

Research Article

Spectral Quantitative Analysis Model with Combining Wavelength Selection and Topology Structure Optimization

Qian Wang,¹ Boyan Cai,² Yajie Yu,² and Hui Cao²

¹School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China

²Shaanxi Key Laboratory of Smart Grid and State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

Correspondence should be addressed to Hui Cao; huicao@mail.xjtu.edu.cn

Received 12 June 2016; Accepted 15 September 2016

Academic Editor: K. S. V. Krishna Rao

Copyright © 2016 Qian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Spectroscopy is an efficient and widely used quantitative analysis method. In this paper, a spectral quantitative analysis model with combining wavelength selection and topology structure optimization is proposed. For the proposed method, backpropagation neural network is adopted for building the component prediction model, and the simultaneousness optimization of the wavelength selection and the topology structure of neural network is realized by nonlinear adaptive evolutionary programming (NAEP). The hybrid chromosome in binary scheme of NAEP has three parts. The first part represents the topology structure of neural network, the second part represents the selection of wavelengths in the spectral data, and the third part represents the parameters of mutation of NAEP. Two real flue gas datasets are used in the experiments. In order to present the effectiveness of the methods, the partial least squares with full spectrum, the partial least squares combined with genetic algorithm, the uninformative variable elimination method, the backpropagation neural network with full spectrum, the backpropagation neural network combined with genetic algorithm, and the proposed method are performed for building the component prediction model. Experimental results verify that the proposed method has the ability to predict more accurately and robustly as a practical spectral analysis tool.

1. Introduction

Spectral quantitative analysis is a nondestructive and fast measurement technique and has been used in a variety of chemical fields [1–3]. The method measures the chemical composition dependent absorption of light that occurs at different wavelengths [4]. Based on the obtained wavelength signals, the spectral quantitative analysis model is built to predict the component concentrations by the regression algorithms [5].

Partial least squares (PLS) is a classical multivariate regression approach for spectroscopy quantitative analysis, and it could handle the multiple correlation among the input wavelength signals [6]. Nevertheless, PLS is a linear regression algorithm essentially [7], and the nonlinearity of wavelength signals may be generated by the instrument variation and the analyte characteristics [8]. To deal with the nonlinear factors, neural network is always adopted for spectral model. Neural network could approximate any function by some

simple interconnected processing units whose structure is inspired by animal brains [9, 10]. Backpropagation neural network (BPNN), as a popular neural network, uses the mean square error and the gradient descent for modifying the connection weights of the neurons [11]. The topology structure of BPNN is usually determined by the human experience [12] and may affect the model effectiveness. That may be one reason why three-layer BPNN is widely used [13–16].

Moreover, the spectral instrument usually records a large number of spectral wavelength signals and the regression model is generally performed based on the obtained wavelengths. However, not all of the obtained wavelengths have the useful information, and the wavelengths without any critical information would corrupt the prediction model [17, 18]. Therefore, wavelengths selection is a vital process for spectral quantitative analysis, and the goal of wavelengths selection is to determine a subset of the obtained spectral wavelengths that could generate the smallest possible errors

of the regression models [19, 20]. Some statistical techniques have been adopted for the wavelength selection, and the importance of each wavelength could be estimated according to the statistical features of the prediction model [21, 22]. Uninformative variable elimination (UVE) is proposed to eliminate the wavelengths that do not contain much information for analyte prediction than random variables [23]. Although UVE is better than the statistical wavelength selection method [24, 25], the effectiveness of UVE would be affected by the quality of the random variables and the selection result is scattered throughout the spectrum [26].

BPNN could be optimized by the heuristic algorithm, and BPNN based on genetic algorithm (GA-BPNN) is proposed for determining the initial connection weights and the thresholds in a fixed topology structure [27, 28]. Furthermore, the wavelength selection could seem as a combinatorial problem; the genetic algorithm combined with PLS (GA-PLS) is presented, where GA finds the optimal subset of wavelengths associated with the PLS model [29, 30]. Because the model structure should be determined based on the number of the selected wavelengths, wavelength selection and the topology structure of BPNN would be optimized simultaneously, that is, a hybrid optimization problem. Evolutionary programming (EP) having no fixed structure outperforms with GA and is suitable for the hybrid optimization problem [31, 32]. Like GA, EP has the crossover operation and the mutation operation. However, the crossover operation of EP is limited by the chromosome form for the hybrid optimization problem, that may result in the side-effect, and EP without the crossover process would not reduce the search efficiency [33]. Furthermore, EP generally has the static mutation probability, and EP may fall into the local minima, which is similar to other searching algorithms [34].

In this paper, a spectral quantitative analysis model with combining wavelength selection and topology structure optimization is proposed. For the proposed method, BPNN is adopted for building the component prediction model, and the simultaneous optimization of the wavelength selection and the topology structure of BPNN is realized by the nonlinear adaptive evolutionary programming (NAEP). The hybrid chromosome in binary scheme of NAEP has four fragments, which represent the number of the hidden layers of BPNN, the number of neurons in each hidden layer, the selection of spectral wavelengths, and two adaptive parameters of the mutation probability of NAEP, respectively. Hence, a chromosome represents an optimization plan. NAEP only has the mutation operation for the next generation, and the mutation probability of each chromosome is updated by a nonlinear equation with considering two adaptive parameters and the fitness values. For the initial generation of NAEP, each chromosome is encoded randomly. BPNN is performed on the calibration set based on different optimization plans represented by different chromosomes. The root-mean-squares error of cross-validation (RMSECV) is the fitness function; namely, the lower the RMSECV, the better the chromosome. The better parent chromosomes would be put into the next generation. The mutation probabilities of other chromosomes are updated according to the latest evaluation results, and the chromosomes are evolved only by the mutation operation.

The evolution process of NAEP terminates based on the stop condition. The chromosome with lowest fitness value is the final result; namely, the selected wavelength and the corresponding topology structure of BPNN are determined. Two real flue gas datasets are employed in the experiments. The effectiveness of PLS, BPNN, GA-BPNN, UVE, GA-PLS, and the proposed method is compared.

The remainder of this paper is organized as follows. In Section 2, The related methods are demonstrated. In Section 3, the proposed method is presented. In Section 4, the experimental results are discussed. Section 5 concludes the paper.

2. The Related Methods

2.1. PLS. For PLS, X represents the input wavelength signals, and the component Y can be expressed by

$$Y = XM + \varepsilon, \quad (1)$$

where M is the matrix of regression coefficients and ε is the error vector.

It assumes that a small number of the latent variables are refined by linear combinations of the vectors of X . Then (1) can be transformed to

$$Y = TC + \varepsilon, \quad (2)$$

$$T = XQ(U^TQ)^{-1},$$

where the matrix T is corresponding to the latent variables and C is the regression coefficients vector.

For T , X is the input matrix, Q is the matrix of weight loading representing the correlation between Y and X , and P is the matrix indicating the influence of X .

2.2. BPNN. BPNN connects the input layer and the output layer by one or more hidden layers. For spectroscopy quantitative analysis, the wavelengths are the signals of the input layer and the component concentration is the signal of the output layer [35]. A neuron is an activation function which is described by the tansig function, and the transfer function of the output layer is a purelin function [36]. The training process BPNN has the information forward-propagation algorithm and the error backpropagation training algorithm [11]. For the information forward-propagation algorithm, the values of each layer are calculated based on the activation function and the values of the previous layer. For the error backpropagation training algorithm, the error is propagated from the output layer to the input layer, and the weights are regulated by feedback. The modification of the weights and the offset values makes the actual output be closer to the expected output.

