UV-Visible Spectrophotometric Quantitative Analysis of Ternary Mixture Using Multivariate Calibration Methods Optimized by a Genetic Algorithm

DURMUS ÖZDEMIR[1], **ERDAL DINÇ**[2,*], **DUMITRU BALEANU**[3,4]

[1]Department of Chemistry, Faculty of Science, Izmir Institute of Technology, Gülbahce köyü 35430 Urla, Izmir, Turkey
[2]Department of Analytical Chemistry, Faculty of Pharmacy, University of Ankara, 06100 Tandogan, Ankara, Turkey
[3]Department of Mathematics and Computer Sciences, Faculty of Arts and Sciences, Çankaya University, 06530, Balgat, Ankara, Turkey
[4]National Institute for Laser, Plasma and Radiation Physics, Institute of Space Sciences, Magurele-Bucharest, P.O. Box,MG-23, R 76911, Romania

Simultaneous determination of ternary mixtures of caffeine, paracetamol and metamizol in commercial tablet formulations using UV-visible spectrophotometry combined with classical least squares (CLS) and genetic algorithm (GA) based multivariate calibration methods were demonstrated. The three genetic multivariate calibration methods are named as Genetic Classical Least Squares (GCLS), Genetic Inverse Least Squares (GILS) and Genetic Regression (GR). The GR method is based on a genetic algorithm based wavelength selection followed by a simple linear regression step whereas the GCLS and GILS are multivariate calibration methods modified by a wavelength selection principle using a genetic algorithm. The sample data set contains the UV-visible spectra of 47 synthetic mixtures (4 to 48 μg/mL) and 16 tablets containing these components from two different producers. The spectra cover the range from 200 to 330 nm in 0.1 nm intervals. Several calibration models were built with the four methods for the three components. Overall, the standard error of calibration (SEC) and the standard error of prediction (SEP) for the synthetic data were in the range of 0.04 and 2.34 μg/mL for all the four methods. Predictive ability of the calibration models generated with synthetic samples was tested with actual tablet samples and results obtained from four methods were compared. The SEP values for the tablets were in the range of 0.31 and 15.44 mg/tablets.

Keywords: UV-visible spectrophotometry, multivariate calibration, Genetic algorithms

The resolution of mixtures containing two or more compounds without any separation in the presence of excipients existing in samples is one of the main issues of simultaneous quantitative determination. Among the various analytical techniques available for the simultaneous determination of caffeine (CAF), paracetamol (PAR), metamizol (MET) and their mixtures with other compounds are spectrophotometry [1–5], HPLC [6, 7], and voltammetry [8].

Modern spectroscopic instruments produce hundreds of spectra in a few minutes for a given sample that may contain multiple components. Unfortunately, univariate calibration methods are not suitable for this type of data, as they require an interference free system. Multivariate calibration deals with data containing instrument responses measured at multiple wavelengths for a sample that usually contains more than one component. In recent years, advances in chemometrics and computers have lead to the development of several multivariate calibration methods [9–12] for the analysis of complex chemical mixtures.

Genetic regression (GR) is a calibration technique that optimizes linear regression models using a genetic algorithm (GA) and has been applied to a number of multi-instrument calibration and wavelength selection problems [13–17]. GAs are non-local search and optimization methods that are based upon the principles of natural selection [18–22]. For a given set of full spectrum data, GR selects an optimum linear combination of wavelengths and simple mathematical operators to build a linear calibration model using simple least squares method.

Classical Least Squares (CLS) extends the classical Beer's Law model in which the absorbance at each wavelength is directly proportional to the component concentrations. Inverse Least Squares (ILS) is based on the inverse of Beer's Law where concentration of an analyte is modeled as a function of absorbance measurements. Genetic Classical Least Squares (GCLS) and Genetic Inverse Least Squares (GILS) are modified versions of original CLS and ILS methods in which a small set of wavelengths are selected from a full spectral data matrix and evolved to an optimum solution using a genetic algorithm.

In this work, CLS and three different genetic algorithms based calibration methods GCLS, GILS and GR were tested with the aim of establishing calibration models that have a high predictive capacity for the simultaneous determination of caffeine (CAF), paracetamol (PAR), metamizol (MET) and their mixtures in commercial tablet formulations using the UV-visible spectrophotometry.

Experimental Part
Materials
Stock solutions containing 100 mg/100 mL caffeine (CAF), paracetamol (PAR) and metamizol (MET) were prepared in 0.1 M HCl. A calibration set consisting of 31 sample solutions containing 4–40μg/mL CAF, 8–40 μg/mL PAR and 12–48 μg/mL MET were obtained from the stock solutions. A prediction set consisting of 16 synthetic mixture solutions in the concentration range of 4–40 μg/mL CAF, 8–40 μg/mL APAP and 12–48 μg/mL MET was prepared by using the same stock solutions. All the solutions were prepared freshly and protected from light.

