Orbital: The Electronic Journal of Chemistry
journal homepage: www.orbital.ufms.br
ISSN 1984-6428

| Vol 7 | | No. 2 | | April-June 2015 |

Full Paper

Application Self-organizing Map Type in a Study of the Profile of Gasoline C Commercialized in the Eastern and Northern Parana Regions

Lívia Ramazzoti Chanan Silva, Karina Gomes Angilelli, Hágata Cremasco, Érica Signori Romagnoli, Aline Regina Walkoff, Dionisio Borsato*

State University Of Londrina, Chemistry Department, Fuels Analyses and Research Laboratory, P.O. BOX 10.001, 86.057-970, Londrina, Parana, Brazil.

Article history: Received: 27 April 2015; revised: 07 June 2015; accepted: 26 June 2015. Available online: 29 June 2015. DOI: http://dx.doi.org/10.17807/orbital.v7i2.732

Abstract: Artificial neural networks self-organizing map type (SOM) was used to classify samples of automotive gasoline C marketed in the eastern and northern regions of the state of Paraná, Brazil. The input order of parameters in the network were the values of temperature of the first drop, the 10, 50 and 90% distilled bulk, the final boiling point, density, residue content and alcohol content. A network with a topology of 25x25 and 5000 training epochs was used. The weight maps of input parameters for the trained network identified that the most important parameters for classifying samples were the temperature of the first drop and the temperature of the 10% and 50% of the distilled fuel.

Keywords: gasoline; weight map; topological map; neural network

1. INTRODUCTION

Automotive gasoline consists of a complex fuel composition, constituted mostly of saturated hydrocarbons, olefins and, in smaller amounts, aromatics and mercaptans. Depending on the designated use and oil refining process, hydrocarbons may have 5 to 13 carbon atoms with boiling points ranging from 35 to 215°C, which is appropriate for use in internal combustion engines with spark ignition. The gasoline provided by Brazilian refineries, is a mixture of petroleum distillates from diverse sources, so they may have, depending on the region of origin, different hydrocarbons compositions and therefore have different characteristics of volatility and performance. Under the current legislation, ethyl alcohol can be added to automotive gasoline, within limits, as an antiknock agent [1, 2].

The distillation test of gasoline aims to assess the volatility and performance characteristics as well as identify possible tampering. This test sets temperatures for the 10, 50, 90% distillates and the final boiling point, as well as the maximum amount of waste generated during distillation.

To solve sorting and grouping similar problems, especially when having a very large number of samples described by independent variables, several researchers have been using Artificial Neural Networks (ANN), which is a tool based on the human brain that attempts to reproduce its logical operations [3-6].

Among the different types of ANNs, there is the Self-organizing Map (SOM), of unsupervised learning, that uses the spatial locations on a topological map as indicative of the features contained in its input patterns [7-9]. Thus, samples that share similarities between them form classes or groups called clusters; the longer the distance between the groups, the greater the difference between the samples. SOM have been successfully applied for solve several types of problems of a general nature such as approximation, classification, categorization and prediction [10-12]. In addition, it covers several areas such as food [13], engineering [7, 14, 15] and health [16, 17].

This study aims to apply and adapt the ANN methodology, using the SOM type, for the classification automotive gasoline C samples

*Corresponding author. E-mail: dborsato@uel.br

marketed in the eastern and northern regions of Paraná state, in Brazil.

2. MATERIAL AND METHODS

2.1. Gasoline samples

During the period between January 1st and May 31st 2014, 191 samples of gasoline C were collected; 114 samples were marketed in the northern region and analyzed at the Laboratory of Research and Analyses Fuels at State University of Londrina, and 77 samples commercialized in the eastern region of Paraná state and analyzed at Laboratory Chronion Chemical Analysis and Trade in the city of Quatro Barras. Samples were subjected to distillation tests, alcohol content and specific mass.

2.2. Gasoline distillation

The gasoline distillation test was performed according to ASTM D 86 standard [18], using a Engler flask with 125 ml of capacity, a distiller, a ASTM 7C/IP 5C thermometer with a range of -2 to 300°C and a graduated beaker. 100 mL of sample was transferred to the flask, then the thermometer was attached to a stopper and introduced into the flask, which was installed in the heat source. The beaker was installed at the output of the condensing tube to collect the distillate. The distillation temperature values were recorded for the first drop, distillation of 10, 50 and 90% of the total volume, and the final boiling point. Additionally, the value of the final distillation residue was recorded.

2.3. Alcohol content in gasoline

To determine the alcohol content in gasoline samples, a 100 mL beaker was used, containing 50 ml of the sample to be analyzed; the remaining 50 mL was composed of an aqueous solution of 10% sodium chloride. The beaker was capped and gently agitated, then allowed to stand for a few minutes.

The formation of two phases was observed, and the percentage of alcohol present in the sample was determined from the bottom phase volume by multiplying the volume change by two.

2.4. Density

The determination of gasoline density was performed according to standard ASTM D-1298 [19].

2.5. Artificial neural networks

The ANN module was used in MATLAB R2007 software and the parameters of input order were the temperature of the first drop and 10, 50 and 90% distilled, the final boiling point, the density, residue and alcohol.

2.6. Processing

All results of the experiments were processed using an Intel Core i7-4790 3.60 GHz computer and 32 GB of RAM.

3. RESULTS AND DISCUSSION

The current legislation establishes compliance parameters for regular gasoline C, where in the distillation test the maximum temperature is 65°C for the first 10% distillate and 80°C for the 50% distillate. To confirm the absence of contaminants, the temperature for the 90% of distillate must be neither higher than 190°C nor lower than 145°C. The final boiling point must be at most 215°C, and the residue content cannot exceed 2%. The alcohol content should be 25%. For the density and temperature of the first drop of the distillation test, current legislation does not establish reference values.

