JOTCS Journal of the Turkish Chemical Society Section: A Chemistry

TURKISH CHEMICAL SOCIETY

Finding Exact Number Of Peaks in Broadband UV-Vis Spectra Using Curve Fitting Method Based On Evolutionary Computing

Fatih Mehmet AVCU[1,*] and Mustafa KARAKAPLAN[2]

[1]University of İnönü, Department of Informatics, 44280, Malatya, Turkey.
[2]University of İnönü, Faculty of Arts and Sciences, Department of Chemistry, 44280, Malatya, Turkey.

Abstract: High performance calculations are needed in order to resolve analytic signals of the day. However, it requires very long periods of time to perform these calculations with single processor systems. In order to reduce these calculation times, there is a need to turn to parallel programming algorithms that share more than one processor. Recently, solving complex problems with genetic algorithms has been widely used in computational sciences. In this work, we show a new method of curve fitting via genetic algorithm based on Gaussian functions, for deconvolution of the overlapping peaks and find the exact number of peaks in UV-VIS absorption spectroscopy. UV-VIS spectra are different than other instrumental analysis data. The resolution of UV-VIS spectra are complicated since the absorption bands are strongly overlapped. Useful information about molecular structure and environment can often be obtained by resolving these peaks properly. The algorithm was parallelized with the island model in which each processor computes a different population. This method has been used for resolving of the UV-VIS overlapping spectrum. The method particular algorithm is robust against bad resolution or noise. The results clearly show the effectiveness of the proposed method.

Keywords: UV-Vis spectroscopy, Data fitting, Genetic algorithm, Parallel computing, Peak Numbers.

Submitted: June 28, 2019. **Accepted:** November 15, 2019.

Cite this: Avcu F, Karakaplan M. Finding Exact Number Of Peaks in Broadband UV-Vis Spectra Using Curve Fitting Method Based On Evolutionary Computing. JOTCSA. 2020;7(1):117–24.

DOI: https://doi.org/10.18596/jotcsa.583632.

***Corresponding author.** E-mail: fatih.avcu@inonu.edu.tr.

INTRODUCTION

UV-VIS spectroscopy is a common qualitative and quantitative evaluation of samples applied to instrumental analysis. The analysis of the UV-VIS spectroscopy which contains the overlapped peaks reveals information about the molecule's structure and its environment. Each individual band in the UV-VIS spectra is complicated because the UV-VIS absorption spectra are generally strongly overlapped. The resolution of individual bands has different half-height width, and peaks numbers in the overlapped spectra are difficult to estimate. This method is used for analytical signal measurement because of its speed and simplicity, and it has been increasing in importance due to decreasing cost. Because of these developments there is a need for more intelligent design of software in order to decrease computation time.

A genetic algorithm is a part of the evolutionary computation influenced by the evolutionary theory of Darwin. It is a stochastic method and optimization algorithm. If we simply explain the aim, at the end of the process it gives the best results and this problem using the evolutionary process. The result is that individuals who survive have developed capabilities that will lead to finding the best individuals. The best result is the convergence that managed to stay alive. Method for optimizing is developed by L. A. Rastrigin in 1963 (1). First, evolutionary computation is

defined in the work of I. Rechenberg's "Strategy of Evolution". After that J. Holland developed genetic algorithms to imitate the process of evolution in the computer (2). Holland uses a bit sequence to code complex structures. David E. Goldberg, who was a student of Holland, has published books that genetic algorithm, can be used on various topics (3).

GA gained popularity in almost every branch in science and engineering disciplines (4-7). GA has more solutions for many problems. The GA is widely used in complex, parameter optimization and multi-extrema problems. The fields where the genetic algorithm is most commonly used are applications of the GA to molecular energy optimizations (8-13), signal processing (14,15), curve fitting (5,16), and linear and non-linear function optimization (17,18). The genetic algorithm is robust and highly efficient at the same time (19).

We propose the theory of curve-fitting based on the Gaussian model and both the minimum separable peak-peak interval and the curve-fitting error to resolve overlapping spectra with GA. Finally, we used synthetic spectral and real UV-VIS spectra to verify the performance of our method.