* email: dinc@pharmacy.ankara.edu.tr; Tel.: +90 312 215 4886

Two commercial tablet formulations (Remidon® tablets produced by Deva Pharm., Turkey, Batch no. 306 1582, consisting of 50 mg CAF, 200 mg PAR and 200 mg MET per tablet and Pirosal® tablets produced by Saba Pharm., Turkey, Batch no. 48, consisting of 30 mg CAF, 160 mg PAR and 220 mg MET per tablet) were investigated. Deva Pharm. Ind. and Saba Pharm. Ind. kindly donated the active compounds. In order to test the new methods for the real samples, a total of 16 tablets (8 for the Remidon® and 8 for the Pirosal®) were accurately weighed and powdered in a mortar. An amount equivalent to one tablet was dissolved in 0.1 M HCl in a 100 mL calibrated flask with the aid of mechanical shaking for 20 min. This solution was filtered into a 100 mL calibrated flask through Whatman No. 42 filter paper. The residue was washed three times with 0.1 M HCl and the solutions for the Remidon® and Pirosal® tablets were diluted to 1:100 and 1:62.5, respectively with the same solvent.

Methods

Sample spectra were measured on a Shimadzu UV-1600 double beam UV-visible spectrophotometer (Shimadzu, JAPAN) from 200 to 330 nm with 0.1 nm intervals. Quartz cells with 1 cm path lengths were used. The CLS and the three new genetic algorithms based multivariate calibration methods (GR, GCLS and GILS) were written in MATLAB programming language using Matlab 5.3 (MathWorks Inc, Natick, MA). The text files for calibration, validation and prediction sets were generated with the use of Microsoft Excel (MS Office 97, Microsoft Corporation, CA).

Results and Discussion
Genetic Regression

Genetic Algorithms (GA) are global search and optimization methods based upon the principles of natural evolution and selection as developed by Darwin. Computationally, the implementation of a typical GA is quite simple and consists of five basic steps including initialization of a gene population, evaluation of the population, selection of the parent genes for breeding and mating, crossover and mutation, and replacing parents with their offspring. These steps have taken their names from the biological foundation of the algorithm.

Genetic Regression (GR) is an implementation of a GA for selecting wavelengths and mathematical operators to build linear calibration models. GR is a hybrid calibration between univariate and multivariate calibration techniques in which it optimizes simple linear regression models through an evolving selection of wavelengths and simple mathematical operators $(+, -, *, /)$. GR follows the same basic initialize/ breed/mutate/evaluate algorithm as other GAs but is unique in the way it encodes genes. A gene is a potential solution to a given problem and the exact form may vary from application to application. Here, the term gene is used to describe the collection of instrumental response pairs combined with the above mentioned operators. These pairs, called 'base pairs', are then combined with an addition operator to produce a score, which relates the instrumental response to component concentration. The term 'population' is used to describe the collection of individual genes in the current generation.

In the initialization step, the first generation of genes is created randomly with a fixed population size. Although random initialization helps to minimize bias and maximize the number of possible recombinations, GR is designed to select initial genes in a somewhat biased random fashion in order to start with genes better suited to the problem than those that would be randomly selected. Biasing is done with a correlation coefficient by plotting the scores of initial genes against the component concentrations. The size of the gene pool is a user defined even number in order to allow breeding of each gene in the population. It is important to note that the larger the population size, the longer the computation time. The number of base pairs in a gene is determined randomly between a fixed low limit and high limit. The lower limit was set to 2 in order to allow single point crossover whereas the higher limit was set to eliminate over fitting problems and reduce the computation time. Once the initial gene population is created, the next step is to evaluate and rank the genes using a fitness function, which is the inverse of the standard error of calibration (SEC).

The third step is where the basic principle of natural evolution is put to work for GR. This step involves the selection of the parent genes from the current population for breeding using a roulette wheel selection method according to their fitness values. The goal is to give a higher chance to those genes with high fitness so that only the best performing members of the population will survive in the long run and will be able to pass their information to the next generations. Because of the random nature of the roulette wheel selection method, however, genes with low fitness values will also have some chance to be selected. Also, there will be genes that are selected multiple times and some genes will not be selected at all and will be thrown out of the gene pool. After the selection procedure is completed, the selected genes are allowed to mate top-down in pairs whereby the first gene mates with the second gene and the third one with the fourth one and so on as illustrated in the following example:

Parents

$$S_1 = (A_{347} * A_{251}) \# + (A_{379} + A_{218}) \qquad (1)$$

$$S_2 = (A_{225} * A_{478}) \# + (A_{343} / A_{250}) + (A_{451} - A_{358}) + (A_{231} - A_{458}) \qquad (2)$$

The points where the genes are cut for mating are indicated by #.

Offspring

$$S_3 = (A_{347} * A_{251}) + (A_{343} / A_{250}) + (A_{451} - A_{358}) + (A_{231} - A_{458}) \qquad (3)$$

$$S_4 = (A_{225} * A_{478}) + (A_{379} + A_{218}), \qquad (4)$$

where A_{347} represents the absorbance at the wavelength given in subscript, S_1 and S_2 represent the first and second parent gene scores and S_3 and S_4 are the corresponding gene scores for the offspring. Here the first part of S_1 is combined with the second part of the S_2 to give the S_3, likewise the second part of the S_1 is combined with the first part of the S_2 to give S_4. This process is called the single point crossover and is common in GR. Single point crossover will not provide different offspring if both parent genes are identical, which may happen in roulette wheel selection, when both genes are broken at the same point. Also note that mating can increase or decrease the number of base pairs in the offspring genes. After crossover, the parent genes are replaced by their offspring and the offspring are evaluated. The ranking process is based on their fitness values following the evaluation step. Then, the selection for breeding/mating starts all over again. This is repeated until a predefined number of iterations are reached. Since GR is an iterative procedure it is possible

that the model can easily be over fitted for the calibration samples but fails for the independent validation samples. In order to avoid this problem, the algorithm is set to run with a leave one out cross validation step for the samples in the calibration set.

