We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Contributions of Multivariate Statistics in Oil and Gas Industry

Leandro Valim de Freitas, Ana Paula Barbosa Rodrigues de Freitas,
Fernando Augusto Silva Marins, Estéfano Vizconde Veraszto,
José Tarcísio Franco de Camargo, J. Paulo Davim and Messias Borges Silva

Additional information is available at the end of the chapter

1. Introduction

This study aims to develop and validate multivariate mathematical models in order to monitor in real time the quality processing of derivatives in an oil refinery.

Methods heavily based on statistical and artificial intelligence as multivariate or chemometric methods have been widely used in the oil industry (KIM; LEE, KIM, 2009). Several articles have been written about applications of multivariate analysis to predict properties of oil derivatives (Santos Junior et al., 2005; Chung, 2007).

Pasadakis, Sourligas and Foteinopoulos (2006) have used the first six principal components of Principal Component Analysis (PCA) as input variables in nonlinear modeling of oil properties.

Pasquini and Bueno (2007) have proposed a new approach to predict the true boiling point of oil and its degree API (American Petroleum Institute) - a measure of the relative density of liquids by Partial Least Squares (PLS) and Artificial Neural Networks (ANN). Samples of mixtures oil were obtained from various producing regions of Brazil and abroad. In this application, the models obtained by the PLS method were superior to neural networks. The short time required for prediction the properties justifies the proposed of characterization the oil quicker to monitor refining processes.

Teixeira et al. (2008) in work with Brazilian gasoline used the multivariate algorithm Soft Independent Modeling of Class Analogy (SIMCA) for clusters analysis. Aiming to quantify the amount of adulteration of gasoline by other hydrocarbons, the PLS method was applied. Finally, the models were validated internally by cross-validation algorithm and externally with an independent set of samples.

Bao and Dai (2009) studied different multivariate methods, including linear and nonlinear techniques in order to minimize the error of prediction by models developed for quality control of gasoline. Lira et al. (2010) applied the PLS method for inference of the quality parameters: density, sulfur concentration and distillation temperatures of the mixture diesel / bio-diesel, providing great savings in time compared with the traditional methods by laboratory equipment.

Aleme, Corgozinho and Barbeira (2010) have conducted a study of classification of samples using the PCA method for discrimination of diesel oil type and the prediction of their origin.

Paiva Ferreira and Balestrassi (2007) have combined the Response Surface Method (RSM) of Design of Experiments (DOE) with Principal Component Analysis in optimizing multiple correlated responses in a manufacturing process.

Huang, Hsu and Liu (2009) have used Mahalanobis-Taguchi integrated with Artificial Neural Networks in data mining to look for patterns and modeling in manufacturing. Pal and Maiti (2010) have adopted the Mahalanobis-Taguchi algorithm to reduce the dimensionality of multivariate data and for optimization with Metaheuristics in the sequence.

Liu et al. (2007) have made inferences about quality parameters of jet fuel using Multiple Linear Regression (MLR) and ANN. The work showed that the performance of modeling by ANN was superior.

In optimization of multivariate models, there are applications combined with Multivariate Analysis of Metaheuristics, such as simulated annealing (SAUNIER, et al., 2009), genetic algorithm (GA) (Roy, Roy, 2009) tabu search (QI; SHI; KONG, 2010), particle swarm (Pal; Mait, 2010), and ant colony (Goodarzi; Freitas; Jensen, 2009; Allegrini; Oliveri, 2011).

With the objective of optimizing the dimensionality of multivariate models and avoid the overfitting phenomenon in determining principal components, Xu and Liang (2001) have used the Monte Carlo Simulation on simulated data sets and two real cases. Gourvénec et al. (2003) compared Monte Carlo cross-validation with the traditional method of cross validation to determine the appropriate number of latent variables.

Adler e Yazhemsky (2010) have combined the Monte Carlo Simulation, PCA and Data Envelopment Analysis (DEA) in a context where there is a relatively large number of variables related to the number of observations for decision making. Llobet et al. (2005), by means a Multiple Criteria Decision-Making (MCDM) model, have used Fuzzy classification of samples of chips. For prediction oxidative and hydrolytic properties, was used an electronic nose based on PLS models, with prior selection of input variables by a GA Metaheuristic.

Wu, Feng and Wen (2011), in studies related to Botany, compared the performance of the growth of a tree species - Carya Cathayensis Sarg by PCA methods and Analytic Hierarchy Process (AHP), identifying the advantages and the disadvantages of each method, although the results obtained by both have been essentially identical.

Zhang et al. (2006) have combined the method Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE), from the Elimination et la Choix Traduisant Réalité (ELECTRE) and Geometrical Analysis for Interactive Assistance (GAIA) with PCA and PLS methods to classify 67 oils and determine an indicator of product quality. Purcell, O'Shea and Kokot (2007) also combined PROMETHEE and GAIA with PCA and PLS in studies related to cloning of sugarcane.