THEORETICAL

Approach to Curve Fitting with Genetic Algorithm

The aim of curve fitting is finding the mathematical model coefficients which minimize the total error over the set of data points being considered. If a function form has been selected from the data, curve fitting becomes an optimization problem. GAs have been used as global optimization techniques problems. They are suitable for curve fitting problems when it is structured as a parameter selection problem. Generally, the point values are data obtained from experimental results. There is no continuous definition between data and functions.

Typically, sum of complex functions must be used; otherwise, a lot of functions must be defined point-by-point. This is the curve fitting process. Since the genetic algorithms for continuous functions global optimization technique as being successful so this technique is suitable for parameter selection problems to curve fitting.

Parallel Genetic Algorithm

Parallel genetic algorithms (PGA)(20-22) are programs which parallelize, distribute, and break into pieces the basic program. Multiple processors then solve simultaneously.

Hardware and Software

Our GA implementation used C++ class libraries and Linux systems. Coarse-grained parallelism, utilizing AMD Opteron 2435 processors with six cores to perform sets of 100 runs, and the OpenMPI (23) parallel programming library was also used.

UV-VIS Peak Shape

UV-VIS spectroscopy is an inexpensive technique applicable to mixtures of organic and inorganic compounds for qualitative and quantitative analysis. Depending on the type and number of components, spectra could become complicated. Location, height and width of overlapping peaks gives knowledge about spectral bands. The spectral bandwidth of UV-VIS is larger than other spectroscopic techniques' spectra. This is because rotation and vibration transitions are overlaid on and overlap the electronic transition energy levels. UV-VIS produces spectra that are simpler, and containing a smaller number of narrower, more than other spectra e.g. NMR, IR, etc. overlapped peaks. Also, these absorption bands may have different functions from each other. In the analysis of spectral bands Lorentzian, Pearson VII and Gaussian functions are preferred to the use of Gaussian functions because UV-VIS spectra contain sum of broad peaks.

In order to simulate overlapping UV-VIS spectra, three type of functions were used as follows;

$$Lorentzian: A(\nu)=\sum_{i=1}^{n}\frac{A_{i,0}H_i^2}{H_i^2+4(\nu-\nu_{i,0})^2} \quad (1)$$

$$Pearson\,VII: A(\nu)=\sum_{i=1}^{n}\frac{A_{i,0}}{[1+4Z_i^2(2^{1/m_i}-1)]} \quad (2)$$

$$Gaussian: A(\nu)=\sum_{i=1}^{n}A_{i,0}e^{(-(\nu-\nu_{i,0})^2/H_i)} \quad (3)$$

where parameter of Eq. (1), (2) and (3) are that $A_{i,0}$ is the absorbance in the center of peak i, $Zi=(\nu-\nu_{i,0})/H_i$ and H_i are the half width of peak i, $\nu_{i,0}$ is the peak position of peak i, m_i is the tailing factor of peak i, and n is the number of the peak.

EXPERIMENTAL

After synthesized nanoparticles of ZrO_2 disperse in solvent and surface was coated with 2-AAEM. After drying and cooling modified nano particles are prepared, analyzed by a Varian Cary 5000 UV-VIS-NIR to measure absorbance at the solid surface. Powdered sample (1 mm) was set in the sample

container. After inserting the relevant part of the DRA device absorbance between 200-800 mm was measured (see Figure 1).

