Scientific Research Publishing

Open Journal of Applied Sciences, 2018, 8, 422-430
http://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

LSSVM Combined with SPA Applied to Near-Infrared Quantitative Determination of the Octane in Fuel Petrol Samples

Lili Xu[1], Jie Gu[2], Huazhou Chen[2,3*], Jiangbei Wen[3], Gaili Xu[2]

[1]Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation (Qinzhou University), Qinzhou, China
[2]College of Science, Guilin University of Technology, Guilin, China
[3]Guangdong Spectrastar Instruments Co. Ltd., Guangzhou, China
Email: *hzchengut@foxmail.com

How to cite this paper: Xu, L.L., Gu, J., Chen, H.Z., Wen, J.B. and Xu, G.L. (2018) LSSVM Combined with SPA Applied to Near-Infrared Quantitative Determination of the Octane in Fuel Petrol Samples. *Open Journal of Applied Sciences*, **8**, 422-430.
https://doi.org/10.4236/ojapps.2018.89032

Received: August 29, 2018
Accepted: September 26, 2018
Published: September 29, 2018

Abstract

Least square support vector machine (LSSVM) combined with successive projection algorithm (SPA) method was applied for near-infrared (NIR) quantitative determination of the octane number in fuel petrol. The NIR spectra of 87 fuel petrol samples were scanned for model establishment and optimization. First order derivative Savitzky-Golay smoother (1st-d'SG) was utilized to improve the NIR predictive ability. Its pretreatment effect was compared with the raw data. SPA was applied for the extraction of informative wavelengths. Considering the linear and non-linear training mechanism, LSSVM regression was employed to establish calibration models. The correlation coefficient (R) and root mean square error (RMSE) were used as the model evaluation indices; accordingly the octane number in fuel petrol was quantitatively determined with the prospective predictive indices. Results showed that after pretreated by 1st-d'SG, 8 SPA-selected wavelengths was generated as the inputs of LSSVM, so that the calibration models were optimized in the way of combining the SPA-LSSVM regression with the SG smoother. The prediction results were quite satisfactory, with the calibrating correlation coefficient of 0.951 and the RMSE of 3.282. An independent testing sample set was used to evaluate the optimal model, the testing correlation coefficient was 0.903 and the RMSE was 4.128. We conclude that NIR spectrometry is feasible to determine the octane in fuel petrol by establishing SPA-LSSVM models. The 1st-d'SG pretreatment has the advantage of selecting wavelengths containing the implicit information. The combination of 1st-d'SG pretreatment and SPA-LSSVM show its applicable potential to predict the octane number in fuel petrol.

Keywords

Near-Infrared, Fuel Petrol, The Octane Number, SPA, LSSVM

1. Introduction

Near Infrared (NIR) spectroscopy is a physical, rapid, non-destructive method to measure the combination and overtone vibration of chemical bonds in molecules, requiring minimal or no sample preparation and, in contrast with traditional chemical analysis, does not require reagents, nor produces wastes [1]. NIR spectroscopy has extensive application in the analytical area of materials, environment, food, and biomedicine [2] [3] [4].

Petrol (or gasoline) is a petroleum-derived, transparent liquid primarily used as a fuel in internal combustion engines. It contains over 300 different chemical compounds, mainly hydrocarbons (compounds that contain only carbon and hydrogen) [5] [6]. How good a petrol at resisting engine knock is an important indicator of the quality of petrol. Fuel petrol is compared to a test mixture of octane (which is good at resisting engine knock) and heptane (which is poor at resisting engine knock). The octane number is a quantitative indicator (valued from 0 to 100) to grade the resisting ability according to its performance in a test engine. Higher value of octane number indicates a better resistance to engine knock [7].

NIR spectroscopy can be used for quantitative determination of the octane in the fuel petrol because the spectral information of main structural components of fuel petrol is distinctly reflected in near-infrared region, and the information is quite stable. Brouillette and his coworkers proposed a method for detecting 22 properties (including the octane) in diesel, petrol, and jet fuels. They calculated the octane by NIR models according to the intensity of C=O peak [8]. Lee and his team investigated the random for est algorithm to establish quantitative analytical model for octane in the complex petrol and naphtha samples. They reported that the model predicted the octane coincide with the values by instrumental tester [9]. However, it cannot output satisfactory predictions if these methods directly utilized for the analysis of the octane in fuel petrol. Other acceptable predictions were obtained by Mass Spectrometry in qualitative analysis [10].

