doi:10.3788/gzxb20174609.0930002

三维荧光光谱结合 HGA-RBF 神经网络在多环芳烃浓度检测中的应用

王书涛,郑亚南,王志芳,马晓晴,王昌冰,程琪

(燕山大学 电气工程学院 河北省测试计量技术及仪器重点实验室,河北 秦皇岛 066004)

摘 要:采用 FS920 荧光光谱仪分析了苯并[k]荧蒽(BkF)、苯并[b]荧蒽(BbF)和两者混合物的荧光特性.结果表明 BkF 的两个荧光峰分别位于 306 nm/405 nm 和 306 nm/430 nm,BbF 的两个荧光峰分别位于 306 nm/410nm 和 306 nm/435 nm.BkF 和 BbF 不同浓度配比及其相互间的荧光干扰,使得混合物荧光特性差异较大,荧光强度和浓度间关系变得复杂.为准确测定混合物中 BkF 和 BbF 的浓度,采用递阶算法优化的径向基神经网络对其进行检测,结果表明 BkF 和 BbF 的平均回收率分别为98.45%和97.71%.该方法能够实现多环芳烃类污染物共存成分的识别和浓度预测.

关键词:光谱学;三维荧光光谱;递阶算法优化的径向基神经网络;多环芳烃;浓度检测

中图分类号:O657.3 **文献标识码**:A **文章编号**:1004-4213(2017)09-0930002-7

Concentration Detection of Polycyclic Aromatic Hydrocarbon Combining Three-dimensional Fluorescence Spectroscopy with HGA-RBF Neural Network

WANG Shu-tao, ZHENG Ya-nan, WANG Zhi-fang, MA Xiao-qing, WANG Chang-bing, CHENG Qi

(*Institute of Electrical Engineering, Measurement Technology and Instrumentation Key Lab of Hebei Province, Yanshan University, Qinhuangdao, Hebei* 066004, *China*)

Abstract: Three-dimensional excitation-emission matrix fluorescence spectroscopy of Benzo [k] Fluoranthene (BkF), Benzo [b] Fluoranthene (BbF), and a mixture of these two substances were analyzed with FS920 fluorescence spectrometer. The results show that the fluorescence peaks of BkF can be observed at 306 nm/405 nm and 306 nm/430 nm, and the fluorescence peaks of BbF locate at 306 nm/410 nm and 306 nm/435 nm. In the mixture of BkF and BbF, concentration ratio and fluorescence interferences make excitation-emission matrix spectra of mixture change largely. Hence, the relationship between fluorescence intensity and concentration is complicated. In order to determine the concentration of BkF and BbF in mixture, radial basis function neural network optimized by hierarchical genetic algorithm was applied and the average recovery of BkF and BbF are 98.45% and 97.71%, respectively. The results showed that the possibility of the identification and concentration prediction of different components in mixed sample of polycyclic aromatic hydrocarbons.

Key words: Spectroscopy; Three-dimensional fluorescence spectroscopy; Hierarchical Genetic Algorithm Radial Basis Function (HGA-RBF) neural network model; Polycyclic aromatic hydrocarbons;

Foundation item: The National Natural Science Foundation of China (No. 61471312), Natural Science Foundation of Hebei (No. F2017203220)

First author: WANG Shu-tao (1978—), male, professor, Ph.D. degree, mainly focuses on photoelectric detection, environmental multi-parameter detection. Email:wangshutao@ysu.edu.cn

Contact author: ZHENG Ya-nan (1991—), female, M.S. degree candidate, mainly focuses on photoelectric-environmental monitoring and signal processing technology. Email:460480254@qq.com

Received: Mar.8, 2017; **Accepted**: Jun.9, 2017

Concentration detection
OCIS Codes: 300.2530; 300.6280; 070.4790;040.1880

0 Introduction

Polycyclic Aromatic Hydrocarbons (PAHs) as a large group of organic contaminants are composed of a plurality of benzene rings, and they are characterized by strong stability, wide distribution and various isomers. PAHs exist widely in air, water and soil, which is principally associated with coal combustion, vehicle emission, industrial complexes and coke production[1]. BkF and BbF, two kinds of the most common PAHs, can cause cancer by cell distortion and mutation[2], which poses a serious threat to human health. Therefore, the concentration of PAHs in food and environment must be strictly controlled.

