Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 675–685
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Short communication

Robust fuzzy mappings for QSAR studies

Mohit Kumar [a,*], Kerstin Thurow [a], Norbert Stoll [b], Regina Stoll [c]

[a] Center for Life Science Automation, F.-Barnewitz-Street 8, D-18119 Rostock, M.V., Germany
[b] Institute of Automation, University of Rostock, Richard-Wagner-Street 31, D-18119 Rostock, Germany
[c] Institute of Occupational and Social Medicine, University of Rostock, St.-Georg-Street 108, D-18055 Rostock, Germany

Received 6 December 2005; received in revised form 9 April 2006; accepted 7 December 2006
Available online 18 January 2007

Abstract

This study presents a new robust method of developing quantitative structure–activity relationship (QSAR) models based on fuzzy mappings. An important issue in QSAR modelling is of robustness, i.e., model should not undergo overtraining and model performance should be least sensitive to the modelling errors associated with the chosen descriptors and structure of the model. We establish robust input–output mappings for QSAR studies based on fuzzy ''if-then'' rules. The identification of these mappings (i.e. the construction of fuzzy rules) is based on a robust criterion that the maximum possible value of energy-gain from modelling errors to the identification errors is minimum. The robustness of proposed approach has been illustrated with simulation studies and QSAR modelling examples. The method of robust fuzzy mappings has been compared with Bayesian regularized neural networks through the QSAR modelling examples of (1) carboquinones' data set, (2) benzodiazepine data set, and (3) predicting the rate constant for hydroxyl radical tropospheric degradation of 460 heterogeneous organic compounds.

Keywords: Fuzzy clustering; Fuzzy model identification; QSAR; Bayesian regularized neural networks

1. Introduction

The QSAR methods developed by Hansch and Fujita [12] identify relationship between chemical structure of compounds and their activity and have been applied to chemistry and drug design [11,18,15]. The QSAR modelling is based on the principle that molecular properties like lipophilicity, shape, and electronic properties modulate the biological activity of the molecule. Mathematically, biological activity is a function of molecular properties' descriptors:

$$\mathrm{BA} = f(d_1, d_2, \ldots),$$

where BA is a biological response (e.g. IC_{50}, ED_{50}, LD_{50}) and d_1, d_2, … are mathematical descriptors of molecular properties. The unknown function f is identified using linear methods (e.g. multiple linear regression, partial least-squares) or nonlinear models (e.g. neural networks). During the last years, the applications of neural networks in chemistry and drug design have dramatically increased. A review of the field can be found, e.g., in Refs. [25,33]. Here, we don't want to discuss the promising role of neural networks in drug lead discovery and development, rather we address the fundamental problems associated with QSAR analysis.

While developing a QSAR model for the design and discovery of bioactive agents, we may come across the situation that descriptors don't accurately capture the molecular properties relevant to the biological activity or the chosen model structure (i.e. number of adjustable model parameters) is not optimal. In such situations, there exist modelling errors. The common problems associated with QSAR modelling can be summarized as follows:

(1) For the chosen structure of the model and descriptors, there may exist modelling errors. The commonly used

* Corresponding author. Tel.: +49 381 4949956; fax: +49 381 4949952.
E-mail addresses: mohit.kumar@uni-rostock.de (M. Kumar), kerstin.thurow@uni-rostock.de (K. Thurow), norbert.stoll@uni-rostock.de (N. Stoll), regina.stoll@uni-rostock.de (R. Stoll).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.12.017

nonlinear model training algorithms (e.g. gradient-descent based backpropagation techniques) are not robust towards modelling errors.

(2) The model identification process may result in the overtraining. This leads to a loss of ability of the identified model to generalize. Although overtraining can be avoided by using validation data sets, but the computation effort to cross-validate identified models can result in large validation times for a large and diverse training data set.

To address these issues, the use of Bayesian regularized neural networks has been suggested in Refs. [6–8]. The construction of a QSAR model using input–output data is an ''ill-posed'' problem and regularization suggested is a general method to convert the model identification problem into a ''well-posed'' problem. However, the choice of regularization parameters is usually not obvious and application dependent. Bayesian regularization provides an optimal value of regularization parameters by applying Bayes' theorem [24]. However, Bayesian regularization considers the model identification problem in a stochastic framework, where assumptions have been made on the nature of distribution of signals. If these assumptions are not met, the identification performance may not be optimal. Thus, a robust method for building QSAR models (that does not make any statistical assumptions) is required.

Fuzzy inference systems based on fuzzy set theory of Zadeh [35] are considered suitable for dealing with many real-world problems, characterized by complexities, uncertainties, and a lack of knowledge of the governing physical laws. Fuzzy logic provides a conceptual framework for dealing with the problem of knowledge representation in the environment of uncertainty and imprecision. The most important application of fuzzy set theory is the fuzzy rule-based models, where the relationships among system variables are modelled using *linguistically interpretable* rules.

We believe that QSAR studies involve complexities and uncertainties (due to the lack of complete knowledge of underlying physical laws), and thus fuzzy rule-based models have much to offer in this field. A major difficulty in exploiting the fuzzy modelling techniques to address the fundamental issues of QSAR studies (listed above) is that the robust methods for the automatic construction of fuzzy models using data sets provided by combinatorial chemistry/HTS (high-throughput screening) are relatively unknown. In literature, various fuzzy identification techniques have been developed using ad hoc approaches, neural networks, genetic algorithms, clustering techniques, and Kalman filtering [1,4,9,13,16,17,23,26–32,34]. Most of the recursive (on-line) fuzzy identification methods use gradient-descent based algorithms (such as backpropagation) for the identification of nonlinear fuzzy model parameters. However, in the presence of data uncertainties and modelling errors, gradient-descent based techniques are not suitable due to their non-robust nature.

Recently, we have started working on robust construction of fuzzy models in presence of data uncertainties and modelling errors [20–22]. To design a robust identification method, a first-principle based approach is to minimize the sensitivity of the identification method towards modelling errors. The sensitivity of an identification method can be assessed by measuring a gain in energy from modelling errors to the identification errors. That is, a sensitive (non-robust) identification method will result in a higher value of identification errors' energy for a given modelling error energy. Therefore, for designing a robust identification method, we try to minimize the maximum possible value of energy-gain from modelling errors to the identification errors. The maximum value of energy-gain (that will be minimized) is calculated over all possible finite disturbances without making any statistical assumptions about the nature of signals. This is what we call as *energy-gain bounding approach* to model identification. In this article, we will illustrate the development of QSAR models based on energy-gain bounding approach. It will be shown that energy-gain bounding approach leads to the development of robust QSAR fuzzy models. The method of QSAR models' development using Bayesian regularized neural networks has been taken as a reference method for comparing the performance of proposed robust QSAR fuzzy models.

We start with the introduction of clustering based fuzzy mappings followed by an energy-gain bounding approach to the identification of these fuzzy mappings. Afterwards, simulation studies and three examples of QSAR studies (concerned with carboquinones, benzodiazepine, and tropospheric degradation data set of volatile organic compounds) have been presented. Finally, the concluding remarks are provided.

2. Theory

2.1. Clustering based fuzzy mappings

We introduce a model that maps the clusters of input parameters' space into real output values based on some predefined fuzzy rules. If input space has been divided into K different clusters, then the model consists of K fuzzy rules:

$$\begin{aligned}&\textbf{If } x \text{ belongs to a cluster having centre } c_1 \textbf{ then } y = \alpha^1,\\ &\qquad\qquad\vdots\\ &\textbf{If } x \text{ belongs to a cluster having centre } c_K \textbf{ then } y = \alpha^K,\end{aligned}$$

where x is an input vector, c_i is the centre of ith cluster, and the values $\alpha^1, \ldots, \alpha^K$ are real numbers. This type of fuzzy model will be characterized by cluster centres $(c_1, \ldots, c_K)$ and numbers $\alpha^1, \ldots, \alpha^K$. If we define a parameter vector θ to characterize cluster centres and a parameter vector α to characterize the numbers $\alpha^1, \ldots, \alpha^K$, then the model output y to input x could be expressed as

$$y = G^{\mathrm{T}}(x, \theta)\alpha,$$

where G is a nonlinear function of x and θ is a K-dimensional vector such that ith element of G represents the degree of membership of input x to the ith cluster. The exact mathematical details have been given in Appendix A. The identification

of above defined fuzzy model means to identify the parameters α and θ.

