Near Infrared Spectroscopic Determination of Diesel Fuel Parameters Using Genetic Multivariate Calibration

D. Özdemir

Near Infrared Spectroscopic Determination of Diesel Fuel Parameters Using Genetic Multivariate Calibration

D. Özdemir[1]

[1]Department of Chemistry, Izmir Institute of Technology, Urla/Izmir, Turkey

Abstract: The use of full spectral region from near infrared spectroscopic analysis does not always end up with a good multivariate calibration model as many of the wavelengths do not contain necessary information. Due to the complexity of the spectra, some of the wavelengths or regions may, in fact, disturb the model-building step. Genetic algorithms are one of the useful tools for solving wavelength selection problems and may improve the predictive ability of conventional multivariate calibration methods. This study demonstrates application of genetic algorithm-based multivariate calibration to near infrared spectroscopic determination of several diesel fuel parameters. The parameters studied are cetane number, boiling and freezing point, total aromatic content, viscosity, and density. Multivariate calibration models were generated using genetic inverse least squares (GILS) method and used to predict the diesel fuel parameters based on their near infrared spectra. For each property, a different data set was used and in all cases the number of samples was around 250. Overall, percent standard error of prediction (%SEP) values ranged between 2.48 and 4.84% for boiling point, total aromatics, viscosity, and density. However, %SEP results for cetane number and freezing point were 11.00% and 14.86%, respectively.

Keywords: diesel fuel, genetic algorithms, multivariate calibration, near infrared spectroscopy

INTRODUCTION

Advances in computers and automation technology have made today's instruments incredibly fast, so that they can produce hundreds of spectra in a few minutes for a given sample that contains multiple components. Unfortunately, univariate calibration methods are not suitable for this type of data, as they require an interference-free system. Thanks to the chemometrics, multivariate

Address correspondence to Durmus Ozdemir, Department of Chemistry, Faculty of Science, Izmir Institute of Technology, Gülbahce 35430 Urla/Izmir, Turkey. E-mail: durmusozdemir@iyte.edu.tr

calibration methods make it possible to relate instrument responses that consist of several predictor variables to a chemical or physical property of a sample. Several classical multivariate calibration methods have been developed in last couple of decades (Lindberg et al., 1983; Geladi and Kowalski, 1986; Haaland and Thomas, 1988; Wentzell et al., 1997) for the analysis of complex chemical mixtures, and the choice of the most suitable calibration method is very important in order to generate calibration models with high predictive ability for future samples. In some cases, conventional methods may not offer a satisfactory solution to a given problem due to the complexity of the data, and it may be necessary to apply some sort of variable selection. There have been many mathematical methods of variable selection (Lindgren et al., 1994; Centner et al., 1996; Forina et al., 1999), and genetic algorithm is one of them offering a fast and effective solution for large scale problems (Leardi et al., 1992; Lucasius and Kateman, 1993; Hörchner and Kalivas, 1995).

Inverse Least Squares (ILS) is based on the inverse of Beer's Law, where concentrations of an analyte are modeled as a function of absorbance measurements. Genetic Inverse Least Squares (GILS) is modified versions of original 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 (GA), and has been applied to a number of wavelength selection problems (Özdemir and Dinç, 2004; Özdemir and Öztürk, 2004; Özdemir, 2005). GAs are non-local search and optimization methods that are based upon the principles of natural selection (Hibbert, 1993; Paradkar and Williams, 1997; Pizarro et al., 1998; Mosley and Williams, 1998; Özdemir and Williams, 1999).

Near infrared (NIR) spectroscopy (McClure, 1994) has become a popular method for simultaneous chemical analysis and is being studied extensively in a number of different fields, such as process monitoring (DeThomas et al., 1994), biotechnology (Arnold et al., 2000), and pharmaceutical and food industry (Wählby and Skjöldebrand, 2001; Tran et al., 2004) because of the potential for on-line, nondestructive and noninvasive instrumentation. The NIR portion of the electromagnetic spectrum covers the range from 780 nm to 2,500 nm, and most of the absorption bands observed in this region are due to overtones and combinations of the fundamental mid-IR molecular vibrational bands. Although all the fundamental vibrational modes can have overtones, the most commonly observed bands arise from the C−H, O−H, and N−H bonds in the molecules.