As the technology develops, a large number of measurement techniques and instruments for PAHs have been developed, such as High Performance Liquid Chromatography (HPLC), Gas Chromatography-Mass Spectrometry (GC-MS), Liquid Chromatography Mass Spectrometry (LCMS)[3-5], etc. For example, Clara Naccaria used HPLC determine the concentration of PAHs in heat-treated milk samples[6] and Amal Al-Rashdan concluded that the best recovery of PAHs in Toasted Bread is 99.06% by GC-MS[7]. However, there are some shortcomings in those methods, for example the long analysis time, the high cost, and the complicated operation. Three-dimensional Excitation-Emission Matrix (EEM) fluorescence spectroscopy is a rapid and sensitive technology based on Lambert-Beer Law. Thus, because of its high sensitivity, rapid detection, good selectivity, convenient operation[8], EEM fluorescence spectroscopy is widely applied in many research fields involving characteristic analysis, substance identification and trace determination, etc[9].

Radial Basis Function (RBF) neural network is a novel and efficient feed forward neural network which is widely used in pattern recognition, error diagnosis and data prediction. However, RBF algorithm has the disadvantage of being easy to fall into local optimum and low training efficiency[10]. In order to overcome these shortcomings, Hierarchical Genetic Algorithm (HGA) is proposed to optimize the RBF neural network meanwhile use HGA-RBF for the detection of PAHs.

1 Experiment

1.1 Equipment and samples

In this study, BkF and BbF are provided by the China National Institute of Metrology. The concentrations of BkF and BbF are both 4.000 μg/mL in Methanol solvent. Firstly, we dilute BkF and BbF into ten samples by adding methanol, respectively, as a consequence the concentrations of the solution are 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000 ng/L. After this, the mixture containing 45 groups can be obtained by mixing BkF and BbF with different concentration ratios.

All EEM spectra are measured by a fluorescence spectrometer (FS920, UK Edinburgh) at a scanning step of 2.0 nm, and the light path structure is shown in Fig.1. EEM spectra are a collection of series of emission spectra over a range of excitation wavelengths. The range of the excitation wavelength and the emission wavelength is based on the fluorescence spectra characteristics of the measured material, metrology, BkF, BbF and the mixture. EEM spectra are collected by scanning emission spectra from 250 nm to 600 nm while varying the excitation wavelength from 200 nm to 450 nm at 2.0 nm increments. And then the fluorescence spectra of BkF, BbF and the mixture are recorded when induced by xenon lamp with the power of 450 W.

BkF and BbF are isomers whose molecular formula is C20H12. As shown in Fig.2, they are composed of four benzene rings. The rigid planar structures prove they have the characteristic of strong fluorescence materials.