Mutation which introduces random deviations into the population was also introduced into the GR during the mating step at a rate of 1% as is typical in GAs. This is usually done by replacing one of the base pairs in an existing gene with a randomly generated new base pair. Mutation allows the GR to explore the search space and incorporate new material into the genetic population. It helps to keep the search moving and can eject GR from a local minimum on the response surface. However, it is important not to set the mutation rate too high since it may keep the GA from being able to exploit the existing population. In the end, the gene with the lowest SEC obtained from cross validation is selected for the model building. This model is used to predict the concentrations of component being analyzed in the independent validation (test) sets. The success of the model in the prediction of the validation set is evaluated using standard error of prediction (SEP). The termination of the algorithm can be done in many ways. The easiest way is to set a predefined iteration number for the number of breeding/mating cycles.

GR has some major advantages over classical univariate and multivariate calibration methods. It is a hybrid calibration method that uses the full spectral information and reduces it to a single score upon which simple calibration models are built. First of all, it is as simple as univariate calibration in terms of the mathematics involved in the model building and prediction steps, but at the same time it has the advantages of the multivariate calibration methods since it uses the full spectrum to extract genetic scores. It automatically corrects baseline fluctuations with the use of simple mathematical operators while forming the base pairs. Note that no data pretreatment is necessary before calibration, which saves time in the data processing.

Genetic Classical Least Squares

The classical least squares (CLS) method extends the classical Beer's Law model in which the absorbance at each wavelength is directly proportional to the component concentrations. Model errors are assumed to be in the measurement of the instrument responses as it was in the classical univariate method. In matrix notation, the CLS model for m calibration samples containing l chemical components whose spectra contain n wavelengths is described as:

$$A = CK + E_A \qquad (5)$$

Where A is the $m \times n$ matrix of the calibration spectra, C is the $m \times l$ matrix of the component concentrations, K is the $l \times n$ matrix of absorptivity-pathlength constants and E_A is the $m \times n$ matrix of the spectral errors or residuals not fit by the model. Here the K matrix represents the first order estimates of the pure component spectra at unit concentration and unit pathlength. The method of least-squares can be used to estimate the K matrix. The least-squares estimate of the K is defined as:

$$\hat{K} = (C'C)^{-1}C'A \qquad (6)$$

Once the estimated $\hat{K}$ matrix obtained, the concentrations of an unknown sample can be predicted from its spectrum by:

$$\hat{c} = (\hat{K}\hat{K}')^{-1}\hat{K}a \qquad (7)$$

Where a is the spectrum of the unknown sample and $\hat{c}$ is the vector of the predicted component concentrations. Because CLS uses full spectral information however, concentrations of all constituents must be known to be able to generate robust calibration models. For real samples, this is almost impossible since there are always impurities in the samples where no concentration information is available. Although selection of a sub set of wavelengths from full spectra may improve the prediction of calibration models, it may be the most time consuming part of the calibration depending on the method used to select wavelengths.

Genetic Classical Least Squares (GCLS) is a modified version of the original CLS method in which a small set of wavelengths are selected from full spectral data using a genetic algorithm. The algorithm used to select the optimum number of wavelengths in GCLS is quite similar to the GR algorithm, but differs in the way it encodes the gene. In GCLS, the term 'gene' describes a vector whose elements are selected wavelengths. The size of the vector is also determined in a random fashion with an upper limit to reduce computation time.

In the initialization step, an even number of genes are formed from full a spectral data matrix and each gene is used to form a CLS model. These models are then evaluated and ranked using the fitness function described in GR. The roulette wheel method is then used to select the gene population for breeding. After the selection procedure is completed, the selected genes are allowed to mate top-down without ranking whereby the first gene mates with second gene and third one with fourth one and so on as described in above with one difference. Since the genes used in GCLS are only vectors of wavelengths and contain no base pairs as described in GR, for each gene a random number is generated between 1 and one less than the length of the gene and the single point crossover process is performed using this number. After crossover, the parent genes are replaced by their offspring and the offspring are evaluated. The ranking process is based on their fitness values and follows the evaluation step. Then the selection for breeding/mating starts all over again. This is repeated until a predefined number of iterations are reached. In each iteration, the best gene with the lowest SEC is stored in order to compare it with the best gene of the next generation. If the next generation produces a better gene then it is replaced with the older one; otherwise the old one is kept for further iterations. At the end, the gene with the lowest SEC is selected for model building. This model is used to predict the concentrations of component being analyzed in the validation (test) sets as described in GR.

Genetic Inverse Least Squares

The major drawback of the CLS is that all of the interfering species must be known and their concentrations included in the model. This need can be eliminated by using the inverse least squares (ILS) method which uses the inverse of Beer's Law. In the ILS method, concentration of a component is modeled as a function of absorbance measurements. Because modern spectroscopic instruments are very stable and provide excellent signal-to-noise (S/N) ratios, it is believed that the majority of errors lie in the reference values of the calibration sample, not in the measurement of their spectra. In fact, in many cases the concentration data of calibration set are generated from another analytical technique that already have its inherent errors which might be higher than those of the spectrometer (for example, Kjeldahl protein analysis used to calibrate NIR spectra).

