

This article was downloaded by:

On: 25 January 2011

Access details: Access Details: Free Access

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Wood Chemistry and Technology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597282>

Prediction of Lignin and Extractive Content of *Pinus nigra Arnold. var. Pallasiana* Tree Using Near Infrared Spectroscopy and Multivariate Calibration

B. Uner^a; İ. Karaman^b; H. Tanriverdi^a; D. Özdemir^b

^a Faculty of Forestry, Forest Products Engineering Department, Suleyman Demirel University, Isparta, Turkey ^b Faculty of Science, Department of Chemistry, İzmir Institute of Technology, Urla, Izmir, Turkey

To cite this Article Uner, B. , Karaman, İ. , Tanriverdi, H. and Özdemir, D.(2009) 'Prediction of Lignin and Extractive Content of *Pinus nigra Arnold. var. Pallasiana* Tree Using Near Infrared Spectroscopy and Multivariate Calibration', Journal of Wood Chemistry and Technology, 29: 1, 24 – 42

To link to this Article: DOI: 10.1080/02773810802607567

URL: <http://dx.doi.org/10.1080/02773810802607567>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Prediction of Lignin and Extractive Content of *Pinus nigra* Arnold. var. *Pallasiana* Tree Using Near Infrared Spectroscopy and Multivariate Calibration

B. Uner,¹ İ. Karaman,² H. Tanriverdi,¹ and D. Özdemir²

¹Suleyman Demirel University, Faculty of Forestry, Forest Products Engineering Department, Isparta, Turkey

²İzmir Institute of Technology, Faculty of Science, Department of Chemistry, Urla, Izmir, Turkey

Abstract: Determination of quality parameters such as lignin and extractive content of wood samples by wet chemistry analyses takes a long time. Near infrared (NIR) spectroscopy coupled with multivariate calibration offers a fast and nondestructive alternative to obtain reliable results. However, due to the complexity of the spectra obtained from NIR, some wavelength selection is generally required to improve the predictive ability of multivariate calibration methods. *Pinus nigra* Arnold. Var. *pallasiana* is the second most growing pine species in Turkey. Even though its rotation period is very high, around 120 years, the forest products industry has widely accepted the use of *Pinus nigra* because of its ability to grow on a wide range of sites and its suitability to produce desirable products. In this study, 51 samples of *Pinus nigra* trees were collected and their lignin and extractive content were determined with standard reference (TAPPI) methods. Then, the same samples were scanned with near infrared spectrometer between 1000 and 2500 nm in diffuse reflectance mode. Multivariate calibration models were built with genetic inverse least squares method for both lignin and extractive content using the concentration information obtained from wet standard reference method. Overall, standard error of calibration (SEC) and standard error of prediction (SEP) were ranged between 0.35% (w/w) and 2.4% (w/w).

Keywords: Extractive, genetic algorithms, lignin, multivariate calibration, near infrared spectroscopy, *Pinus nigra* Arnold. Var. *pallasiana*

This project was funded by TUBITAK (The Scientific and Technological Research Council of Turkey, Project No. 105O524). The authors also acknowledge Isparta Regional Forest Service and Keçiörlü Forest Service for their support in providing wood samples and to İzmir Institute of Technology for the financial support of purchasing the near infrared spectrometer.

Address correspondence to D. Özdemir, İzmir Institute of Technology, Faculty of Science, Department of Chemistry, Gülbahçe 35430, Urla, İzmir/Turkey. E-mail: durmusozdemir@iyte.edu.tr

INTRODUCTION

Wood is a natural composite that contains cellulose, lignin, hemicelluloses, and extractive components. These components are utilized in different industries after processing wood with chemical or mechanical methods such as in the pulping industry. In order to obtain high pulp yield in chemical pulping condition, wood should have high cellulose, low lignin, and extractive contents.^[1,2] Tree breeding programs employ different silvicultural treatments to improve tree growth. These trees may have different chemical properties than natural grown trees. Moreover, wood samples obtained from the trees in different locations shows different properties (chemical compositions, morphology, juvenile wood, mature wood, etc.).^[3] These differences play an important role on finished products such as brightness of bleached paper and board mechanical properties. Therefore, it is important to determine these contents. Determination of chemical content of wood with traditional chemical methods is a costly and time-consuming process. In recent years, spectroscopic methods are being used to characterize wood properties. The utilization of spectroscopic methods is fast and provides more information.

Near infrared (NIR) spectroscopy^[4,5] has become a popular method for simultaneous chemical analysis and is being studied extensively in a number of different fields such as process monitoring,^[6] biotechnology,^[7] and the pharmaceutical industry^[8] because of the potential for on-line, nondestructive, and noninvasive instrumentation. Traditionally, NIR spectroscopy has found its widest application area in agriculture and the food industry.^[9–17] The NIR portion of the electromagnetic spectrum covers the range from 780 nm to 2500 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.