2.2. An energy-gain bounding approach

The aim is to identify fuzzy model parameters (α, θ) using input–output data set $\{x(j), y(j)\}_{j=0}^{k}$. Here, $x(j)$ is the jth-indexed input data (i.e. molecular properties' descriptors) and $y(j)$ is the corresponding biological response. In the literature of neural networks and fuzzy modelling, instantaneous-gradient-based algorithms (such as backpropagation) are the most commonly used techniques for the identification of nonlinear model parameters. In the situation of modelling errors, any identification method (e.g. gradient descent) may lead to the errors in the identification of model parameters. The modelling errors arise due to the non-optimal choice of either descriptors or model structure. An ideal identification method is the one that results in no identification errors, no matter what the modelling errors are. However, in practical problems, it is not easy to design an ideal identification method since the modelling errors' signals are unknown. Therefore, we are concerned to design a robust identification method that is least sensitive to the modelling errors. Our approach to the design of robust identification method is based on an energy-gain criterion that

$$\max \frac{\text{energy of identification errors}}{\text{energy of modelling errors}} \rightarrow \min.$$

In other words, the identification method minimizes the maximum possible value of energy-gain from modelling errors to the identification errors. Such an identification method will guarantee that *small modelling errors cannot lead to large identification errors*. The maximum value of energy-gain (that will be minimized) is calculated over all possible finite disturbances without making any statistical assumptions about the nature of signals. This is what we call as "energy-gain bounding approach" to fuzzy model identification.

The energy-gain bounding approach to the identification of fuzzy model parameters has been studied in Ref. [19]. Here, we present the final results of energy-gain bounding approach. It was shown in Ref. [19] that the identification of fuzzy model parameters using data set $\{x(j), y(j)\}_{j=0}^{k}$ follows by performing for $j = 0, \ldots, k$, the recursions

$$\theta_j = \arg\min_{\theta} \left[\frac{\left[y(j) - G^{\mathrm{T}}(x(j),\theta)\alpha_{j-1}\right]^2}{1+\mu(j)\left\|G(x(j),\theta)\right\|^2} + (\mu_\theta(j))^{-1}\left\|\theta - \theta_{j-1}\right\|^2 \right], \tag{1}$$

$$\alpha_j = \alpha_{j-1} + \frac{\mu(j)G\big(x(j),\theta_j\big)\left[y(j) - G^{\mathrm{T}}\big(x(j),\theta_j\big)\alpha_{j-1}\right]}{1+\mu(j)\left\|G(x(j),\theta_j)\right\|^2}, \ \alpha_{-1} = 0, \tag{2}$$

where θ_{-1} corresponds to the initial guess of cluster centres, and $\mu(j)$, $\mu_\theta(j)$ are positive values which reflect the weightage given to the data pair ($x(j)$, $y(j)$) in the identification of fuzzy parameters. The optimal values of parameters ($\mu(j)$, $\mu_\theta(j)$), which result in fast convergence and low misadjustment error are given as

$$\mu(j) = \frac{\left\|\widehat{p}_j\right\|^2}{C\left\|G(x(j),\theta_j)\right\|^2}, \quad \mu_\theta(j) = s_\theta \frac{\left\|\widehat{p}_j\right\|^2}{C\left\|G(x(j),\theta_j)\right\|^2},$$

$$\widehat{p}_j = \omega\widehat{p}_{j-1} + (1-\omega)\frac{y(j) - G^{\mathrm{T}}\big(x(j),\theta_j\big)\alpha_{j-1}}{\left\|G(x(j),\theta_j)\right\|^2} G\big(x(j),\theta_j\big),$$

where s_θ is a predefined positive constant, ω $(0 < \omega < 1)$ is a smoothing factor, and C is a positive constant that should be chosen proportional to the magnitude of modelling errors.

2.3. Simulation studies

The aim of this section is to verify the robustness of energy-gain bounding approach for the fuzzy identification of unknown input–output mappings in presence of disturbance signals. For this purpose assume a training data set of 49 input–output pairs

$$\left\{ \begin{bmatrix} x_1(j) \\ x_2(j) \end{bmatrix}, y(j) \right\}_{j=0,\ldots,48},$$

generated according to

$$y(j) = f(x_1(j), x_2(j)) + nj,$$

$$f(x_1,x_2) = (1-x_1x_2)\mathrm{e}^{-(x_1+x_2)^2} - \cos(4x_1x_2) + \log(1+x_1x_2),$$

where input points

$$\left\{ \begin{bmatrix} x_1(j) \\ x_2(j) \end{bmatrix} \right\}_{j=0,\ldots,48}$$

are distributed uniformly on two-dimensional space $[-0.9, 0.9] \times [-0.9, 0.9]$, and n_j is a random number chosen from a uniform distribution on the interval $[-0.2, 0.2]$. The goal is to identify the unknown function $f(x_1, x_2)$ using a model. Let us try to identify the unknown function $f(x_1, x_2)$ using fuzzy mappings based on energy-gain bounding method. The identification method (1) and (2) can be implemented using a Gauss–Newton based algorithm suggested in Appendix B. Let us choose the number of clusters K equal to 7 that makes total number of adjustable model parameters equal to 21 (14 cluster centres' parameters and $\alpha^1, \ldots, \alpha^7$). The initial guess about seven different cluster centres is taken as

$$[c_1 \cdots c_7] = \begin{bmatrix} -0.9 & -0.9 & -0.9 & 0 & 0.9 & 0.9 & 0.9 \\ -0.9 & 0 & 0.9 & 0 & -0.9 & 0 & 0.9 \end{bmatrix}, \ \theta_{-1} = \left[c_1^{\mathrm{T}} \cdots c_7^{\mathrm{T}}\right]^{\mathrm{T}}.$$

The identification parameters were chosen as $\tilde{m} = 2$ (for expression (A2)), $C = 0.05$, $\omega = 0.99$, and $s_\theta = 0.05$. The performance of the identified model is evaluated by defining generalization error as

$$\mathrm{GE} = \frac{1}{100}\sum_{l=1}^{100} |f(x_1(l), x_2(l)) - \mathrm{MO}(x_1(l), x_2(l))|^2,$$

where the points $\{[x_1(l)\ x_2(l)]^{\mathrm{T}}\}_{l=1,\ldots,100}$ are uniformly distributed on two-dimensional input space and MO(x_1, x_2) is model output for input vector $[x_1\ x_2]^{\mathrm{T}}$. Fig. 1 shows the learning curve of fuzzy model parameters where generalization error has been plotted with the length of training (i.e. number of epochs). The generalization error after 100 epochs has been reduced to 0.0853.

Regarding the identification parameters, $\tilde{m} = 2$, $\omega = 0.99$ and $s_\theta = 0.05$ provide in general good results. However, C should be chosen proportional to the magnitude of modelling errors. A practical method of adjusting the value of C is to choose a value that ensures a smooth learning of model parameters, i.e., the curve of Fig. 1 is smooth.

We also employed a Bayesian regularized neural network for the identification of unknown function using the same training data set. We consider a two-layer feed-forward network. The first layer has five neurons and the second layer has one neuron so that total number of adjustable network parameters is equal to 21 (same as that of fuzzy mappings). The training algorithm has been run until the number of effective parameters has converged. To eliminate the effects caused by random initialization of weights at the start of training, the network has been trained independently 50 times and the value of generalization error is calculated for each trained network. The average value of generalization error over 50 independently trained networks was calculated equal to 0.2444 (also shown in Fig. 1).

The above example has clearly shown the better performance (in terms of generalization error) of the proposed approach than the Bayesian regularized neural network in presence of disturbances. We will now further verify the robustness properties of proposed approach through examples of QSAR modelling.