The ILS model for m calibration samples with n wavelengths for each spectrum is described by:

$$C = AP + E_c \qquad (8)$$

where **C** and **A** are the same as in CLS, **P** is the $n \times l$ matrix of the unknown calibration coefficients relating l component concentrations to the spectral intensities and E_c is the $m \times l$ matrix of errors in the concentrations not fit by the model. In the calibration step, ILS minimizes the squared sum of the residuals in the concentrations. The biggest advantage of ILS is that equation 8 can be reduced for the analysis of single component at a time since analysis is based on an ILS model is invariant with respect to the number of chemical components included in the analysis. The reduced model is given as:

$$c = Ap + e_c \qquad (9)$$

where **c** is the $m \times 1$ vector of concentrations for the analyte that is being analyzed, **p** is $n \times 1$ vector of calibration coefficients and e_c is the $m \times 1$ vector of concentration residuals not fit by the model. During the calibration step, the least-squares estimate of **p** is:

$$\hat{p} = (A'A)^{-1} A' \cdot c \qquad (10)$$

where $\hat{p}$ is the estimated calibration coefficients. Once $\hat{p}$ is calculated, the concentration of the analyte of interest can be predicted with the equation below.

$$\hat{c} = a' \cdot \hat{p} \qquad (11)$$

where $\hat{c}$ is the scalar estimated concentration and **a** is the spectrum of the unknown sample. The ability to predict one component at a time without knowing the concentrations of interfering species has made ILS one of the most frequently used calibration methods.

The major disadvantage of ILS is that the number of wavelengths in the calibration spectra should not be more than the number of calibration samples. This is a big restriction since the number of wavelengths in a spectrum will generally be much more than the number of calibration samples and the selection of wavelengths that provide the best fit for the model is not a trivial process. Several wavelength selection strategies, such as stepwise wavelength selection and all possible combination searches, are available to build an ILS model that fits best the data. Here we used the same genetic algorithm described in GCLS to build genetic inverse least squares (GILS) models with one difference. This difference is in the way the mating and single point crossover operations are carried out. Because the number of wavelengths is restricted in absorbance data matrix in the ILS, the size of the largest gene is restricted to one less than the number of calibration samples in the concentration vector. However, if the single point crossover is set to take place in any point of a gene, then the mating step could produce new genes that have a larger number of wavelengths than the number of calibration samples even though all the genes in the initial gene pool were set to have smaller number of wavelengths than the size of the concentration vector. In order to avoid this problem, the crossover operation is performed around the middle of each gene in GILS so that the new generations will have a small chance to have larger sizes than the number of calibration samples. The rest of the algorithm is the same as the one used in GCLS.

Table 1 and table 2 show concentration profiles of calibration and validation samples, respectively. In addition, 16 samples of commercial tablets having each constituent were used to build prediction set. UV spectra of pure metamizol (20 µg/mL), paracetamol (16 µg/mL), and caffeine (19 µg/mL) in 0.1 M HCl along with a ternary mixture of the three components between 200 and 320 nm wavelength range are shown in figure 1. As seen from the figure, paracetamol and metamizol show maximum absorbance around 242 and 259nm, respectively whereas pyridoxine has a maximum absorbance around 272 nm. Their mixture spectrum, however, indicates strong overlap over the entire region which indicates that the use of multivariate methods would be needed to resolve these components. Through the genetic multivariate calibration process, it is expected that these overlaps will be resolved and reveal the information necessary to build successful calibration models otherwise almost impossible with univariate calibration methods.

Several calibration models were generated with the four methods and table 3 shows the results of ternary mixtures for calibration and validation sets. Here, the CLS method was applied to whole spectrum data set and in the case of

Table 1

COMPOSITION OF THE CALIBRATION SET CONTAINING CAF, PAR AND MET COMPOUNDS

Sample Number	µg/mL CAF	µg/mL PAR	µg/mL MET	Sample Number	µg/mL CAF	µg/mL PAR	µg/mL MET
1	12.0	0.0	0.0	17.0	20.0	16.0	16.0
2	20.0	0.0	0.0	18.0	20.0	20.0	16.0
3	32.0	0.0	0.0	19.0	20.0	24.0	16.0
4	40.0	0.0	0.0	20.0	20.0	32.0	16.0
5	48.0	0.0	0.0	21.0	20.0	40.0	16.0
6	12.0	16.0	16.0	22.0	0.0	0.0	4.0
7	20.0	16.0	16.0	23.0	0.0	0.0	12.0
8	32.0	16.0	16.0	24.0	0.0	0.0	19.0
9	40.0	16.0	16.0	25.0	0.0	0.0	32.0
10	48.0	16.0	16.0	26.0	0.0	0.0	40.0
11	0.0	8.0	0.0	27.0	20.0	16.0	4.0
12	0.0	16.0	0.0	28.0	20.0	16.0	12.0
13	0.0	20.0	0.0	29.0	20.0	16.0	19.0
14	0.0	32.0	0.0	30.0	20.0	16.0	32.0
15	0.0	40.0	0.0	31.0	20.0	16.0	40.0
16	20.0	8.0	16.0				

Table 2

COMPOSITION OF THE PREDICTION SET CONTAINING CAF, PAR AND MET COMPOUNDS

Sample Number	CAF	PAR	MET	Sample Number	CAF	PAR	MET
	μg/mL				μg/mL		
1	12.0	16.0	16.0	9.0	20.0	24.0	16.0
2	20.0	16.0	16.0	10.0	20.0	32.0	16.0
3	32.0	16.0	16.0	11.0	20.0	40.0	16.0
4	40.0	16.0	16.0	12.0	20.0	16.0	4.0
5	48.0	16.0	16.0	13.0	20.0	16.0	12.0
6	20.0	8.0	16.0	14.0	20.0	16.0	19.0
7	20.0	16.0	16.0	15.0	20.0	16.0	32.0
8	20.0	20.0	16.0	16.0	20.0	16.0	40.0