NIR spectroscopy is also used for the chemical and mechanical characterizations of different wood species.^[18–23] In the Southern United States, *Pinus taeda* is a very important species mainly utilized in pulping and wood using industry. Therefore, many NIR researches were carried out on this tree. Transmittance NIR spectra was used to determine chemical composition of solid wood of loblolly pine.^[24] Strong correlation was found between wet chemistry and the NIR spectra determination. In the same research, sample preparation and quantity were also discussed. It was found that the source of error is mostly coming from the wet chemistry. NIR spectra along with the multivariate analysis were used to determine chemical and mechanical properties of loblolly pine.^[25] Samples were selected from different locations and the height of the tree. The correlation coefficient for lignin and extractives was 0.80. This measurement was carried out between 500 nm and 2400 nm. When the spectrum was narrowed (650 nm–1150 nm), the correlation coefficient was not changed

a lot. This makes NIR spectroscopy a relevant and quick method to determine chemical constituents of wood without destruction. Mechanical properties of loblolly pine were determined with bending strength. MOE and MOR were calibrated with NIR. Their correlation coefficient was 0.88 and 0.92, respectively. The chemical composition of *Pinus taeda* was investigated with NIR spectroscopy.^[26] Diffuse reflectance NIR spectroscopy and radial strips from seven different locations were used. Strips represented the juvenile wood, transition zone between mature and juvenile wood, and mature wood. Prediction of error was high. This is reasoned for the different origins of the samples. The decay resistance of larch wood (*Larix decidua* Mill. and *Larix sibirica* Lebed.) was tested with NIR.^[27] Samples were taken from logs that were coming from industry. This means that each log was coming from different sites and soil conditions. A partial least square regression calibration model based on the FT-NIR spectra and reference data for standardized wood decay tests successfully predicted decay resistance of two larch species. NIR spectroscopy is also utilized to determine Kappa number in the paper industry. Usually a wet chemical analysis method is performed to obtain this value. It was suggested that NIR spectroscopy could be used directly in mills.^[28]

Advances in computers and automation technology have made today's instruments incredibly fast; 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 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 the last couple of decades for the analysis of complex chemical mixtures.^[29–32] 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.^[33–35] Genetic algorithm is one of them, offering a fast and effective solution for large-scale problems.^[36–38]

The inverse least squares (ILS) method 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 a modified version of the original ILS method 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) that has been applied to a number of wavelength selection problems.^[39–41] GAs are non-local search and optimization methods that are based on the principles of natural selection.^[42–46]

Pinus nigra Arnold. *Var. pallasiana* is the second most growing pine species in Turkey. Even though its rotation period is very high (around 120 years), the

forest products industry has widely accepted the use of *Pinus nigra* because of its ability to grow on a wide range of sites and its suitability to produce desirable products. Therefore, the objective of this article is to determine lignin and extractive content with NIR spectroscopy. NIR spectroscopy will help to determine chemical content quickly and reduce chemical cost. It will also provide more information to producers.

MATERIALS AND METHODS

Genetic Inverse Least Squares

The inverse least squares (ILS) method, which uses the inverse of Beer's Law, models the concentration 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 the calibration set is generated from another analytical technique that already has its inherent errors, which might be higher than those of the spectrometer (e.g., Kjeldahl protein analysis used to calibrate NIR spectra).

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

$$\mathbf{C} = \mathbf{A}\mathbf{P} + \mathbf{E}_C \quad (1)$$

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 because analysis is based on an ILS model invariant with respect to the number of chemical components included in the analysis. The reduced model is given as:

$$\mathbf{c} = \mathbf{A}\mathbf{p} + \mathbf{e}_c \quad (2)$$

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

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

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 following equation:

$$\hat{c} = \mathbf{a}' \cdot \hat{\mathbf{p}} \quad (4)$$

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 because 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 step-wise wavelength selection and all possible combination searches, are available to build an ILS model that fits the data best.

Genetic algorithms (GA) are global search and optimization methods based on 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 a reduced data set. 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 response 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 GILS. 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

$$S_1 = (A_{1147}, A_{951}, \#A_{2179}, A_{2218}) \quad (5)$$

$$S_2 = (A_{1225}, A_{1478}, \#A_{1343}, A_{950}, A_{1451}, A_{2358}, A_{931}, A_{1158}) \quad (6)$$

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

Offspring

$$S_3 = (A_{1147}, A_{951}, A_{1343}, A_{950}, A_{1451}, A_{2358}, A_{931}, A_{1158}) \quad (7)$$

$$S_4 = (A_{2179}, A_{2218}, A_{1225}, A_{1478}) \quad (8)$$

where A_{1147} 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 that 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 GILS 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 because 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 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 and 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 GILS as in all the GAs, the program has been set to run several times for each component in 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 multivariate 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 because 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.

Materials

Pinus nigra Arnold. var pallasiana trees used for the study were collected from the Söğütadağı region, Keçiöbolu, Isparta, Turkey. The trees were sampled from

plots ranged in the same elevation of 1230 m with average precipitation of 51.5 cm and average July maximum temperature of 30.3°C and average January temperature of 1.8°C. Trees aged from 17 to 22 years were used in this study and diameter ranged from 15 to 25 cm. Tree selection was based on good form trees and eccentric piths were not used. Samples were taken from the breast height section to determine chemical contents. A total of 51 trees were examined. Among these trees, 23 were collected from a thinning applied plot and 28 were from a control plot. All tree samples were kept in bed for 3 years and were then planted in the field. Twelve years later thinning was applied.

Methods

Extractive and lignin content of wood were determined according to TAPPI Standard test method T 204 om-88 and T222 om-88. Wood meal samples were prepared using a Wiley mill and ground to pass various mesh screens. In order to determine extractive content, ethanol-benzene solution (1:2 v/v) was used. Extraction was carried out in soxhlet for 6 hours. In order to determine acid insoluble lignin content of wood, the carbohydrates in wood were hydrolyzed and dissolved with sulfuric acid according to the TAPPI Standard test method. Acid insoluble lignin is filtered off, dried, and the content is measured. All tests were repeated three times.

