CHEMICAL ENGINEERING TRANSACTIONS
VOL. 43, 2015
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Chief Editors: Sauro Pierucci, Jiří J. Klemeš

**ISBN** 978-88-95608-34-1; **ISSN** 2283-9216
DOI: 10.3303/CET1543261

# Optimizing the Polynomial to Represent the Extended True Boiling Point Curve from High Vacuum Distillation Data Using Genetic Algorithms

Astrid L. Ceron Rodriguez*, Laura Plazas Tovar, Maria Regina Wolf Maciel, Rubens Maciel Filho

School of Chemical Engineering, University of Campinas, UNICAMP, Zipcode 13083–852, Campinas–SP, Brazil
astridceron@feq.unicamp.br

Molecular distillation process has been used to obtain an extended true boiling point (TBP) curve (above 565 °C up to 700 °C), compared with the results offered by traditional methodologies like ASTM D 2892 and ASTM D 5236. This separation process has the advantage of operating under conditions that reduce the thermal decomposition of the oil. In this paper, polynomials to represent the extended true boiling point curve up to 565 °C from molecular distillation data are proposed. The development is based on molecular distillation experimental results of 14 atmospheric and 5 vacuum oil residues obtained in Pilot and lab scale distiller in other works of the research group at Separation Processes Development Laboratory (UNICAMP). The experimental data were classified in seven different classes based on API density. In first instance, a database was built to perform an extension of the TBP curve of each oil, using the DESTMOL correlation to find the atmospheric temperatures that correspond to the distiller operation temperatures. The results of the three methodologies (ASTM D 2892, ASTM D 5236 and molecular distillation) were adjusted to a 3rd order polynomial as function of the accumulated mass percent. The coefficients were optimized using genetic algorithms. Finally, a variable analysis procedure was developed in order to determine the influence of the genetic algorithm parameters (population size and number of generations) in the obtained response and improve the average absolute deviation percent (%AAD). As a result, a third order fitting polynomial was found for every oil class, presenting %ADD lower than 3%.

## 1. Introduction

The true boiling point curve (TBP) represents a characterization process of petroleum or crude oil, mostly used in refining to determine the sub-products yield and provide information about the operating conditions of oil separation (Argirov et al., 2012). The TBP curve describes the mass (or volume) distilled fraction while increasing temperature. ASTM D 2892 (2010) and ASTM D 5236 (2003) methodologies can be used in order to obtain the TBP curve of any oil up to 565 °C (Behrenbruch and Dedigama, 2007). In order to overcome this limitation, the research group at the Separation Processes Development Laboratory (UNICAMP) developed a procedure that allows oil fraction separation up to 700 °C, by using Molecular Distillation. The outcome of that research was the generation of the DESTMOL correlation Eq(1) (Sbaite et al., 2006), which converts Molecular Distiller operation Temperature ($T_{DM}$) at low pressure (0.001-0.0001 mmHg) (Zuñiga et al., 2009) to equivalent atmospheric temperatures (TAE). This information is then used to calculate an extension of TBP curve.

$$TAE = 456.4 + 0.1677T_{DM} + 1.64 \cdot 10^{-4} T_{DM}{}^{2} + 4.13 \cdot 10^{-6} T_{DM}{}^{3} \quad (1)$$

In this work, characterization results for various oils obtained using ASTM D 2892 (2010), ASTM D 5236 (2003) and Molecular Distillation methodologies were used to generate a correlation that represents an extended TBP curve as a function of cumulative mass percent distilled. The equation parameters were

Please cite this article as: Ceron Rodriguez A., Plazas Tovar L., Wolf Maciel M.R., Maciel Filho R., 2015, Optimizing the polynomial to represent the extended true boiling point curve from high vacuum distillation data using genetic algorithms, Chemical Engineering Transactions, 43, 1561-1566 DOI: 10.3303/CET1543261

estimated and optimized for each oil using a genetic algorithm and a Design of Experiments (DOE) technique, trying to minimize the average absolute deviation. Figure 1 illustrates the implemented methodology.