Table 3

RESULTS OF TERNARY MIXTURES IN CALIBRATION AND PREDICTION SETS CONTAINING CAFFEINE, PARACETAMOL AND METAMIZOLE OBTAINED WITH FOUR MULTIVARIATE CALIBRATION METHODS (CLS, GCLS, GILS AND GR)

Component	Data Sets	Parameters	Methods			
			CLS	GCLS	GILS	GR
Caffeine	Calibration	SEC (μg/ml)	1.97	0.35	0.29	0.37
		Recovery (%)	107.52	101.25	100.5	100.78
		RSD	13.13	3.42	2.01	3
	Prediction	SEP (μg/ml)	2.27	0.35	0.25	0.32
		Recovery (%)	105.53	100.96	99.99	100.39
		RSD	14.08	3.57	1.3	3.05
Paracetamol	Calibration	SEC (μg/ml)	0.63	0.21	0.04	0.68
		Recovery (%)	100.07	100.22	100.04	100.06
		RSD	3.76	1.28	0.34	3.33
	Prediction	SEP (μg/ml)	0.59	0.24	0.03	0.48
		Recovery (%)	100.75	100.19	99.98	99.12
		RSD	4.02	1.46	0.2	2.93
Metamizol	Calibration	SEC (μg/ml)	2.34	0.32	0.15	0.32
		Recovery (%)	99.99	100.15	100.04	100.29
		RSD	8.17	1.11	0.66	1.36
	Prediction	SEP (μg/ml)	1.78	0.32	0.13	0.37
		Recovery (%)	97.58	99.84	100.01	100.43
		RSD	8.1	1.05	0.67	1.43

SEC: Standard Error of Calibration, SEP: Standard Error of Prediction, RSD: Relative Standard Deviation, Recovery (%) : Percent Average Recovery

genetic algorithm based methods (GCLS, GILS and GR) the algorithms were set to run with 20 gene and 100 iterations. The results given in table for GCLS, GILS and GR are from the runs that generate the lowest SEC. Then these models were used later to predict the actual tablet samples and compared with each other based on their success of predicting actual samples as shown in table 4–6.

A close examination of the results given in table 3 indicates that GR, GCLS and GILS methods generates very similar results for caffeine and metamizol in which the standard error of calibration (SEC) and standard error of prediction values ranging between 0.04 and 0.37 μg/mL. On the other hand, the CLS method generated relatively high SEC and SEP results ranging between 1.78 and 2.34 μg/mL for the same components. The same conclusion can also be made for the average recovery and relative standard deviation (RSD) of average recovery values. The reason why CLS generated relatively larger results even for the synthetic samples could be the fact that the whole spectra of the ternary samples are strongly overlapped. On the other hand, the rest of the three methods are based on the wavelength selection which gives an advantage of possibility to improve the prediction ability of calibration models through the evolutionary process. The results for the component paracetamol were quite different for all four methods in which the GILS method generated the lowest calibration and prediction errors.

In general, it seems like GILS produces better results than other three. However, this could be a very misleading conclusion if one considers the results given in table 4–6 where the results of metamizol, paracetamol and caffeine

Sample No.	Predicted Metamizol (mg/ tablet)							
	Remidon®tablet				Pirosal® tablet			
	CLS	GCLS	GILS	GR	CLS	GCLS	GILS	GR
1	214.98	202.2	202.06	199.5	209.56	218.42	221.09	222.4
2	215.43	203.9	204.53	205.43	209.06	219.49	220.49	214.59
3	215.77	202.3	203.1	201.15	208.91	220.24	219.87	213.98
4	216.53	202.5	203.05	201.31	208.9	217.73	220.25	219.65
5	216.55	202.7	203.88	203.6	209.42	219.72	219.34	216.69
6	218.63	208	208.4	207.25	209.69	220.6	220.73	213.9
7	214.3	198.8	201.18	197.57	209.84	219.81	222.14	218.89
8	209.83	194.1	196.45	189.11	211.09	222.18	222.39	220.12
mean	215.25	201.8	202.82	200.62	209.56	219.77	220.79	217.53
SD	2.55	4	3.36	5.6	0.71	1.35	1.06	3.2
RSD	1.18	1.98	1.66	2.79	0.34	0.61	0.48	1.47
SEP	15.44	4.16	4.23	5.28	10.46	1.28	1.26	3.88

Table 4
RESULTS OF TWO DIFFERENT COMMERCIAL PHARMACEUTICAL DOSAGE FORMS CONTAINING METAMIZOL (200 mg REMIDON® TABLET) AND (220 mg PIROSAL® TABLET) OBTAINED WITH FOUR MULTIVARIATE CALIBRATION METHODS (CLS, GCLS, GILS AND GR)

SD: Standard Dviation, RSD: Relative Standard Deviation, SEP: Standard Error of Prediction.