Near infrared spectroscopic measurements of the sample described earlier were carried out in diffuse reflectance mode between 1000 and 2500 nm wavelength interval against gold surface background. Among the 51 samples, 21 were randomly selected and initial wet chemical analyses were performed on these samples, then the remaining 30 were analyzed as a second party. Although wet chemical analysis of the samples in the first and the second set were carried out at different times, NIR analysis of the all samples (51 of them) were analyzed at the same time in a day.

Software

After collection of raw spectra, all data were transferred into a Microsoft Excel worksheet where text files of calibration sets and independent validation sets were prepared. Multivariate calibration models were built with the GILS method for extractive and lignin content for the first 21 samples and the remaining 30 samples separately and also for the combination of both data sets. The GILS method was implemented in MATLAB programming language Version 7.0 (MathWorks Inc., Natick, MA).

RESULTS AND DISCUSSION

The results of wet chemical analysis for *Pinus nigra* Arnold. var. *Pallasiana* were given in Table 1. Wood is an organic living material and its properties

Table 1. The result of wet chemical analysis for *Pinus nigra Arnold. var. Pallasiana* tree samples

Sample	Extractive content (wt%)	Lignin content (wt%)	Treatment	Sample	Extractive content (wt%)	Lignin content (wt%)	Treatment
1	4.60	22.90	thinning	27	10.31	22.47	thinning
2	7.30	24.61	thinning	28	6.81	22.98	thinning
3	7.40	23.88	no thinning	29	7.63	26.61	thinning
4	6.80	17.53	no thinning	30	6.92	23.82	thinning
5	10.00	19.19	no thinning	31	5.52	30.85	thinning
6	5.30	22.16	thinning	32	9.10	22.41	thinning
7	6.30	21.60	no thinning	33	9.31	18.84	thinning
8	6.60	16.87	no thinning	34	8.24	21.17	thinning
9	8.80	19.46	no thinning	35	8.74	21.07	thinning
10	7.60	23.59	no thinning	36	8.70	28.27	thinning
11	9.40	19.44	no thinning	37	6.17	19.65	no thinning
12	13.00	18.00	no thinning	38	8.01	21.84	no thinning
13	10.60	20.71	no thinning	39	7.22	26.15	no thinning
14	8.30	20.80	no thinning	40	8.70	24.68	no thinning
15	6.90	20.28	thinning	41	8.77	24.48	no thinning
16	8.80	17.96	thinning	42	5.81	23.06	no thinning
17	4.60	14.30	thinning	43	6.74	20.86	no thinning
18	5.10	20.99	thinning	44	8.68	26.32	no thinning
19	7.80	24.23	thinning	45	8.50	21.06	no thinning
20	7.60	26.46	thinning	46	10.52	21.21	no thinning
21	4.30	23.03	thinning	47	7.26	28.44	no thinning
22	8.64	22.37	no thinning	48	9.70	34.47	no thinning
23	7.66	22.59	thinning	49	8.53	23.88	no thinning
24	6.17	26.51	thinning	50	5.76	19.94	no thinning
25	7.33	24.89	no thinning	51	9.04	16.70	no thinning
26	8.46	19.42	thinning				

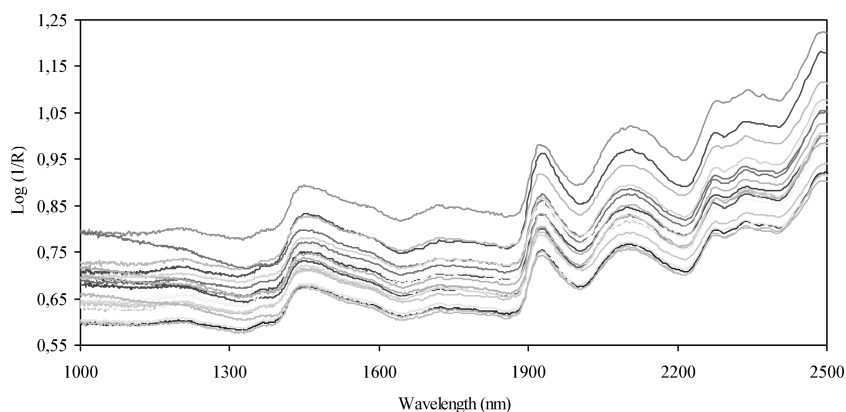


Figure 1. Near infrared diffuse reflectance spectra of 20 wood samples collected from *Pinus nigra* Arnold. var. *pallasiana* trees.

change from tree to tree depending on growth conditions. This change can be seen in Table 1 and average lignin content was $22.45 \pm 3.68\%$ and extractive content was $7.77 \pm 1.66\%$. This measurement takes a long time and consumes a lot of chemicals. NIR spectroscopy has become a popular method for simultaneous chemical analysis.

Near infrared diffuse reflectance spectra of about 20 wood samples among 51 are shown in Figure 1. It is evident that the samples yield high absorbance values around 1450, 1900, and 2100 nm wavelength regions. However, because they are all pine wood samples, their spectral characteristics are very much alike except for the baseline differences among the samples. This type of baseline shift in the absorbance scale is quite common in diffuse reflectance spectroscopy and part of it is due to sample composition differences and part of it is due to sample inhomogeneities. Because the GILS method used here is a genetic algorithm based on multivariate calibration techniques, it was expected that it can select certain combinations of wavelengths that have maximum correlation with extractive and lignin content of the samples.