2.3. GA-BPNN. For GA-BPNN, the chromosome is encoded for the initial connection weights and the thresholds of a fixed topology structure. The individuals of the father generation are generated randomly. Then, BPNN is performed based on the information represented by each individual, and the fitness value of each individual is evaluated. Some individuals

are reserved for the next generation during the selection operation and their fitness values have a great impact on the reserve probability. Some new individuals are obtained by the crossover operation and the mutation operation. The reserved individuals and the new individuals form the next generation, and the iteration procedure is running constantly until the program satisfies its requirements. After the initial weights are determined, the backpropagation training method is used to adjust the final weights of BPNN.

2.4. UVE. For UVE, an auxiliary matrix containing random noise is generated firstly and it has the same size as the input matrix. Then, the input matrix is combined with the auxiliary matrix to form the combination matrix, which has twice as many wavelength signals as the input matrix. PLS is performed on the combination matrix with the leave-one-out procedure. The criterion value of each column of the combination matrix is estimated by the average of its regression vector and its standard deviation. The original wavelength signal whose criterion value is not larger than a threshold is the uninformative wavelength and would be eliminated, where the threshold is set as the maximum value of the ratio of coefficient to the standard deviation of the auxiliary matrix region. Hence, UVE selects the wavelengths swiftly and practically.

2.5. GA-PLS. In the GA-PLS method, the chromosome is coded by a binary string, and the length of a chromosome is equal to the number of all the wavelengths. Each gene of the chromosome is 1 or 0, which indicates that the wavelength is selected or dropped. For GA-PLS, a random population including a number of chromosomes is initialized, and the PLS model is built for each chromosome, where each chromosome represents a solution of wavelength selection. The prediction precision of PLS model is adopted as the fitness values. A new population is generated by the selection, the crossover, and the mutation. The iteration process is repeated and terminates with reaching the condition, which is the number of iterations or a predefined fitness value. Then, the chromosome with the smallest fitness value is the final result of the wavelength selection.

3. The Proposed Method

For the proposed method, BPNN is adopted for building the component prediction model, and NAEP simultaneously optimizes the wavelength selection and the topology structure of BPNN.

For the new individuals of the next generation, NAEP has not the crossover operation and only has the mutation operation in evolving process. The mutation probability (P_m) is updated by

$$P_m = p_1 + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{NTV^2}{2\sigma^2}\right), \quad (3)$$

where p_1 and σ are two adaptive parameters and NTV is the normalized fitness value.

The hybrid chromosome in binary scheme of NAEP has four fragments, which is shown in Figure 1. Fragment

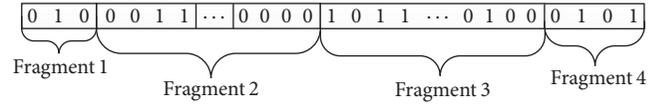


FIGURE 1: The structure of the chromosome.

1 represents the number of the hidden layers (N_{hl}). With considering the model complexity of BPNN, fragment 1 has three genes; namely, the maximum value of N_{hl} is 7. Fragment 2 has twenty-eight genes and every four genes are used for representing the number of neurons in the hidden layer; namely, the maximum value of neurons in each hidden layer is 15. If N_{hl} is the number of hidden layers determined by fragment 1, the values of the genes from the $(4N_{hl} + 1)$ position to the end position of fragment 2 are all zero. Fragment 3 is used for the wavelength selection. The length of fragment 3 is equal to the number of all the wavelengths. Each gene of fragment 3 is 1 or 0, which represents that the corresponding wavelength is selected or dropped. Fragment 4 adopts two parts for representing p_1 and σ , respectively, and each part has two genes. The binary value of part 1 is 00, 01, 10, or 11, which represents that p_1 is 0.05, 0.1, 0.15, or 0.2, respectively. In the same way, the different binary values of part 2 represent that σ is 0.35, 0.45, 0.55, or 0.65.

The steps of the proposed method are described in the following.

Step 1 (initialization and evaluation of fitness values). Each chromosome is initialized randomly. BPNN is performed on the calibration set according to the information represented by each chromosome. For BPNN, the number of the hidden layers and the number of neurons in the hidden layer are determined by fragment 1 and fragment 2, respectively. The number of neurons in the input layer of BPNN is the number of selected wavelengths, which is represented by fragment 3. The number of neurons in output layer of BPNN is 1, which is the prediction value of a component concentration. The fitness function is RMSECV. The lower the RMSECV, the better the chromosome.

Step 2 (selection). The elitists strategy is adopted in the selection operation of the proposed method. The chromosomes in the parent generation are ranked based on the ascending order of the corresponding RMSECV values. The top 10% ranked chromosomes are retained for the next generation. For the elitists strategy, the quality information of population is preserved in the iteration process; namely, the search scope could be guided in the optimum direction and the convergence speed would be improved.

Step 3 (mutation). The mutation operation realizes the diversity of the population. The roulette wheel mechanism is used to perform the proportionally choice for the chromosomes being mutated. The mutation probability value of each chosen chromosome is calculated by (3), where p_1 and σ are determined by fragment 4.

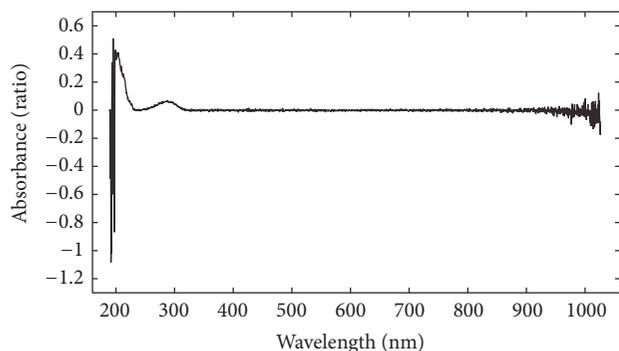


FIGURE 2: The spectrum of Dataset 1.

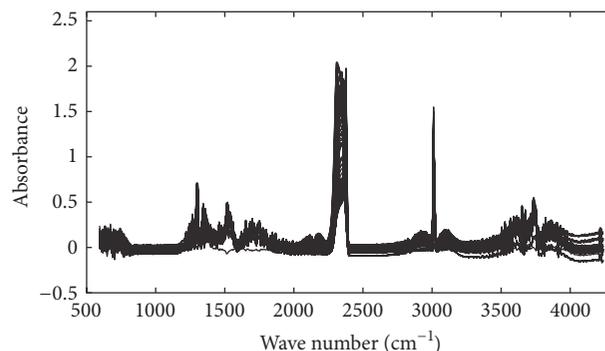


FIGURE 3: The spectrum of Dataset 2.

Step 4 (termination and output). When the number of iterations equals the predefined limit, the proposed algorithm is stopped and the chromosome with the smallest fitness value is output.

According to the best individual obtained by the proposed method, the topology structure of BPNN is determined and the wavelengths are selected. The spectral quantitative analysis model built by the optimized BPNN with the selected wavelengths would have higher accuracy. In the next section, the experiments will be performed to further verify the effectiveness of the proposed method.

4. Experimental Results

Two real flue gas datasets are employed in the experiments, and they are obtained during a combustion process.