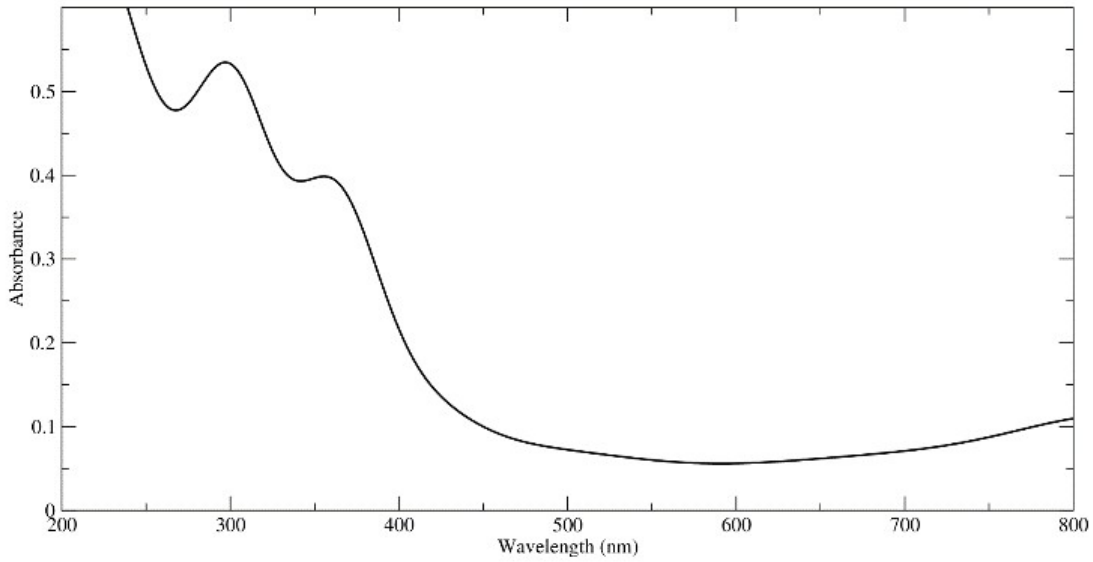

Figure 1: Full UV-Vis Spectrum of ZrO_2 Compound.

RESULTS AND DISCUSSION

Optimization of PGA-Parameters

Our previous studies show that a centralized type of island model is the most effective one among other types of island models (24). In this study, the shape of the island model is shown in Figure 2(a) and the flowchart for the algorithm is shown in Figure 2(b).

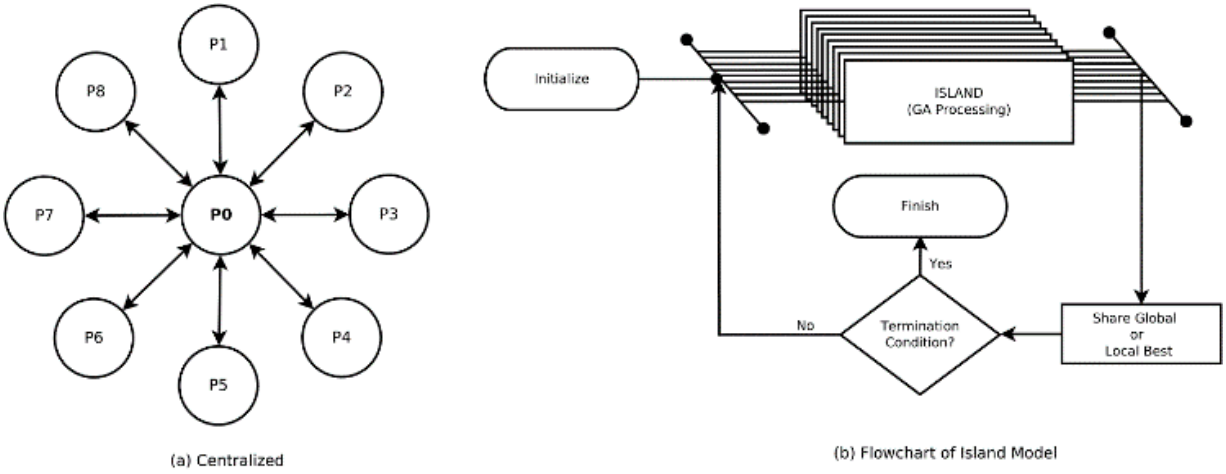

Figure 2: Selected Island Model(a) and Flowchart of Island Model(b).

Search space shrinking ratio that is narrowing the search space with this factor is important factor in GA and PGA. If this value is too low, the best solution can be missed or the algorithm called show a local minimum as the best solution. In the case of wide search space, the algorithm will be inefficient. In the present work a search space factor of 1.9 was adequate to good convergence as shown as in Figure 3.