In order to construct NIR spectroscopic multivariate calibration models for extractive and lignin content three different calibration sets were prepared. The first calibration set was generated from the aforementioned 21 samples. This data set is called the first data set in which 14 of them were randomly selected for the calibration set including the samples having minimum and maximum extractive and lignin contents. The remaining 7 samples were reserved for independent test samples. As can be seen from Table 1, the minimum and maximum extractive and lignin contents were not the same samples, and, therefore different samples were used in both calibration and validation sets for extractive and lignin content. Reference extractive and lignin contents versus predicted values based on NIR spectra using the GILS method are shown in Figure 2

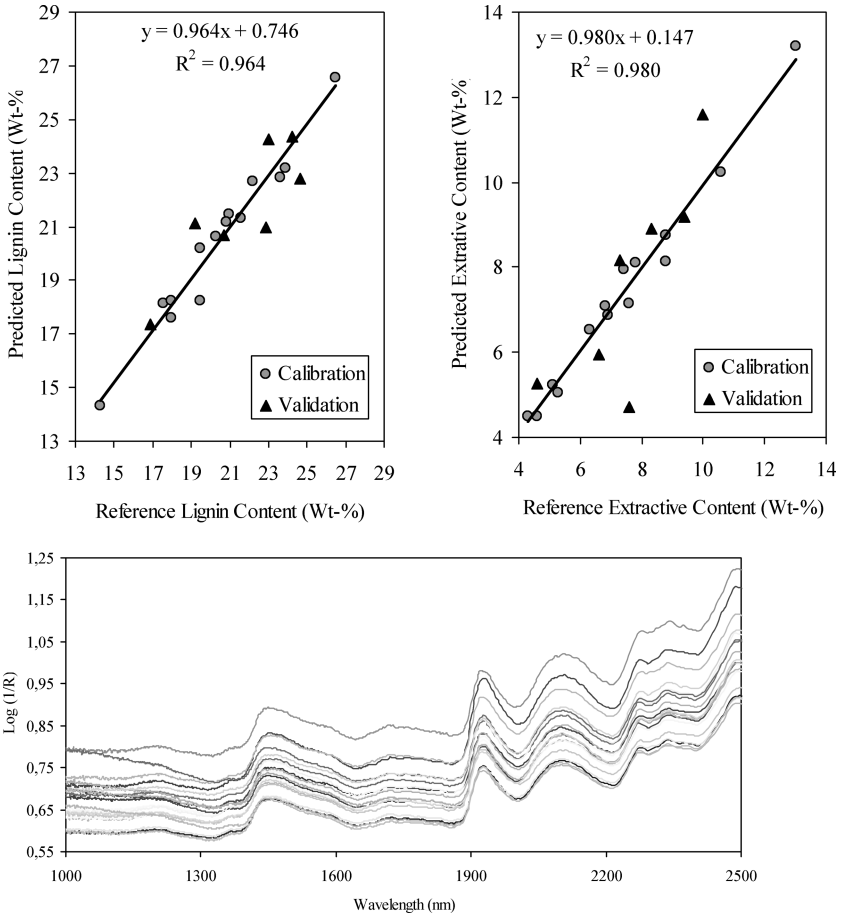


Figure 2. Reference versus predicted lignin and extractive contents for the first data set.

for the first data set. Calibration models for lignin content determination gave standard error of calibration (SEC) and standard error of prediction (SEP) values as 0.61% (w/w) and 1.33% (w/w) for calibration and independent test sets, respectively. In the case of extractive content determination, the SEC and SEP values were 0.35% (w/w) and 1.36% (w/w) for calibration and prediction sets, respectively. The R^2 value of regression lines for lignin was 0.964 and that for extractive content was 0.980.

When these SEC and SEP errors are examined, it is seen that the values for lignin content were comparable even though the SEP value is about twice that of the SEC. It must be realized that the GILS method is an iterative procedure due to the genetic algorithm used to select a subset of wavelengths from the whole

spectral range. As mentioned earlier, NIR spectra of these samples suffer from somewhat large baseline fluctuations. This causes the GILS to model this effect while preparing calibration models even though the cross validation approach is used during the model-building step. Since independent test samples in the prediction set do not have the same baseline trends as in the calibration set, predictions result in larger SEP values. Yet, when the overall calibration performance of the model is examined, it is possible to state that the NIR spectra do contain quantitative information that is correlated with extractive and lignin content of the wood samples studied here. It is also worth mentioning that the results of GILS can only be as good as the results of wet chemical analysis because the model building is based on the reference values obtained from it.

Figure 3 shows the reference lignin and extractive contents versus GILS predicted values for the second data set with 30 samples, of which 20 were used for model building in the calibration set and the remaining 10 samples were reserved for the prediction set. While the concentrations of lignin content were ranging between 14% (w/w) and 27% (w/w) for the first data set, the upper level of lignin content in the second data set was around 35% (w/w). On the other hand, the extractive content of the samples in the second data set were distributed in a narrower range between 5% (w/w) and 11% (w/w) when compared with the first data set. The SEC values for lignin and extractive content were 1.02% (w/w) and 0.34% (w/w), respectively, whereas the SEP values were ranged between 1.46% (w/w) and 0.66% (w/w) for lignin and extractive content. The R^2 value of regression lines for lignin was 0.941 and that for extractive content was 0.929.