Regarding to the control charts designed to monitor the mean vector, Machado and Costa (2008) have studied the performance of T^2 charts based on principal components for monitoring multivariate processes. Lourenço et al. (2011) have used the principles of Process Analytical Technology (PAT) in the construction of control charts based on the scores of the first principal component versus time for the on-line monitoring of pharmaceutical processes.

Moreover, Multivariate Analysis is an important technique in various areas of knowledge such as Data Mining (Kettaneh; Berglund; Wold, 2005); Econometrics (Mackay, 2006); Marketing (Ahn; Choi; Han, 2007) and Supply Chain Management (Pozo et al., 2012).

2. Application: Oil refining

The first process in a refinery is atmospheric distillation or direct distillation, where components of crude oil are separated into different sections using different boiling points. The main products obtained in this process are: liquefied petroleum gas (LPG), naphtha - precursor of gasoline, jet fuel, diesel and fuel oil.

Additionally, refineries usually have a second tower, vacuum distillation, to produce diesel cuts. These intermediate streams feeding a chemical process called Fluid Catalytic Cracking (FCC). In this, two noble streams are generated: LPG, and gasoline. It is a refining scheme much more flexible, but though modern, may also present difficulties for framing products stricter specifications.

The production scheme level 3 is more flexible and cost effective than the previous one, because it uses the chemical process of Coking, which transforms a fraction of lower value - vacuum residue of distillation towers, in the noblest products like LPG, gasoline, naphtha and diesel oil.

This final refining scheme incorporates the process Hydrotreating of middle fractions generated in the Coker Unit, enabling increased supply of diesel with good quality. This scheme allows a more balanced supply of gasoline and diesel oil, producing more diesel and less gasoline than the previous settings.

Of course, there are other macro-processes and auxiliary processes such as water treatment plant, effluent disposal, sulfur recovery units, units of hydrogen generation and consequently other interconnections, details of which are not subject of this work (ANP, 2012).

3. Methods

3.1. Acquisition database: Infrared radiation

In the oil industry, signs of infrared radiation generated by sensors are associated with the prediction of the quality of distillates such as naphtha, gasoline, diesel and jet fuel (Kim, Cho; Park, 2000).

Freitas et al. (2012) and Pasquini (2003) explain this instrumentation (Figure 1): the polychromatic radiation emitted by the source has a wavelength selected by a Michelson interferometer. The beam splitter has a refractive index such that approximately half of the radiation is directed to the fixed mirror and the other half is reflected, reaching the movable mirror and is therefore reflected by them. The optical path differences occur due the movement of the movable mirror that promotes wave interference.

An interferogram is obtained as a result of a graph of the signal intensity received by the detector versus the difference in optical path traveled by the beams. By calculating the Fourier Transform (FT) the interferogram can be written as a sum of sines and cosines (Tarumi et al, 2005) and in this case, happens to be called transmittance spectra, T (Forato; Filho; Colnago, 1997). Finally, the spectrum of transmittance, T, is converted to absorbance spectra, A, by co-logarithm of T (Suarez et al. May 2011). The absorbance can be interpreted as the amount of radiation that the sample absorbs and the transmittance, the fraction of radiation that the sample does not absorb. These phenomena occur depending on their chemical composition (Kramer; Small, 2007).
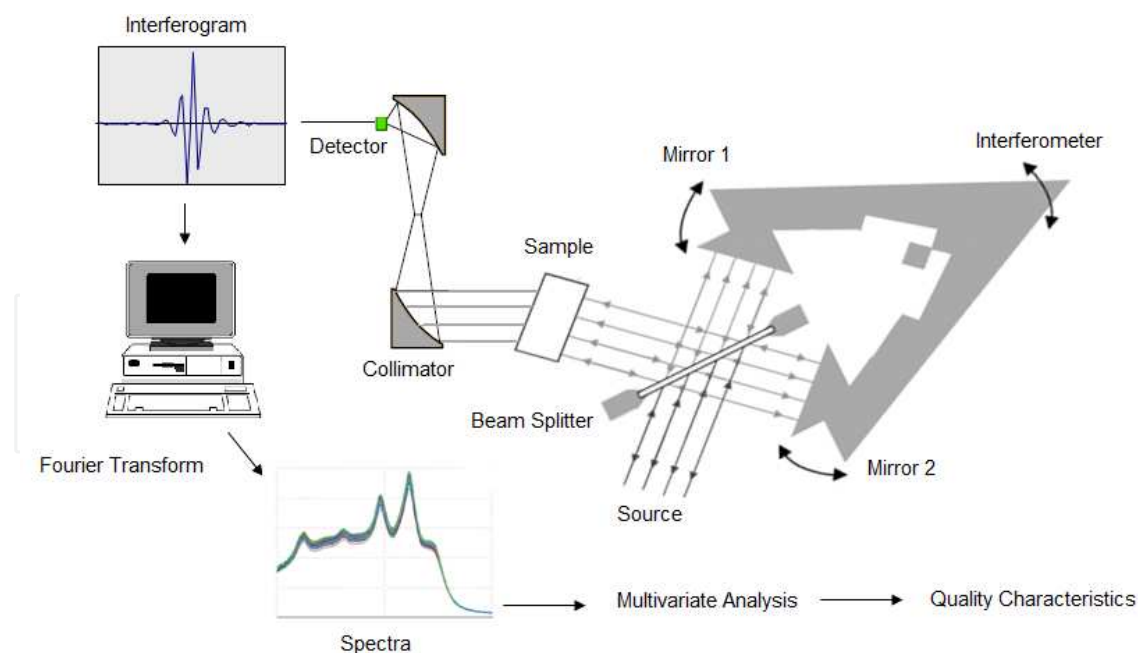
Figure 1. Scheme for technology acquisition database (Adapted from Pasquini, 2003)