A broad description of the six properties of the diesel fuel data set studied here was given elsewhere (Soyemi et al., 2000), and detailed information about how these properties are determined by standard reference methods was provided. In this work, genetic algorithm-based calibration method genetic inverse least squares (GILS) was tested with the aim of establishing calibration models that have a high predictive ability for the near infrared spectroscopic determination of several chemical and physical parameters of diesel fuels.

GENETIC INVERSE LEAST SQUARES

The major drawback of the classical least squares (CLS) method 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 is generated from another analytical technique that already has 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:

$$\mathbf{C} = \mathbf{AP} + \mathbf{E_C} \tag{1}$$

where $\mathbf{C}$ is the $m \times l$ matrix of the component concentrations, $\mathbf{A}$ is the $m \times n$ matrix of the calibration spectra, $\mathbf{P}$ is the $n \times l$ matrix of the unknown calibration coefficients relating l component concentrations to the spectral intensities, and $\mathbf{E_C}$ 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 Eq. (1) can be reduced for the analysis of a 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:

$$\mathbf{c} = \mathbf{Ap} + \mathbf{e_c} \tag{2}$$

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

$$\hat{\mathbf{p}} = (\mathbf{A'A})^{-1}\mathbf{A'} \cdot \mathbf{c} \tag{3}$$

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

$$\hat{c} = \mathbf{a'} \cdot \hat{\mathbf{p}} \tag{4}$$

where $\hat{c}$ is the scalar estimated concentration and $\mathbf{a}$ 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 which fits the data best.

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 inverse least squares (GILS) is an implementation of a GA for selecting wavelengths to build multivariate calibration models with reduced data sets. GILS follows the same basic initialize/breed/mutate/evaluate algorithm as other GAs to select a subset of wavelengths 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 responses at the wavelength range given in the data set. 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, GILS 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 predicted results of initial population against the actual 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 instrumental responses 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 overfitting 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

$$S_1 = (A_{347}, A_{251}, \#A_{379}, A_{218}) \tag{5}$$

$$S_2 = (A_{225}, A_{478}, \#A_{343}, A_{250}, A_{451}, A_{358}, A_{231}, A_{458}) \tag{6}$$

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

Offspring

$$S_3 = (A_{347}, A_{251}, A_{343}, A_{250}, A_{451}, A_{358}, A_{231}, A_{458}) \tag{7}$$

$$S_4 = (A_{379}, A_{218}, A_{225}, A_{478}) \tag{8}$$

where A_{347} represents the instrument response at the wavelength given in subscript, S_1 and S_2 represent the first and second parent genes, and S_3 and S_4 are the corresponding genes 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 GILS. 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 instrument responses 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 is reached.

Mutation, which introduces random deviations into the population, was also introduced into the GILS during the mating step at a rate of 1%, as is typical in GAs. This is usually done by replacing one of the responses in an existing gene with a randomly selected new one. Mutation allows the GR to explore the search space and incorporate new material into the genetic population. It helps keep the search moving and can eject GILS 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. Also, the GILS method is an iterative algorithm, and, therefore, there is a high possibility that the method may easily over-fit the calibration data so that the predictions for independent sets might be poor. To eliminate possible over-fitting problems, cross validation is used in which one spectrum is left out of the calibration set and the model is constructed with $m - 1$ sample. Then this model is used to predict the concentration of left-out sample. This process is continued until all samples are left out at least once in each iteration. As long as the number of spectra in the calibration set is not too large, cross validation is an effective method of eliminating over-fitting. If the number of calibration spectra is very large, then the GILS method has the option of a half-validation approach in which the half of the spectra in the calibration set is used to validate the model in each iteration.

In the end, the gene with the lowest SEC (highest fitness) is selected for the model building which is done by simple least squares. This model is used to predict the concentrations of component being analyzed in the prediction (test) sets. The success of the model in the prediction of the test sets is evaluated using standard error of prediction (SEP). Because random processes are heavily involved in GR, as in all the GAs, the program has been set to run several times for each component in a given multi-component mixture during the course of this study. The best run (i.e., the one generating the lowest SEC for the calibration set and at the same time producing SEPs for prediction sets that are in the same range with the SEC) is subsequently selected for evaluation and further analysis. 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.

GILS has some major advantages over classical univariate and multi-variate calibration methods. First of all, it is quite simple 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 with a reduced data set since it uses the full spectrum to extract genes. By selecting a subset of instrument responses, it is able to eliminate nonlinearities that might be present in the full spectral region.