When SEC and SEP values are examined in the second data set, it is seen that the agreement between these values is better than those obtained for the first data set. One possible explanation of this improvement could be attributed to increased number of calibration and prediction samples. On the other hand, the R^2 of calibration lines were now lower than those obtained for the first data set. This is also an expected outcome of calibration models with larger data set as variability increases with the larger number of samples in the calibration set.

The third data set analyzed in this study was formed by combining the first and the second data sets into a single set as given in Table 1. The calibration and prediction sets are formed by adding the corresponding spectra in the first data set to the data in the second data set. The calibration plots for lignin and extractive content are given in Figure 4.

The wet chemical reference analysis of the first and the second data sets were carried out at different times and consequently some variations in lignin and extractive contents of the samples are revealed, as can be seen in Table 1. This variability can also be seen in Figures 2 and 3 for both lignin and extractive contents. For example, in Figure 2, the dynamic range of lignin was in the range of 13 to 29% (w/w) whereas this was between 13 and 37%

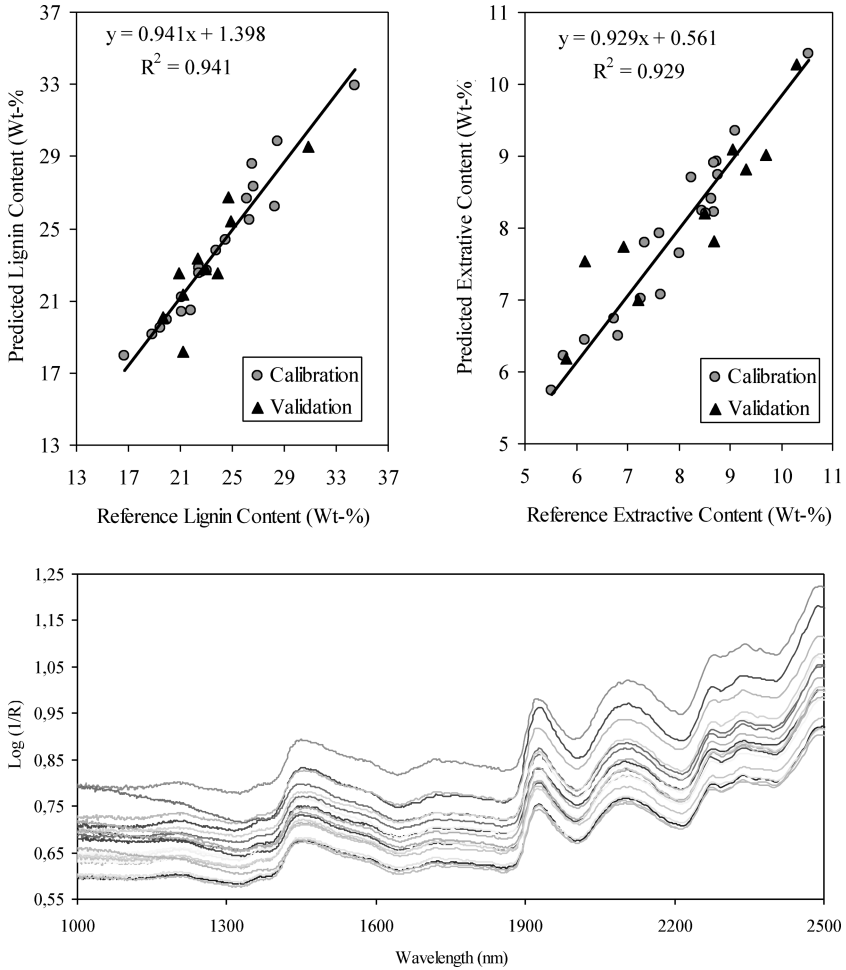


Figure 3. Reference versus predicted lignin and extractive contents for the second data set.

(w/w) in Figure 3 for the second data set. A similar but reverse trend about the extractive content can also be seen between Figures 2 and 3 where a larger dynamic range was seen in Figure 2. On the other hand NIR analysis of the samples was performed at the same time in a day and calibration models were generated by combining both data sets. We suspect that this could be the reason for the decrease in the predictive ability of the models generated with GILS in the third data set. It is worth mentioning, however, that this decrease is mostly due to the reference analysis and due to the NIR-based GILS modeling. Both SEC and SEP values were somewhat higher in the third data set compared to the first and second data sets. For the determination of lignin content, SEC