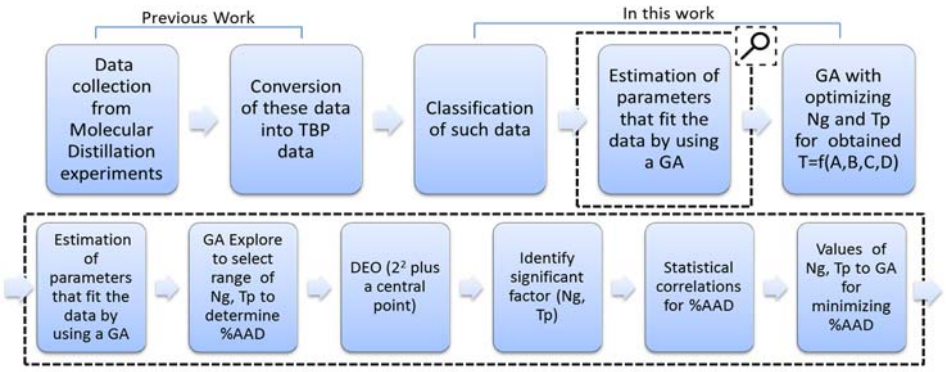


*Figure 1. Block diagram of the implemented methodology.*

## 2. Extended TBP curve parameters optimization

To estimate the correlation parameters of each characterized petroleum, the PIKAIA sub-routine was implemented in Fortran-90 language (Compiler Visual Studio 2008). This is a free access genetic algorithm, developed by High Altitude Observatory (Metalfe and Charboneau, 2003).

### 2.1 Sampling classification

The API density of each crude oil, as well as the TBP curves obtained by standard methodologies (ASTM D 2892 (2010) and ASTM D 5236 (2003)) were used as a criteria to classify the different vacuum and atmospheric residues, in order to establish a single trend in the curve extension, which corresponds to Molecular Distillation. As a result, seven different groups were obtained, as shown in Table 1 and Figure 2. In order to distinguish the atmospheric (400-420 °C+) and vacuum (540-565 °C+) residues, they were named with different coded names composed of one letter and the cut temperature.

*Table 1: Atmospheric and vacuum residues classified according to API gravity of original crude oil.*

| Residue | Petroleum °API | Residue | Petroleum °API | Residue | Petroleum °API | Residue | Petroleum °API |
|---|---|---|---|---|---|---|---|
| A 400 °C+ | | F 420 °C+ | | N 420 °C+ | | | |
| B 400 °C+ | | G 540 °C+ | 19.2 | O 565 °C+ | 24.2 | P 400 °C+ | 25.6 |
| C 400 °C+ | 16.9 | H 400 °C+ | | K 400 °C+ | | Q 420 °C+ | |
| D 400 °C+ | | I 400 °C+ | 20.0 | L 400 °C+ | 22.4 | R 400 °C+ | 33.7 |
| E 565 °C+ | | J 400 °C+ | | M 550 °C+ | | S 550 °C+ | |

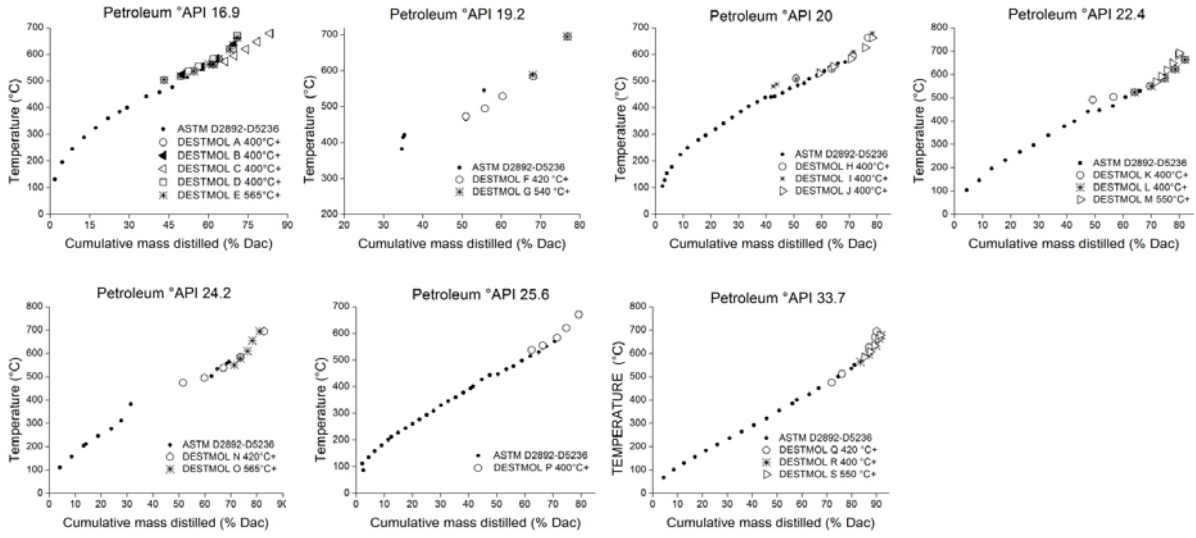


*Figure 2. TBP data obtained by ASTM D 2892 - D 5236 and Molecular Distillation for each group in Table 1.*