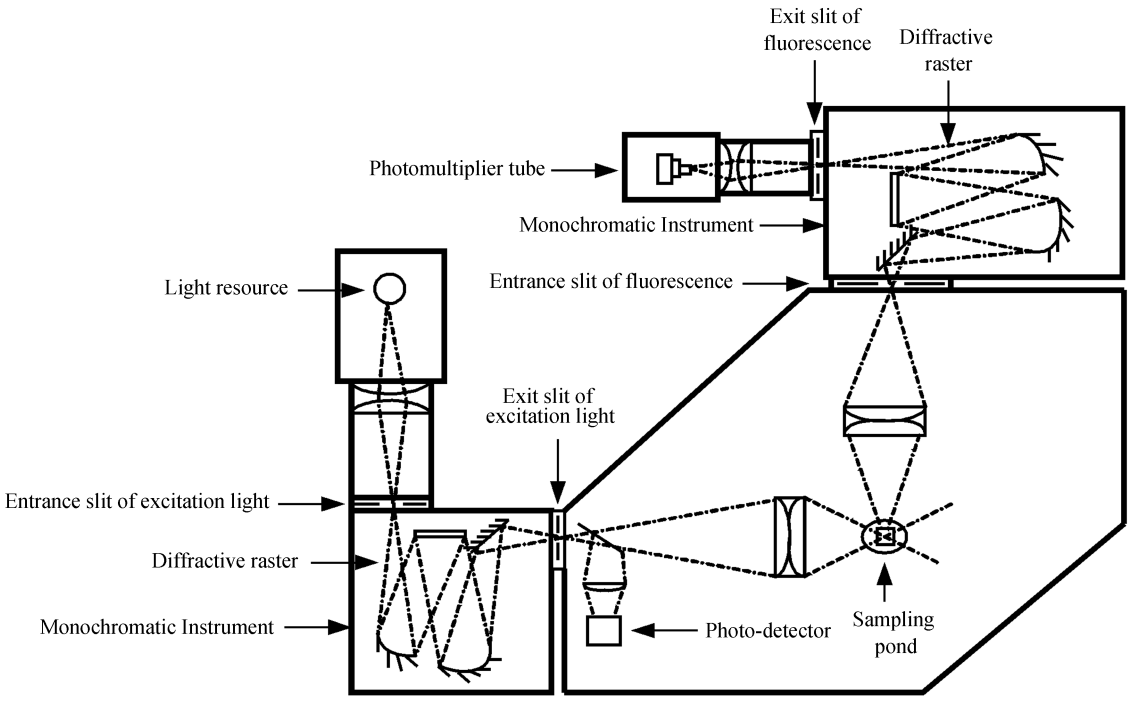
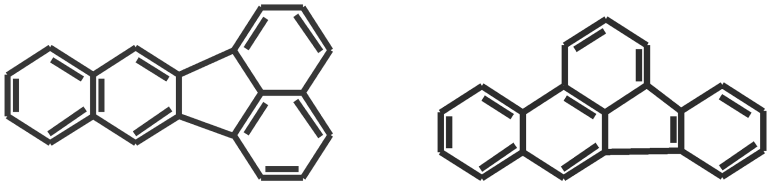


Fig.1 Light path structure of FS920 fluorescence spectrometer



(a) BkF (b) BbF

Fig.2 Molecular structure of BkF and BbF

1.2 Analysis of fluorescence spectra

In order to overcome the problem that the small and medium fluorescence peaks of the three dimensional fluorescence spectra are masked by the large fluorescence peaks, we use contour fluorescence spectra which is connected to a line with the same strength point to represent the fluorescence information. The color of the vertical bar on the right side of the graph represents the intensity of the fluorescence peak.

The contour fluorescence spectra of Methanol are shown in Fig. 3. The main fluorescence peak of methanol is identified at $\lambda_{ex}/\lambda_{em}=300$ nm/350 nm. When the emission wavelength ranges from 500 nm to 600 nm, there is serious Rayleigh scattering which will give the fluorescence spectra a great influence and even distort the fluorescence spectra . In order to avoid the interference from Rayleigh scattering , the

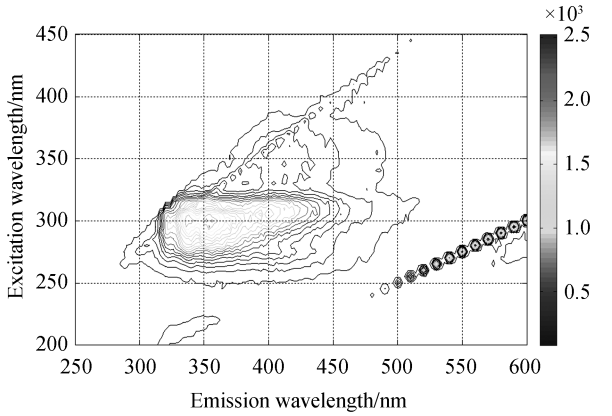


Fig.3 Contour fluorescence spectra of methanol

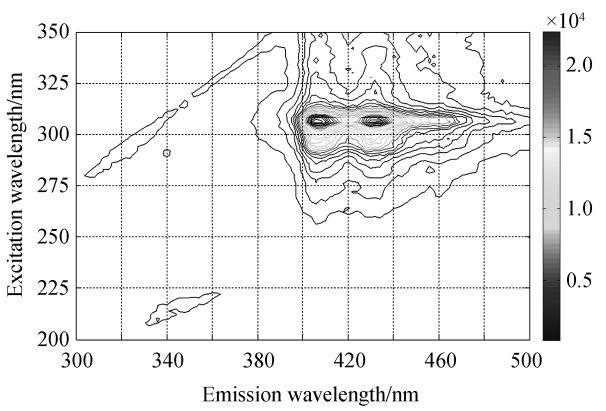


Fig.4 Contour fluorescence spectra of BkF

emission wavelength of the measured materials is limited to 500 nm.