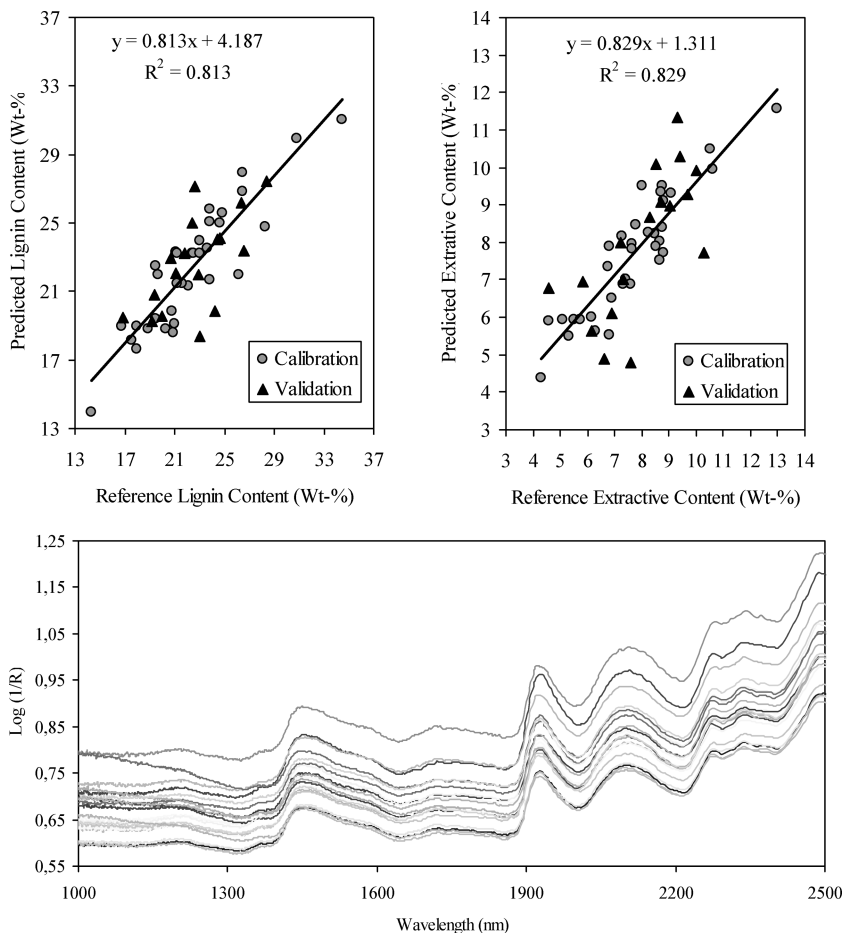


Figure 4. Reference versus predicted lignin and extractive contents for the third data set.

and SEP values were 1.77% (w/w) and 2.40% (w/w), respectively. In the case of extractive content determination similar results were obtained in which the SEC was 0.75% (w/w) and the SEP was 1.40% (w/w). This increase in calibration and prediction results are also reflected in R^2 values of regression as the values went down to 0.813 for the lignin content and 0.829 for the extractive content.

Because the 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 5 illustrates the frequency distribution of selected wavelengths in 100 runs with 20 gene and 50 iteration for the calibration set that contains 14 samples in the first data set.

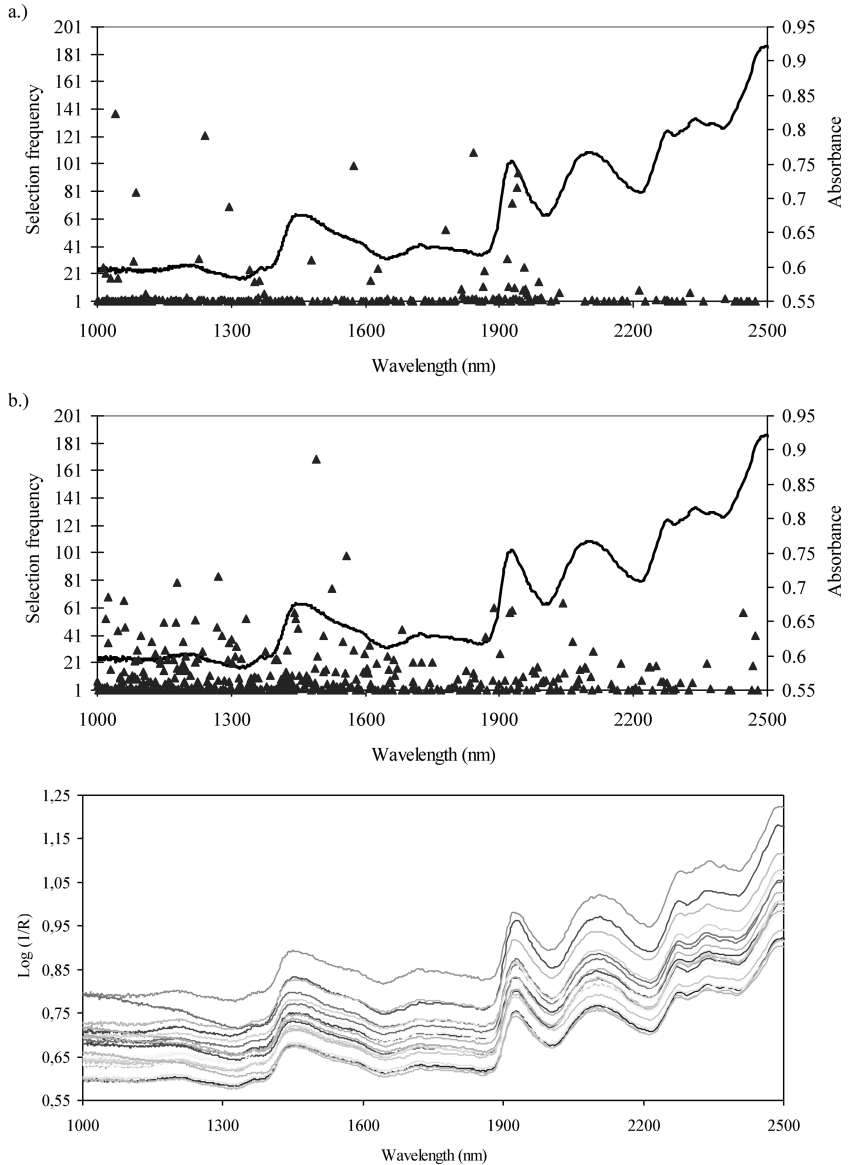


Figure 5. Wavelength selection frequency distribution of GILS method for a.) lignin and b.) extractive content.