### 2.2 Genetic Algorithm application: obtaining the equation parameters.

The genetic algorithms are computational methods used to find optimized solutions emulating a process of natural evolution (Kothari, 2012). These algorithms are very sensitive to variations in their configuration parameters: Population size, number of generations, crossover probability and mutation rate (Azadeh et al. 2010). As mentioned before, in this case the PIKAIA sub-routine was implemented, using the predefined values for mutation rate and crossover probability (0.005 and 0.85) and varying the population size and number of generations, aiming to find the coefficients of polynomial that fits each dataset. The optimized equation was a 3rd order expression in function of cumulative mass percent Eq(2).

$$T = A + B \cdot \%D_{ac} + C \cdot \left(\%D_{ac}\right)^2 + D \cdot \left(\%D_{ac}\right)^3 \quad (2)$$

Where:
*T* is the= Boiling temperature (°C)
% $D_{ac}$ is the = cumulative mass percent distilled

### 2.3 Design of Experiments to estimate the algorithm parameters.

The STATISTIC 7 software from Statsoft Inc. was used to develop a DOE in order to determine how the population size (Tp) and number of generations (Ng) influence on the results, quantified by the average absolute deviation Eq(3).

$$\%AAD = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left[\frac{T_{i,cal} - T_{i,ref}}{T_{i,ref}}\right] \cdot 100 \quad (3)$$

Where:
*n* is the= number of data points
$T_{i,cal}$ is the= Ti calculated (°C)
$T_{i,ref}$ is the= Ti experimental (°C)

Previously, a sensitive analysis was performed, in order to define a valid range for Tp and Ng parameters. As a result, the ranges that exhibited an %AAD less than 10 % were preselected to feed the DOE. Table 2 shows the DOE input ranges. The central composite design ($2^2$ plus a central point) was the experiment design chosen to do the statistical analyses.

*Table 2. Preselected parameter ranges*

| Petroleum °API | Factor | Codified Factor | Level | | | | |
|---|---|---|---|---|---|---|---|
| | | | $-\sqrt{2}$ | -1 | 0 | 1 | $-\sqrt{2}$ |
| 16.9 | Population Size | X1 | 113 | 115 | 120 | 125 | 127 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 19.2 | Population Size | X1 | 72 | 80 | 100 | 120 | 128 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 20,0 | Population Size | X1 | 52 | 63 | 90 | 117 | 128 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 22.4 | Population Size | X1 | 113 | 115 | 120 | 125 | 127 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 24.2 | Population Size | X1 | 113 | 115 | 120 | 125 | 127 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 25.6 | Population Size | X1 | 62 | 70 | 90 | 110 | 118 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |
| 33.7 | Population Size | X1 | 72 | 80 | 100 | 120 | 128 |
| | Number of Generations | X2 | 300 | 358 | 500 | 642 | 700 |

## 3. Results

From the P test results, the appropriate factors and combinations were established for each scenario, taking into account the maximum confidence levels (Table 3). The selected results (in bold, Table 3) satisfied the following conditions: An Effect less than 0.05 at 95 % confidence, less than 0.1 at 90% confidence or less than

0.15 at 85 % confidence. In all cases, the codified factor that corresponds to the size of population (X1) is the parameter that presents a bigger influence, as expected for any genetic algorithm (Roeva et al., 2013).