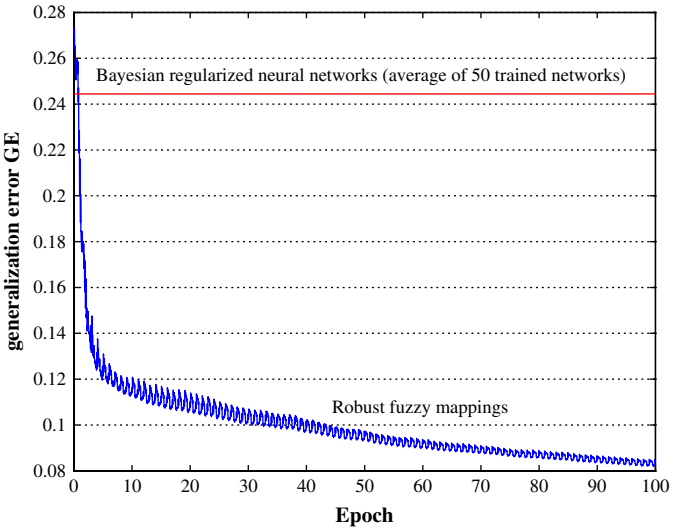

Fig. 1. Learning curve of fuzzy model parameters.

3. QSAR modelling examples

In the literature, Bayesian regularized neural networks have been considered as a robust method of QSAR modelling [6–8]. Therefore, we compare the performance of proposed robust fuzzy mappings with the Bayesian regularized neural networks. The first example is concerned with carboquinones exhibiting anticarcinogenic activity. The anticarcinogenic activity of carboquinones depends upon substituents R^1 and R^2. For the QSAR study of carboquinones, we choose the input–output data set of Refs. [3,14], shown in Table 1. The influence of R^1 and R^2 on activity is described using physicochemical parameters: molecular refractivity constants (MR), hydrophobicity constant (π), substituent constants (F and R), as well as $\mathrm{MR}_{1,2}$ and $\pi_{1,2}$. The biological data consist of *minimum effective dose* (that gives a 40% increase in lifespan of test animals) on a chronic treatment schedule. The minimum effective dose (MED) is expressed in log(1/conc), where conc is the concentration of the substance necessary to give the desired effect. We assume that MED is a function of physicochemical parameters:

$$\mathrm{MED} = f(\mathrm{MR}_{1,2}, \pi_{1,2}, \pi_2, \mathrm{MR}_1, F, R).$$

The data of Table 1 have been divided, as illustrated in Table 2, into four different pairs of training and testing sets.

The data sets have been normalized on a scale of [−1, 1] using following equations

$$\mathrm{MR}_{1,2}^N = \frac{\mathrm{MR}_{1,2}}{6.14},\ \pi_{1,2}^N = \frac{\pi_{1,2}}{5},\ \pi_2^N = \frac{\pi_2}{3.16},\ \mathrm{MR}_1^N = \frac{\mathrm{MR}_1}{3.07},$$
$$F^N = \frac{F}{0.52},\ R^N = \frac{R}{1.05},\ \mathrm{MED}^N = \frac{(\mathrm{MED} - 5.7551)}{1.4251}.$$

A Bayesian regularized feed-forward neural network that consists of two layers (four neurons in first and one neuron in second layer) is used to identify the relationships between MED^N and ($\mathrm{MR}_{1,2}^N$, $\pi_{1,2}^N$, ${\pi_2}^N$, ${\mathrm{MR}_1}^N$, F^N, R^N). To develop a fuzzy QSAR model, the six-dimensional input space is partitioned into five different clusters. The initial guess about cluster centres is taken by performing fuzzy c-means clustering on training data. The identification parameters were chosen as $\tilde{m} = 2$, $\omega = 0.99$, $s_\theta = 0.05$, and $C = 0.001$. Table 3 makes a comparison of the two techniques on different training/testing data sets. Table 3 shows the root mean square error (RMSE) in predicting MED values. The better performance of robust fuzzy mappings on the data sets can be seen in the table. To further illustrate the better performance of fuzzy mappings, consider the case that about 50% of the data serve as training data and rest 50% as testing data. All the odd numbered data in Table 1 are used for the training and even numbered for the testing. A fuzzy model of 10 rules is trained taking $\tilde{m} = 2$, $\omega = 0.99$, $s_\theta = 0.05$, and $C = 0.001$. The initial guess about clusters is that the cluster centres are equally spaced in the input space. Fig. 2 shows the performance of QSAR models on training and testing data. RMSE of Bayesian regularized neural network on training data is 0.1426 and on testing data is 0.3528. Robust fuzzy mappings show a better

Table 1
Carboquinones' data set

No.	R^1	R^2	$MR_{1,2}$	$\pi_{1,2}$	π_2	MR_1	F	R	MED
1	C_6H_5	C_6H_5	5.08	3.92	1.96	2.54	0.16	−0.16	4.33
2	CH_3	$(CH_2)_3C_6H_5$	4.50	3.66	3.16	0.57	−0.08	−0.26	4.47
3	C_5H_{11}	C_5H_{11}	4.86	5.00	2.50	2.43	−0.08	−0.26	4.63
4	$CH(CH_3)_2$	$CH(CH_3)_2$	3.00	2.60	1.30	1.50	−0.08	−0.26	4.77
5	CH_3	$CH_2C_6H_5$	3.57	2.51	2.01	0.57	−0.12	−0.14	4.85
6	C_3H_7	C_3H_7	3.00	3.00	1.50	1.50	−0.08	−0.26	4.92
7	CH_3	$CH_2OC_6H_5$	3.79	2.16	1.66	0.57	−0.04	−0.13	5.15
8	$CH_2CH_2OCON(CH_3)_2$	$CH_2CH_2OCON(CH_3)_2$	6.14	0.72	0.36	3.07	−0.08	−0.26	5.16
9	C_2H_5	C_2H_5	2.06	2.00	1.00	1.03	−0.08	−0.26	5.46
10	CH_3	$CH_2CH_2OCH_3$	2.28	1.03	0.53	0.57	−0.08	−0.26	5.57
11	OCH_3	OCH_3	1.58	−0.04	−0.02	0.79	0.52	−1.02	5.59
12	CH_3	$CH(CH_3)_2$	2.07	1.80	1.30	0.57	−0.08	−0.26	5.60
13	C_3H_7	$CH(OCH_3)CH_2OCONH_2$	4.24	0.98	−0.52	1.50	−0.04	−0.13	5.63
14	CH_3	CH_3	1.14	1.00	0.50	0.57	−0.08	−0.26	5.66
15	H	$CH(CH_3)_2$	1.60	1.30	1.30	0.10	−0.04	−0.13	5.68
16	CH_3	$CH(OCH_3)C_2H_5$	2.75	1.53	1.03	0.57	−0.04	−0.13	5.68
17	C_3H_7	$CH_2CH_2OCONH_2$	3.56	1.45	−0.05	1.50	−0.08	−0.26	5.68
18	$CH_2CH_2OCH_3$	$CH_2CH_2OCH_3$	3.42	1.03	0.53	1.71	−0.08	−0.26	5.69
19	C_2H_5	$CH(OC_2H_5)CH_2OCONH_2$	4.23	0.98	−0.02	1.03	−0.04	−0.13	5.76
20	CH_3	$CH_2CH_2OCOCH_3$	2.78	1.23	0.73	0.57	−0.08	−0.26	5.78
21	CH_3	$(CH_2)_3$-dimer	1.96	2.00	1.50	0.57	−0.08	−0.26	5.82
22	CH_3	C_2H_5	1.60	1.50	1.00	0.57	−0.08	−0.26	5.86
23	CH_3	$CH(OCH_2CH_2OCH_3)CH_2OCONH_2$	4.45	0.01	−0.49	0.57	−0.04	−0.13	6.03
24	CH_3	$CH_2CH(CH_3)OCONH_2$	3.09	0.75	0.25	0.57	−0.08	−0.26	6.14
25	C_2H_5	$CH(OCH_3)CH_2OCONH_2$	3.77	0.48	−0.52	1.03	−0.04	−0.13	6.16
26	CH_3	$CH(C_2H_5)CH_2OCONH_2$	3.55	1.25	0.75	0.57	−0.08	−0.26	6.18
27	CH_3	$CH(OC_2H_5)CH_2OCONH_2$	3.77	0.48	−0.02	0.57	−0.04	−0.13	6.18
28	CH_3	$(CH_2)_3OCONH_2$	3.09	0.95	0.45	0.57	−0.08	−0.26	6.18
29	CH_3	$(CH_2)_2OCONH_2$	2.63	0.45	−0.05	0.57	−0.08	−0.26	6.21
30	C_2H_5	$(CH_2)_2OCONH_2$	3.09	0.95	−0.05	1.03	−0.08	−0.26	6.25
31	CH_3	CH_2CH_2OH	1.78	0.34	−0.16	0.57	−0.08	−0.26	6.39
32	CH_3	$CH(CH_3)CH_2OCONH_2$	3.09	0.75	0.25	0.57	−0.08	−0.26	6.41
33	CH_3	$CH(OCH_3)CH_2OCONH_2$	3.31	−0.02	−0.52	0.57	−0.04	−0.13	6.41
34	H	$N(CH_2)_2$	1.66	0.18	0.18	0.10	0.10	−0.92	6.45
35	CH_2CH_2OH	CH_2CH_2OH	2.42	−0.32	−0.16	1.21	−0.08	−0.26	6.54
36	CH_3	$N(CH_2)_2$	2.13	0.68	0.18	0.57	0.06	−1.05	6.77
37	CH_3	$CH(OCH_3)CH_2OH$	2.47	−0.13	−0.63	0.57	−0.04	−0.13	6.90