EXPERIMENTAL

The near infrared spectra of diesel fuel samples, together with six properties that were measured at the Southwest Research Institute, are obtained from the web site of Eigenvector Research Corporation (http://software.eigenvector. com/Data/index.html). It contains six different data sets for the parameters: cetane number, boiling point, freezing point, total aromatic content, viscosity, and density. Each set has a different number of samples (around 250), and the above physical and chemical properties of the samples were determined

independently using standard reference methods before the near infrared spectra were recorded. The diesel fuel data set investigated here has also been used in a number of previous studies (Leardi, 2000; Boger, 2003; Esteban-Diez et al., 2004) for testing new variable selection and calibration algorithms. Each spectrum in all data sets contains 401 distinct wavelengths, but the wavelength range was not indicated. The genetic inverse least squares (GILS) method was written in MATLAB programming language using Matlab 5.3 (MathWorks Inc, Natick, MA).

RESULTS AND DISCUSSION

The diesel fuel data set used in this study was selected to demonstrate the applicability of genetic multivariate calibration for the determination of some physical and chemical properties of diesel fuels that are very important in assigning the quality of the fuel using near infrared spectroscopy in place of expensive and time-consuming standard reference methods. Several different properties were investigated, and a different data set is used for each one of them. Table 1 shows these properties along with the number of calibration and prediction samples in each set. The composition of calibration and prediction sets was decided based on the information given in the description of the data set at Eigenvector Research database archive. It was reported that each data set contains 20 high-leverage samples which were included in the calibration set of our study. The remaining samples were split into two sets with the names ending with an extension of $-a$ and $-b$ for calibration and prediction sets, respectively. Figure 1 shows near infrared spectra of six different diesel fuel samples, one from each set. Because the main composition of all samples are similar, the spectral features of these fuel samples are very much alike, and only minute differences exist in some parts of the whole spectrum. Throughout the multivariate calibration process, it is expected that these differences will reveal the information necessary to build successful calibration models otherwise almost impossible with univariate calibration methods.

Table 1. The properties of the diesel fuel data sets investigated in this study along with the number of calibration and prediction samples in each set

| Properties	Cetane number	Boiling point, °C	Freezing point, °C	Total aromatics, w/w%	Viscosity, cSt	Density, g/mL
Number of calibration sample	133	133	136	138	136	142
Number of prediction sample	112	113	115	118	116	121
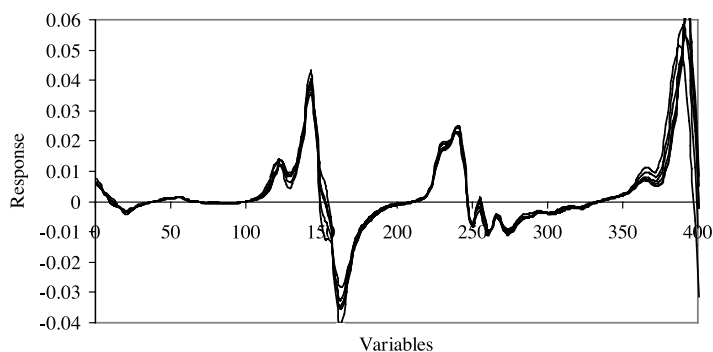
Figure 1. Near infrared spectra of six diesel fuel samples taken from six different data sets.

In order to prepare calibration models, six separate calibration and pre-diction sets were prepared for all the properties, as shown in Table 1. Because of the random nature of the GILS method, the program was set to run 30 times with 20 genes and 50 iterations. Since the GILS program is iterated 50 times in each run, it is possible that the algorithm may end up with an over-fit model which produces very poor results for the prediction set. On the other hand, the number of calibration samples for each property is around 135, and this number is too large to apply full cross validation methods to avoid possible over-fitting. To eliminate the problem, a half-validation approach was used in GILS method in which all the odd num-bered samples in the original calibration set were selected for the model building step, and the even numbered samples were reserved for model validation in each iteration. This approach not only eliminates the over-fitting but also significantly reduces the iteration time. Table 2 shows the

Table 2. Results of genetic inverse least squares (GILS) method for the six parameters of diesel fuel data