*Table 3: Tp and Ng effects in %AAD.*

| Petroleum °API | Confidence | Factor | Effect | P | Petroleum °API | Confidence | Factor | Effect | P |
|---|---|---|---|---|---|---|---|---|---|
| 19.2 | 85% | Mean | 2.3648 | 0.00003 | 16.9 | 90% | Mean | 3.2516 | 0.00152 |
| | | X1(L) | 0.2068 | **0.13665** | | | X1(L) | -1.9775 | **0.02570** |
| | | X1(Q) | 0.3149 | 0.16454 | | | X1(Q) | 2.0952 | **0.04546** |
| | | X2(L) | -0.0735 | 0.71969 | | | X2(L) | -0.7160 | 0.27792 |
| 20 | 85% | Mean | 2.9754 | 0.00002 | | | X1(L)*X2(L) | 1.5609 | 0.12523 |
| | | X1(L) | -0.3194 | **0.10986** | 25.6 | 95% | Mean | 3.0909 | 0.00016 |
| | | X1(Q) | 0.5384 | **0.05075** | | | X1(L) | -0.1421 | **0.05146** |
| | | X2(L) | -0.0675 | 0.69855 | | | X1(Q) | -0.6593 | 0.06569 |
| 22.4 | 90% | Mean | 2.1952 | 0.05251 | | | X2(L) | -0.1683 | 0.80559 |
| | | X1(L) | -2.7004 | **0.00701** | 33.7 | 95% | Mean | 2.1105 | 0.00083 |
| | | X1(Q) | 1.7667 | 0.14296 | | | X1(L) | -2.0679 | **0.00032** |
| | | X2(L) | 0.5273 | 0.62638 | | | X1(Q) | 2.3641 | **0.00096** |
| 24.2 | 90% | Mean | 4.0795 | 0.00016 | | | X2(L) | -0.0006 | 0.99958 |
| | | X1(L) | 0.7271 | **0.05146** | | | X2(Q) | -0.0722 | 0.71543 |
| | | X1(Q) | -1.1117 | **0.06569** | | | X1(L)*X2(L) | 0.0028 | 0.98669 |
| | | X2(L) | -0.1229 | 0.80559 | | | | | |

In each case a prediction model was generated, in order to represent mathematically the interaction between the Tp and Ng parameters and the %ADD. These models were verified using a variance analysis (ANOVA), where the valid models are chosen depending on the F-test (i.e., models which the Calculated F is greater than the Critical F) (Table 4). Figure 3 shows the surface responses for the different groups of petroleum.

*Table 4: Valid Statistic Models for %AAD as function of Tp and Ng and F-Test results.*

| Petroleum °API | Prediction Model | %Explained Variation | F Test | |
|---|---|---|---|---|
| | | | Calculated F | Critical F |
| 16.9 | $\%AAD = 697.61 - 10.8T_p + 0.042T_p^2 - 0.1344N_g + 0.0011T_pN_g$ | 86.47 | 6.39 | $F_{4;4;0.010}$=4.10 |
| 19.2 | $\%AAD = 6.3182 - 0.0856T_p + 0.000454T_p^2$ | 64.68 | 3.05 | $F_{3;5;0.015}$=2.79 |
| 20.0 | $\%AAD = 6.6178 - 0.07239T_p + 0.000369T_p^2 - 0.000238N_g$ | 67.70 | 3.49 | $F_{3;5;0.015}$=2.79 |
| 22.4 | $\%AAD = 445.17 - 7.1052T_p + 0.0285T_p^2$ | 82.12 | 7.66 | $F_{3;5;0.010}$=3.62 |
| 24.2 | $\%AAD = -302.72 + 5.039T_p - 0.021T_p^2$ | 74.44 | 4.85 | $F_{3;5;0.005}$=3.62 |
| 25.6 | $\%AAD = -2.97 + 0.1448T_p - 0.00082T_p^2 - 0.00059N_g$ | 85.73 | 10.01 | $F_{3;5;0.005}$=5.40 |
| 33.7 | $\%AAD = 36.41 - 0.6429T_p + 0.00296T_p^2 - 0.00174N_g$ | 99.55 | 132.75 | $F_{5;3;0.005}$=9.07 |

Finally, based on the surface responses, the critical points that correspond to the minimization of the %ADD value were defined. These parameter values were used to run the genetic algorithm again. The %ADD values calculated under these conditions were very close to the prediction model results (Table 5), validating the whole analysis.

The appropriate equation parameters (optimized equations) that describe the extended TBP are listed in Table 5. Profiles of the extended TBP equations are presented in Figure 4, where the different trends presented in each equation proves the different compositions or fraction distributions which depends on the type of crude oil.