performance (RMSE of 0.1210 on training and 0.2598 on testing data).

The identified fuzzy model consists of following 10 rules:

if x belongs to a cluster having centre c_1 then $MED^N = 0.6080$,
if x belongs to a cluster having centre c_2 then $MED^N = 0.9102$,
if x belongs to a cluster having centre c_3 then $MED^N = 2.4835$,
if x belongs to a cluster having centre c_4 then $MED^N = 0.1435$,
if x belongs to a cluster having centre c_5 then $MED^N = -1.1957$,
if x belongs to a cluster having centre c_6 then $MED^N = -1.5532$,
if x belongs to a cluster having centre c_7 then $MED^N = -1.5840$,
if x belongs to a cluster having centre c_8 then $MED^N = -0.6234$,
if x belongs to a cluster having centre c_9 then $MED^N = -0.0356$,
if x belongs to a cluster having centre c_{10} then $MED^N = -0.0844$,

Table 2
Training/testing data sets of carboquinones

	Testing data	Training data
Set 1	No. 4, 8, …, 36 in Table 1	Rest
Set 2	No. 3, 7, …, 35 in Table 1	Rest
Set 3	No. 2, 6, …, 34 in Table 1	Rest
Set 4	No. 1, 5, …, 37 in Table 1	Rest

Table 3
Performance of fuzzy mappings and Bayesian regularized neural networks (BRNN) on carboquinones' data

Data set	Training RMSE (BRNN)	Training RMSE (fuzzy mappings)	Testing RMSE (BRNN)	Testing RMSE (fuzzy mappings)
Set 1	0.2051	0.1844	0.2717	0.2280
Set 2	0.2080	0.1663	0.4596	0.3459
Set 3	0.1602	0.1272	0.3222	0.2969
Set 4	0.1946	0.1597	0.2405	0.2454

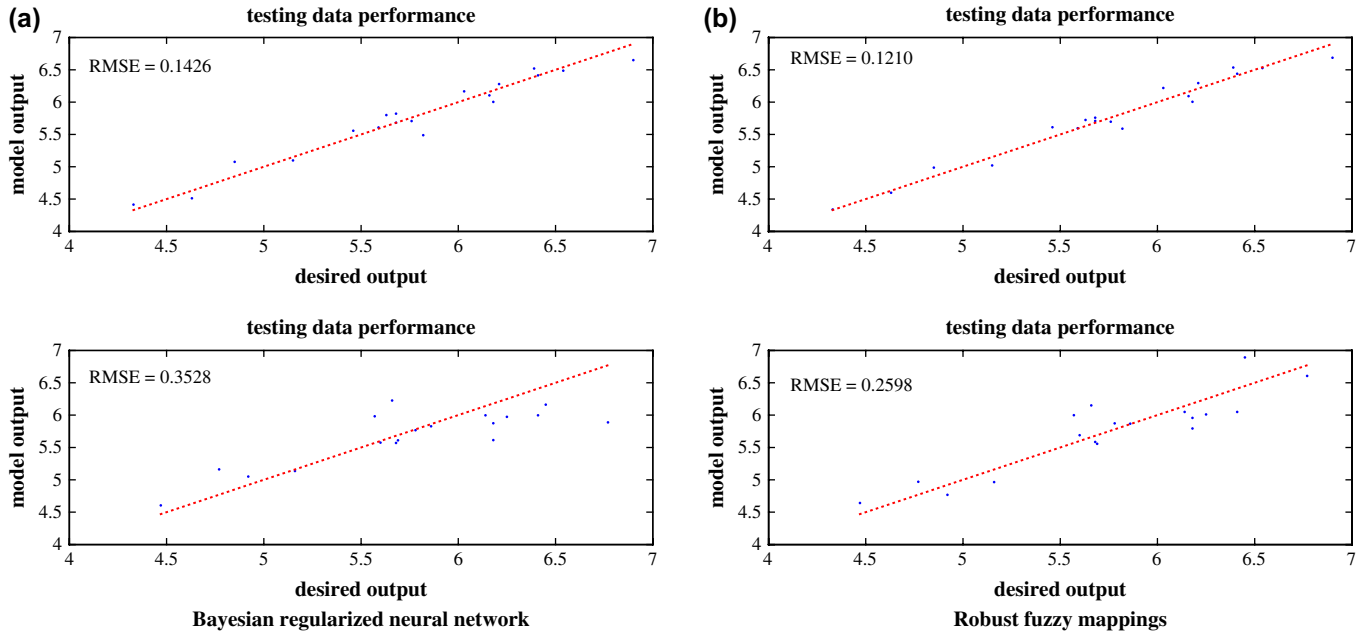

Fig. 2. Performance of QSAR models on carboquinones' data set.

where $x = [\mathrm{MR}^N_{1,2}\ \pi^N_{1,2}\ {\pi_2}^N\ {MR_1}^N\ F^N\ R^N]^{\mathrm{T}}$ and

$$c_1 = [0.1736\ \ -0.0627\ \ -0.1962\ \ 0.0323\ \ -0.2331\ \ -1.0268]^{\mathrm{T}},$$

$$c_2 = [0.2560\ \ 0.0406\ \ -0.0643\ \ 0.1354\ \ -0.0813\ \ -0.9299]^{\mathrm{T}},$$

$$c_3 = [0.2231\ \ -0.0715\ \ -0.1555\ \ 0.2250\ \ 0.0375\ \ -0.8202]^{\mathrm{T}},$$

$$c_4 = [0.4346\ \ 0.2497\ \ 0.1976\ \ 0.2683\ \ 0.1545\ \ -0.6980]^{\mathrm{T}},$$

$$c_5 = [0.5276\ \ 0.3653\ \ 0.2337\ \ 0.5358\ \ 0.4536\ \ -0.7102]^{\mathrm{T}},$$

$$c_6 = [0.6057\ \ 0.5292\ \ 0.4574\ \ 0.5369\ \ 0.3695\ \ -0.4676]^{\mathrm{T}},$$

$$c_7 = [0.7160\ \ 0.5964\ \ 0.5558\ \ 0.6512\ \ 0.5558\ \ -0.4197]^{\mathrm{T}},$$

$$c_8 = [0.8120\ \ 0.7737\ \ 0.7479\ \ 0.7687\ \ 0.6573\ \ -0.3236]^{\mathrm{T}},$$

$$c_9 = [0.8843\ \ 0.8627\ \ 0.8403\ \ 0.8469\ \ 0.7602\ \ -0.2396]^{\mathrm{T}},$$

$$c_{10} = [0.9851\ \ 1.0140\ \ 0.9941\ \ 1.0044\ \ 1.0052\ \ -0.1188]^{\mathrm{T}}.$$

The second example deals with the data set of benzodiazepines shown in Table 4. The data set has been taken from Refs. [2,3]. It is assumed that *anti*-pentylenetetrazole effect (ape) is a function of input parameters MR-3, π-3, MR-7, σ_m-3, F-4, R-4, and I-1. Table 5 illustrates the generation of three pairs of training/testing data sets from the data of Table 4. Please note in Table 5 that about 75% of data in set 1 are training data, about 66% of data in set 2 are training data, and about 50% of data in set 3 are training data. We compare a two-layer feed-forward Bayesian regularized neural network (five neurons in first layer and one neuron in second layer) with a fuzzy model of six rules. The initial guess about clusters is such that the six cluster centres are equally spaced in the input space. The fuzzy model parameters were identified taking $\tilde{m} = 2$, $\omega = 0.99$, $s_\theta = 0.05$, and $C = 0.05$. Table 6 summarizes the performance of two methods on the data sets of Table 5. Again, a better performance of robust fuzzy mappings could be seen in Table 6.