When the concentration of BkF is 4.000 ng/L, the contour map of fluorescence is shown in Fig.4. The X-axis represents the emission spectra from 300 nm to 500 nm, whereas the Y-axis is the excitation wavelength from 200 nm to 350 nm. And the contour lines are shown for each EEM spectra to represent fluorescence intensity. The two fluorescence peaks can be identified at $\lambda_{ex}/\lambda_{em}=306$ nm/405 nm and $\lambda_{ex}/\lambda_{em}=306$ nm /430 nm, so the optical excitation wavelength of BkF is 306 nm.

EEM spectra of BbF which the concentration is 4.000 ng/L is described in Fig. 5. When the excitation wavelength is 240 ~ 360 nm and the emission wavelength is 300 ~ 500 nm, the two fluorescence peaks are located at $\lambda_{ex}/\lambda_{em}=306$ nm /410 nm and $\lambda_{ex}/\lambda_{em}=306$ nm/435 nm. So the emission maximum of BbF can be selected as 440 nm. Then we can identify BkF and BbF easily according to the fluorescence spectra of the two.

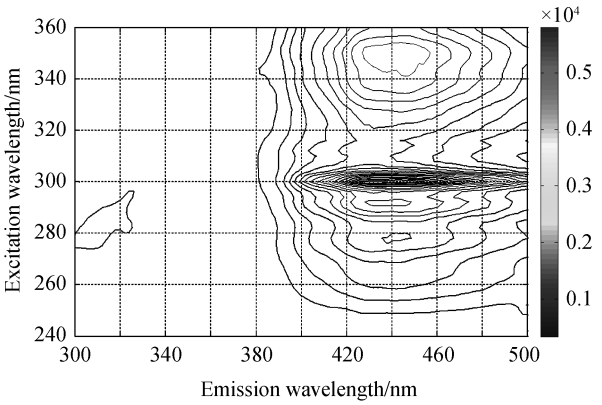
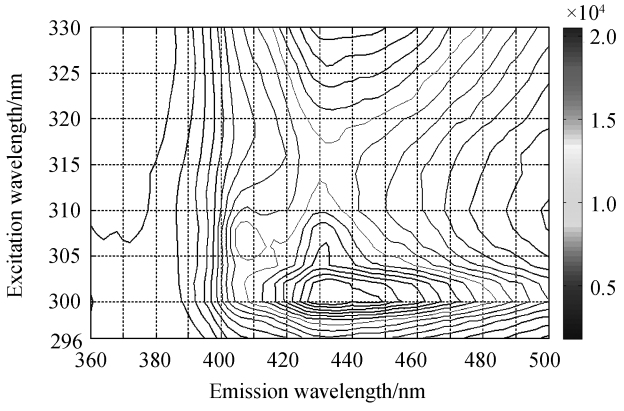
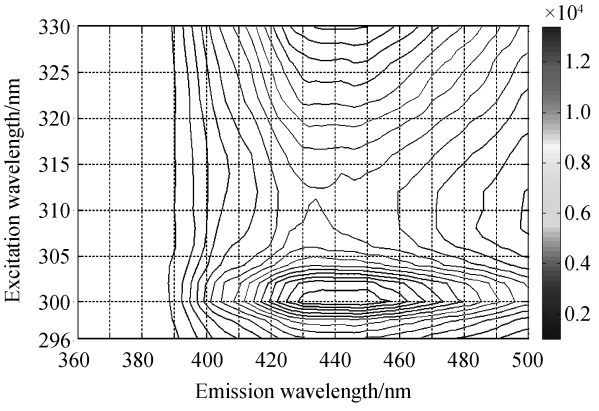