Fuel petrol is a complex multi-component material. Its near-infrared spectrum contains information of all components as well as measurement noise. The prospective precise of calibration model will be difficult to improve if the full-range NIR spectrum is used. Wavelength selection is quite necessary for model optimization [11] [12]. Successive projection algorithm (SPA) is an effective method to search wavelengths (or wavelength combinations) which are most informative and with least colinearity [13]. The selected wavelengths (or wavelength combinations) are used to establish calibration models so that the modeling parameters are expected to be optimized and the predictive results to be improved.

Calibration model establishment requires a set of spectra with reference concentrations of the target component. According to Beer-Lambert law, the NIR spectrum theoretically is the linear combination of the pure absorbance of every

single component. However, the target is always one of all components, so that nonlinear methods should be employed if the linear model cannot meet the relationship between the spectra and the concentration. The least squares support vector machine (LSSVM) regression method was proved owning the capability to handle ill-posed problems and lead to unique global models [14] [15].

Our study is concern with the wavelength selection for NIR quantitative determination of fuel petrol samples. For the modeling (calibration-validation) & testing sample division system, we randomly selected out a certain number of independent samples to be the test set, and the calibration-validation partitioning for the modeling set is achieved by the method of sample set partitioning based on joint x-y distances (SPXY) [16] [17]. In modeling process, we firstly reduced the data dimension by using SPA to select the contributive wavelengths of the octane, and then established LSSVM models to optimize the NIR regression procedure. Take the correlation coefficient (R) and the root mean square error (RMSE) as the evaluating indices of parameter optimization. The optimized LSSVM model with its parameters is expected to provide theoretical references for accurate determination of the octane in fuel petrol and for the rapid detection of oil-change period.

All the computational works in this study, such as the data pretreatment, model establishment and optimization, etc., were archived by MATLAB R2014a in a PC with a 8-core CPU and a 16 GB memory.

2. Material and Methods

2.1. Experiment and Measurement

A total of 87 fuel petrol samples were collected. An oil filter (TY-II/1000) was dedicated to evaporate samples until fully dewatering. The dewatering samples were considered as pure petrol. The octane number of the pure petrol samples can be detected by using SKY2102-VII Gasoline Octane Number Tester (Shanghai Shenkai Petroleum Instrument Co. Ltd.). The 87 values range from 83.4 to 89.6. The arranged pure petrol samples were placed in an ultrasonic cleaning machine for oscillation beforehand.

NIR spectra of the 87 pure petrol samples were recorded on the NIR-Quest 256 spectrometer (Ocean Optics, USA) fitted with In GaAs detector and quartz-halogen light source. The scanning spectral range was set 800 - 2498 nm with 4 nm resolution, so that we have 850 wavelengths per spectrum. The experiment temperature was controlled at 25˚C ± 1˚C and the relative humidity was at 47% ± 1% RH throughout the scanning process. Each sample was measured thrice and the mean value was calculated for model establishment.

NIR analytical process requires a modeling-testing division for samples. And modeling optimization process should have the samples partitioned for calibration and validation. Firstly, 22 samples were randomly selected for testing, which were not subjected to the modeling process. The remaining 65 samples were used for modeling. We are planning to have 45 samples for calibration and 20

for validation. To obtain representative sample sets, the calibration-validation partition should be carried out based on both the absorbance values and the measured values from SKY2102-VII, so the SPXY method [16] was utilized.

2.2. Pretreatment for Spectral Data

The NIR spectra of 87 pure petrol samples were showed in **Figure 1**. As can be seen in **Figure 1**, a sharp absorption peak appeared around 1400 nm and another around 1900 nm, which respectively corresponds to the first overtone of the chemical bonds of H-O and C=O. Because of these two peaks and the reasons of high frequency random noise, baseline drift, light scattering in raw spectra, the multivariate calibration models cannot smoothly give out prospective prediction results, thus pretreatment methods should be employed to reduce the interferences in raw data.