Finally, consider the problem of predicting the rate constant for hydroxyl radical tropospheric degradation of 460 volatile organic compounds considered in Ref. [10]. The data set of Ref. [10] contains hydroxyl radical degradation rate constants of 460 heterogeneous organic compounds, measured in gas phase at 25 °C and 1 atm. In Ref. [10], QSAR models have been developed based on Genetic-algorithm selection of theoretical molecular descriptors and the data set was split into training and test sets using the methods of statistical D-optimal experimental design and Kohonen artificial neural network. Let us assume that

$$-\log k(\mathrm{OH}) = f(\mathrm{HOMO}, n\mathrm{X}, \mathrm{CIC0}, n\mathrm{CaH}),$$

where k(OH) is the hydroxyl radical rate constant, HOMO is the energy of highest occupied molecular orbital, nX is the number of halogen atoms, CIC0 is the complementary information content index, and nCaH is the number of unsubstituted aromatic carbon (sp2). The data set of Ref. [10], split into training and testing using D-optimal algorithm, has been used to develop QSAR models based on robust fuzzy mappings and Bayesian regularized neural network. Again, the

Table 4
Benzodiazepines' data set

No.	MR-3	π-3	MR-7	σ_m-3	*F*-4	*R*-4	*I*-1	*anti*-Pentylenetetrazole effect (ape)
1	0.6	0.71	2.42	0.37	0	0	0	4.99
2	0.09	0.14	0.57	0.34	0.43	−0.34	0	3.33
3	1.84	1.07	0.1	0.15	0	0	1	3.57
4	0.6	0.71	1.92	0.37	0	0	0	5.83
5	2.77	2.07	0.1	0.15	0	0	1	3.79
6	0.74	−0.28	2.42	0.71	0	0	0	4.80
7	1.56	0.18	0.10	−0.15	0	0	1	3.84
8	0.60	0.71	0.10	0.37	0.26	−0.51	1.0	4.60
9	0.60	0.71	2.95	0.37	0	0	0	3.38
10	0.60	0.71	1.65	0.37	0	0	0	4.34
11	0.74	−0.28	1.92	0.71	0	0	0	5.29
12	0.60	0.71	3.41	0.37	0	0	0	5.06
13	0.60	0.71	2.39	0.37	0	0	0	5.35
14	0.10	0	0.57	0	0.43	−0.34	0	4.53
15	0.60	0.71	3	0.37	0	0	0	4.64
16	0.60	0.71	2.48	0.37	0	0	0	4.73
17	0.60	0.71	2.95	0.37	0.43	−0.34	0	4.76
18	0.50	0.88	1.92	0.43	0	0	0	5.06
19	1.38	0.61	0.10	0.15	0	0	1	4.15
20	0.60	0.71	1.44	0.37	0	0	0	5.07
21	1.37	−1.58	0.10	0.52	0	0	1	4.08
22	0.60	0.71	0.10	0.37	−0.04	−0.13	1	4.57
23	1.56	0.18	0.57	−0.15	0	0	0	4.69
24	0.60	0.71	3.41	0.37	0.43	−0.34	0	5.38
25	1.56	0.18	0.57	0.37	0.41	−0.15	0	5.82
26	0.74	−0.28	2.95	−0.15	0	0	0	4.28
27	0.74	−0.28	2.48	0.71	0	0	0	4.70
28	0.60	0.71	0.10	0.37	0.41	−0.15	1	5.85
29	0.60	0.71	1.82	0.37	0	0	0	4.90
30	0.63	−0.57	0.10	0.56	0	0	1	5.30
31	0.74	−0.28	0.10	0.71	0.38	0.19	1	5.70
32	0.60	0.71	0.10	0.37	0	0	1	4.65
33	0.63	−0.57	0.10	0.56	0.43	−0.34	1	5.63
34	0.60	0.71	1.03	0.37	0	0	0	4.90
35	1.38	0.61	0.57	0.15	0	0	0	3.60
36	0.60	0.71	1.51	0.37	0	0	0	5.28
37	0.60	0.71	0.10	0.37	0.44	−0.17	1	5.77
38	0.60	0.71	0.10	0.37	0.43	−0.34	1	6.46
39	0.74	−0.28	0.57	0.71	0.38	0.19	0	5.71
40	0.50	0.88	2.48	0.43	0	0	0	4.76
41	0.60	0.71	1.45	0.37	0	0	0	5.35
42	0.60	0.71	0.57	0.37	0.41	−0.15	0	6.03
43	0.60	0.71	0.57	0.37	0	0	0	5.31
44	0.74	−0.28	0.57	0.71	0.41	−0.15	0	6.92
45	0.5	0.88	0.10	0.43	0.38	0.19	1	4.97
46	0.89	0.86	0.10	0.39	0	0	1.0	5.20
47	0.63	−0.57	0.57	0.56	0	0	0	5.44
48	0.74	−0.28	0.10	0.71	0	0	1	5.60
49	0.74	−0.28	0.57	0.71	0	0	0	5.62
50	0.50	0.88	0.10	0.43	0	0	1	5.48
51	0.60	0.71	0.57	0.37	0	−0.34	0	5.88
52	0.50	0.88	1.45	0.43	0	0	0	5.03
53	0.74	−0.28	0.1	0.71	0.41	−0.15	1	6.30
54	0.74	−0.28	0.54	0.71	0	0	0	5.77
55	0.74	−0.28	0.10	0.71	0.67	0.16	1	5.97
56	0.74	−0.28	0.10	0.71	0.43	−0.34	1	6.00
57	0.74	−0.28	0.57	0.71	0.43	−0.34	0	6.50

data have been normalized on [−1, 1]. The considered Bayesian regularized network consists of two layers (13 neurons in first and one neuron in second layer). To develop a robust fuzzy QSAR model, the training data were partitioned into 15 different clusters using fuzzy *c*-means clustering. This served as the initial guess about cluster centres. The identification parameters were chosen as $\tilde{m} = 2$, $\omega = 0.99$, $s_\theta = 0.05$, and $C = 0.003$. Fig. 3 shows the performance of QSAR models

Table 5
Training/testing data sets of benzodiazepines

	Testing data	Training data
Set 1	No. 4, 8, …, 56 in Table 4	Rest
Set 2	No. 3, 6, …, 57 in Table 4	Rest
Set 3	No. 2, 4, …, 56 in Table 4	Rest

to predict the $-\log k(\mathrm{OH})$ values. Here, robust fuzzy mappings (having RMSE value equal to 0.4291 on training and 0.3704 on testing data) show a slightly better performance than the Bayesian network (having RMSE value equal to 0.4472 on training and 0.3763 on testing data).

We have seen in different examples of QSAR studies (Figs. 2 and 3 and Tables 3 and 6) that robust fuzzy mappings outperform the Bayesian regularized neural networks.