The chemical bonds of the type carbon-hydrogen (CH), oxygen-hydrogen (OH) and nitrogen-hydrogen (NH), present in petroleum products (Pasquini; Bueno, 2007), are responsible for the absorption of infrared radiation, however, are not very intense and

overlap. The broad spectral bands formed are difficult to interpret (Skoog; Holler; Crouch, 2007) due to the phenomenon of collinearity (Naes; Martens, 1984). The origin of this phenomenon is associated with the manner in which the infrared radiation interacts with matter and can be demonstrated by Quantum Mechanics at work Pasquini (2003).

These input variables (radiation absorbed), called X_i are correlated, so are said collinear or multicollinear (NAES et al., 2002). To illustrate the collinearity, X is a dummy matrix a_{ij} with i rows and j in terms columns, where a_{ij} is the radiation absorption of three samples i (i = 1, 2, 3) at two wavelengths j (j = 1, 2).

$$\mathbf{X} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

The columns of X are linearly dependent, so the variables column j_1 and j_2 are colinear, that is, when increases j_1, j_2 increases proportionally. This causes the determinant of X'X to be zero, where X' is the transpose of matrix X.

$$X^T X = \begin{bmatrix} 14 & 28 \\ 28 & 56 \end{bmatrix}$$

Then, the det (X'X) = (14.56) - (28.28) = 0 and this according to Naes et. al (2002) means that there is a singular error matrix and that those erros are propagated when the dependent properties, Y, are determined by regression methods which are not based on the principal components, such as the MLR.

However, the multivariate approaches such as Principal Component Regression (PCR) and PLS have been quite appropriate due to dimensionality reduction, which creates a new set of variables called principal components (Rajalahti; Kvalheim, 2011). So with data mining for Multivariate Analysis, it is possible to relate the physicochemical properties (quality characteristics) of products with the chemical composition of the sample reflected by the absorption spectra. So once modeled a property, just a sample is subjected to infrared radiation to predict their properties.

3.1. Acquisition data base: Reference properties

In this work were modeled properties of gasoline, diesel and jet fuel. For gasoline, the octane number and for diesel oil and jet fuel, the kinematic viscosity property.

According to Freitas (2012), kinematic viscosity of the diesel oil and jet fuel products is an important property in terms of its effect on power system and in fuel injection. Both high and low viscosities are undesirable since they can cause, among others, problems in fuel atomization. The formation of large and small droplets (low viscosity), can lead to a poor distribution of fuel and compromise the mixture air – fuel resulting in an incomplete combustion followed by loss power and greater fuel consumption.

The octane number of a gasoline is an important characteristic which is related to their ability to burn in spark-ignition engines. It is determined by comparing its tendency to detonate with the reference fuel with octane known under standard operating conditions.

When it comes to defining the octane required by engines, many countries use anti-knock index (I), defined by Equation 1:

$$I = \frac{MON + RON}{2} \tag{1}$$

where MON is the Motor Octane Number and RON is the Research Octane Number. The method MON measures the resistance to detonation when gasoline is being burned in the most demanding operating conditions and at higher rotations. The test is done in motors CFR (Cooperative Fuel Research), single-cylinder with variable compression ratio equipped with the necessary instrumentation in a stationary base, as shown in Figure 2.

Figure 2. CFR engine for MON octane (WAUKESHA, 2012)

The RON method evaluates the resistance of the gasoline to detonation under milder conditions and work in less rotation than that measured by octane number MON. The test is done in similar engines to those used for testing in MON octane.

It takes two hours and half to run the test MON and it is spent the same time for the test RON.

4. Results and discussion

Samples of gasoline, diesel and jet fuel, collected during 1 year, were subjected to laboratory tests, to determine the input variables, X_i, which are the infrared radiation absorbed, and the response variables, Y_i, that are physicochemical properties. The physicochemical properties will be predicted by PLS models.

The Table 1 summarizes the validation results of each model for products gasoline, diesel and jet fuel, where RMSEP (Root Mean Square Error of Prediction) corresponds to the standard deviation of the residuals (differences between measured and predicted values by the model).