According to the physical properties of the NIR absorption band of H-O and C=O, we utilized the 1st-derivative Savitzky-Golay smoother (1st-d'SG) for data pretreatment, and the pretreated spectra were showed in **Figure 2**. We can see from **Figure 2** that the 1st-derivative + SG pretreatment well reserve the spectral features of the peak signals at 1400 nm and 1900 nm, and all spectral curves obviously became smoother.

The NIR spectral data after 1st-d'SG pretreatment were used for quantitative determination for the octane number in petrol samples. Based on the spectral in full scanning range, wavelength selection models were established for the calibration and validation samples by SPA method, so that the informative wavelengths can be selected. The selected wavelengths will be taken as the input variables of LSSVM regression and the NIR analytical models can be further established for optimization.

3. Results and Discussion

3.1. Wavelength Selection by SPA

SPA was used to select wavelengths for NIR quantitative analysis of fuel petrol. The NIR data were beforehand pretreated by first derivative SG smoother (1st-d'SG). According to Balabin's reports [18] [19], a good wavelength selection model provided by SPA always includes 2 - 20 variables (*i.e.* wavelengths). We tried to expand the number to 25, and respectively established models for the number changed from 1 to 25. The RMSE corresponding to each number of variables by SPA was showed in **Figure 3**. As **Figure 3** shows, the RMSE curve sharply declines when the number changes from 1 to 5, which indicated that the proper number should be larger than 5 in order to avoid over-fitting. And also, the curve seems to be stable when the number goes from 4 to 9, and rises when larger than 9. Thus we concluded that the number of informative wavelengths for the best NIR calibration model of fuel petrol should be no less than 4 and no more than 9.

Next, we reset the number of variables changed from 4 to 9 and repeat wave-

length selection by SPA. Then we have the best 8 variables (the most informative wavelengths) for the NIR analysis of the octane in fuel petrol. The specific 8 wavelengths were 1040, 1316, 1392, 1476, 1800, 1856, 1904, 2298 (the 8 circles in Figure 4). These wavelengths distributed at the peak or trough locations of the 1st-d'SG spectrum. This result reflected the merit of SPA that the selected informative variables do not appear at the flat region but gather at the special locations of the 1st-derivative curve, which point to some dedicated functional groups. The 8 wavelengths selected by SPA based on the 1st-d'SG pretreated data reserved the main features of spectra. Thus we use these 8 SPA-selected wavelengths to establish calibration models for NIR determination of the octane in fuel petrol.

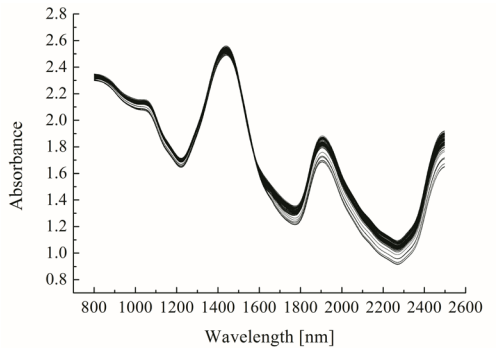

Figure 1. Near infrared absorbance spectra of 87 petrol samples.

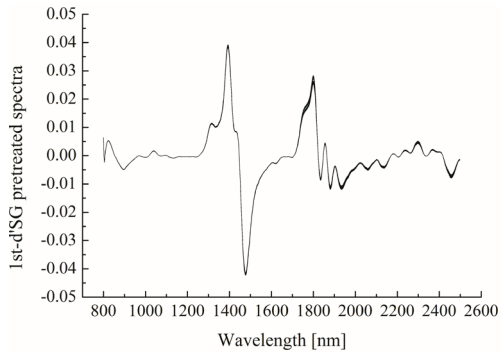

Figure 2. First derivative and SG pretreated near infrared spectra of 87 petrol samples.

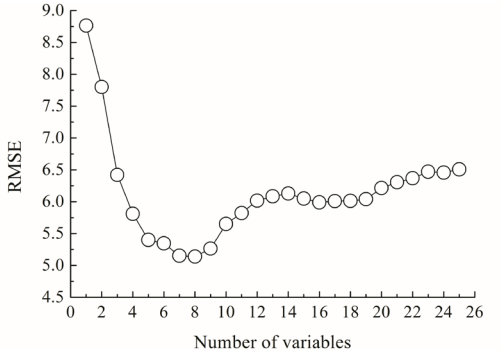

Figure 3. RMSE corresponding to each number of variables by SPA.