4. Concluding remarks

This study has presented a new robust method of developing QSAR models using fuzzy mappings. We have addressed the fundamental issues involved in QSAR studies related to modelling errors. The robust construction of fuzzy mappings is based on an energy-gain bounding approach. Simulation studies and QSAR modelling examples have been provided to illustrate that in presence of modelling errors (a common problem in QSAR studies), the proposed robust method of fuzzy modelling is more suitable than the Bayesian regularized neural networks. This is due to the fact that energy-gain bounding approach mathematically takes into account the issue of modelling errors in a sensible manner without making any assumption about the nature of signals.

Table 6
Performance of fuzzy mappings and Bayesian regularized neural networks (BRNN) on benzodiazepines' data

Data set	Training RMSE (BRNN)	Training RMSE (fuzzy mappings)	Testing RMSE (BRNN)	Testing RMSE (fuzzy mappings)
Set 1	0.5721	0.5011	0.5462	0.4500
Set 2	0.5444	0.4451	0.7647	0.7102
Set 3	0.4067	0.4417	0.7695	0.6620

Acknowledgements

This study was supported by Bundesministerium für Bildung und Forschung (BMBF), Germany. The tropospheric degradation data set has been taken from http://ecb.jrc.it/.

Appendix A. Mathematical formulation of clustering based fuzzy mappings

Let us consider a Sugeno type fuzzy inference system $(F_s : X \to Y)$, mapping n-dimensional input space $(X = X_1 \times X_2 \times \cdots \times X_n)$ to one-dimensional real line, consisting of K different following rules:

If x belongs to a cluster having centre c_1 **then** $y = \alpha^1$,
⋮
If x belongs to a cluster having centre c_K **then** $y = \alpha^K$,

where $x \in R^n$ is n-dimensional input vector, $c_i \in R^n$ is the centre of ith cluster, and the values $\alpha^1, \ldots, \alpha^K$ are real numbers. Let $A_i(x)$ denote a multivariable membership function $A_i : X \to [0, 1]$ that represents the degree of membership of

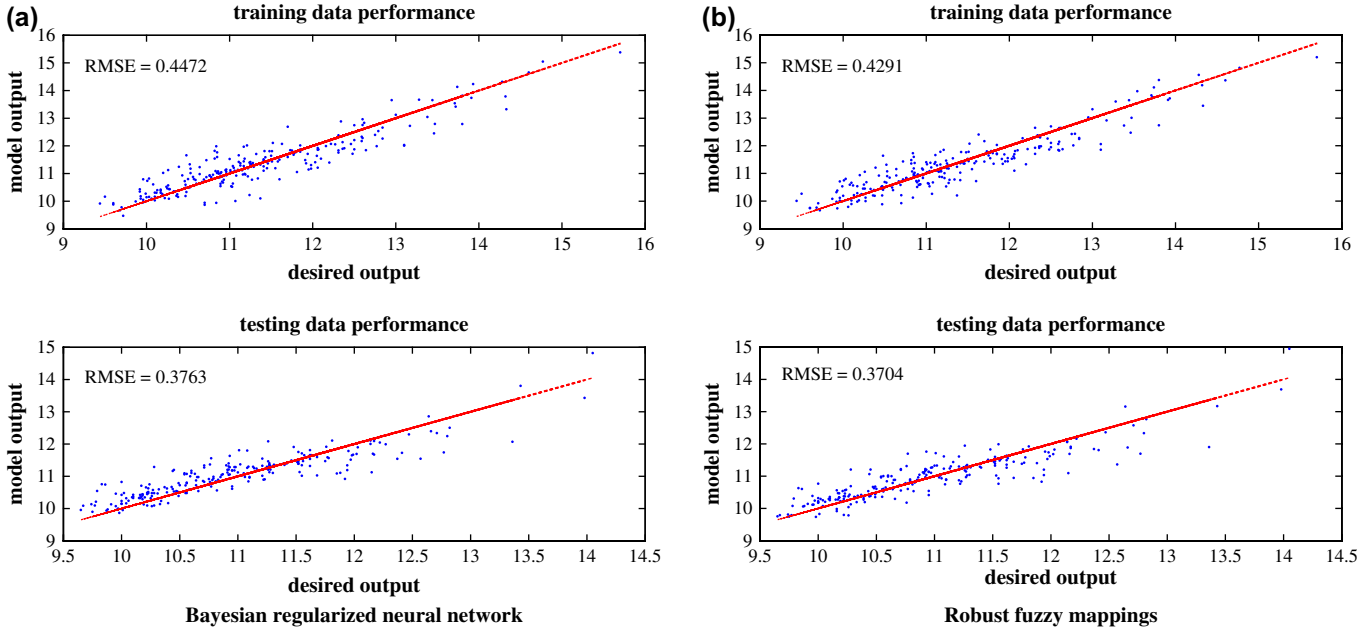

Fig. 3. Performance of QSAR models to predict the $-\log k(\mathrm{OH})$ values.

input vector $x \in X$ to the ith cluster. The different fuzzy rules can be aggregated as

$$F_s(x) = \frac{\sum_{i=1}^{K} \alpha^i A_i(x)}{\sum_{i=1}^{K} A_i(x)}. \tag{A1}$$

For the fuzzy partition of input space X into K different clusters by the method of fuzzy c-means (FCM), the membership function $A_i(x)$ must ensure (see, Ref. [5]):

$$\sum_{x \in X} \sum_{i=1}^{K} A_i^{\tilde{m}}(x) \|x - c_i\|^2 \rightarrow \text{Minimum}, \quad \sum_{i=1}^{K} A_i(x) = 1,$$

where $\tilde{m} > 1$ is the *fuzzifier* and $\|\cdot\|$ denotes the Euclidean norm. The membership function that minimizes above objective function for a given choice of cluster centres $\{c_i\}_{i=1}^{K}$ follows as

$$\mathrm{FCM}_i(x, c_1, \ldots, c_K) = \begin{cases} \dfrac{1}{\sum_{j=1}^{K} \left(\dfrac{\|x - c_i\|^2}{\|x - c_j\|^2} \right)^{1/(\tilde{m}-1)}} & \text{for } x \in X \setminus \{c_j\}_{j=1,\ldots,K}, \\ 1 & \text{for } x = c_i, \\ 0 & \text{for } x \in \{c_j\}_{j=1,\ldots,K} \setminus \{c_i\}. \end{cases} \tag{A2}$$

Fig. A1(a) shows a one-dimensional example of membership function $\mathrm{FCM}_i(\cdot)$ for three different clusters with cluster centres at 20, 50 and 80 for $\tilde{m} = 2$. The membership functions constructed using Eq. (A2), have an inherent limitation of being non-convex, as seen from Fig. A1(a). That is, the points lying far away from cluster centre may be assigned more membership to the cluster than the points lying closer. The origin of the problem is the constraint that membership values sum is equal to unity.

To overcome the problem of non-convexity, we choose a robust clustering criterion by assuming that there is a noise cluster outside each data cluster. We seek to minimize

