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

4,100
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the 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
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. Gourvénece 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).

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 (Skog; 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 \( j (j = 1, 2) \).

\[
X = \begin{bmatrix}
1 & 4 \\
2 & 5 \\
3 & 6 \\
\end{bmatrix}
\]

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 \( XX \) to be zero, where \( X^\top \) is the transpose of matrix \( X \).

\[
X^\top X = \begin{bmatrix}
14 & 28 \\
28 & 56 \\
\end{bmatrix}
\]

Then, the \( \det (XX) = (14.56) - (28.28) = 0 \) and this according to Naes et. al (2002) means that there is a singular error matrix and that those errors 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 (Rajalhti; 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:

$$ I = \frac{\text{MON} + \text{RON}}{2} $$

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.

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.

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

Table 1. Summary of results of modeling and validation.

![Figure 3. Normality Test for the property MON](image)
Figure 4. Normality Test for the property RON

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


Contributions of Multivariate Statistics in Oil and Gas Industry