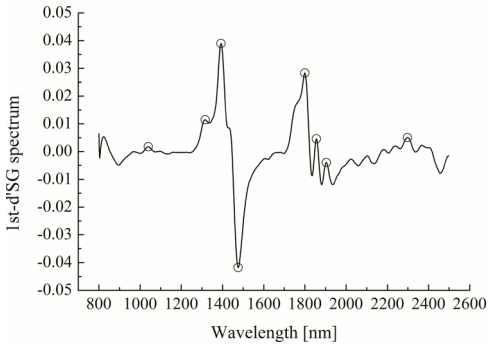

Figure 4. The 8 most informative variables selected by SPA.

3.2. LSSVM Model Establishment

NIR calibration models were established by using LSSVM regression with kernel function. Theradial basis function (RBF) kernel is common used in chemometrics [14]. RBF kernel includes two tunable parameters γ and σ. We utilized genetic algorithm (GA) to optimize the parameters. We have the GA iteration error of 0.02 at the 30th iteration, and the optimal value of γ and σ can be found. According to the abovementioned SPA wavelength-selected results, we established NIR calibration models by LSSVM regression, and further to predict the octane number for the prediction samples. The prediction results were listed in **Table 1**. Here we used correlation coefficient (R) and root mean square error (RMSE) as the model indicators.

The results in **Table 1** indicated that, on one hand, the raw spectra and the 1st-d'SG pretreated data let out the calibration models with relatively low prediction bias, while the 1st-d'SG pretreatment provided a noise-reduced data to give out an optimal model with higher R and minor RMSE. On the other hand, SPA wavelength selection method performs well for LSSVM models. The SPA-selected 8 informative wavelengths did improve the LSSVM modeling results, with the GA-optimized LSSVM modeling parameters (γ, σ) were (89.34, 32.45), and calibrating R and RMSE were 0.951 and 3.282, respectively, which was quite satisfactory for the NIR quantitative analysis of the octane number in fuel petrol.

For model evaluation, the 22 independent testing samples, not subjected to the modeling process, were used to examine the optimal SPA-LSSVM models with or without 1st-d'SG pretreatment. The predictive values of the 22 testing samples can be calculated respectively for raw data and the pretreated data, and the correlation charts were showed in **Figure 5**. The correlation coefficient for the raw data modeling was 0.876 and RMSE was 4.416 (**Figure 5(a)**). The correlation coefficient for the 1st-d'SG data modeling was 0.903 and RMSE was 4.128 (**Figure 5(b)**), which was a little better than the results on raw data. We can see from **Figure 5(b)** that the testing samples evenly distributed on both sides of the regression line and the distance to the regression line is relatively closer. It showed that the SPA-LSSVM model based on 1st-d'SG pretreated data gave out the best model testing results.

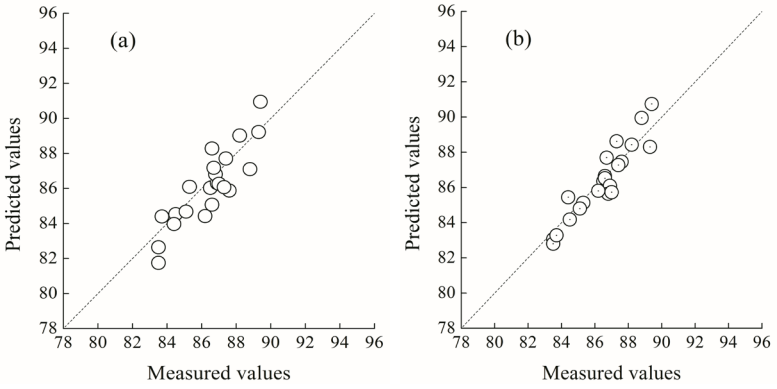

Figure 5. Prediction of the octane number based on the raw spectra (a) and the 1st-d'SG pretreated spectra (b).

Table 1. Prediction results by SPA-LSSVM models.