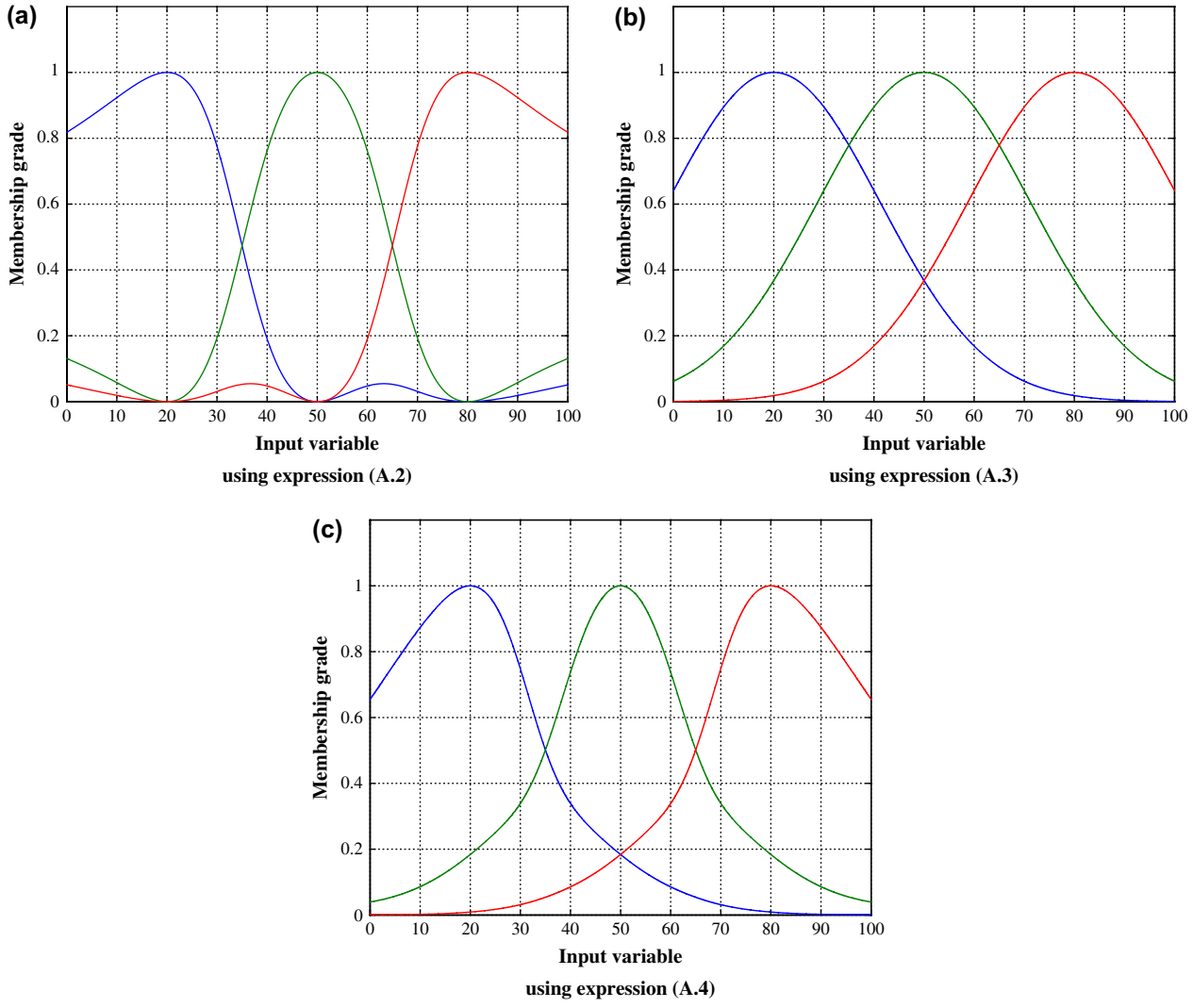

Fig. A1. Membership functions of three clusters in one-dimensional input space. (a) Using expression (A2); (b) using expression (A3); (c) using expression (A4).

$$J_c(A_i(x), c_1, \ldots, c_K) = \sum_{x \in X} \sum_{i=1}^{K} \left[A_i(x) \|x - c_i\|^2 + \{1 + A_i(x) \log A_i(x) - A_i(x)\} \delta_i \right],$$

where the second term in the objective function is intended to be a noise cluster. The term $\{1 + A_i(x) \log A_i(x) - A_i(x)\}$ may be interpreted as the degree to which x does not belong to the ith cluster and thus the membership of x to the noise cluster. If the distance of x to the cluster centre c_i is greater than $\sqrt{\delta_i}$, then the minimization of $J_c(\cdot)$ enforces a small value of $A_i(x)$ and a large value of membership of x_i to the noise cluster. Therefore, one of the strategy may be to set δ_i equal to the distance of nearest cluster centre from c_i, i.e.,

$$\delta_i = \min_j \left\| c_j - c_i \right\|^2.$$

Setting $\partial J_c(A_i(x), c_1, \ldots, c_K)/\partial A_i(x) = 0$, leads to following expression for optimal membership function:

$$RC_i(x, c_1, \ldots, c_K) = \exp\left(- \frac{\|x - c_i\|^2}{\delta_i} \right). \tag{A3}$$

Fig. A1(b) shows the membership functions (A3) for three different clusters with centres at 20, 50 and 80. Although the membership functions shape is convex, but the assignment of membership values to a cluster is totally independent of the location of other cluster centres. As an illustration, consider a point 50 (the centre of second cluster) in Fig. A1(b), that has been assigned the membership values more than 0.30 to the first and third clusters. This point, however, in case of FCM was assigned zero memberships to the first and third clusters (that was obviously more natural assignment).

The advantages of both methods can be combined by adopting a mixed clustering criterion. Assume that the membership function A_i has two components A_{1i} and A_{2i} such that

$$A_i = \frac{A_{1i}^{\tilde{m}}}{2} + \frac{A_{2i}}{2},$$

where A_{1i}, A_{2i} minimizes following constrained objective function:

$$\sum_{x \in X} \sum_{i=1}^{K} \left[\left(A_{1i}^{\tilde{m}}(x) + A_{2i}(x) \right) \|x - c_i\|^2 + \{1 + A_{2i}(x) \log A_{2i}(x) - A_{2i}(x)\} \delta_i \right], \ \sum_{i=1}^{K} A_{1i}(x) = 1.$$

Now, A_{1i} will be given by Eq. (A2) and A_{2i} by Eq. (A3). Thus,

$$A_i(x, c_1, \ldots, c_K) = \frac{|\mathrm{FCM}_i(x, c_1, \ldots, c_K)|^{\tilde{m}}}{2} + \frac{RC_i(x, c_1, \ldots, c_K)}{2}. \tag{A4}$$

Fig. A1(c) shows a one-dimensional example of above membership function for three clusters (with centres at 20, 50 and 80) for $\tilde{m} = 2$. As seen from Fig. A1(c), expression (A4) provides a compromise between non-convex nature of Eq. (A2) and independent behavior of Eq. (A3) with respect to the location of cluster centres. If we denote

$$G_i(x, c_1, \ldots, c_K) = \frac{A_i(x, c_1, \ldots, c_K)}{\sum_{i=1}^{K} A_i(x, c_1, \ldots, c_K)}, \tag{A5}$$

then the output of Sugeno type fuzzy inference system for this shape of membership functions follows from Eq. (A1) as

$$F_s(x) = \sum_{i=1}^{K} \alpha^i G_i(x, c_1, \ldots, c_K).$$

Introduce following notations: $\alpha = [\alpha^i]_{i=1,\ldots,K} \in R^K$, $\theta = [c_1^{\mathrm{T}}, \ldots, c_K^{\mathrm{T}}]^{\mathrm{T}} \in R^{Kn}$, and $G(x, \theta) = [G_i(x, \theta)]_{i=1,\ldots,K} \in R^K$. Now, the above expression can be rewritten as

$$F_s(x) = G^{\mathrm{T}}(x, \theta)\alpha. \tag{A6}$$

As indicated by above expression, the output of a clustering based fuzzy model is linear in consequents α and nonlinear in cluster centres θ.