Fig.5 Contour fluorescence spectra of BbF

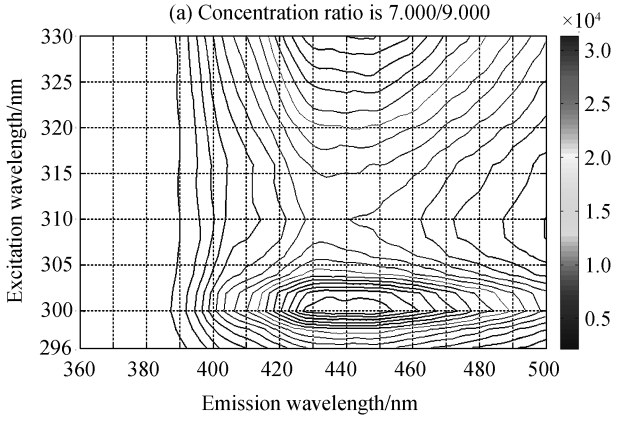
In the mixture of BkF and BbF, the EEM spectra of mixture at various concentration ratio are illustrated in Fig.6. It can be seen from Fig.6 that the effective excitation and emission wavelength are 296~330 nm and 400 ~ 500 nm, respectively. The great changes of spectra can be observed from the samples at various concentration ratios. The mutual interferences coming from the EEM spectra of BkF and BbF and those between the excitation spectra and the emission spectra of the two have serious impacts on the fluorescence characteristics. Hence, the fluorescence spectra of the mixture are very complicated and the relationship between fluorescence intensity and concentration is nonlinear. It's very difficult to calculate the concentration of BkF and BbF in the mixture on the basis of the fluorescence characteristics.



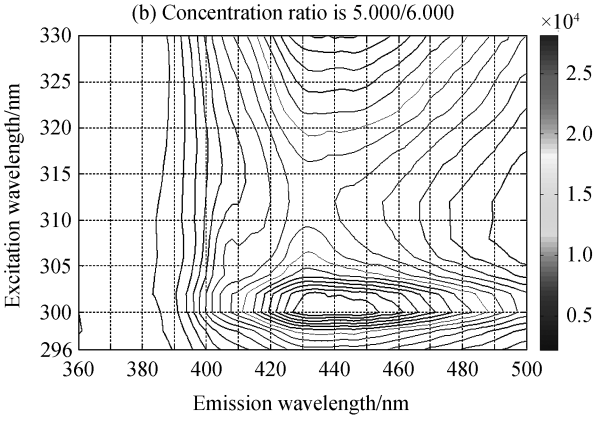
(a) Concentration ratio is 7.000/9.000



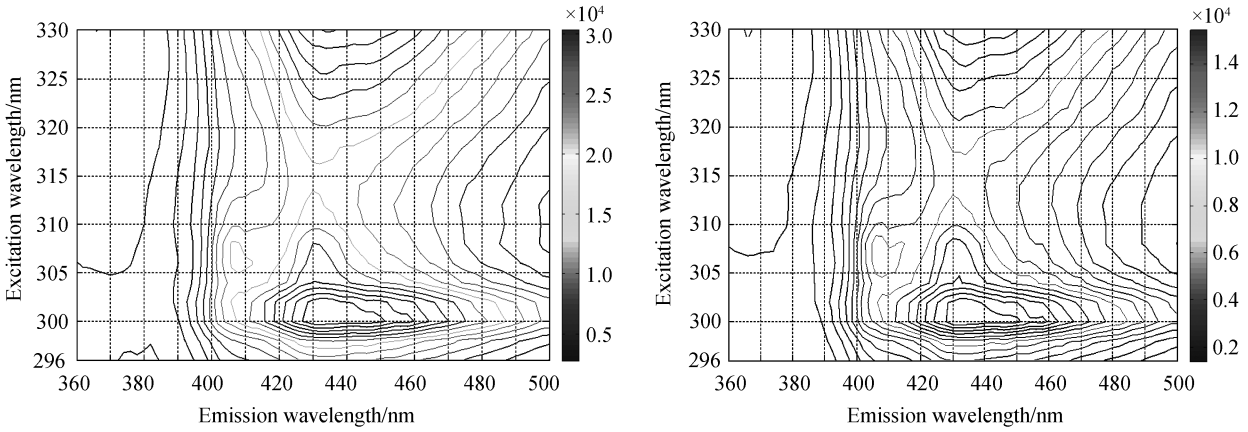
(b) Concentration ratio is 5.000/6.000



(c) Concentration ratio is 3.000/4.000



(d) Concentration ratio is 8.000/10.000



(e) Concentration ratio is 2.000/5.000 (f) Concentration ratio is 6.000/7.000

Fig.6 Contour fluorescence spectra of mixtures with different concentration ratio of BkF to BbF