As can be seen from the figure there are a number of regions where selection frequencies are very high compared to the rest of the spectrum. The wavelength region around 1900 nm for lignin content indicates a strong tendency for GILS method to select while for extractive content, around 1500 nm

is the most frequently selected region. A significant difference between the frequency distribution of lignin content and extractive content is that the selected wavelengths are more distributed in the former and much more wavelengths are selected. On the other hand, extractive content determination was achieved with less number of wavelengths in the final best gene of each iteration.

CONCLUSIONS

The near infrared spectroscopic determination of lignin and extractive content of *Pinus nigra* Arnold. Var. *pallasiana* trees was carried out with both standard reference methods and near infrared spectroscopy. Spectroscopic measurements were performed on solid milled samples of raw wood in diffuse reflectance mode of NIR and offer much faster analysis for the screening purpose of these analyses. However, it is important to realize that the performance of the NIR-based multivariate methods relies on good and reliable reference analysis results. Results are demonstrated that NIR spectroscopy coupled with multivariate calibration could be used for routine and fast analysis of these quality parameters of wood samples.

REFERENCES

1. Poke, F.; Raymond, C.A. Predicting extractives, lignin, and cellulose contents using near infrared spectroscopy on solid wood in *Eucalyptus globulus*. *J. Wood Chem. Technol.* **2006**, *26*, 187–199.
2. Sykes, R.; Li, B.; Hodge, G.; Goldfarb, B.; Kadla, J.; Chang, H.-M. Prediction of loblolly pine wood properties using transmittance near-infrared spectroscopy. *Can. J. For. Res.* **2005**, *35*, 2423–2431.
3. Zobel, B.J.; van Buijtenen, J.P. *Wood Variation: Its Causes and Control*; Springer-Verlag: Berlin, Germany, 1989.
4. Wesley, I.J.; Larroque, O.; Osborne, B.G.; Azudin, N.; Allen, H.; Skerritt, J.H. Measurement of gliadin and glutenin content of flour by NIR spectroscopy. *J. Cereal Sci.* **2001**, *34*, 125–133.
5. McClure, W.F. Near infrared spectroscopy—The giant is running. *Anal. Chem.* **1994**, *66*, 43A–53A.
6. DeThomas, F.A.; Hall, J.W.; Monfre, S.L. Real-time monitoring of polyurethane production using near infrared spectroscopy. *Talanta* **1994**, *41*, 425–431.
7. Arnold, S.A.; Crowley, J.; Vaidyanathan, S.; Matheson, L.; Mohan, P.; Hall, J.W.; Harvey, L.M.; McNeil, B. At-line monitoring of a submerged filamentous bacterial cultivation using near infrared spectroscopy. *Enzyme Microb Tech.* **2000**, *27*, 691–697.
8. Tran, C.D.; Oliveira, D.; Grishko, V.I. Determination of enantiomeric compositions of pharmaceutical products by near-infrared spectrometry. *Anal. Biochem.* **2004**, *325*, 206–214.
9. Delwiche, S.R. Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* **1998**, *27*(3), 241–254.

10. Ferré, J.; Rius, F.X. Selection of the best calibration sample subset for multivariate regression. *Anal. Chem.* **1996**, *68*, 1565–1571.
11. Ferrioa, J.P.; Villegasb, D.; Zarcob, J.; Apariciob, N.; Arausc, J.L.; Royob, C. Assessment of durum wheat yield using visible and near-infrared reflectance spectra of canopies. *Field Crop Res.* **2005**, *94*(2–3), 126–148.
12. Hareland, G.A. Evaluation of flour particle size distribution by laser diffraction, sieve analysis and near-infrared reflectance spectroscopy. *J. Cereal Sci.* **1994**, *20*(2), 183–190.
13. Kalivas, J.H. Two data sets of near infrared spectra. *Chem. Intell. Lab. Syst.* **1997**, *37*(2), 255–259.
14. McCaig, T.N. Extending the use of visible/near-infrared reflectance spectrophotometers to measure colour of food and agricultural products. *Food Res. Int.* **2002**, *35*(8), 731–736.
15. Miralbés, C. Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chem.* **2004**, *88*, 620–628.
16. Puchwein, G.; Eibelhuber, A. Outlier detection in routine analysis of agricultural grain products by near-infrared spectrometry. *Anal. Chim. Acta* **1989**, *223*, 95–103.
17. Sorvaniemi, J.; Kinnunen, A.; Tsados, A.; Mälkki, Y. Using partial least squares regression and multiplicative scatter correction for FT-NIR data evaluation of wheat flours. *Food Sci. Techn.* **1993**, *26*(3), 251–258.
18. Hedrick, S.E.; Bennett, R.M.; Rials, T.G. Correlation of near-infrared spectroscopy measurements with the properties of treated wood. *J. Mater. Civil Eng.* **2007**, *19*(4), 279–285.
19. So, C.L.; Eberhardt, T.L.; Rapid analysis of inner and outer bark composition of Southern Yellow Pine bark from industrial sources. *Holz Als Roh-Und Werkstoff* **2006**, *64*(6), 463–467.
20. Schimleck, L.R.; Kube, P.D.; Raymond, C.A. Extending near infrared reflectance (NIR) pulp yield calibrations to new sites and species. *J. Wood Chem. Technol.* **2006**, *26*(4), 299–311.
21. Tsuchikawa, S.; Hirashima, Y.; Sasaki, Y. Near-infrared spectroscopic study of the physical and mechanical properties of wood with meso- and micro-scale anatomical observation. *Appl. Spect.*, **2005**, *59*(1), 86–93.
22. Cogdill, R.P.; Schimleck, L.R.; Jones P.D. Estimation of the physical wood properties of *Pinus taeda* L. radial strips using least squares support vector machines. *J. Near Infrared Spect.* **2004**, *12*(4), 263–269.
23. Jonsson, P.; Sjostrom, M.; Wallbacks, L. Strategies for implementation and validation of on-line models for multivariate monitoring and control of wood chip properties. *J. Chemometr.* **2004**, *18*(3–4), 203–207.
24. Yeh, T.; Yamada, T.; Capanema, E.; Chang, H.; Chiang, V.; Kadla J.F. Rapid screening of wood chemical component variations using transmittance near-infrared spectroscopy. *J. Agric. Food Chem.* **2005**, *53*, 3328–3332.
25. Kelley, S.S.; Rials, T.G.; Snell, R.; Groom, L.H.; Sluiter, A. Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. *Wood Sci. Technol.* **2004**, *38*, 257–276.
26. Jones, P.D.; Schimleck, L.R.; Peter, G.F.; Daniels, R.F.; Clark III, A. Nondestructive estimation of wood chemical composition of sections of radial wood strips by