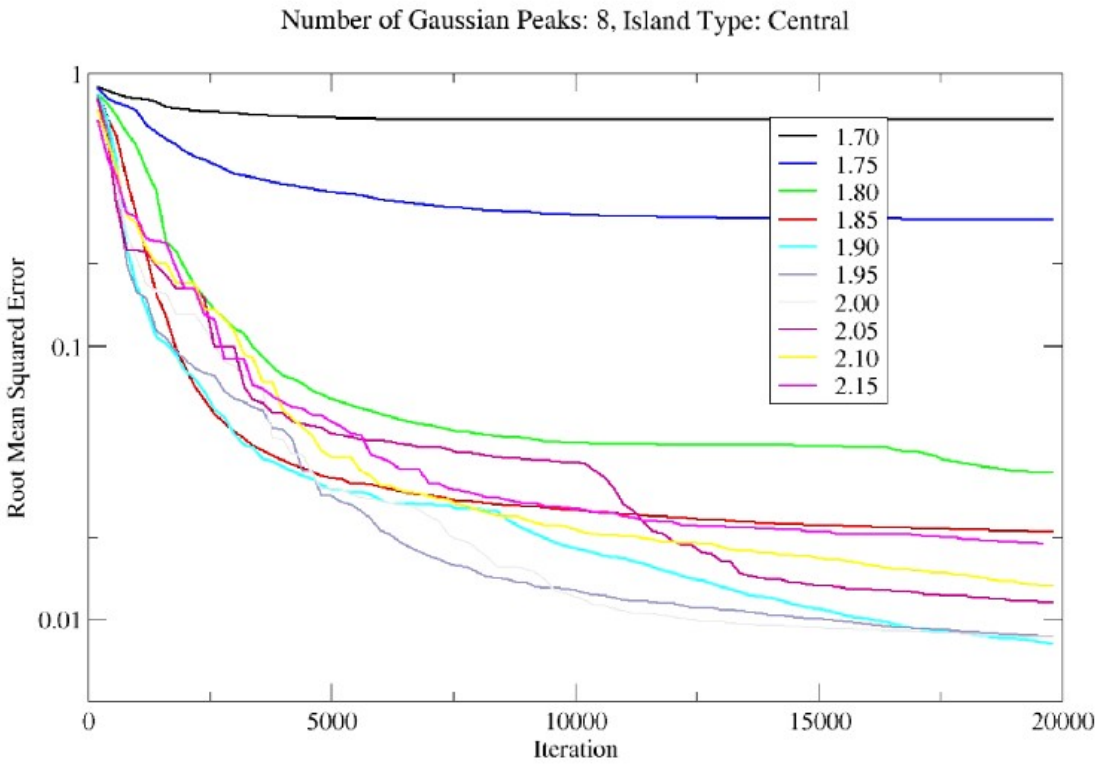

Figure 3: Effect of Search Space Factor on The Convergence of Parallel Genetic Algorithm.

The main problem is the choice of the analytical function describing the contour of the individual spectral band. Although there are many shape functions, the procedure for analysis of the overlapping bands is based on the alternative use of a Gaussian function, a Lorentzian function, or Pearson VII function. Figure 4 showed that Gaussian function most effective in UV-Vis spectra studied.

The algorithm was tested on theoretical data set. This theoretical set showed Figure 5.The spectra in Figure 5 have been calculated with 8 Gaussian functions which include wide and narrow absorption bands. Calculated data and experimental data fully coincide. Automatic detection of peaks enough success in here.

The method is based on minimizing the deviation between calculated data and real model of data. The quality of result was tested by Eq. (4).

$$RMSE = \sqrt{\frac{\sum v \sum_{j=1}^{Np} [A_j(v) - \bar{A}_j(v)]^2}{N_d}} \quad (4)$$

Where this formula's parameters are that RMSE (Root Mean Squared Error) is the square root of mean sum of squared residuals, N_p is the number of the peak, N_d is the number of the data point is data points number, A_j is the observed absorbance is observed absorbance, and $\bar{A}_j$ is the calculated absorbance.