The Figures 3-6 illustrate that the residues of models follow normal distribution, since in all cases the p-value was greater than 0.05.

Product	Property	Number of samples	Latent Variables	RMSEP	Correlation
Diesel Oil	Viscosity (40ºC)	180	8	0.116 cSt	0.9368
Gasoline	MON	350	6	0.22	0.8723
Gasoline	RON	350	7	0.22	0.9891
Jet Fuel	Viscosity (-20ºC)	279	7	0.01 cSt	0.8836

Table 1. Summary of results of modeling and validation.
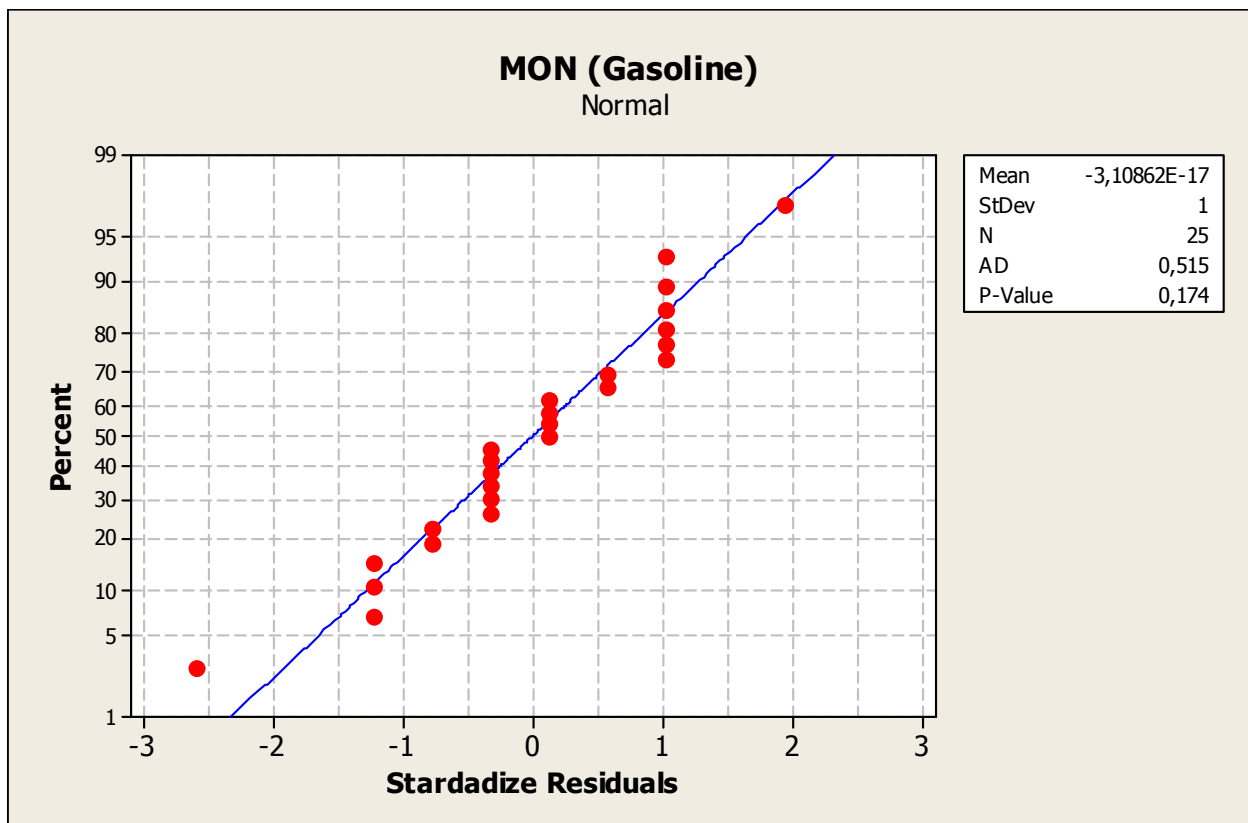
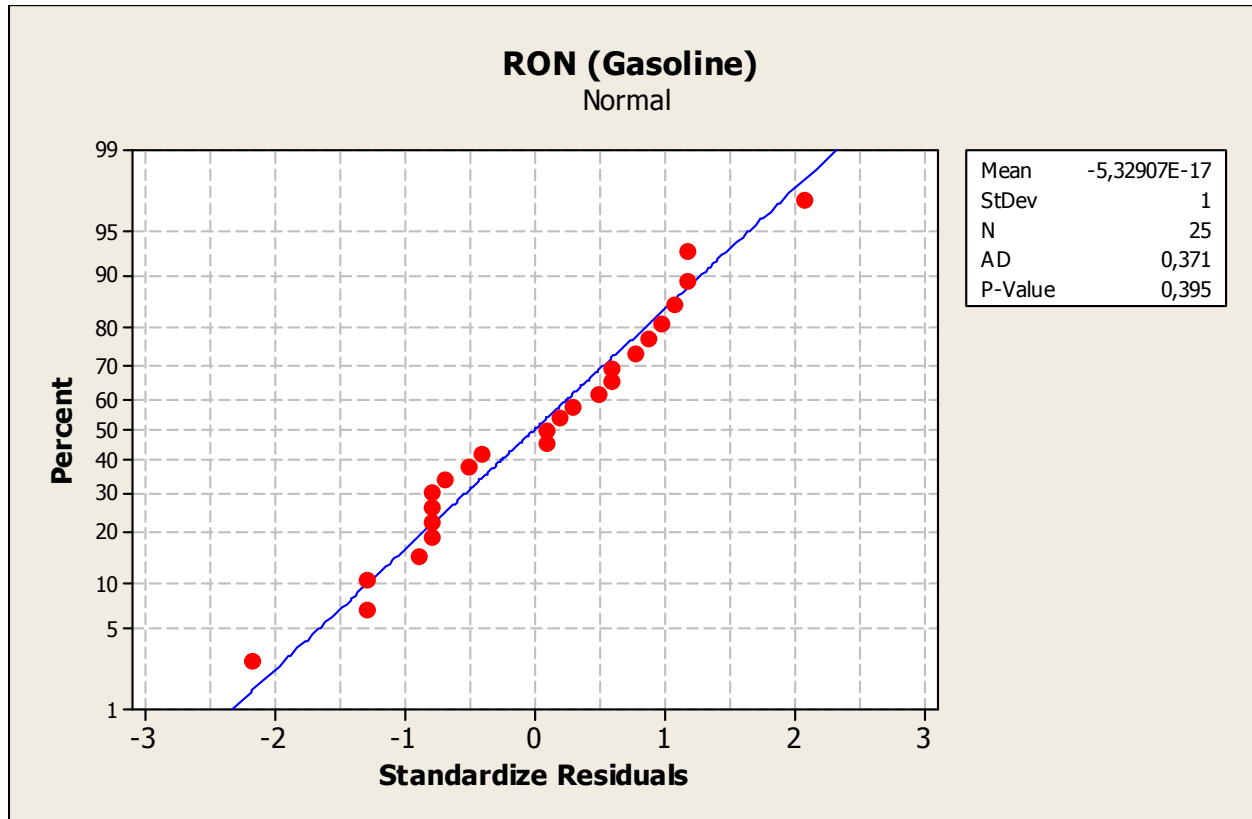**Figure 3.** Normality Test for the property MON