Sample No.	Predicted Paracetamol (mg/ tablet)							
	Remidon® tablet				Pirosal® tablet			
	CLS	GCLS	GILS	GR	CLS	GCLS	GILS	GR
1	211.19	209.4	208.37	205.33	165.58	160.34	159.17	154.93
2	214.39	212.2	210.37	208.66	166.97	161.26	159.23	156.88
3	213.59	211.7	210.77	207.1	168.14	163.27	160.08	156.87
4	209.79	208.1	205.28	204.54	166.2	160.6	158.66	154.57
5	210.3	208.4	206.28	204.19	168.02	163.17	160.86	156.53
6	216.93	214.3	211.85	211.32	167.95	162.86	161.43	158.57
7	204.48	203.1	201.07	200.7	166.53	161.33	157.3	156.08
8	207.41	204.3	201.35	202.44	166.83	162.39	160.53	155.63
mean	211.01	208.9	206.92	205.53	167.03	161.9	159.66	156.26
SD	3.98	3.86	4.16	3.41	0.94	1.16	1.34	1.26
RSD	1.89	1.85	2.01	1.66	0.56	0.72	0.84	0.81
SEP	11.62	9.64	7.94	6.39	7.08	2.19	1.3	3.92

Table 5
RESULTS OF TWO DIFFERENT COMMERCIAL PHARMACEUTICAL DOSAGE FORMS CONTAINING PARACETAMOL (200 mg REMIDON® TABLET) AND (160 mg PIROSAL® TABLET) OBTAINED WITH FOUR MULTIVARIATE CALIBRATION METHODS (CLS, GCLS, GILS AND GR)

SD: Standard Dviation. RSD: Relative Standard Deviation. SEP: Standard Error of Prediction.

in commercial tablet formulations are given, respectively. It can be seen that the three genetic algorithm based methods generate comparable results. However, it is evident that the CLS method is unable to predict the composition of actual tablets as good as the genetically modified multivariate methods. The reason for this result is that the composition of the actual tablet samples is different than the synthetic calibration samples. In addition, the GILS method produced relatively high SEP values for the tablets even though it generated the lowest SEC values for the synthetic mixtures which might be the indication of slightly over fitted models. In fact, while GR method performs worst than GILS in the determination of synthetic samples, it generated better prediction results for the actual tablets. One possible explanation of this is that GR is able to do automatic background correction by using subtraction operator in base pairs whereas GCLS and GILS are not able to do this. Another reason could be the possible nonlinearities in the spectra that GR may be able to correct but others are not. The regression coefficients (R^2) were ranged between 0.998 and 1.000 for the GR, GCLS and

GILS methods indicating very good fit between actual and predicted concentrations. In terms of the overall performance of the four methods it can be said that the genetically modified methods perform better than CLS for actual tablet samples.

In order to determine whether the genetic algorithm selected wavelengths corresponds the particular component absorbance region, the GCLS, GILS and GR methods were also set to run 50 times with 20 genes and 100 iterations. The overall distribution of the selected wavelengths for each component along with a pure component spectrum is shown in figure 2 for GR method. As can be seen from the figure, the genetic regression method select the wavelengths that correspond the each component absorption region even though the algorithm starts with the whole spectrum information at the beginning of each run and each wavelength has equal chance of being selected. The explanation of this is in the evolutionary nature of GR method where the wavelength that is suited for the particular component survives in the long run of iterations and others do not. This gives an

Table 6

RESULTS OF TWO DIFFERENT COMMERCIAL PHARMACEUTICAL DOSAGE FORMS CONTAINING
CAFFEINE (50 mg REMIDON® TABLET) AND (30 mg PIROSALÂ TABLET) OBTAINED WITH FOUR
MULTIVARIATE CALIBRATION METHODS (CLS, GCLS, GILS AND GR)

Sample No.	Predicted Caffeine (mg/ tablet)							
	Remidon® tablet				Pirosal® tablet			
	CLS	GCLS	GILS	GR	CLS	GCLS	GILS	GR
1	64.33	51.08	51.69	49.3	33.57	30	31.42	30.13
2	65.73	52.14	52.72	50.57	33.54	30.27	30.78	29.93
3	64.4	51.4	52.49	49.89 ·	33.46	30.26	31.26	30.25
4	65.97	52.47	52.98	50.8	33.21	29.69	31.01	29.85
5	63.95	50.52	51.09	49.03	33.14	30.42	31.43	30.35
6	65.89	52.66	53.18	51.05	33.56	30.83	31.65	30.69
7	62.02	49.12	49.6	47.55	33.19	30.13	31.29	30.21
8	64.33	50.42	50.78	48.94	32.56	30.14	31.14	29.93
mean	64.58	51.22	51.82	49.64	33.28	30.22	31.25	30.17
SD	1.31	1.2	1.26	1.17	0.34	0.33	0.27	0.28
RSD	2.03	2.34	2.43	2.36	1.02	1.09	0.87	0.91
SEP	14.63	1.66	2.16	1.15	3.29	0.38	1.27	0.31

SD: Standard Dviation, RSD: Relative Standard Deviation, SEP: Standard Error of Prediction.
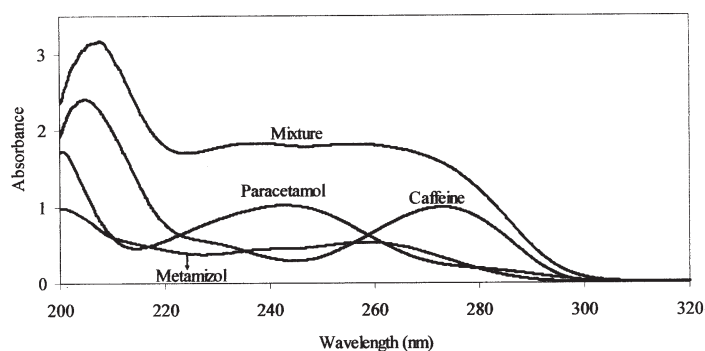