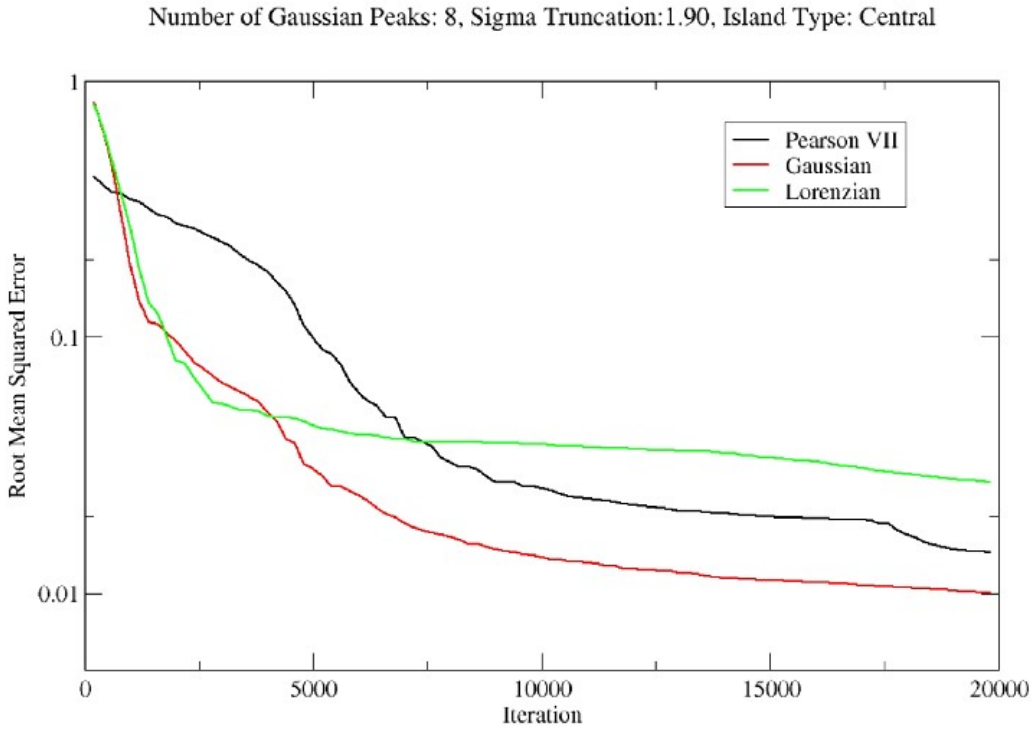

Figure 4: Comparison of Convergence of Three Peak Shapes on Parallel Genetic Algorithm.