2 HGA-RBF neural network

Artificial Neural Network (ANN), or called Parallel Distributed Processing Model, is a complicated network system that conforms to the neural structure of the brain and their modular structures. ANN is able to learn complex non-linear relationships between the input and the output, characterized by self-adaptive and self-learning. RBF neural network as one kind ANN is a feed-forward network which is generally composed of three layers, the input layer, the hidden layer and the output layer [11-12]. In the input layer, the number of nodes is the dimension of input data, and the input is

$$\boldsymbol{X}=[X_1,X_2,\cdots,X_N]^{\mathrm{T}} \tag{1}$$

where $\boldsymbol{X}_i(1\leqslant i\leqslant N)$ is the input vector, and N is the number of nodes in the input layer.

The Gaussian function is usually used as the radial basis function, and the radial basis vector of RBF neural network is defined as

$$\boldsymbol{H}=[h_1,h_2,\cdots,h_j,\cdots,h_m]^{\mathrm{T}} \tag{2}$$

$$h_j=\exp\left[-\frac{\|X-C_j\|^2}{2b_j^2}\right]\quad(j=1,2,\cdots,m) \tag{3}$$

where h_j is Gauss function, m is the number of hidden nodes of the network, $\|\cdot\|$ is the Euclidean Norm, b_j is the base width parameter of node j, C_j is the center vector of node j. The transform from the hidden layer to the output layer is a linear map, and the weight vector is

$$\boldsymbol{W}=[w_1,w_2,\cdots w_j\cdots w_m]^{\mathrm{T}} \tag{4}$$

The network output can been obtained using

$$y_m=\sum_{j=1}^{m}w_jh_j \tag{5}$$

Genetic Algorithms (GAs) are well-known search techniques used in computing to find a reasonable solution for optimization problems. GAs inspired by birds foraging behavior are based on principles of natural selection and population genetics[13]. To avoid RBF neural network falling into local optimum, all parameters of the hidden layer are performed by a HGA. HGA is proposed on the basis of the hierarchical structure of chromosomes, including control genes and parameter genes[14-15]. Control genes are composed by the binary numbers, which determine the number of the hidden layer nodes, m. The function

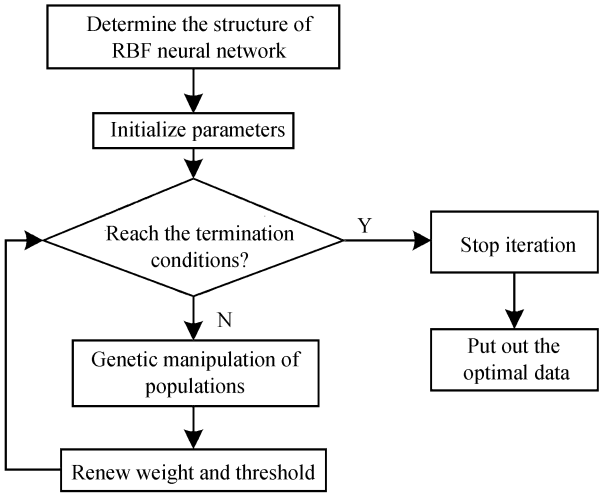


Fig.7 Procedure of HGA-RBF neural network

of parameter genes controlled by control genes is to optimize b_j, c_j and w_j. The procedure of HGA-RBF Neural Network is schematically depicted in Fig.7 [16-17].

The input of RBF neural network is the EEM spectra and it is expressed by

$$\boldsymbol{X}=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ X_{N1} & X_{N2} & \cdots & X_{Np} \end{bmatrix} \tag{6}$$

where $X_{ij}(i=1,2,\cdots,N,j=1,2,\cdots,p)$ is the fluorescence intensity when the excitation wavelength is 301 nm, N is 51, the number of excitation wavelength ranging 400 nm to 500 nm at 2 nm increments, and p is 45, the number of samples. The 30 groups of samples are randomly selected from 45 groups of mixture samples as the training samples and the remaining 15 samples are used for network testing. Therefore the input nodes of the training samples constitute a 51×30 matrix and the testing sample is a 51×15 matrix. The output values of the network are the concentrations of BkF and BaP.

