Pharm. Biomed. Analysis 17 (1998) 69–76.
[6] C.J. Richardson, A. Mbanefo, R. Aboofazeli, M.J. Lawrence, D.J. Brown, J. Colloid Interface Sci. 187 (1997) 296–303.
[7] L.J. Erb, Pharm. Res. 10 (2) (1993) 165–170.
[8] J. Zupan, J. Gasteiger, Anal. Chim. Acta 248 (1991) 1–30.
[9] B.W. Corrigan, P.R. Mayo, F. Jamali, Ther. Drug Monit. 19 (1997) 25–28.
[10] N.J. Kale, L.V. Allen Jr, Int. J. Pharm. 57 (1989) 87–93.
[11] D.J. Mitchell, G.J.T. Tiddy, L. Waring, T. Bostock, M.P. McDonald, J. Chem. Soc. Faraday Trans. 1 (79) (1983) 975–1000.