4.1. Experimental Datasets

Dataset 1. The dataset is collected during the coal combustion process. It includes 98 samples and each sample consists of a spectrum for a mixture of sulfur dioxide (SO_2), nitrogen monoxide (NO), and nitrogen dioxide (NO_2). The concentration ranges are 0–1500 ppm, 0–3000 ppm, and 0–500 ppm for SO_2 , NO, and NO_2 , respectively. The absorbance spectra are measured by the USB2000t fiber optic spectrometer. The range of spectral number is from 187.87 nm to 1026.97 nm with the resolution of 0.35 nm. Each spectrum contains 2048 wavelengths. There is some noise in wavelengths less than 200 nm. To investigate the robustness of the proposed method, these noise wavelengths are still a part of input data. The spectrum of Dataset 1 is shown in Figure 2.

Dataset 2. The dataset is recorded during the nature gas combustion process by the GASMET DX4000 Fourier transform infrared (FTIR) gas analyzer and includes 106 samples. Each sample consists of different densities of methane (CH_4), carbon monoxide (CO), and carbon dioxide (CO_2). The wavelength range is from 549.44 cm^{-1} to 4238.28 cm^{-1} with an interval of 7.72 cm^{-1} ; namely, each sample has 473 wavelength signals. The concentration ranges of CH_4 , CO, and CO_2 are 0–0.0459 ppm, 0–0.4083 ppm, and 0–0.3818 ppm, respectively. The spectrum of Dataset 2 is shown in Figure 3.

4.2. Experimental Procedure. In the experiments, PLS, BPNN, GA-BPNN, UVE, GA-PLS, and the proposed method are performed on the datasets. Each dataset is separated into a calibration set and a validation set with the shutter grouping strategy [11]. A fifth of the total samples would be put into the validation dataset and the rest of the samples are put into calibration dataset. The calibration set is used to build the prediction model, and the validation set is used for evaluating the effectiveness of the model.

For BPNN, three layers are used and they are the input layer, the hidden layer, and the output layer. The number of neurons in the input layer equals the number of all wavelengths. The number of neurons in the hidden layer is 15. The number of neurons in the output layer is 1; namely, the output is the component concentration. For PLS, UVE, and GA-PLS, the numbers of latent variables with the smallest RMSECV value are determined [37]. For GA-PLS, the latent variables for each individual in the population need to be redetermined at every iteration. For GA-BPNN, GA-PLS, and the proposed method, the population size is 40 and the fitness function is the RMSECV value. For GA-BPNN and GA-PLS, the crossover probability is 0.6, and the mutation probability is 0.01, which are empirically determined by experiences from the series of the GA-BPNN studies. In the experiments, 10-fold cross-validation is employed for the RMSECV value. Furthermore, the root mean-squared error of prediction (RMSEP), the squared cross-validation correlation coefficient (R_{cv}^2), the squared correlation coefficient of calibration (R_c^2), the squared correlation coefficient of prediction (R_p^2), and the compression ratio (CR) would be taken into account for comparing the predictive ability of different models. CR equals $(N_t - N_s)/N_t \times 100\%$, where N_t is the number of total wavelengths and N_s is the number of the selected wavelengths.

4.3. Results Analysis. The experimental results of Dataset 1 for SO_2 are shown in Table I. Although the CR value of the proposed method is smaller than that of UVE, the RMSEP value of the proposed method is smallest; namely, it is 40.57%, 69.74%, 34.59%, 21.66%, and 18.3% lower than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. UVE may ignore the wavelengths with useful information. The CR value of the proposed method is larger than that of GA-PLS.

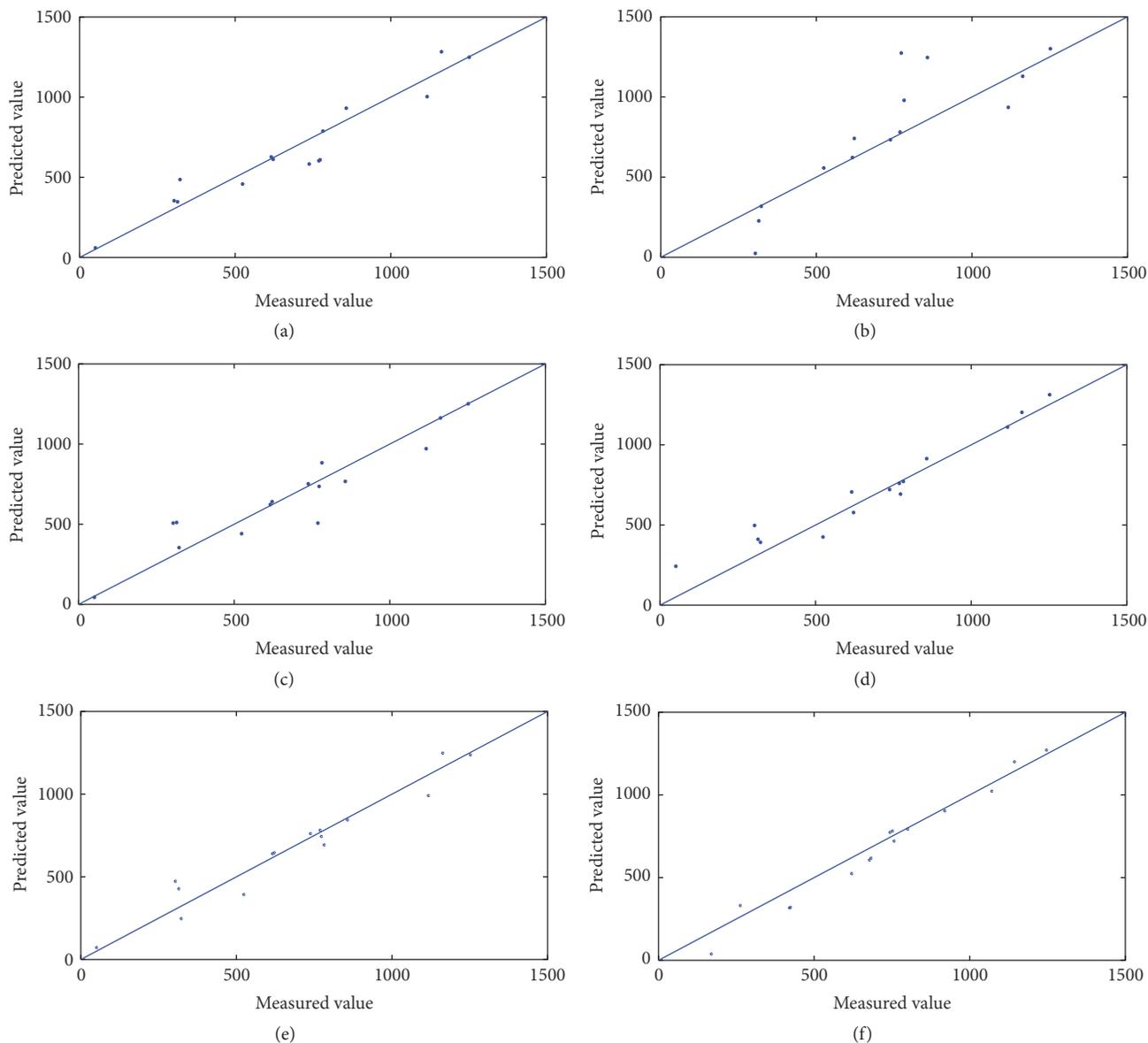


FIGURE 4: The predicted value versus the measured value scatter diagram of different methods for SO_2 . (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

TABLE 1: The experimental results of Dataset 1 for SO_2 .