Parameters	Cetane number	Boiling point, °C	Freezing point, °C	Total aromatics, w/w%	Viscosity, cSt	Density, g/mL
Range	40–60	198–286	(−24)–(−0.2)	13–47	1.3–3.6	0.80-0.87
R^2	0.865	0.986	0.897	0.997	0.973	0.995
SEC	1.31	2.03	1.45	0.33	0.07	<0.01
%SEC	6.94	2.23	6.28	1.14	2.99	1.60
SEP	2.11	3.62	2.77	0.67	0.10	<0.01
%SEP	11.00	3.92	14.86	2.48	4.82	3.36
APR	100.15	100.01	105.19	100.32	99.93	100.03
SD	4.39	1.40	104.17	2.37	4.52	0.19
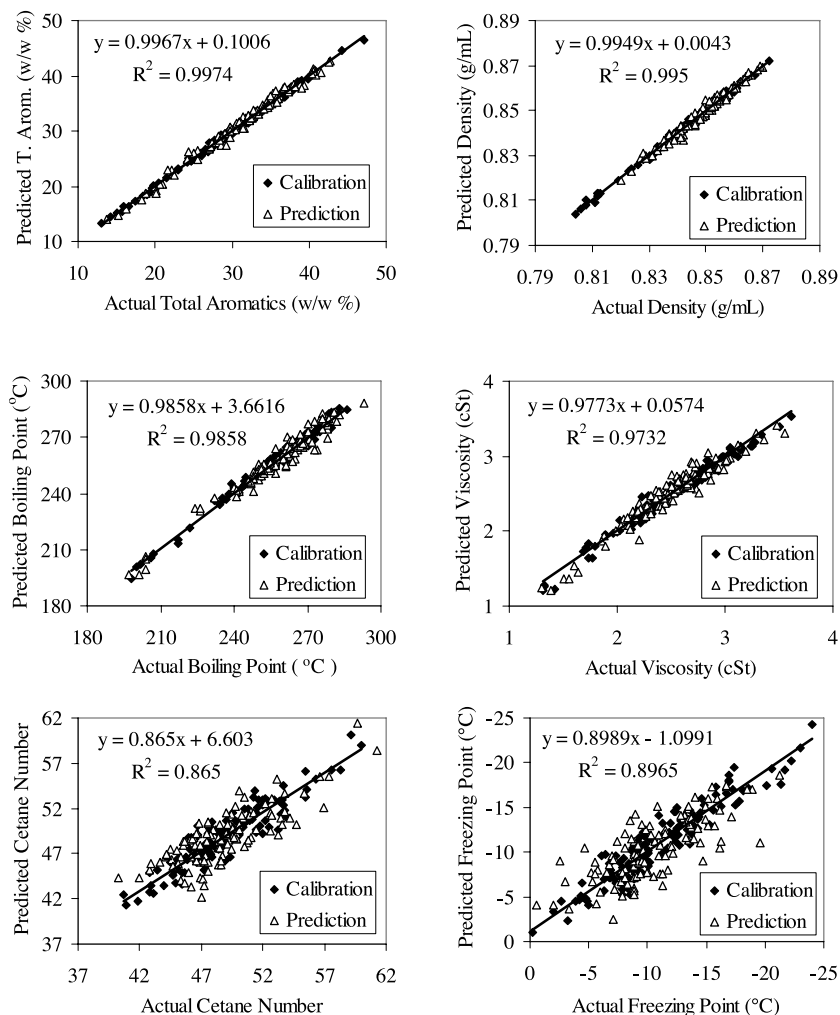
Figure 2. Actual versus GILS predicted properties of diesel fuel data.