Due to the large difference in the order of magnitude of the fluorescence intensity, the raw data are normalized by

$$X_i'=\frac{X_{\max}-X_i}{X_i-X_{\min}} \tag{7}$$

where X_i' is the normalized value, $X_{\max}$ and $X_{\min}$ are the maximum and minimum value of the X_i, respectively. After the normalization the input values of network are transformed into [0, 1].

3 Prediction of concentration

In the RBF neural network training process, weights and thresholds of neurons are renewed continually with iteration. The network training is not finished until the error of prediction value and actual value reaches a set minimum value, which is set to 10^{-3}. Fig.8 plots the number of epochs and Mean Squared Error (MSE) of network training. The plots suggest that HGA-RBF neural network is much better than RBF neural network. Then network testing is performed.

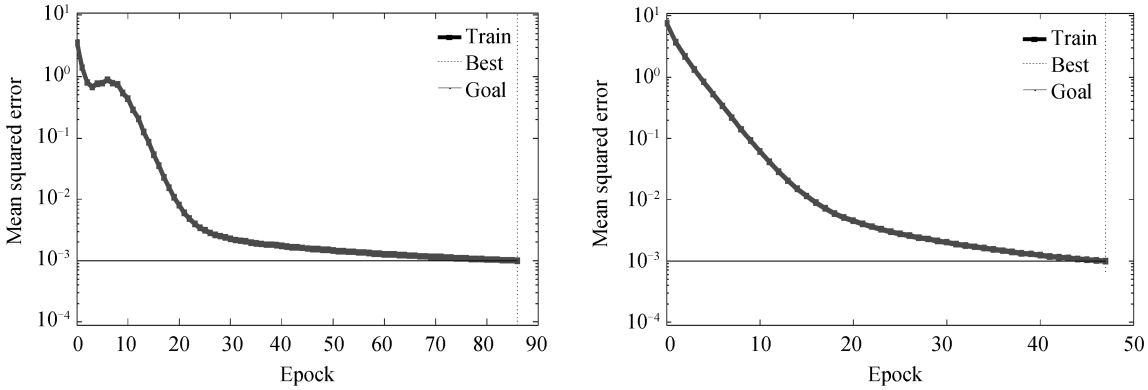


Fig.8 Error curve of the training process

Applying the trained network, we get the prediction results of the mixture, listed in Table 1. The

Table 1 Prediction results of HGA-RBF network

Prediction samples	Actual value/(ng·L^{-1})		Prediction value /(ng·L^{-1})		Recovery rate/%	
	BkF	BbF	BkF	BbF	BkF	BbF
1	1.000	5.000	0.982	4.921	98.20	98.42
2	2.000	7.000	1.963	6.849	98.15	97.84
3	3.000	8.000	2.981	7.795	99.37	97.43
4	4.000	1.000	3.874	0.976	96.85	97.60
5	5.000	6.000	4.898	5.891	99.18	98.18
6	6.000	4.000	5.767	3.954	97.96	98.85
7	7.000	2.000	6.943	1.899	99.19	94.95
8	8.000	3.000	7.887	2.953	98.59	98.43

prediction results can be estimated by the recovery of the measured materials. The average recoveries of BkF and BbF are 98.45% and 97.71%, respectively.

4 Conclusion

In this study, EEM fluorescence spectroscopy is found to be appropriate for identifying BkF and BbF. Two fluorescence peaks are identified in EEM fluorescence spectra of BkF and BbF, respectively. Two fluorescence peaks of BkF are close to those of BbF. The experiment results demonstrate that concentration ratio and fluorescence interferences are significant on EEM fluorescence spectra of the mixture of BkF and BbF. In the work, RBF neural network optimized by HGA is adopted to finish the concentration determination. When compared with RBF neural network, the performance of HGA-RBF neural network is much better than that of RBF neural network. The results indicate that the developed RBF neural network is able to determine the concentration of BkF and BbF in mixture, effectively. As such, it is a useful alternative to PAHs in concentration determination. Furthermore, it is a promising way to determine other fluorescence materials.