SO_2	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	88.1731	67.2346	0.9600	0.9810	0
BPNN	219.4631	132.0324	0.9358	0.7939	0
GA-BPNN	66.3185	61.0886	0.9726	0.8700	0
UVE	57.9015	51.0014	0.9853	0.9860	0.8365
GA-PLS	55.1286	48.9536	0.9799	0.9855	0.5149
The proposed method	55.0993	39.9556	0.9834	0.9883	0.5509

The RMSECV value of the proposed method is 55.0993 and is also the smallest. Figure 4 shows the predicted value versus the measured value scatter diagram of different methods for

SO_2 . PLS, UVE, and GA-PLS have better performance than BPNN. GA-BPNN is distributed as close to the diagonal line as PLS, UVE, and GA-PLS. The proposed method has the best result and is distributed more closed to the diagonal line on both sides. Therefore, the prediction ability of the proposed method is higher than the other methods for SO_2 of Dataset 1.

The experimental results of Dataset 1 for NO_2 are shown in Table 2. PLS has the worst performance and its RMSEP value is 406.1298. The RMSECV value of the proposed method is smallest, and the RMSEP value of the proposed method is 68.98%, 68.92%, 40.04%, 36.69%, and 52.08% lower than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. Although the CR value of UVE-PLS is larger than that of the proposed method, other indicators

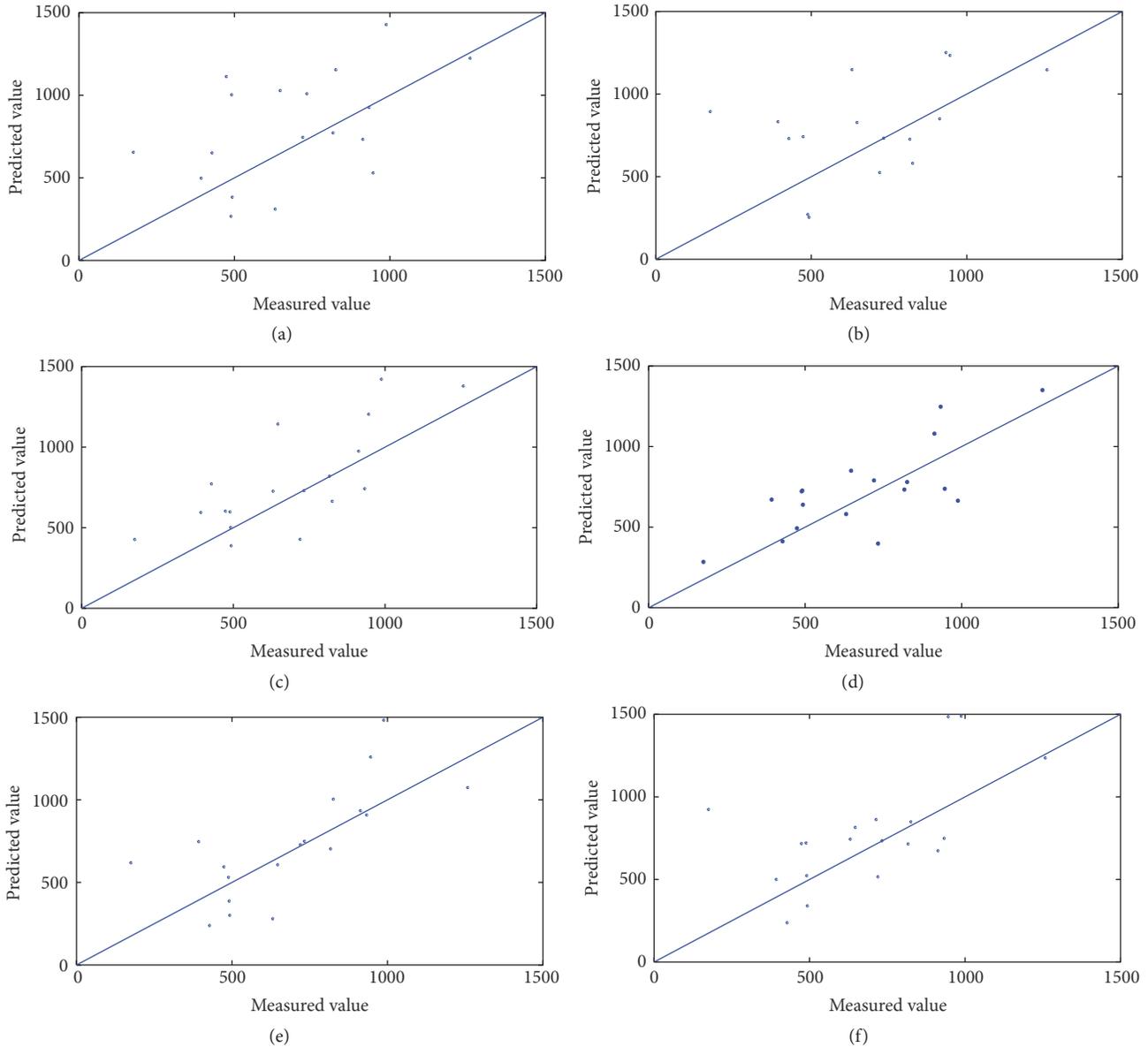


FIGURE 5: The predicted value versus the measured value scatter diagram of different methods for NO_2 . (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

show the proposed method has better performance. Moreover, the CR value of the proposed method is higher than that of GA-PLS. Figure 5 shows the predicted value versus the measured value scatter diagram of different methods for NO_2 . The points of PLS and BPNN spread on both sides of the diagonal line, while the points of GA-PLS, UVE, and GA-BPNN are more close to the line. The proposed method has the best performance as the points are closest to the line. Furthermore, the proposed method is the most robust of all as the points are plotted with roughly the same distance to the line. Hence, the proposed method has higher prediction precision for NO_2 of Dataset 1.

Table 3 demonstrates the experimental results of Dataset 1 for NO. The RMSECV value of the proposed method is smallest. The RMSEP value of BPNN is largest. The RMSEP

TABLE 2: The experimental results of Dataset 1 for NO_2 .

NO_2	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	424.2092	406.1298	0.4910	0.2198	0
BPNN	488.1834	405.3482	0.6651	0.6515	0
GA-BPNN	213.2712	210.1223	0.8314	0.7028	0
UVE	183.9214	198.9831	0.9001	0.7713	0.9818
GA-PLS	249.2341	262.9014	0.8199	0.6310	0.5039
The proposed method	128.5434	125.9855	0.8813	0.8854	0.5218

value of the proposed method is 33.34%, 85.76%, 8.17%, 7.20%, and 41.93% smaller than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. R_{cv}^2 and R_p^2 of the