	γ	σ	Calibration set		Validation set	
			R	RMSE	R	RMSE
Raw spectra	87.51	28.16	0.946	3.746	0.913	4.048
1st-d'SG	89.34	32.45	0.951	3.282	0.927	3.841

Combined considering the modeling results showed in Figure 5 and Table 1, we conclude that SPA method can effectively extract informative wavelengths from the full NIR scanning range. The 1st-d'SG pretreatment has the advantage of selecting wavelengths containing the implicit information. LSSVM regression can establish calibration models by linear and non-linear training mechanism. The combination of 1st-d'SG pretreatment and SPA-LSSVM show the potential to predict the octane number for fuel petrol samples.

4. Conclusions

NIR spectrometry was applied to quantitative determination of the octane number in fuel petrol samples. LSSVM models were established with its merit on linear and non-linear training mechanism, and the informative wavelengths were extracted by SPA method. Additionally, we noted that pretreatment is another way to improve the prediction results of calibration models.

The NIR spectral data keep the octane number going through the 1st-d'SG pretreatment. The main feature of octane in fuel petrol was reserved and obviously appeared at the peaks and troughs in the 1st-d'SG pretreated spectra. The noise interference was reduced and the pretreated data become much smooth. Additionally, SPA wavelength selection method was discussed respectively for the raw spectra and the 1st-d'SG data. Results showed that the SPA-selected wavelengths contained the main information of the full spectrum, and also the data dimension was effectively reduced. The 8 SPA-selected wavelengths all distribute at the peak or trough locations of the 1st-d'SG spectrum. It meant that the 1st-derivative peaks physically reflect the spectral information of

octane.

LSSVM models were established for determination of the octane number. Results show that SPA-LSSVM modeling based on the 1st-d'SG pretreated data gave out high correlation coefficient and relatively low RMSE for validation samples. And, for further testing, the 22 independent samples evenly distributed on both sides of the regression line and the distance to the regression line is relatively closer. It meant that the best testing results was obtained by SPA-LSSVM model based on 1st-d'SG data.

Acknowledgements

The research was funded by the National Natural Scientific Foundation of China (No. 61505037), the Natural Scientific Foundation of Guangxi (No. 2015GXNSFBA139259, No. 2016GXNSFBA380077) and the Scientific Research Project of Guangxi Education Office (No. KY2015LX538).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Burns, D.A. and Ciurczak, E.W. (2008) Handbook of Near-Infrared Analysis. 3rd Edition, Taylor and Francis, New York.

[2] Allouche, Y., López, E.F., Maza, G.B. and Márquez, A.J. (2015) Near Infrared Spectroscopy and Artificial Neural Network to Characterise Olive Fruit and Oil Online for Process Optimization. *Journal of Near Infrared Spectroscopy*, **23**, 111-121. https://doi.org/10.1255/jnirs.1155

[3] Alishahi, A., Farahmand, H., Prieto, N. and Cozzolino, D. (2010) Identification of Transgenic Foods Using NIR Spectroscopy: A Review. *Spectrochim Acta Part A*: *Molecular and Biomolecular Spectroscopy*, **75**, 1-7. https://doi.org/10.1016/j.saa.2009.10.001

[4] Chen, H.Z., Xu, L.L., Song, Q.Q., Feng, Q.X. and Tang, G.Q. (2016) Rapid Detection on Surface Color of Shatian Pomelo Using Vis-NIR Spectrometry for the Identification of Maturity. *Food Analytical Methods*, **9**, 192-201. https://doi.org/10.1007/s12161-015-0188-5

[5] Stauffer, E., Dolan, J.A. and Newman, R. (2008) Flammable and Combustible Liquids. In: *Fire Debris Analysis*, Academic Press, Burlington, 199-233. https://doi.org/10.1016/B978-012663971-1.50011-7

[6] Alinezhad, K., Hosseini, M. and Movagarnejad, K. (2010) Experimental and Modeling Approach to Study Separation of Water in Crude Oil Emulsion under Non-Uniform Electrical Field. *Korean Journal of Chemical Engineering*, **27**, 198-205. https://doi.org/10.1007/s11814-009-0324-2

[7] Foong, T.M., Morganti, K.J., Brear, M.J., Da Silva, G., Yang, Y. and Dryer, F.L. (2014) The Octane Numbers of Ethanol Blended with Gasoline and Its Surrogates. *Fuel*, **115**, 727-739. https://doi.org/10.1016/j.fuel.2013.07.105

[8] Brouillette, C., Smith, W., Shende, C., Gladding, Z., Farquharson, S., Morris, R.E., Cramer, J.A. and Schmitigal, J. (2016) Analysis of Twenty-Two Performance