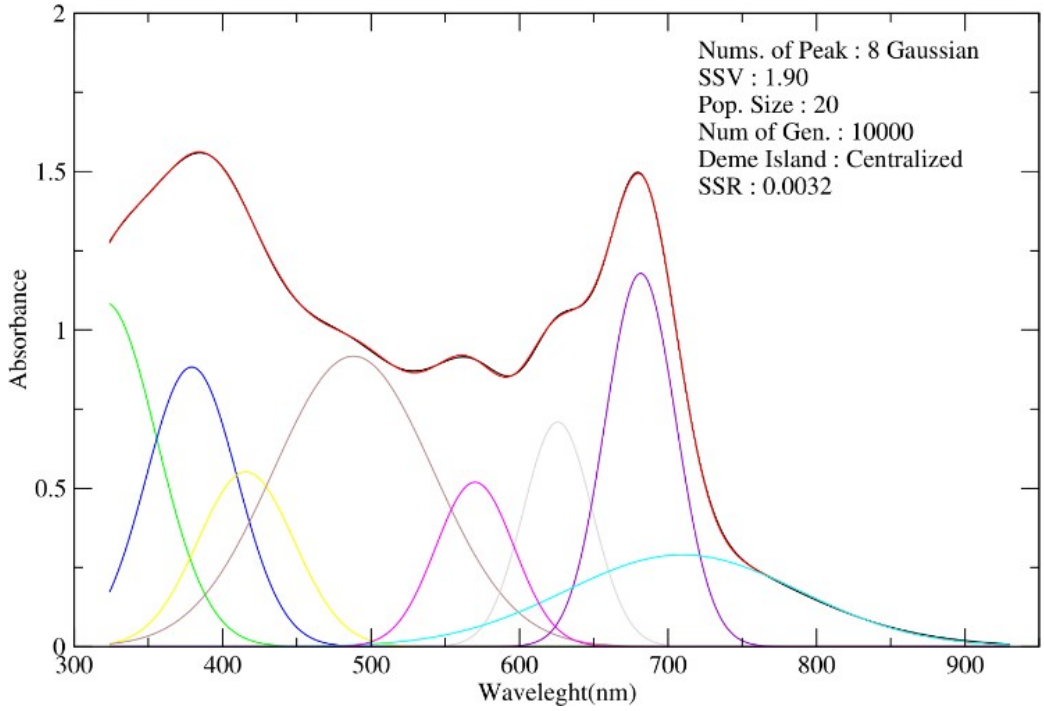

Figure 5: Deconvolution of UV-Vis spectrum of $C_{36}H_{44}FeN_4O_2$ compound.

A spectrum can be expressed with sum of peaks, which is more than the necessary number of peaks, but it cannot be represented by fewer peaks. The synthetic experimental spectrum that includes 8 peaks was recalculated with different peak numbers and the results in Figure 6 were obtained. When it was calculated again with 8 peaks, a sudden decrease in SSR value was

observed. The minimum number of peaks that must be present in a spectrum can be easily found by using this comparison. In calculations with more peaks than necessary, you can reach the lower SSR values, but it requires more time for computation. In addition, the positions of the peaks can be changed in every calculation.

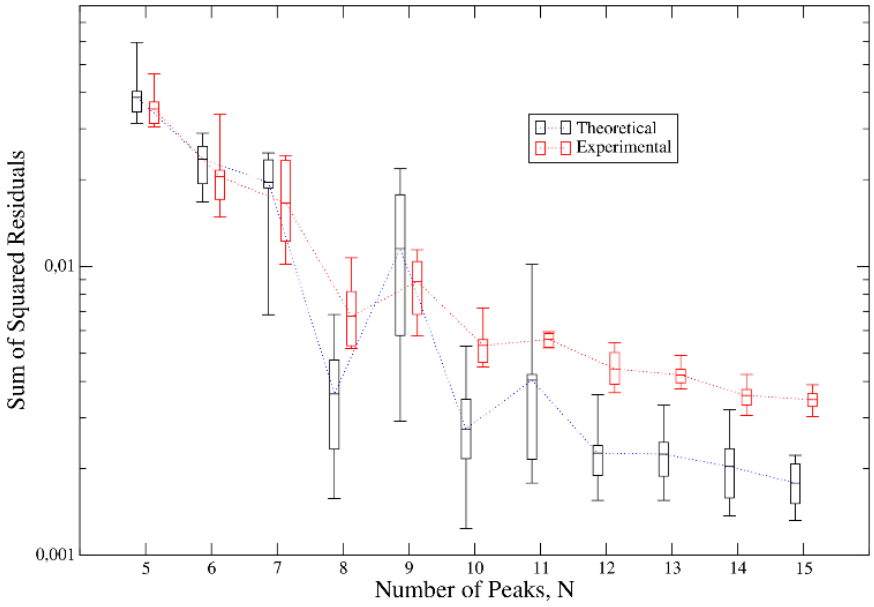

Figure 6: Number of peaks and SSR values in $C_{36}H_{44}FeN_4O_2$ compound.

Experimental Data

The experimental spectrum in Figure 12, with high levels of overlap and noise, was investigated by the proposed method.

Although the spectrum consists of 4 peaks, it is difficult to resolve these with 4 Gaussian peaks. It was resolved with 8 Gaussian peaks. If the number of iterations is increased, value of SSR will decrease. However, the value of 0.0008 sum of squared residuals was observed to be significant enough. This result shows that a reasonable resolution can be obtained for the experimental data. In addition, this method seems to be a good tool for the resolution of the components with very similar UV-Vis spectra.