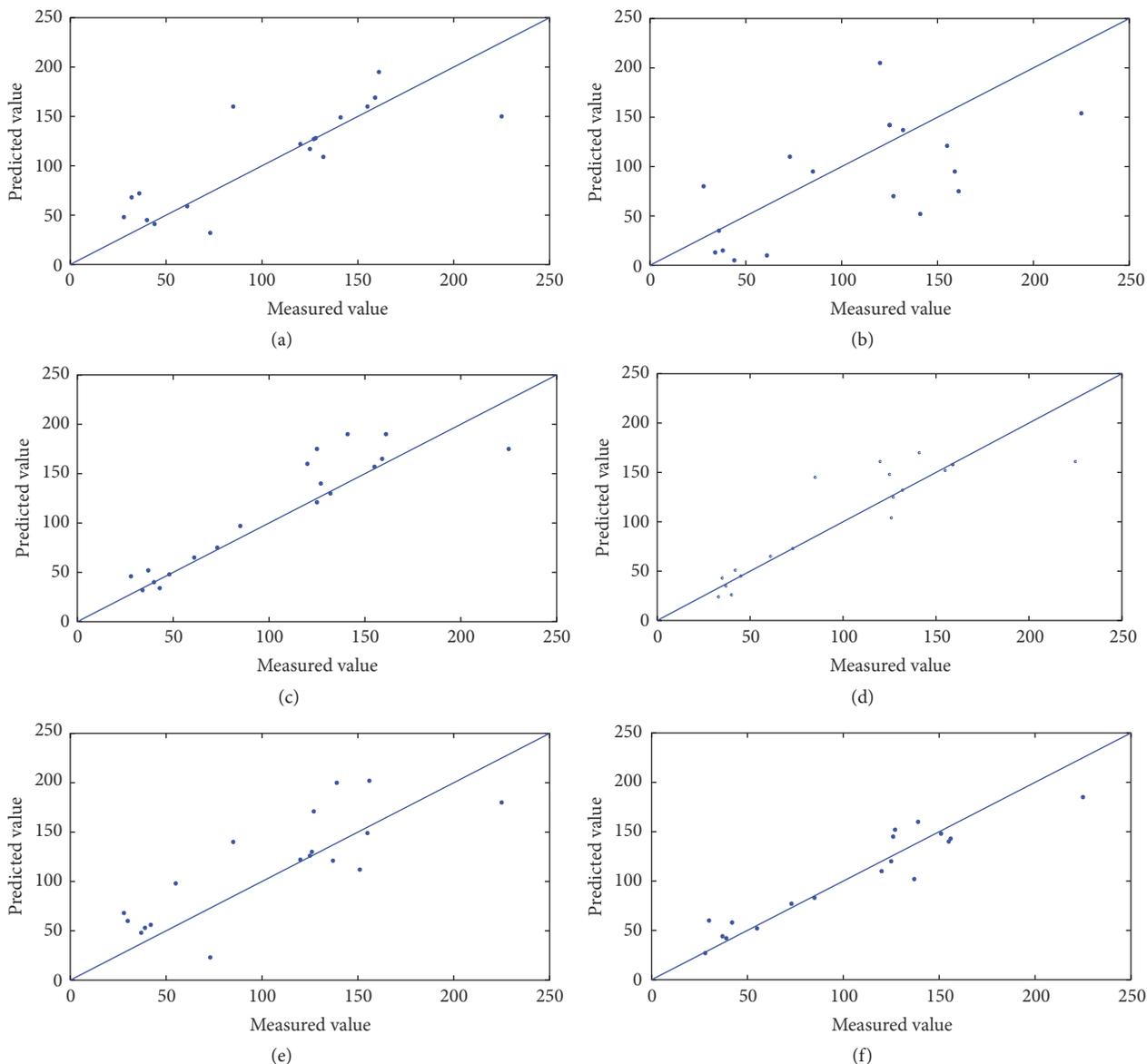


FIGURE 6: The predicted value versus the measured value scatter diagram of different methods for NO. (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

TABLE 3: Analytical results for NO.

NO	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	45.2984	33.1924	0.5319	0.6906	0
BPNN	145.7634	155.3482	0.6651	0.6165	0
GA-BPNN	31.8127	24.0923	0.8545	0.7314	0
UVE	25.6659	23.8409	0.8383	0.8790	0.8545
GA-PLS	24.7231	38.1008	0.8804	0.6172	0.5059
The proposed method	21.9856	22.1235	0.9882	0.9888	0.5926

proposed method are the largest. The predicted value versus the measured value scatter diagram of different methods for NO is shown in Figure 6. The PLS, BPNN, and GA-PLS do not obtain the good results as many points are far away from

the diagonal line. The points of GA-BPNN, UVE, and the proposed method are more close to the line. Some of points of the proposed method are on the line, and the average distance between the points and the line of the proposed method is smaller than those of GA-BPNN and UVE. Thus, the performance of the proposed method is best for the NO of Dataset 1.

In the same way, the analytical results for Dataset 2 are discussed in the following. The experimental results of Dataset 2 for CH_4 are shown in Table 4. Although the CR value of UVE is largest, the performance of UVE is worst because the RMSEP value of UVE is largest. The RMSEP value of the proposed method is 23.39%, 37.01%, 20.95%, 44.70%, and 21.68% smaller than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. The RMSECV value

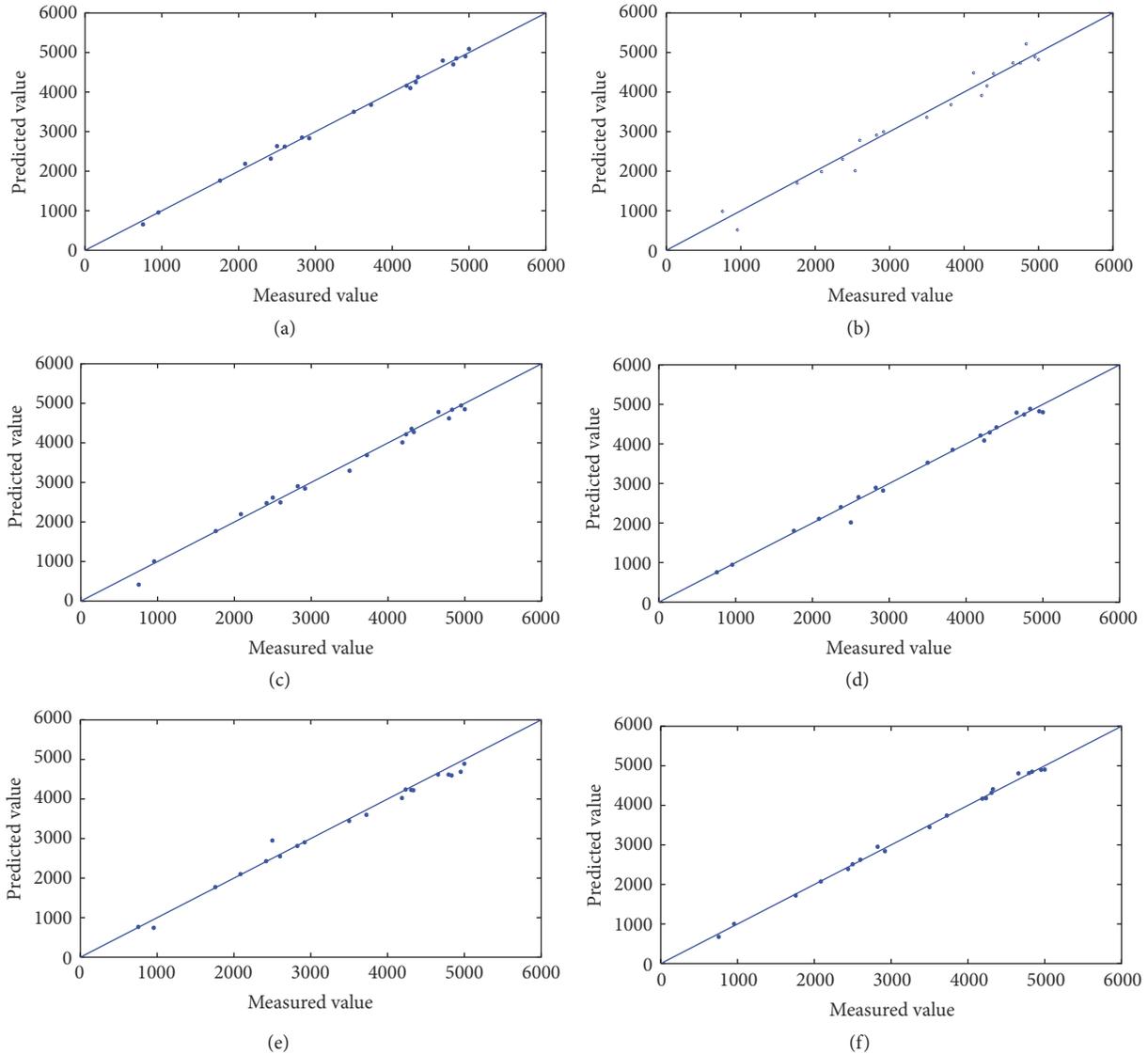


FIGURE 7: The predicted value versus the measured value scatter diagram of different methods for CH_4 . (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