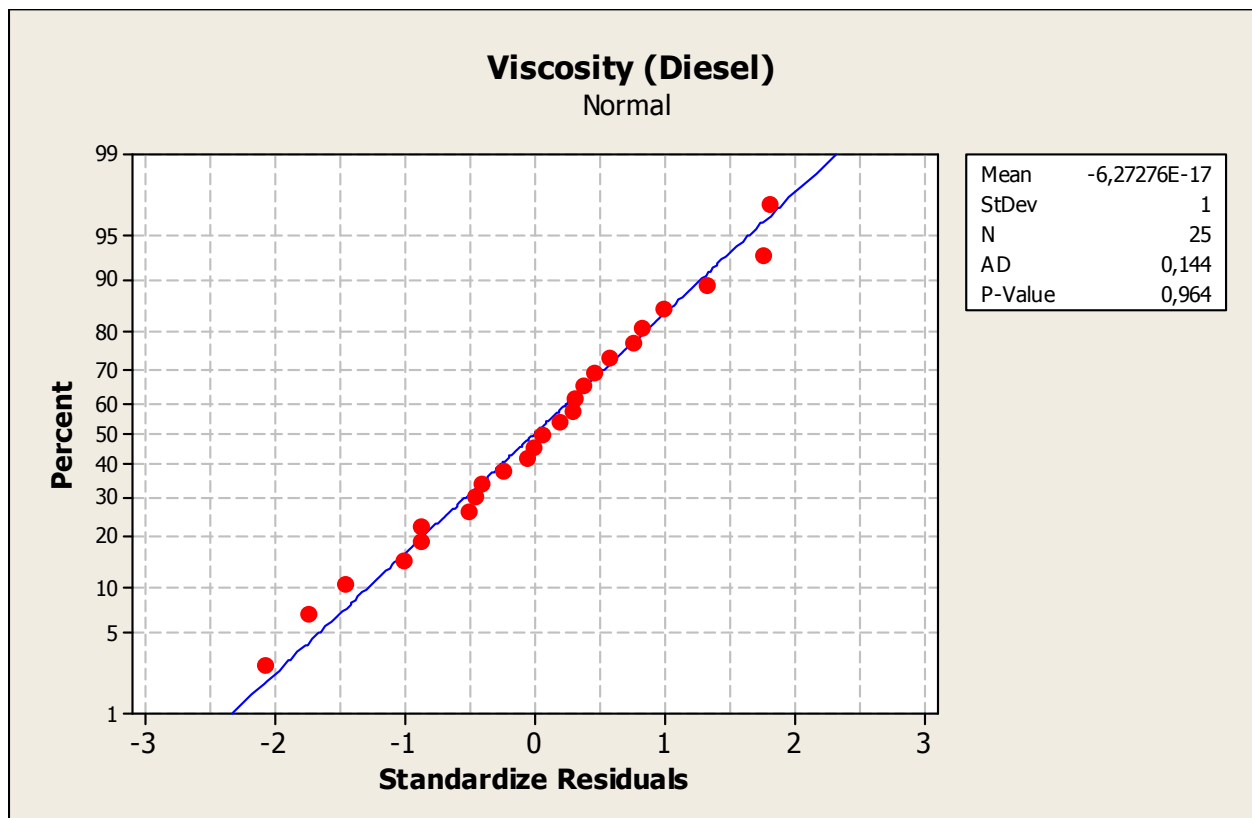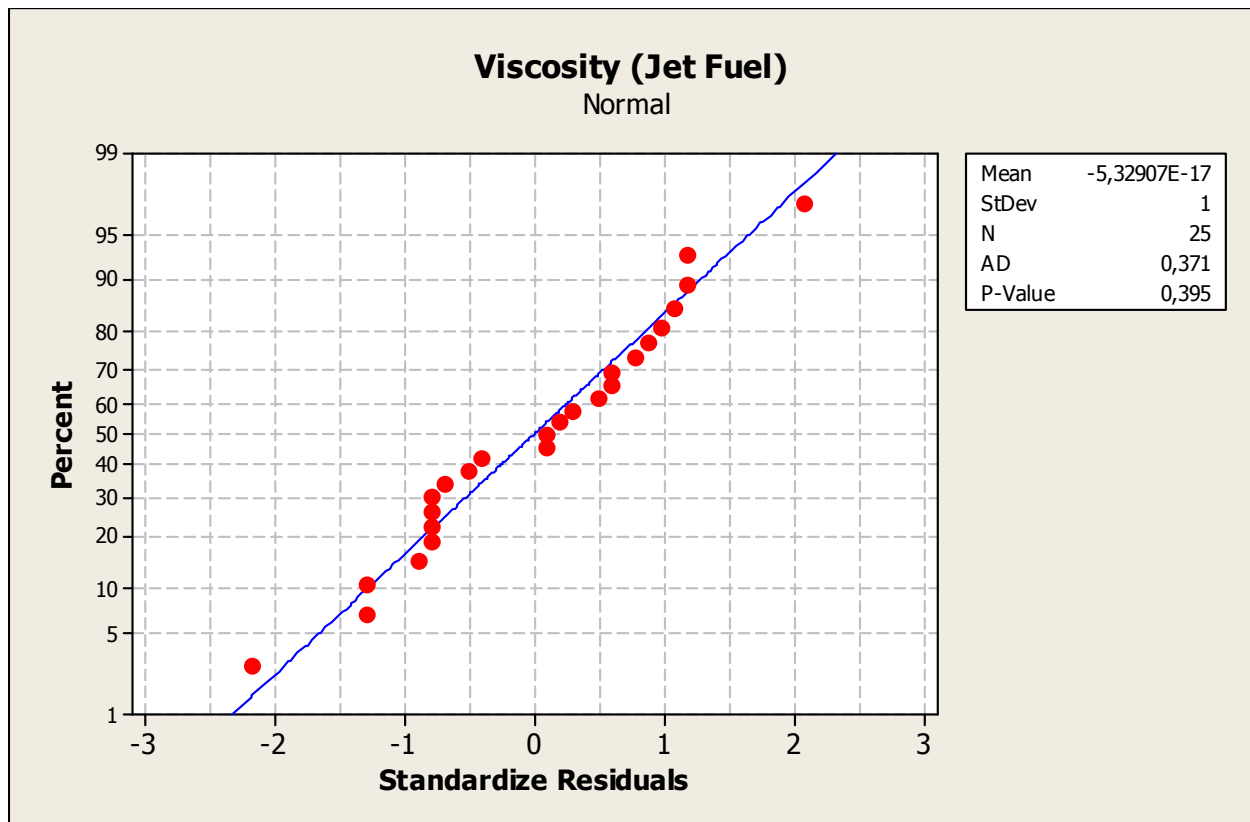**Figure 4.** Normality Test for the property RON