References

[1] MELYMUK L, ROBSON M, HELM P A, *et al*. PCBs, PBDEs, and PAHs in toronto air: spatial and seasonal trends and implications for contaminant transport[J].*Science of the Total Environment*, 2012, **429**(6): 272-280.

[2] YIN Dan-dan, WU-Jing, XIE Chao-bo, *et al*. Comparison of three-dimensional fluorescence characteristics of two isomers: phenanthrene and anthrancen[J]. *Spectroscopy and Spectral Analysis*, 2013, **33**(12): 3263.

[3] AYLA D. Factorial design optimization of solid-phase microextraction for high-performance liquid chromatography-ultraviolet spectrometry analysis of polycyclic aromatic hydrocarbons in cigarette filter tar[J]. *Polycyclic Aromatic Compounds*, 2014, **34**(2): 115-134.

[4] ZHAO Yue, HONG Bo, FAN Yue-qing, *et al*. Accurate analysis of polycyclic aromatic hydrocarbons (PAHs) and alkylated PAHs homologs in crude oil for improving the gas chromatography/mass spectrometry performance[J]. *Ecotoxicology & Environmental Safety*, 2014, **100**(1): 242-250.

[5] GROVA N, SALQUEBRE G, SCHRODER H, *et al*. Determination of PAHs and OH-PAHs in rat brain by gas chromatography tandem (triple quadrupole) mass spectrometry[J].*Chemical Research in Toxicology*, 2011, **24**(10): 1653-1667.

[6] NACCARIA C, CRISTANIA M, GIOFREB F, *et al*. PAHs concentration in heat-treated milk samples[J]. *Food Research International*, 2011, **44**(3): 716-724.

[7] ALRASHDAN A, HELALEH M, NISAR A, *et al*. Determination of the levels of polycyclic aromatic hydrocarbons in toasted bread using gas chromatography mass spectrometry[J].*International Journal Analytical Chemistry*, 2010, **2010**(7): 1-8.

[8] MASHIN N I, RAZUVAEV A G, CHERNYAEVA E A, *et al*. Mutual influence of elements in x-ray fluorescence analysis of thin bilayer Ni/Ge-systems[J]. *Journal of Applied Spectroscopy*, 2013, **80**(1): 1-7.

[9] NAGL S, SCHAEFERLING M, WOLFBERIES O S. Fluorescence analysis in microarray technology[J]. *Microchimica Acta*, 2005, **151**(1-2): 1-21.

[10] ALIB H, ADELNM, AMIR H M. A reliable radial basis function neural network model (RBF-NN) for the prediction of densities of ionic liquids[J]. *Journal of Molecular Liquids*, 2017, **231**(4): 462-473.

[11] FATHI V, MONTAZER G A. An improvement in RBF learning algorithm based on PSO for real time applications[J]. *Neurocomputing*, 2013, **111**(6): 169-176.

[12] ZHANG Yi-bo, HE Huan, MENG Qing-fan, *et al*. Application of near infrared reflectance spectroscopy-radial basis function neural network for non-destructive determination of coriolus versicolor[J].*Acta Optica Sinica*, 2010, **30**(12): 3552-3557.

[13] ARAS N. Forecasting residential consumption of natural gas using genetic algorithms[J]. *Energy Exploration & Exploitation*, 2008, **26**(4): 241-266.

[14] ANTONIO C C. A study on synergy of multiple crossover operators in a hierarchical genetic algorithm applied to structural optimisation[J]. *Structural &Multidisciplinary Optimization*, 2009, **32**(2): 117-135.

[15] VASILE M, LESKINEN J, LEE D. Active flow control of airfoil using mesh/meshless methods coupled to hierarchical genetic algorithms for drag reduction design[J]. *Engineering Computations*, 2013. **30**(4): 562-580.

[16] Katrin F, Robert K, Thomas A, *et al*. Normalization and gene p-value estimation: issues in microarray data processing [J].*Bioinformatics & Biology Insights*, 2008, **2**(2): 291.

[17] CHEN C L, HIROSHI I, GADI W, *et al*. Individual A-scan signal normalization between two spectral domain optical coherence tomography devices[J]. *Investigative, Ophthalmology & Visual Science*, 2013, **54**(5): 3463-71.