of the proposed method is also smallest. Figure 7 shows the predicted value versus the measured value scatter diagram of different methods for CH_4 . PLS, BPNN, and GA-BPNN have less points which are right located on the diagonal line. There are one or two points of UVE and GA-PLS that are not close to the line. Most of points of the proposed method are on the diagonal line. Therefore, the accuracy and the robustness of the proposed method for the CH_4 of Dataset 2 are validated.

Table 5 lists the experimental results of Dataset 2 for CO. Although the RMSECV of UVE is smaller than that of the proposed method, the RMSEP of the proposed method is smallest. The RMSEP of the proposed method is 64.69%, 77.61%, 59.74%, 9.44%, and 38.89% smaller than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. Moreover, the CR value of the proposed method is larger than that of GA-PLS. Figure 8 shows the predicted value versus

TABLE 4: Analytical results for CH_4 .

CH_4	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	115.4345	99.5357	0.9939	0.9956	0
BPNN	101.0567	121.0567	0.9938	0.9752	0
GA-BPNN	97.2312	96.4683	0.9861	0.9844	0
UVE	119.4832	137.8875	0.9935	0.9914	0.8647
GA-PLS	81.8355	97.3635	0.9970	0.9919	0.5285
The proposed method	79.4571	76.2553	0.9978	0.9920	0.5327

the measured value scatter diagram of different methods for CO. The points of PLS, BPNN, and GA-BPNN spread on both sides of the diagonal line, while the points of UVE, GA-PLS, and the proposed method are more close to the diagonal

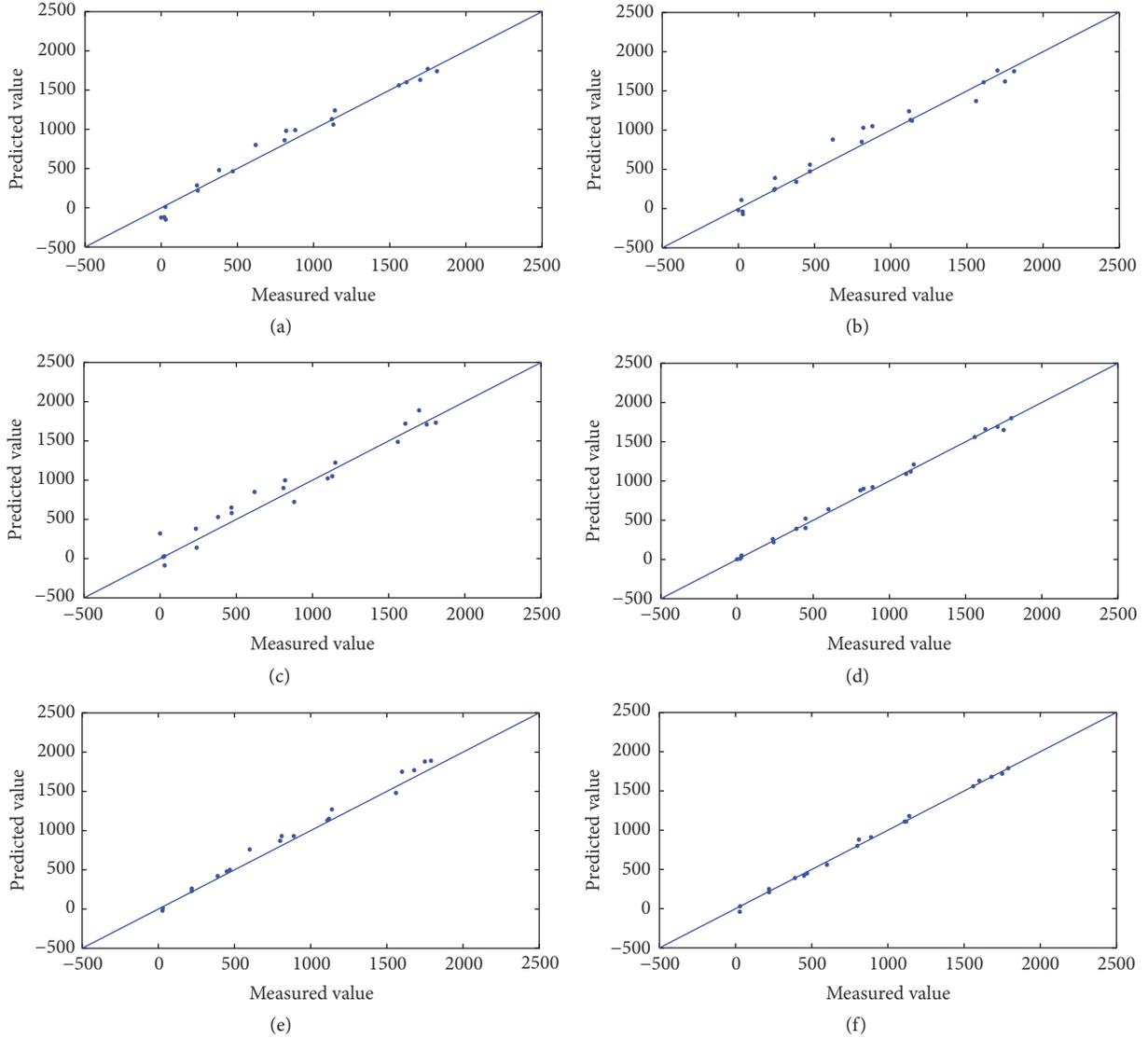


FIGURE 8: The predicted value versus the measured value scatter diagram of different methods for CO. (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

TABLE 5: Analytical results for CO.

CO	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	78.7763	77.12223	0.9842	0.9855	0
BPNN	150.2814	121.6186	0.9315	0.8482	0
GA-BPNN	82.9801	67.6445	0.9732	0.9581	0
UVE	29.0501	30.0698	0.9982	0.9976	0.8879
GA-PLS	41.4748	44.5643	0.9956	0.9950	0.5497
The proposed method	29.1491	27.2319	0.9993	0.9975	0.5530

line. Furthermore, most of points of the proposed method are settled on the diagonal line. Hence, the performance of the proposed method is the best for the CO of Dataset 2.