Figure 5. Normality Test for the property viscosity (diesel)

Figure 6. Normality Test for the property viscosity (jet fuel)

5. Conclusions

The following conclusions can be drawn from the results of this study:

It was possible to model mathematically the properties octane number and viscosity of the products gasoline, diesel and jet fuel.

The developed models were externally validated according to ASTM D-6122 and their predictions have precision equivalent to the reference methods.

The results were used in an oil refinery and contributed immensely to speed up the decision-making in blendings systems. Unlike the laboratory trials, the response time of a property along with the computational time does not exceed three minutes.

Author details

Leandro Valim de Freitas
Petróleo Brasileiro SA (PETROBRAS), Brazil
São Paulo State University (UNESP), Brazil

Ana Paula Barbosa Rodrigues de Freitas and Messias Borges Silva
São Paulo State University (UNESP), Brazil
University of São Paulo (USP), Brazil

Fernando Augusto Silva Marins
São Paulo State University (UNESP), Brazil

Estéfano Vizconde Veraszto
Municipal College Franco Montoro Professor (FMPFM), Brazil
Campinas State University (UNICAMP), Brazil

José Tarcísio Franco de Camargo
Municipal College Franco Montoro Professor (FMPFM), Brazil

J. Paulo Davim
Aveiro University (UA), Portugal

6. References

Adler, N.; Yazhemsky, E. Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. European Journal of Operational Research, 2010, 202, 273-284.

AGENCY OF OIL, NATURAL GAS AND BIOFUELS. Schemes Production in Oil Refining: <http://www.anp.gov.br/?pg=7854&m=esquema+de+refino&t1=&t2=esquema+de+refino&t3=&t4=&ar=0&ps=1&cachebust=1331008874709>. Accessed in 06 March 2012.

Ahn, H.; Choi, E.; Han, I. Extracting underlying meaningful features and canceling noise using independent component analysis for direct marketing. Expert Systems with Applications, 2007, 33, 181-191.

Aleme, H. G.; Corgozinho, C. N. C.; Barbeira, P. J. S. Diesel oil discrimination by origin and type using physicochemical properties and multivariate analysis. Fuel, 2010, 89, 3151-3156.

Bao, X.; Dai, L. Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties. Fuel, 2009, 88, 1216-122.

Forato, L. A.; Filho, R. B.; Colnago, L. A. Estudos de métodos de aumento de resolução de espectros de FTIR para análises de estruturas secundárias de proteínas. Química Nova, 1997, 20, 146-150.

Freitas, L. V.; Freitas, A. P. B. R.; Marins, F. A. S.; Loures, C. C. A.; Silva, M. B. Multivariate modeling in quality control of viscosity in fuel: an application in oil industry. Fuel Injection in Automotive Engineering. Rijeka: InTech, 2012.

Goodarzi, M; Freitas, M.; Jensen, R. Ant colony optimization as a feature selection method in the QSAR modeling of anti-HIV-1 activities of 3-(3,5-dimethylbenzyl) uracil derivatives using MLR, PLS and SVM regressions. Chemometrics and Intelligent Laboratory Systems, 2009, 98, 123-129.

Gourvénec, S.; Pierna, J. A. F.; Massart, D. L.; Rutledge. An evaluation of the PoLiSh smoothed regression and the Monte Carlo Cross-Validation for the determination of the complexity of a PLS model. Chemometrics and Intelligent Laboratory Systems, 2003 68, 48-51.

Huang, C. L.; Hsu, T. S.; Liu, C. M. The Mahalanobis–Taguchi system – Neural network algorithm for data-mining. Expert Systems with Applications, 2009, 36, 5475-5480.

Kettaneh, N.; Berglund, A.; WOLD, S. PCA and PLS with very large data sets. Computational Statistics & Data Analysis, 2005, 48, 69-85.

Kim, D.; Lee, J.; Kim, J. Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties. Korean Journal of Chemical Engineering, 2009; 26; 489-495.

Kim, K.; Cho, I.; Park, J. Use of real-time NIR (near infrared) spectroscopy for the on-line optimization of a crude distillation unit. In: NPRA, Computer Conference. Chicago, 2000.

Kramer, K. E.; Small, G. W. Robust absorbance computations in the analysis of glucose by near-infrared spectroscopy. Vibrational Spectroscopy, 2007, 43, 440-446.

Lira, L. F. B.; Vasconcelos, F. V. C.; Pereira, C. F.; Paim, A. P. S.; Stragevitch, L.; Pimentel M. F. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. Fuel, 2010, 89, 405-409.

Liu, H.; Yu, J.; Xu, J.; Fan, Y.; Bao, X. Identification of key oil refining technologies for China National Petroleum Co. (CNPC). Energy Police, 2007, 35, 2635 -2647.

Llobet, M. V. E.; Brezmes J.; Vilanova, X.; Correig, X. A fuzzy ARTMAP- and PLS-based MS e-nose for the qualitative and quantitative assessment of rancidity in crisps. Sensors and Actuators 2005, 106, 677-686.

Lourenço, V.; Herdling, T.; Reich, G.; Menezes, J. C.; Lochmann, D. Combining microwave resonance technology to multivariate data analysis as a novel PAT tool to improve process understanding in fluid bed granulation. European Journal of Pharmaceutics and Biopharmaceutics, 2011, 78, 513-521.