Properties of Diesel, Gasoline, and Jet Fuels Using a Field-Portable Near-Infrared (NIR) Analyzer. *Applied Spectroscopy*, **70**, 746-755. https://doi.org/10.1177/0003702816638279

[9] Lee, S., Choi, H., Cha, K. and Chung, H. (2013) Random Forest as a Potential Multivariate Method for Near-Infrared (NIR) Spectroscopic Analysis of Complex Mixture Samples: Gasoline and Naphtha. *Microchemical Journal*, **110**, 739-748. https://doi.org/10.1016/j.microc.2013.08.007

[10] Ferreiro-Gonzalez, M., Ayuso, J., Alvarez, J.A., Palma, M. and Barroso, C.G. (2014) New Headspace-Mass Spectrometry Method for the Discrimination of Commercial Gasoline Samples with Different Research Octane Numbers. *Energy Fuels*, **28**, 6249-6254. https://doi.org/10.1021/ef5013775

[11] Chen, H., Feng, Q., Jia, Z. and Song, Q. (2014) Improvement of Partial Least Squares Modelling for Determination of Soil Nitrogen by Fourier Transform Near-Infrared Spectrometry. *Asian Journal of Chemistry*, **26**, 4839-4844.

[12] Chen, H., Liu, Z., Cai, K., Xu, L. and Chen, A. (2018) Grid Search Parametric Optimization for FT-NIR Quantitative Analysis of Solid Soluble Content in Strawberry Samples. *Vibrational Spectroscopy*, **94**, 7-15. https://doi.org/10.1016/j.vibspec.2017.10.006

[13] Araujo, M.C.U., Saldanha, T.C.B., Galvao, R.K.H., Yoneyama, T. and Chame, H.C. (2001) The Successive Projections Algorithm for Variable Selection in Spectroscopic Multicomponent Analysis. *Chemometrics and Intelligent Laboratory Systems*, **57**, 65-73. https://doi.org/10.1016/S0169-7439(01)00119-8

[14] Barman, I., Dingari, N.C., Singh, G.P., Soares, J.S., Dasari, R.R. and Smulko, J.M. (2012) Investigation of Noise-Induced Instabilities in Quantitative Biological Spectroscopy and Its Implications for Non-Invasive Glucose Monitoring. *Analytical Chemistry*, **84**, 8149-8156. https://doi.org/10.1021/ac301200n

[15] Chen, H.Z., Ai, W., Feng, Q.X. and Tang, G.Q. (2015) FT-MIR Modelling Enhancement for the Quantitative Determination of Haemoglobin in Human Blood by Combined Optimization of Grid-Search LSSVR Algorithm with Different Pre-Processing Modes. *Analytical Methods*, **7**, 2869-2876. https://doi.org/10.1039/C5AY00145E

[16] Galvao, R.K.H., Araujo, M.C.U., Jose, G.E., Pontes, M.J.C., Silva, E.C. and Saldanha, T.C.B. (2005) A Method for Calibration and Validation Subset Partitioning. *Talanta*, **67**, 736-740. https://doi.org/10.1016/j.talanta.2005.03.025

[17] Silva, M., Ferreira, M.H., Braga, J.W.B. and Sena, M.M. (2012) Development and Analytical Validation of a Multivariate Calibration Method for Determination of Amoxicillin in Suspension Formulations by Near Infrared Spectroscopy. *Talanta*, **89**, 342-351. https://doi.org/10.1016/j.talanta.2011.12.039

[18] Balabin, R.M., Safieva, R.Z. and Lomakina, E.I. (2011) Support Vector Machine Regression (SVR/LS-SVM)—An Alternative to Neural Networks (ANN) for Analytical Chemistry? Comparison of Nonlinear Methods on Near Infrared (NIR) Spectroscopy Data. *Analyst*, **136**, 1703-1712. https://doi.org/10.1039/c0an00387e

[19] Balabin, R.M. and Smirnov, S.V. (2011) Variable Selection in Near-Infrared Spectroscopy: Benchmarking of Feature Selection Methods on Biodiesel Data. *Analytica Chimica Acta*, **692**, 63-72. https://doi.org/10.1016/j.aca.2011.03.006