results of the GILS method for the six parameters studied. To evaluate the model performance, standard error of calibration (SEC) and standard error of prediction (SEP) were used for calibration and prediction sets, respectively. These parameters can be dimensionally compared to the property studied. To be able to report the result free of dimensionality consideration, percentage of SEC (%SEC) and SEP (%SEP) were also reported. As can be seen from Table 2, the %SEC values were ranged between 1.14% and 6.94% and for the %SEP between 2.48 and 14.86% for all the properties. It is evident that the GILS method was able to generate successful calibration models for all the properties except for the determination of
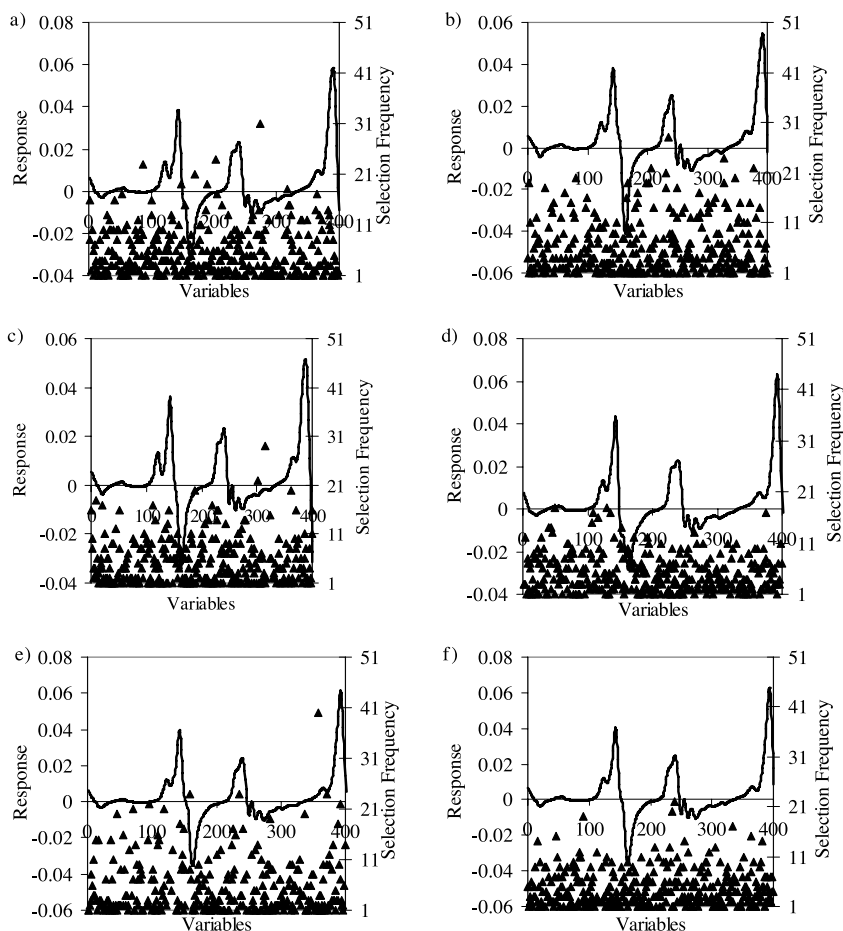
Figure 3. Frequency distribution of the genetic algorithm selected variables for the six properties of diesel fuel data: (a) total aromatics, (b) density, (c) boiling point, (d) viscosity, (e) cetane number, and (f) freezing point.

freezing point and cetane number. For these two properties, predictions were somewhat poor.

Since there are more than 100 samples in the prediction sets, it is not practical to illustrate the GILS predicted results in a table, so average percent recoveries (APR) and associated standard deviation (SD) values for the prediction sets are also given in the last two rows of Table 2. As can be seen, except for the freezing point, all other determinations were resulted with standard deviations ranging from 0.19 to 4.52. The plot of actual versus GILS predicted concentrations based on NIR data for six properties of diesel fuel data is illustrated in Figure 2. The R^2 values were ranged between 0.973 and 0.997 for total aromatics content, density, boiling point, and viscosity,

indicating very good correlation between predicted and actual values. For cetane number and freezing point, the values were 0.865 and 0.897, respectively. As a result, it can be said that NIR spectroscopy in conjunction with multivariate calibration can be used for fast determination of several physical and chemical properties of diesel fuels.

Because GILS is a wavelength selection based method, it is interesting to observe the distribution of selected wavelengths in multiple runs over the entire full spectral region. Figure 3 illustrates the frequency distribution of selected wavelengths in 30 runs for all of the six properties. The distribution of selected wavelengths does not indicate a strong localization of the algorithm but in each run, GILS method was able go generate successful calibration models for most of the properties of diesel fuels studied.

CONCLUSIONS

This study has demonstrated the application of NIR spectroscopy with genetic multivariate calibration to simultaneous determination of diesel fuel properties. Because of the fact that the percent standard error of prediction (%SEP) values are below 5% for total aromatic content, density, boiling point, and viscosity, the NIR spectroscopy can be used for simultaneous determination of chemical and physical properties of diesel fuels. On the other hand, the genetic algorithm used in the GILS method is able to select and extract the most relevant information to build successful calibration models that have high predictive ability for the independent test samples.

REFERENCES

Arnold, S. A., Crowley, J., Vaidyanathan, S., Matheson, L., Mohan, P., Hall, J. W., Harvey, L. M., and McNeil, B. (2000). At-line monitoring of a submerged filamentous bacterial cultivation using near infrared spectroscopy. *Enzyme Microb. Technol.* 27:691–697.

Boger, Z. (2003). Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Anal. Chim. Acta* 490:31–40.