Machado, M. A. G.; Costa, A. F. B. The use of principal components and univariate charts to control multivariate processes. Operational Resarch, 2008, 28, 173-196.

Mackay, D. Chemometrics, econometrics, psychometrics - How best to handle hedonics? Expert Systems with Applications, 2006, 17, 529-535.

Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. Partial Least Squares. In: Multivariate calibration and classification. Chichester. NIR Publications, 2002.

Naes, T.; Martens, H. Multivariate calibration II. Chemometric methods. Trends in analytical chemistry, 1984, 3, 266-271.

Paiva, A. P. Metodologia de superfície de resposta e análise de componentes principais em otimização de processos de manufatura com múltiplas respostas correlacionadas. PhD Thesis; Itajubá Federal University, 2006.

Pal, A.; Maiti, J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis–Taguchi system using Mahalanobis distance and binary particle swarm optimization. Expert Systems with Applications, 2010, 37, 1286-1293.

Pal, A.; Maiti, J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis–Taguchi system using Mahalanobis distance and binary particle swarm optimization. Expert Systems with Applications, 2010, 37, 1286-1293.

Pasadakis, N.; Sourligas, S.; Foteinopoulos, Ch. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. Fuel, 2006, 85, 1131-1137.

Pasquini, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. Journal of Chemical Brazilian Society, 2003, 14, 198-219.

Pasquini, C.; Bueno, A. F. Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. Fuel, 2007, 86, 1927-1934.

Pasquini, C.; Bueno, A. F. Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. Fuel, 2007, 86, 1927-1934.

Pozo, C.; Fermenia, R. R.; Caballero, J.; Gosa, G. G. Jiménez, L. On the use of Principal Component Analysis for reducing the number of environmental objectives in multi-objective optimization: Application to the design of chemical supply chains. Chemical Engineering Science, 2012, 69, 146-158.

Purcell, D. E.; O'shea, M. G.; Kokot, S. Role of chemometrics for at-field application of NIR spectroscopy to predict sugarcane clonal performance. Chemometrics and Intelligent Laboratory Systems, 2007, 87, 113-124.

Qi, S.; Shi, W. M.; Kong, W. Modified tabu search approach for variable selection in quantitative structure–activity relationship studies of toxicity of aromatic compounds. Artificial Intelligence in Medicine, 2010, 49, 61-66.

Rajalahti, T.; Kvalheim, O. M. Multivariate data analysis in pharmaceutics: A tutorial review. International Journal of Pharmaceutics, 2011, 417, 280-290.

Roy, K.; Roy, P. P. Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. European Journal of Medicinal Chemistry, 2009, 44, 2913-2922.

Santos Jr V. O.; Oliveira F. C. C.; Lima, D. G.; Petry, A. C.; Garcia, E.; Suarez, P. A. Z.; Rubim, J. C. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. Analytica Chimica ACTA, 2005, 547, 188-196.

Saunier, O.; Bocquet, M.; Mathieu, A.; Isnard, O. Model reduction via principal component truncation for the optimal design of atmospheric monitoring networks. Atmospheric Environment, 2009, 43, 4940-4950.

Skoog, D. A.; Holler, F. J.; Crouch, S. R. Principles of Instrumental Analysis, 6 ed. Porto Alegre: Bookman, 2007.

Suarez, J. R. C.; Londoño, L. C. P.; Reyes, M. V.; Diem, M.; Tague, T. J.; Rivera, S. P. H. Fourier Transforms - New Analytical Approaches and FTIR Strategies: Open-Path FTIR Detection of Explosives on Metallic Surfaces. Rijeka: InTech, 2011. 520p.

Tarumi, T.; Small, G. W.; Combs, R. J.; Kroutil, R. T. Infinite impulse response filters for direct analysis of interferogram data from airborne passive Fourier transform infrared spectrometry. Vibrational Spectroscopy, 2005, 37, 39-52.

Teixeira, L. S. G.; Oliveira, F. S.; Santos, H. C. S.; Cordeiro, P. W. L.; Almeida, S. Q. Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline. Fuel, 2008, 87, 346-352.

WAUKESHA. Industrial Engine - Industrial Gas Engine: <http://www.dresserwaukesha.com/index.cfm/go/list-products/productline/CFR-F1-F2-octane-category/>. Accessed in 06 March 2012.

Wu, D. S.; Feng, X.; Wen, Q.Q. The Research of Evaluation for Growth Suitability of Carya Cathayensis Sarg. Based on PCA and AHP. Procedia Engineering, 2011, 15, 1879-1883.

Xu, Q. S.; Liang, Y. Z. Monte Carlo cross validation. Chemometrics and Intelligent Laboratory Systems, 2001, 56, 1-11.

Zhang, G.; Ni, Y.; Churchill, J.; Kokot, S. Authentication of vegetable oils on the basis of their physico-chemical properties with the aid of chemometrics. Talanta, 2006, 70, 293-300.