Fig.1. UV spectra of pure Metamizol
(20 µg/mL), Paracetamol (16 µg/mL), and Caffeine
(19 µg/mL) in 0.1 M HCl along with a ternary
mixture of the three components between 200
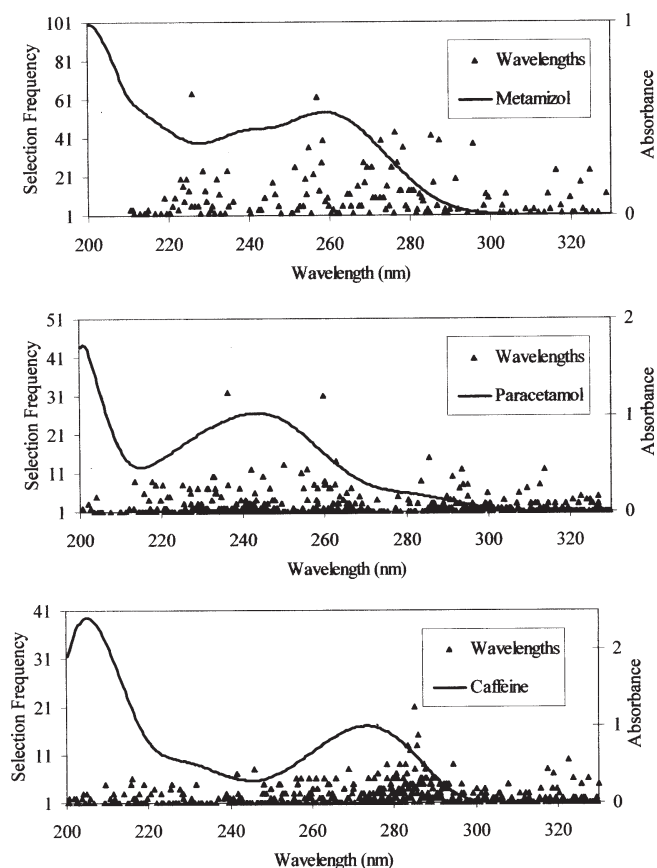and 320 nm wavelength range

Fig.2. Frequency of the wavelengths selected by
GR method in a total of 100 runs for Metamizol,
Paracetamol and Caffeine along with UV spectra of
each component between 200 and 320 nm
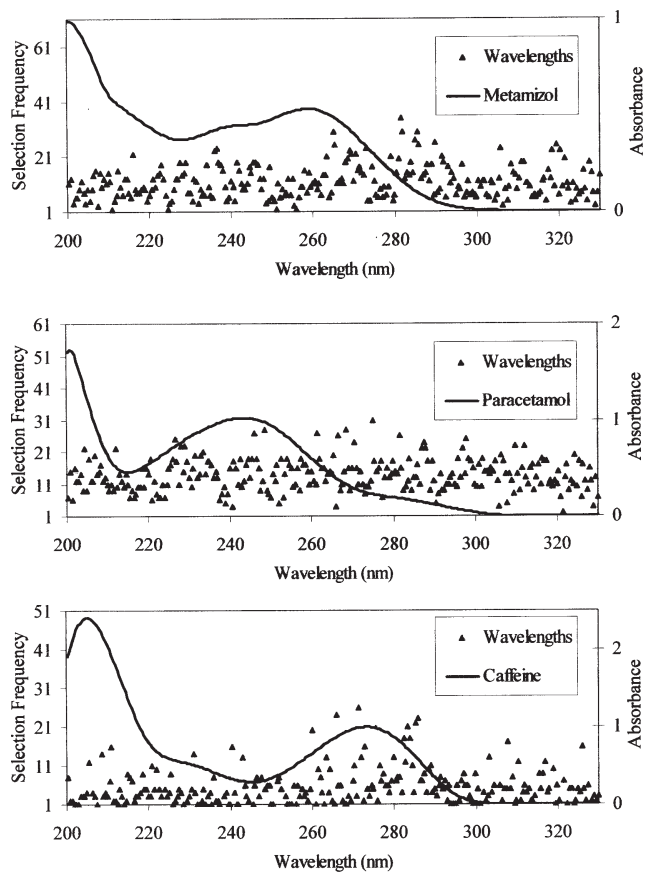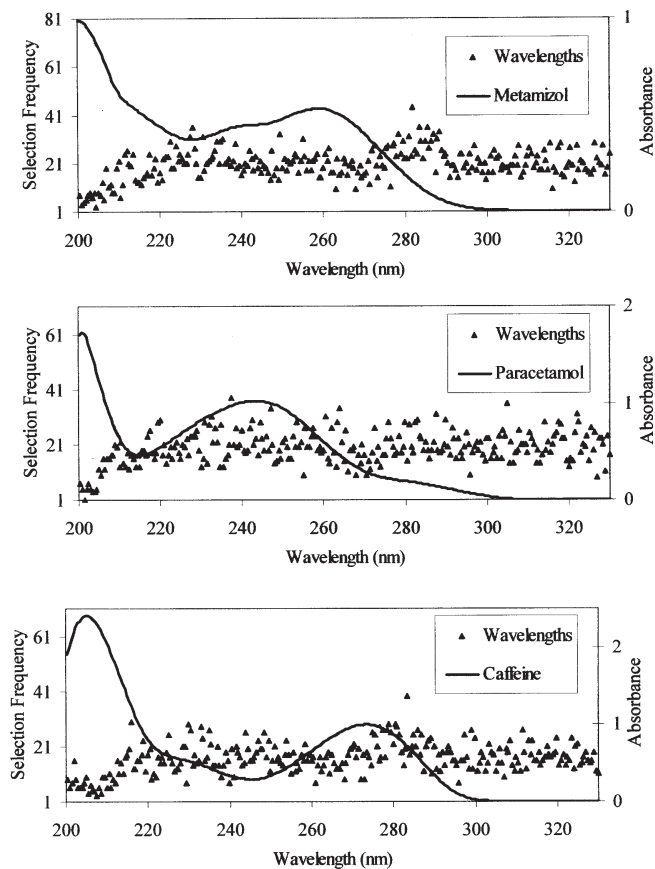wavelength range.