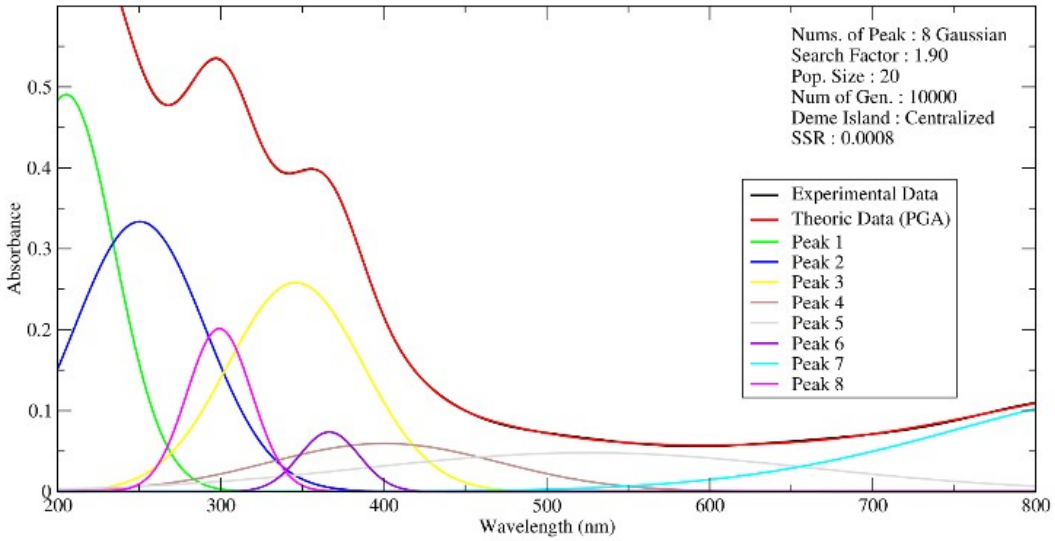

Figure 7: Deconvolution of UV-VIS Spectrum of ZrO_2 Compound.

The individual bands in UV-Vis absorption spectroscopy of ZrO_2 compound are best described with Gaussian functions using PGA. The parameters of individual peaks of ZrO_2 compound are shown in Table 1.

Table 1: Individual Peaks Parameters of UV-VIS Spectrum of ZrO_2 Compound.

Peak Number	X	Y	Half-Height Width
1	0.648942	209.548289	43.992142
2	0.249610	261.542086	38.881243
3	0.359505	300.469008	30.534187
4	0.376847	358.057764	45.094099
5	0.082054	430.533495	48.046265
6	0.039971	504.090613	57.958996
7	0.051891	623.967310	132.833099
8	0.113161	844.744351	130.713092

CONCLUSIONS

In this study, we implemented a PGA to perform curve fitting. The PGA was developed to resolve overlapping spectra with Gaussian peak shapes. The method can be used to represent all peaks of an overlapping spectrum, respectively, by optimizing the parameters of each peak. The performance of PGA was validated by the resolution of the simulated and experimental overlapping spectrum. The results showed that number of individual bands in overlap spectra in UV-VIS spectra that they are best described with Gaussian function estimated easily and the individual peaks can be correctly resolved in very highly overlapped areas. The algorithm shown is unresponsive to poor resolution or noise. In order to perform genetic algorithm accounting more quickly and effectively, we apply parallel computing models to the calculation. This application was shown to improve spectra fit and speed.

ACKNOWLEDGEMENTS

Our article was funded by Inonu University Scientific Research Project Department [Project Number: FBA-2019- 1726].

REFERENCES

1. L. A. Rastrigin, The convergence of the random search method in the extremal control of a many-parameter system, Automat. Remote Control 24 (11), 1963, pp. 1337–1342

2. J. H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, University of Michigan Press, 1975.

3. D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

4. C. B. Lucasius, G. Kateman, Application of genetic algorithms in chemometrics, in: Proceedings of the 3rd International Conference on Genetic Algorithms, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1989, pp. 170–176

5. M. Karakaplan, Fitting lorentzian peaks with evolutionary genetic algorithm based on stochastic search procedure, Analytica Chimica Acta 587 (2), 2007 pp. 235–239