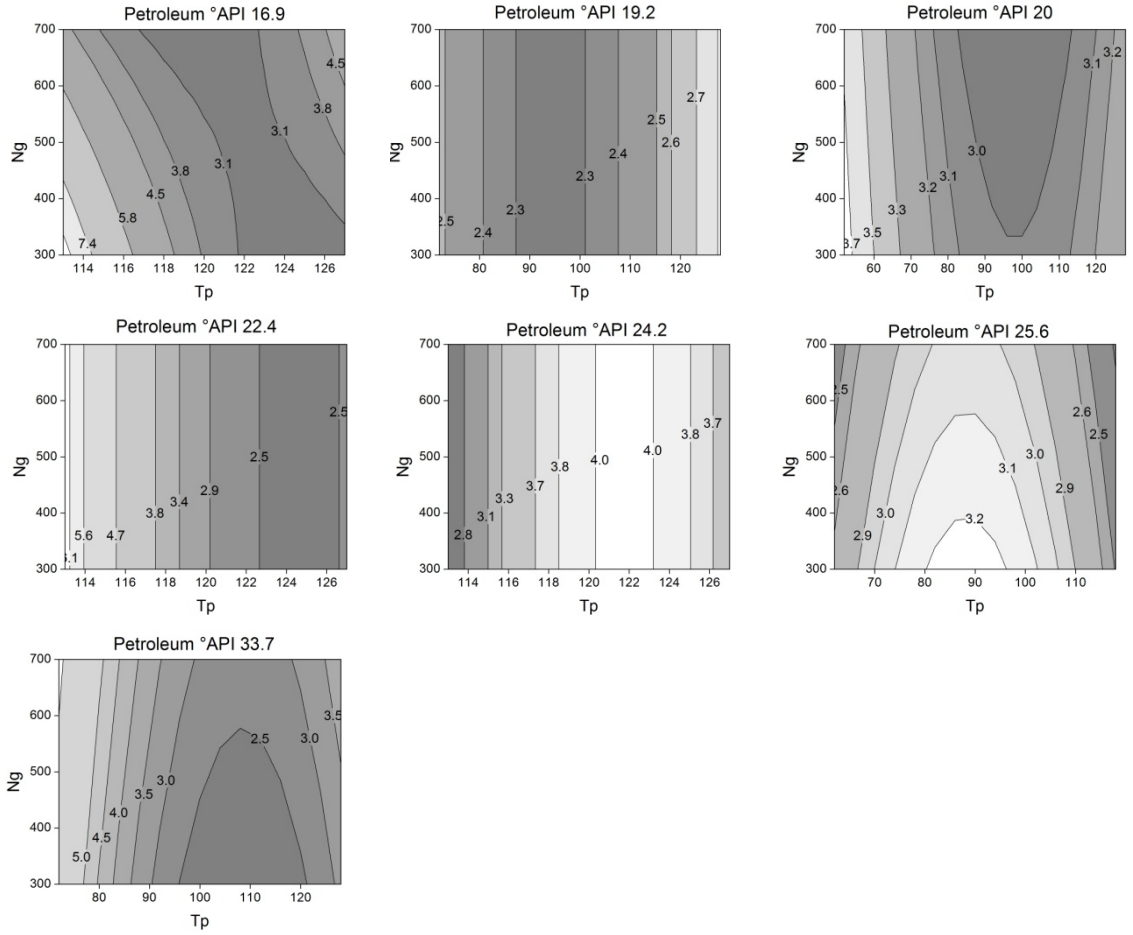


*Figure 3: Surface responses for statistical prediction models as function of Tp and Ng.*

*Table 5: Calculated %AAD: prediction model and genetic algorithm results.*

| Petroleum °API | Parameters | | Optimized Equations | Predicted Value (%ADD) | Algorithm Result (%AAD) |
|---|---|---|---|---|---|
| | Tp | Ng | | | |
| 16.9 | 125 | 449 | $T = 118.01 + 15.99\%D_{ac} - 0.26032\%D_{ac}^{\ 2} + 0.0019003\%D_{ac}$ | 3.1588 | 3.0821 |
| 19.2 | 93 | 512 | $T = 127.23 + 16.99\%D_{ac} - 0.36106\%D_{ac}^{\ 2} + 0.0031000\%D_{ac}$ | 2.2846 | 2.3788 |
| 20.0 | 97 | 562 | $T = 75.605 + 17.02\%D_{ac} - 0.29001\%D_{ac}^{\ 2} + 0.0021517\%D_{ac}$ | 2.9373 | 2.8776 |
| 22.4 | 123 | 490 | $T = 51.260 + 12.40\%D_{ac} - 0.14499\%D_{ac}^{\ 2} + 0.0010379\%D_{ac}$ | 2.1036 | 1.9643 |
| 24.2 | 113 | 500 | $T = 48.035 + 14.172\%D_{ac} - 0.19998\%D_{ac}^{\ 2} + 0.0014737\%D_{ac}$ | 2.4796 | 2.9890 |
| 25.6 | 118 | 500 | $T = 89.999 + 9.9850\%D_{ac} - 0.08999\%D_{ac}^{\ 2} + 0.0006701\%D_{ac}$ | 2.4031 | 2.2501 |
| 33.7 | 110 | 501 | $T = 26.959 + 9.2420\%D_{ac} - 0.09694\%D_{ac}^{\ 2} + 0.0007894\%D_{ac}$ | 2.3112 | 2.3824 |