http://www.revistadechimie.ro

Fig.3. Frequency of the wavelengths selected by GILS method in a total of 100 runs for Metamizol, Paracetamol and Caffeine along with UV spectra of each component between 200 and 320 nm wavelength range.

Fig. 4. Frequency of the wavelengths selected by GCLS method in a total of 100 runs for Metamizol, Paracetamol and Caffeine along with UV spectra of each component between 200 and 320 nm wavelength range.

advantage to the genetic algorithm based methods where only the information related to the particular component are used to construct the model thereby reducing the noise in the overall information.

Figure 3 and Figure 4 show the overall distribution of the selected wavelengths for each component along with a pure component spectrum obtained with GILS and GCLS, respectively. However, these two genetic algorithm based method do not seem to do a component specific wavelength selection as opposed to GR method. The reason for this is the difference in the form of the genes in GR method and GILS and GCLS methods. Genes for GR contains base pairs whereas genes for GILS and GCLS are just vectors of wavelengths.

Conclusions

This study illustrates the application of the CLS and three genetic algorithms based multivariate calibration methods to simultaneous determination of pharmaceuticals in synthetic and actual tablet formulations. It can be said that GR, GCLS and GILS methods generate acceptable results in the given concentration range of the components. These methods coupled with UV-visible spectroscopy could be an alternative to other methods such as chromatography, which is more expensive and time consuming.

References

1.M. MARCELO, R.J. SENA, J. POPPI, J. Pharm. Biomed. Anal., **34,** 2004,p. 27
2.DINÇ, E., J. Pharm. Biomed. Anal., **33,** 2003, p. 605
3.DINÇ, E., SERIN, C. TUGCU-DEMIROZ, F., DOGANAY, T., Int. Journal of Pharm., **250,** 2003, p. 339
4.DINÇ,E., F. Onur, Anal. Chim. Acta., **359**, 998, p. 93
5.DINÇ, E., Talanta, **48**, 1999, p. 1145
6.FRANETA, T., AGBABA,D., ERIC, S., PAVKOV, S., ALEKSIC,M., VLADIMIROV, S., Il Farmaco, **57**, 2002, p. 709
7.KARTAL, M., J. Pharm. Biomed. Anal., **26**, 2001, p. 857
8. ABU-QARE,A.W., ABOU-DONIA, M.B., J. Pharm. Biomed. Anal., **26**, 2001 , p.939
9.HAALAND, D.M. , THOMAS,E.T., Anal. Chem ., **60**, 1988, p. 1193
10.GELADI, P., KOWALSKI,B.R., Anal. Chim. Acta., **185**, 1986, p. 1
11.WENTZELL, P.D., ANDREWS, D.T., KOWALSKI, B.R.,Anal. Chem., **69**, 1997, p. 2299
12.ESBENSEN, K., GELADI, P., WOLD, S., Chem. Intell. Lab. Syst., **2**, 1987, p. 37
13. ÖZDEMIR, D., ÖZTÜRK, B., Turk. J. Chem. ,**28,** 2004, p. 497
14. ÖZDEMIR , D., DINÇ, E., Chem. Pharm. Bull. ,**52**, 7, 2004, p. 810.
15.ÖZDEMIR,D., MOSLEY, R.M , WILLIAMS, R.R., Appl. Spectr., **52**, 1998, p. 599
16.D ÖZDEMIR, D., MOSLEY, , WILLIAMS,R.R., Appl. Spectr., **52**, 1998, p. 1203
17.ÖZDEMIR, D., WILLIAMS, R.R.,Appl. Spectr. ,**53**, 1999, p. 210
18.LUCASIUS, C.B., KATEMAN, G., Chem. Intell. Lab. Syst., **19**, 1993, p. 1
19. HÖRCHNER, U., KALIVAS, J.H., Anal. Chim. Acta., **311**, 1995, p. 1.
20.RIMBAUD, D.J. , LEARDI, R.,DE NOORD, O.E., MASSART, D.L., Anal. Chem., **67,** 1995, p. 4295
21.BANGALORE, A.S., SHAFFER, R.E., ARNOLD, M.A., SMALL, G.S., Anal. Chem., **68**,1996,p. 4200
22.HARTNETT, M., DIAMOND, D.,Anal. Chem., **69**, 1997) p. 1909