- diffuse reflectance near infrared spectroscopy. *Wood Sci. Technol.* **2006**, *40*, 709–720.
27. Sykacek, E.; Gierlinger, N.; Wimmer, R.; Schwanninger, M. Prediction of natural durability of commercial available European and Siberian larch by near-infrared spectroscopy. *Holzforschung*, **2006**, *60*, 643–647.
 28. Tsuchikawa, S. A review of recent near infrared research for wood and paper. *Appl. Spect. Reviews*, **2007**, *42*, 43–71.
 29. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17.
 30. Haaland, D.M.; Thomas, E.V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60*, 1193–1202.
 31. Lindberg, W.; Persson, J.A.; Wold, S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and lignin sulfonate. *Anal. Chem.* **1983**, *55*, 643–648.
 32. Wentzell, P.D.; Andrews, D.T.; Kowalski, B.R. Maximum likelihood multivariate calibration. *Anal. Chem.* **1997**, *69*, 2299–2311.
 33. Centner, V.; Massart, D.L.; De Noord, O.E.; De Jong, S.; Vandeginste, B.M.; Sterna, C. Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* **1996**, *68*, 3851–3858.
 34. Forina, M.; Casolino, C.; Pizarro, M.C. Iterative predictor weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. *J. Chemometr.* **1999**, *13*, 165–184.
 35. Lindgren, F.; Geladi, P.; Rännar, S.; Wold, S. Interactive variable selection (IVS) for PLS. Part 1: Theory and algorithms. *J. Chemometr.* **1994**, *8*, 349–363.
 36. Hörchner, U.; Kalivas, J.H. Further investigation on a comparative study of simulated annealing and genetic algorithm for wavelength selection. *Anal. Chim. Acta*, **1995**, *311*, 1–13.
 37. Leardi, R.; Boggia, R.; Terrile, M. Genetic algorithms as a strategy for feature selection. *J. Chemometr.* **1992**, *6*, 267–281.
 38. Lucasius, C.B.; Kateman, G. Understanding and using genetic algorithms. Part 1. Concepts, properties and context. *Chem. Intell. Lab. Syst.* **1993**, *19*, 1–33.
 39. Özdemir, D. Determination of octane number of gasoline using near infrared spectroscopy and genetic multivariate calibration methods. *Petrol. Sci. Technol.* **2005**, *23*, 1139–1152.
 40. Özdemir, D.; Dinç, E. Determination of thiamine HCl and pyridoxine HCl in pharmaceutical preparations using uv-visible spectrophotometry and genetic algorithm based multivariate calibration methods. *Chem. Pharm. Bull.* **2004**, *52*(7), 810–817.
 41. Özdemir, D.; Öztürk, B. Genetic multivariate calibration methods for near infrared (NIR) spectroscopic determination of complex mixtures. *Turk. J. Chem.* **2004**, *28*, 497–514.
 42. Hibbert, D.B. Genetic algorithms in chemistry. *Chem. Intell. Lab. Syst.* **1993**, *19*, 277–293.
 43. Mosley, R.M.; Williams, R.R. Determination of the accuracy and efficiency of genetic regression. *Appl. Spectrosc.* **1998**, *52*, 1197–1202.
 44. Özdemir, D.; Williams, R.R. Multi-instrument calibration in uv-visible spectroscopy using genetic regression. *Appl. Spectrosc.* **1999**, *53*, 210–217.

45. Paradkar, R.P.; Williams, R.R. Genetic regression as a calibration technique for solid phase extraction of dithizone-metal chelates. *Appl. Spectrosc.* **1997**, *51*, 92–100.
46. Pizarro, M.C.; Forina, M.; Casolino, M.C.; Leardi, R. Extraction of representative subsets by potential functions methods and genetic algorithms. *Chem. Intell. Lab. Syst.* **1998**, *40*, 33–51.