Centner, V., Massart, D. L., de Noord, O. E., de Jong, S., Vandeginste, B. M., and Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* 68:3851–3858.

DeThomas, F. A., Hall, J. W., and Monfre, S. L. (1994). Real-time monitoring of polyurethane production using near infrared spectroscopy. *Talanta* 41:425–431.

Eigenvector Research, Inc. http://www.eigenvector.com/Data/SWRI, Manson, Washington. Retrieved: 2005.

Esteban-Diez, I., González-Sáiz, J. M., and Pizarro, C. (2004). OWAVEC: A combination of wavelet analysis and an orthogonalization algorithm

as a pre-processing step in multivariate calibration. *Anal. Chim. Acta.* 515:31–41.

Forina, M., Casolino, C., and Pizarro Millan, C. (1999). Iterative predictor weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. *J. Chemom.* 13:165–184.

Geladi, P., and Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Anal. Chim. Acta.* 185:1–17.

Haaland, D. M., and Thomas, E. V. (1988). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60:1193–1202.

Hibbert, D. B. (1993). Genetic algorithms in chemistry. *Chemom. Intell. Lab. Syst.* 19:277–293.

Hörchner, U., and Kalivas, J. H. (1995). Further investigation on a comparative study of simulated annealing and genetic algorithm for wavelength selection. *Anal. Chim. Acta.* 311:1–13.

Leardi, R., Boggia, R., and Terrile, M. (1992). Genetic algorithms as a strategy for feature selection. *J. Chemom.* 6:267–281.

Leardi, R. (2000). Application of genetic algorithm—PLS for feature selection in spectral data sets. *J. Chemom.* 14:643–655.

Lindberg, W., Persson, J. A., and Wold, S. (1983). Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and lignin sulfonate. *Anal. Chem.* 55:643–648.

Lindgren, F., Geladi, P., Rännar, S., and Wold, S. (1994). Interactive variable selection (IVS) for PLS. Part 1: Theory and algorithms. *J. Chemom.* 8:349–363.

Lucasius, C. B., and Kateman, G. (1993). Understanding and using genetic algorithms. Part 1. Concepts, properties and context. *Chemom. Intell. Lab. Syst.* 19:1–33.

McClure, W. F. (1994). Near infrared spectroscopy—The giant is running. *Anal. Chem.* 66:43A–53A.

Mosley, R. M., and Williams, R. R. (1998). Determination of the accuracy and efficiency of genetic regression. *Appl. Spectrosc.* 52:1197–1202.

Özdemir, D. (2005). Determination of octane number of gasoline using near infrared spectroscopy and genetic multivariate calibration methods. *Petrol. Sci. Technol.* 23:1139–1152.

Özdemir, D., and Dinç, E. (2004). Determination of thiamine HCl and pyridoxine HCl in pharmaceutical preparations using UV—Visible spectrophotometry and genetic algorithm based multivariate calibration methods. *Chem. Pharm. Bull.* 52:810–817.

Özdemir, D., and Öztürk, B. (2004). Genetic multivariate calibration methods for near infrared (NIR) spectroscopic determination of complex mixtures. *Turk. J. Chem.* 28:497–514.

Özdemir, D., and Williams, R. R. (1999). Multi-instrument calibration in UV-visible spectroscopy using genetic regression. *Appl. Spectrosc.* 53:210–217.

Paradkar, R. P., and Williams, R. R. (1997). Genetic regression as a calibration technique for solid phase extraction of dithizone-metal chelates. *Appl. Spectrosc.* 51:92–100.

Pizarro Millan, C., Forina, M., Casolino, M. C., and Leardi, R. (1998). Extraction of representative subsets by potential functions methods and genetic algorithms. *Chemom. Intell. Lab. Syst.* 40:33–51.

Soyemi, O. O., Busch, M. A., and Busch, K. W. (2000). Multivariate analysis of near infrared spectra using the G-programming language. *J. Chem. Inf. Comput. Sci.* 40:1093–1100.

Tran, C. D., Oliveira, D., and Grishko, V. I. (2004). Determination of enantiomeric compositions of pharmaceutical products by near-infrared spectrometry. *Anal. Biochem.* 325:206–214.

Wåhlby, U., and Skjöldebrand, J. (2001). NIR measurements of moisture changes in foods. *J. Food Eng.* 47:303–312.

Wentzell, P. D., Andrews, D. T., and Kowalski, B. R. (1997). Maximum Likelihood Multivariate Calibration. *Anal. Chem.* 69:2299–2311.