The experimental results of Dataset 2 for CO₂ are shown in Table 6. The RMSEP of BPNN equals 57.7472 which is largest. The RMSEP of the proposed method is 36.17%,

38.68%, 36.66%, 26.08%, and 17.32% smaller than that of PLS, BPNN, GA-BPNN, UVE, and GA-PLS, respectively. Figure 9 shows the predicted value versus the measured value scatter diagram of different methods for CO₂. The points of BPNN, GA-BPNN, and UVE are close to the diagonal line, and some points of PLS and GA-PLS are directly on the line. Almost all the points of the proposed method stay right on the line. Therefore, the prediction capability of the proposed method is the best for CO₂ of Dataset 2.

In summary, the experimental results verify that the proposed method is successfully employed for the spectral quantitative analysis of Dataset 1 and Dataset 2 with higher accuracy.

5. Conclusions

This paper proposes a spectral quantitative analysis model with combining wavelength selection and topology structure

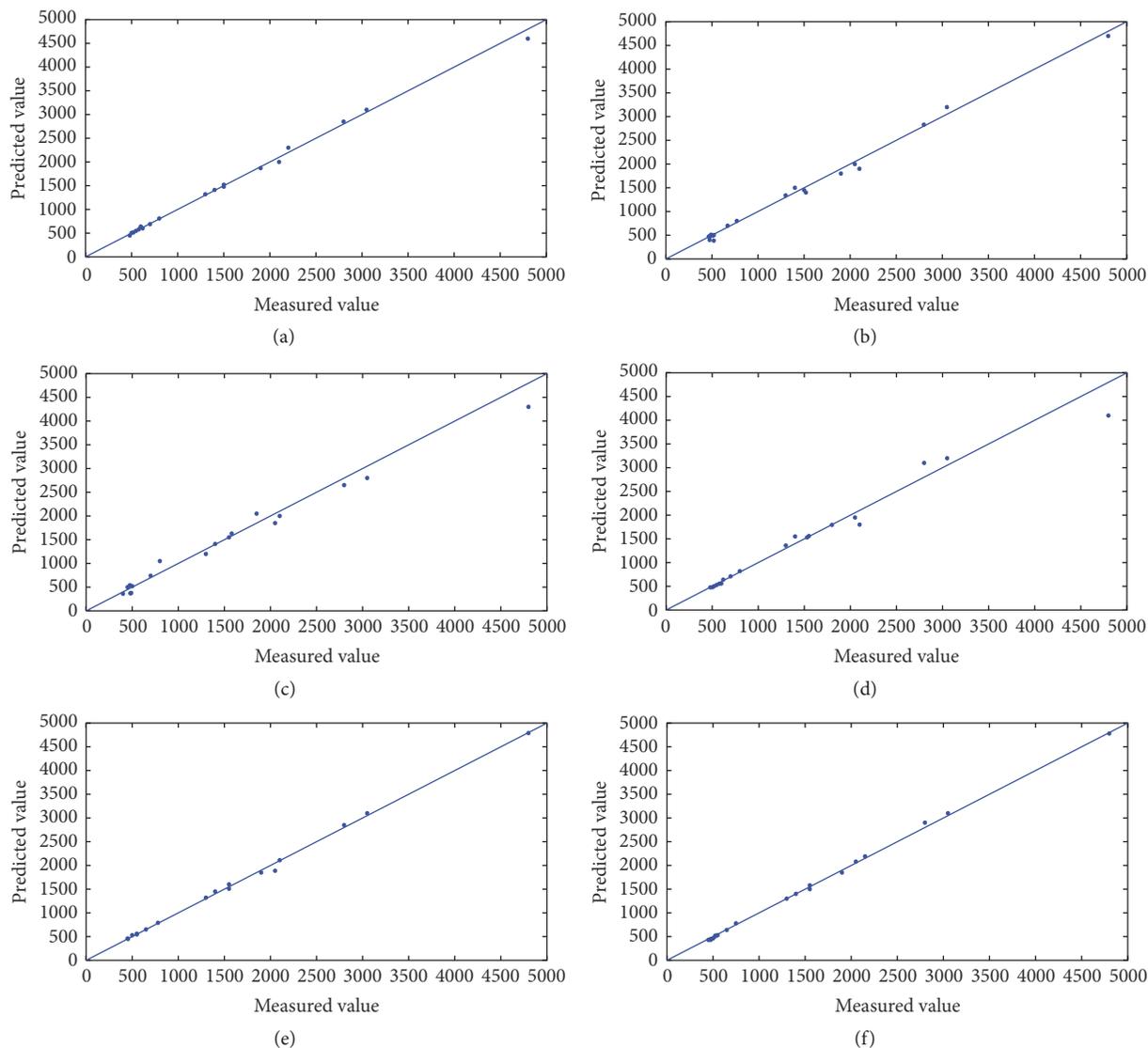


FIGURE 9: The predicted value versus the measured value scatter diagram of different methods for CO₂. (a) PLS. (b) BPNN. (c) GA-BPNN. (d) UVE. (e) GA-PLS. (f) The proposed method.

TABLE 6: Analytical results for CO₂.

CO ₂	RMSECV	RMSEP	R_{cv}^2	R_p^2	CR
PLS	41.5959	55.4754	0.9971	0.9981	0
BPNN	55.8712	57.7472	0.9957	0.9895	0
GA-BPNN	51.1870	55.9048	0.9961	0.9901	0
UVE	45.4300	47.9081	0.9963	0.9910	0.9345
GA-PLS	43.5677	42.8292	0.9959	0.9982	0.5137
The proposed method	41.9192	35.4125	0.9980	0.9991	0.5370

optimization. The proposed method has some advantages as follows. First, the proposed method can be used for the spectral quantitative analysis. Second, the proposed method realizes the simultaneous optimization of the wavelength selection and the topology structure. Third, the proposed method only has the mutation operation which can simplify

the iteration procedure without decreasing the precision. The experiments results verify that the proposed method has higher predicative ability for spectral quantitative analysis and can be applied to different types of spectra.

Competing Interests

The authors declare no conflict of interests.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61375055), the Program for New Century Excellent Talents in University (NCET-12-0447), Natural Science Foundation of Shanxi Province of China (2014JQ8365), State Key Laboratory of Electrical Insulation and Power Equipment (EIPE16313), and the Fundamental

Research Funds for the Central University. The authors are grateful for the grammar modification provided by Yanxia Wang.