Appendix B. A Gauss–Newton based algorithm

Given input–output data pairs $\{x(j), y(j)\}_{j=0}^{N}$, to compute the parameters

$$\theta_j = \arg\min_{\theta} \left[\frac{\left[y(j) - G^{\mathrm{T}}(x(j), \theta)\alpha_{j-1} \right]^2}{1 + \mu \|G(x(j), \theta)\|^2} + \mu_\theta^{-1} \left\| \theta - \theta_{j-1} \right\|^2 \right]$$
$$= \arg\min_{\theta} \|r(\theta)\|^2, \ \text{where } r(\theta) = \begin{bmatrix} \left[\frac{\left[y(j) - G^{\mathrm{T}}(x(j), \theta)\alpha_{j-1} \right]^2}{1 + \mu \|G(x(j), \theta)\|^2} \right]^{1/2} \\ \left(\mu_\theta^{-1} \right)^{1/2} \left(\theta - \theta_{j-1} \right) \end{bmatrix},$$

$$\alpha_j = \alpha_{j-1} + \frac{\mu G\left(x(j), \theta_j\right) \left[y(j) - G^{\mathrm{T}}\left(x(j), \theta_j\right) \alpha_{j-1} \right]}{1 + \mu \left\| G(x(j), \theta_j) \right\|^2},$$

we suggest a Gauss–Newton based algorithm. The algorithm consists of following steps:

(1) Choose initial guess about cluster centres θ_{-1}, number of maximum epochs $E_{\max}$, $\alpha_{-1} = 0$, epoch count $\mathrm{EC} = 0$, and data index $j = 0$.
(2) If $\mathrm{EC} < E_{\max}$,
  (a) if $j \le N$,
    (i) define

$$r(\theta) = \begin{bmatrix} \left[\frac{\left[y(j) - G^{\mathrm{T}}(x(j), \theta)\alpha_{j-1} \right]^2}{1 + \mu \|G(x(j), \theta)\|^2} \right]^{1/2} \\ \left(\mu_\theta^{-1} \right)^{1/2} \left(\theta - \theta_{j-1} \right) \end{bmatrix}$$

and let $s^*(\theta)$ be the unique solution of following linear least-squares problem:

$$s^*(\theta) = \arg\min_{s} \left[\|r(\theta) + r'(\theta)s\|^2 \right],$$

where $r'(\theta)$ is the Jacobian matrix of vector r with respect to θ, determined by the method of finite-differences. The Jacobian $r'(\theta)$ is a full rank matrix, as a result of using regularization.

(ii) Compute $\theta_j = \theta_{j-1} + s^*(\theta_{j-1})$.

(iii) Compute

$$\alpha_j = \alpha_{j-1} + \frac{\mu G\big(x(j),\theta_j\big)\big[y(j) - G^{\mathrm{T}}\big(x(j),\theta_j\big)\alpha_{j-1}\big]}{1+\mu\big\|G(x(j),\theta_j)\big\|^2}.$$

(iv) $j := j+1$ and go to step 2(a).

(b) $\mathrm{EC} := \mathrm{EC}+1$, $\alpha_{-1} := \alpha_N$, $\theta_{-1} := \theta_N$, $j = 0$, and go to step 2.

References

[1] Janos Abonyi, Robert Babŭska, Ferenc Szeifert, Modified Gath–Geva fuzzy clustering for identification of Takagi–Sugeno fuzzy models, IEEE Trans. Syst. Man Cybern. B (October 2002) 612–621.

[2] T. Aoyama, H. Ichikawa, Obtaining the correlation indices between drug activity and structural parameters using a neural network, Chem. Pharm. Bull. 39 (2) (1991) 372–378.

[3] T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to quantitative structure–activity relationship (QSAR) analysis, J. Med. Chem. 33 (1990) 2583–2590.

[4] R. Babuška, Fuzzy Modeling for Control, Kluwer Academic Publishers, Boston, 1998.

[5] J.C. Bezdeck, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[6] Frank R. Burden, Martyn G. Ford, David C. Whitley, David A. Winkler, Use of automatic relevance determination in QSAR studies using Bayesian neural networks, J. Chem. Inf. Comput. Sci. 40 (2000) 1423–1430.

[7] Frank R. Burden, David A. Winkler, New QSAR methods applied to structure–activity mapping and combinatorial chemistry, J. Chem. Inf. Comput. Sci. 39 (1999) 236–242.

[8] Frank R. Burden, David A. Winkler, Robust QSAR models using Bayesian regularized neural networks, J. Med. Chem. 42 (1999) 3183–3187.

[9] A. González, R. Pérez, Completeness and consistency conditions for learning fuzzy rules, Fuzzy Set. Syst. 96 (1998) 37–51.

[10] P. Gramatica, P. Pilutti, E. Papa, Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training–test sets and consensus modeling, J. Chem. Inf. Comput. Sci. 44 (2004) 1794–1802.

[11] Z. Guo, Structure–activity relationships in medicinal chemistry: development of drug candidates from lead compounds, Pharmacochem. Libr. 23 (1995) 299–320.

[12] C. Hansch, T. Fujita, ρ–σ–π Analysis. A method for the correlation of biological activity and chemical structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.

[13] F. Herrera, M. Lozano, J.L. Verdegay, Generating fuzzy rules from examples using genetic algorithms, in: Proceedings of the Fifth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'94), Paris, France, July 1994, pp. 675–680.

[14] H. Ichikawa, T. Aoyama, How to see characteristics of structural parameters in QSAR analysis: descriptor mapping using neural networks, SAR QSAR Environ. Res. 1 (1993) 115–130.

[15] R.C. Jackson, Update on computer-aided drug design, Curr. Opin. Biotechnol. 6 (1995) 646–651.

[16] J.S.R. Jang, C.T. Sun, E. Mizutani, Neuro-Fuzzy and Soft Computing; a Computational Approach to Learning and Machine Intelligence, Prentice-Hall, Upper Saddle River, 1997.

[17] J.-S. Roger, A.N.F.I.S. Jang, Adaptive-network-based fuzzy inference systems, IEEE Trans. Syst. Man Cybern. 23 (May 1993) 665–685.

[18] K.L.E. Kaiser, Quantitative structure–activity relationships in chemistry, Can. Chem. News 51 (1999).

[19] M. Kumar, N. Stoll, R. Stoll, An energy-gain bounding approach to robust fuzzy identification, Automatica 42 (5) (May 2006) 711–721.

[20] M. Kumar, R. Stoll, N. Stoll, Robust solution to fuzzy identification problem with uncertain data by regularization. Fuzzy approximation to physical fitness with real world medical data: an application, Fuzzy Optim. Decis. Making 3 (1) (March 2004) 63–82.

[21] M. Kumar, R. Stoll, N. Stoll, A min–max approach to fuzzy clustering, estimation, and identification, IEEE Trans. Fuzzy Syst. 14 (2) (April 2006).

[22] M. Kumar, R. Stoll, N. Stoll, A robust design criterion for interpretable fuzzy models with uncertain data, IEEE Trans. Fuzzy Syst. 14 (2) (April 2006).

[23] J. Liska, S.S. Melsheimer, Complete design of fuzzy logic systems using genetic algorithms, in: Proceedings of the Third IEEE International Conference on Fuzzy Systems, 1994, pp. 1377–1382.

[24] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (1992) 415–447.

[25] D.T. Manallack, D.J. Livingstone, Neural networks in drug discovery: have they lived up to their promise? Eur. J. Med. Chem. 34 (1999) 195–208.

[26] D. Nauck, R. Kruse. A neuro-fuzzy approach to obtain interpretable fuzzy systems for function approximation, in: Proceedings of the IEEE International Conference on Fuzzy Systems 1998 (FUZZ-IEEE'98), Anchorage, AK, May 1998, pp. 1106–1111.

[27] K. Nozaki, H. Ishibuchi, H. Tanaka, A simple but powerful heuristic method for generating fuzzy rules from numerical data, Fuzzy Set. Syst. 86 (1997) 251–270.

[28] J.J. Shan, H.C. Fu, A fuzzy neural network for rule acquiring on fuzzy control systems, Fuzzy Set. Syst. 71 (1995) 345–357.

[29] D. Simon, Design and rule base reduction of a fuzzy filter for the estimation of motor currents, Int. J. Approx. Reason. 25 (October 2000) 145–167.

[30] D. Simon, Training fuzzy systems with the extended Kalman filter, Fuzzy Set. Syst. 132 (December 2002) 189–199.

[31] P. Thrift, Fuzzy logic synthesis with genetic algorithms, in: Proceedings of the Fourth International Conference on Genetic Algorithms, 1991, pp. 509–513.

[32] L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Trans. Syst. Man Cybern. 22 (6) (1992) 1414–1427.

[33] David A. Winkler, Neural networks as robust tools in drug lead discovery and development, Mol. Biotechnol. 27 (2004) 139–167.

[34] S. Wu, M.J. Er, Dynamic fuzzy neural networks – a novel approach to function approximation, IEEE Trans. Syst. Man Cybern. B 30 (2000) 358–364.

[35] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans. Syst. Man Cybern. 3 (January 1973) 28–44.