6. W. D. Whitley, S. B. Rana, R. B. Heckendorn, Island model genetic algorithms and linearly separable problems, in: Selected Papers from AISB Workshop on Evolutionary Computing, Springer-Verlag, London, UK, UK, , 1997, pp. 109–125

7. R. Wehrens, L. M. Buydens, Evolutionary optimisation: a tutorial, TrAC Trends in Analytical Chemistry 17 (4), 1998, pp. 193 – 203

8. D. M. Deaven, K. M. Ho, Molecular geometry optimization with a genetic algorithm, Physical Review Letters 75 (2), 1995, pp. 288–291

9. J. Zhao, R.-H. Xie, Genetic algorithms for the geometry optimization of atomic and molecular clusters, Journal of Computational and Theoretical Nanoscience 1 (2), 2004, pp. 117–131

10. M. A. Addicoat, A. J. Page, Z. E. Brain, L. Flack, K. Morokuma, S. Irle, Optimization of a genetic algorithm for the functionalization of fullerenes, Journal of Chemical Theory and Computation 8 (5), 2012, pp. 1841–1851

11. R. P. F. Kanters, K. J. Donald, cluster: Searching for unique low energy minima of structures using a novel implementation of a genetic algorithm, Journal of Chemical Theory and Computation 10 (12), 2014, pp. 5729–5737

12. N. Alexandrova, A. I. Boldyrev, Search for the lin0/+1/-1 (n = 57) lowest-energy structures using the ab initio gradient embedded genetic algorithm (gega). elucidation of the chemical bonding in the lithium clusters, Journal of Chemical Theory and Computation 1 (4), 2005, pp. 566–580

13. M. Chen, D. A. Dixon, Tree growthhybrid genetic algorithm for predicting the structure of small (tio2)n, n = 213, nanoclusters, Journal of Chemical Theory and Computation 9 (7), 2013, pp. 3189–3200

14. K. Tang, K. Man, S. Kwong, Q. He, Genetic algorithms and their applications, IEEE Signal Processing Magazine 13 (6), 1996 , pp. 22–37

15. E. Jones, P. Runkle, N. Dasgupta, L. Couchman, L. Carin, Genetic algorithm wavelet design for signal classification, IEEE Trans. Pattern Anal. Mach. Intell. 23 (8), 2001, pp. 890–895

16. M. Gulsen, A. E. Smith, D. M. Tate, A genetic algorithm approach to curve fitting, International Journal of Production Research 33 (7), 1995, pp. 1911–1923

17. P. Vankeerberghen, J. Smeyers-Verbeke, R. Leardi, C. L. Karr, D. L. Massart, Robust regression and outlier detection for non-linear models using genetic algorithms, Chemometrics and Intelligent Laboratory Systems 28 (1), 1995, pp. 73–87

18. H. Ahonen, P. A. de Souza Jnior, V. K. Garg, A genetic algorithm for fitting lorentzian line shapes in mssbauer spectra, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 124 (4), 1997, pp. 633–638

19. C. Lucasius, G. Kateman, Genetic algorithms for large-scale optimization in chemometrics: An application, TrAC Trends in Analytical Chemistry 10 (8), 1991, pp. 254 – 261

20. H. Muhlenbein, Evolution in time and space - the parallel genetic algorithm, in: Foundations of Genetic Algorithms, Morgan Kaufmann, 1991, pp. 316–337

21. H. Muhlenbein, M. Schomisch, J. Born, The parallel genetic algorithm as function optimizer, Parallel Computing 17 (67), 1991, pp. 619–632

22. C. B. Louise, M. C. Tran, T. G. Obrig, Sensitization of human umbilical vein endothelial cells to shiga toxin: involvement of protein kinase c and NF-kappaB., Infection and Immunity 65 (8), 1997, pp. 3337–3344

23. MPICH | high-performance portable MPI. URL http://www.mpich.org/ (2019).

24. M. Karakaplan, F. M. Avcu, A parallel and non-parallel genetic algorithm for deconvolution of NMR spectra peaks, Chemometrics and Intelligent Laboratory Systems 125 (0), 2013, pp. 147 – 152