References

- [1] L. Liang, J. Sun, Q. He, Z. Chen, S. Zhu, and L. Lin, "Semi-quantitative analysis on the content of berberine hydrochloride in compound berberine tablets with the fluorescence spectral imaging method," *Journal of Innovative Optical Health Sciences*, vol. 9, no. 2, Article ID 1650018, 7 pages, 2016.
- [2] Y. Ma, M. Wei, X. Zhang, T. Zhao, X. Liu, and G. Zhou, "Spectral study of interaction between chondroitin sulfate and nanoparticles and its application in quantitative analysis," *Spectrochimica Acta—Part A: Molecular and Biomolecular Spectroscopy*, vol. 153, pp. 445–450, 2016.
- [3] Z. Wu, L. B. Zeng, and Q. S. Wu, "Study of cervical exfoliated cells DNA quantitative analysis based on multi-spectral imaging technology," *Spectroscopy and Spectral Analysis*, vol. 36, no. 2, pp. 496–501, 2016.
- [4] V. Kopačková, "Using multiple spectral feature analysis for quantitative pH mapping in a mining environment," *International Journal of Applied Earth Observation & Geoinformation*, vol. 28, no. 1, pp. 28–42, 2014.
- [5] C. Hua-zhou, S. Qi-qing, S. Kai, and J. Zhen, "Multidimensional scaling linear regression applied to FTIR spectral quantitative analysis of clinical parameters of human blood serum," *Spectroscopy and Spectral Analysis*, vol. 35, no. 4, pp. 914–918, 2015.
- [6] Y. Li and J. Jing, "A consensus PLS method based on diverse wavelength variables models for analysis of near-infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 45–49, 2014.
- [7] C. Koch, A. E. Posch, H. C. Goicoechea, C. Herwig, and B. Lendl, "Multi-analyte quantification in bioprocesses by Fourier-transform-infrared spectroscopy by partial least squares regression and multivariate curve resolution," *Analytica Chimica Acta*, vol. 807, pp. 103–110, 2014.
- [8] D.-S. Cao, Y.-Z. Liang, Q.-S. Xu, Q.-N. Hu, L.-X. Zhang, and G.-H. Fu, "Exploring nonlinear relationships in chemical data using kernel-based methods," *Chemometrics and Intelligent Laboratory Systems*, vol. 107, no. 1, pp. 106–115, 2011.
- [9] V. N. Vapnik, "Statistical learning theory," *Encyclopedia of the Sciences of Learning*, vol. 41, no. 4, pp. 3185–3185, 2010.
- [10] A. Rădulescu, "Neural network spectral robustness under perturbations of the underlying graph," *Neural Computation*, vol. 28, no. 1, pp. 1–44, 2016.
- [11] J. Yuan and S. Yu, "Privacy preserving back-propagation neural network learning made practical with cloud computing," *IEEE Transactions on Parallel & Distributed Systems*, vol. 25, no. 1, pp. 212–221, 2014.
- [12] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Van-Dewalle, "Least squares support vector machines," *Euphytica*, vol. 2, no. 2, pp. 1599–1604, 2002.
- [13] L. Xing and L. Chen, "Quantitative analysis of near-infrared spectroscopy by combined stationary wavelet transform-support vector machine," *Spectroscopy Letters*, vol. 46, no. 1, pp. 47–53, 2013.
- [14] J. C. L. Alves and R. J. Poppi, "Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM)," *Talanta*, vol. 104, no. 2, pp. 155–161, 2013.
- [15] Z.-Y. Deng, B. Zhang, W. Dong, and X.-P. Wang, "Research on prediction method of fatty acid content in edible oil based on raman spectroscopy and multi-output least squares support vector regression machine," *Spectroscopy and Spectral Analysis*, vol. 33, no. 11, pp. 2997–3001, 2013.
- [16] R.-M. Luo, S.-M. Tan, Y.-P. Zhou et al., "Quantitative analysis of tea using ytterbium-based internal standard near-infrared spectroscopy coupled with boosting least-squares support vector regression," *Journal of Chemometrics*, vol. 27, no. 7-8, pp. 198–206, 2013.
- [17] A.-X. Zhao, X.-J. Tang, Z.-H. Zhang, and J.-H. Liu, "The spectral characteristic wavelength selection and parameter optimization based on tikhonov regularization," *Spectroscopy and Spectral Analysis*, vol. 34, no. 7, pp. 1836–1839, 2014.
- [18] L. Ling, L. I. Wei, Z. Rui-li, L. Rui-an, L. Gang, and W. Xiaorong, "Wavelength selection of the oximetry based on test analysis of variance," *Spectroscopy and Spectral Analysis*, vol. 34, no. 7, pp. 2005–2009, 2014.
- [19] X. Shao, G. Du, M. Jing, and W. Cai, "Application of latent projective graph in variable selection for near infrared spectral analysis," *Chemometrics & Intelligent Laboratory Systems*, vol. 114, pp. 44–49, 2012.
- [20] M. Vohland, M. Ludwig, S. Thiele-Bruhn, and B. Ludwig, "Determination of soil properties with visible to near- and mid-infrared spectroscopy: effects of spectral variable selection," *Geoderma*, vol. 223–225, no. 1, pp. 88–96, 2014.
- [21] L. Xu and I. Schechter, "Wavelength selection for simultaneous spectroscopic analysis. Experimental and theoretical study," *Analytical Chemistry*, vol. 68, no. 14, pp. 2392–2400, 1996.
- [22] Y. Dong, B. Xiang, Y. Geng, and W. Yuan, "Rough set based wavelength selection in near-infrared spectral analysis," *Chemometrics & Intelligent Laboratory Systems*, vol. 126, pp. 21–29, 2013.
- [23] V. Centner, D.-L. Massart, O. E. De Noord, S. De Jong, B. M. Vandeginste, and C. Sterna, "Elimination of uninformative variables for multivariate calibration," *Analytical Chemistry*, vol. 68, no. 21, pp. 3851–3858, 1996.
- [24] Q.-J. Han, H.-L. Wu, C.-B. Cai, L. Xu, and R.-Q. Yu, "An ensemble of Monte Carlo uninformative variable elimination for wavelength selection," *Analytica Chimica Acta*, vol. 612, no. 2, pp. 121–125, 2008.
- [25] Y. Hao, X. Sun, H. Zhang, and Y. Liu, "Application of effective wavelength selection methods to determine total acidity of navel orange," *Sensor Letters*, vol. 9, no. 3, pp. 1229–1234, 2011.
- [26] J. Ghasemi, A. Niazi, and R. Leardi, "Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture," *Talanta*, vol. 59, no. 2, pp. 311–317, 2003.
- [27] R. Irani and R. Nasimi, "Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9862–9866, 2011.
- [28] W. Sun and Y. Xu, "Financial security evaluation of the electric power industry in China based on a back propagation neural network optimized by genetic algorithm," *Energy*, vol. 101, pp. 366–379, 2016.
- [29] D. Jouan-Rimbaud, D.-L. Massart, R. Leardi, and O. E. De Noord, "Genetic algorithms as a tool for wavelength selection in multivariate calibration," *Analytical Chemistry*, vol. 67, no. 23, pp. 4295–4301, 1995.

- [30] M. Arakawa, Y. Yamashita, and K. Funatsu, "Genetic algorithm-based wavelength selection method for spectral calibration," *Journal of Chemometrics*, vol. 25, no. 1, pp. 10–19, 2011.
- [31] R. Nordon, D. Lutz, R. Genzel et al., "The impact of evolving infrared spectral energy distributions of galaxies on star formation rate estimates," *The Astrophysical Journal*, vol. 745, no. 2, p. 167, 2011.
- [32] H. Momm and G. Easson, "Evolving spectral transformations for multitemporal information extraction using evolutionary computation," *Journal of Applied Remote Sensing*, vol. 5, no. 1, Article ID 053564, 2011.
- [33] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, Germany, 1992.
- [34] A. P. Engelbrecht, *Evolutionary Programming*, John Wiley & Sons, 2007.
- [35] P. Sinha, A. Chandwani, and T. Sinha, "Algorithm of construction of optimum portfolio of stocks using genetic algorithm," *International Journal of Systems Assurance Engineering and Management*, vol. 6, no. 4, pp. 447–465, 2015.
- [36] Y. Zhang, D. Guo, and Z. Li, "Common nature of learning between back-propagation and hopfield-type neural networks for generalized matrix inversion with simplified models," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 24, no. 4, pp. 579–592, 2013.
- [37] J. A. F. Pierna, B. Lecler, J. P. Conzen, A. Niemoeller, B. Baeten, and P. Dardenne, "Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products," *Analytica Chimica Acta*, vol. 705, no. 1-2, pp. 30–34, 2011.