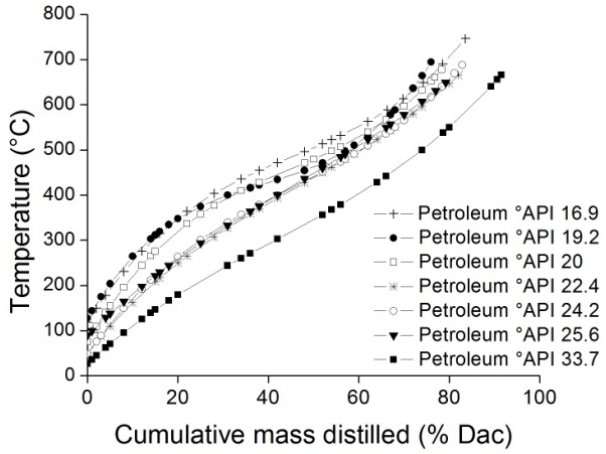


*Figure 3: Calculated Extended True Boiling Point curves by optimized equations.*

## 4. Conclusions

Due to the PEV curves follow a 3rd order trend, previous work developed an algebraic and statistic approaches, producing mathematical expressions statistically consistent. However, the adjustable parameters were not optimized. The scope of this work was the optimization of the parameters aforementioned, using genetic algorithms, exploiting their effectiveness solving search and optimization problems.

The optimal conditions for the genetic algorithm (number of generations (Ng) and population size (Tp) parameters) can be established by implementing a central composite design for each of the oil types, determining the minimum Average Absolute Deviation. As expected, the population size was the algorithm parameter with a greater influence on the genetic algorithm results.

PIKAIA sub-routine was used to optimize the equation parameters that fit the extended TBP curves defined by 3$^{rd}$ order polynomials, finding a single trend for each crude oil with %ADD less than 3%.

## References

Argirov G., Ivanov S., Cholakov G., 2012, Estimation of crude oil TBP from crude viscosity, Fuel, 97, 358-365.

ASTM Standard D 2892, 2010, Standard test method for distillation of crude petroleum, ASTM International, United States.

ASTM Standard D 5236, 2003, Standard test method for distillation of heavy hydrocarbon mixtures (Vacuum pot still method), ASTM International, United States.

Azadeh A., Layegh L., 2010, Optimal model for supply chain system controlled by kanban under JIT philosophy by integration of computer simulation and genetic algorithm, Australian Journal of Basic and Applied Sciences, 4, 370-378.

Behrenbruch P., Dedigama T., 2007, Classification and characterization of crude oils based on distillation properties, Journal of Petroleum Science and Engineering, 57, 166-180.

Celis O.J., Plazas Tovar L., Jardini Munhoz A.L., Siegel C., Maciel Filho R, Wolf Maciel M.R., 2011, Computational approach for studying the laser radiation thermal cracking process of heavy petroleum fraction: optimization of laser operational conditions, Chemical Engineering Transactions, 24, 421-426.

Kothari D., P., 2012, Power System Optimization, CISP Proceedings, 18-21.

Metalfe T., Charbonneau P., 2003, Stellar structure modelling using a parallel genetic algorithm for objective global optimization, Journal of Computational Physics, 185, 176-193.

Nedelchev, A., Stratiev, D., Ivanov, A., Stoilov, G., 2011, Boling Point Distribution of Crude Oils Based on TBP and ASTM D-86 Distillation Data, Petroleum & Coal, 53(4), 275-290.

Roeva O., Fidanova S., Paprzycki M., 2013, Influence of te population size on the genetic algorithm performance in case of cultivation process modeling, Federated Conference on Computer Science and Information Systems, 371-376.

Sbaite P., Batistella C.B., Winter A., Vasconcelos C.J.G, Wolf Maciel M. R., Maciel Filho R., Gomes A., Medina L., Kunert R., 2006, True boiling point extended curve of vacuum residue through molecular distillation, Petroleum Science and Technology, 24, 265-274.

Zuñiga L., Lima N., Wolf Maciel M. R., Maciel Filho R., Batistella C., Manca D., Manenti F., Medina L., 2009, Modeling and simulation of molecular distillation process for a heavy petroleum cut, Chemical Engineering Transactions, 17, 1639-1644.