Figure 1 shows the temperature values of the parameters obtained in the gasoline distillation test (°C), the residual percentage and the density (kg m-3). The horizontal lines indicate the boundaries of the parameters and the vertical line separates the samples by marketing area. The results show that only 10 samples were in disagreement relative to final boiling point. The amounts of alcohol are not shown because the results were between 24 and 25% in all samples.

To study the profile of gasoline C commercialized in the northern and eastern regions of Paraná, the self-organizing map type of ANN was applied, which transforms a pattern of arbitrary dimension incident signals into a two-dimensional discrete map by accomplishing this transformation in a topologically ordered way [10].

This network enables the recognizing of a pattern in a large amount of gasoline samples with different profiles, produced in different distilleries, with different performance characteristics and also allowing verify possible tampering in an automate way, in a short time interval.

The SOM network in the neural network module of MATLAB R2007 was fed with the following parameters: specific mass, alcohol content, the temperature of the first drop, 10%, 50% and 90% of the distillation test, the final boiling point and the residue of 191 samples. Specifications have not been established for the purposes of training the network.

The learning rate of the trained network started at 0.2 and decreased to 0.0013, and the neighborhood relationship had an initial value of 12 which decreased to 0.054.

The amount of training epochs, which is the number of times the network analyzes the input data, must be selected in a way that the average quantization error is stabilized at the end of the learning step [10].

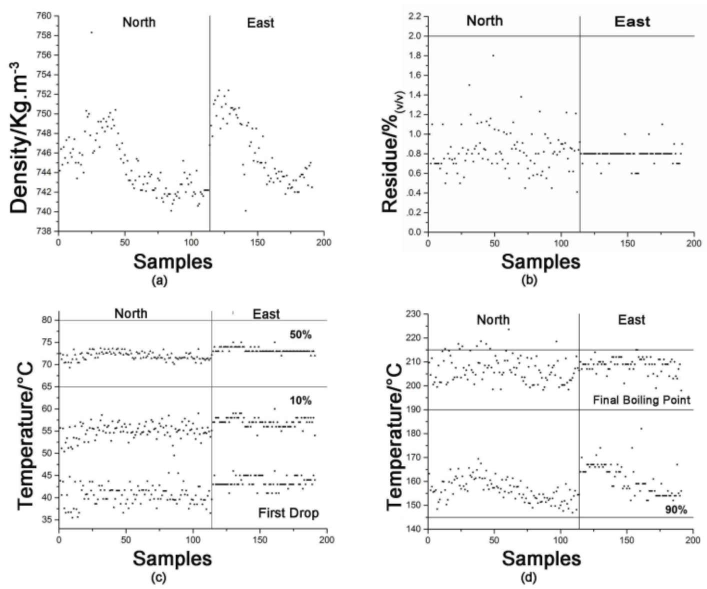

Figure 1. Values obtained for density (a), distillation residue (b), temperature of the first drop, 10 and 50% of the distillate (c), and temperature of 90% of the distillate and the final boiling point (d).

However, the larger the number of epochs, the longer the computational processing time. In a preliminary study, 7000 epochs were used (Figure 2), where it was possible to verify that the stabilization of the error occurred after 5000 epochs, so this was the value used in network training.

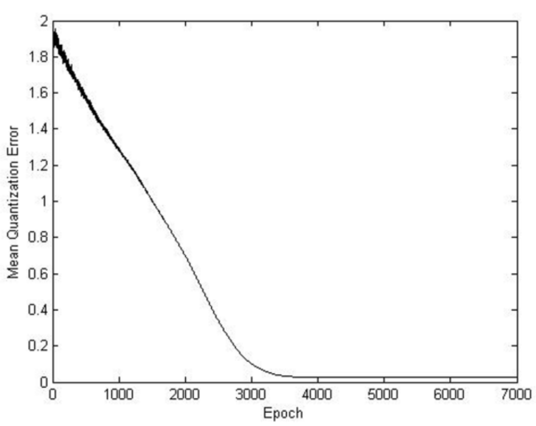

Figure 2. Training error as a function of epochs.

The topology used is important because, if it is too small, the neighborhood relation between neurons is very close and samples end up being classified into one group. If too large, several groups of specialized neurons are formed, and various neighborhood relationships are possible [20].

Topologies were analyzed ranging from 10x10 to 40x40; the one with the best distribution of samples was the 25x25 topology, as depicted in Figure 3.

The formation of a well-defined group of samples from the northern region (N) was located on the left of the map, and two groups in the eastern region (L) were located on the top and bottom right. Between the two eastern groups, the continuation of the northern group was observed. It was found that only three samples of the eastern group were placed in the northern group, corresponding to 98.4% accuracy. For the northern group, accuracy was 100%, demonstrating the ability to discriminate samples by the trained network.

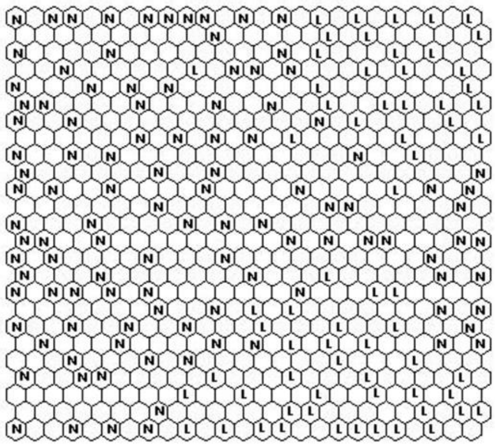

Figure 3. Distribution of samples according to the winner neuron.

The formation of groups in the topological map is justified by the weight maps, obtained from the trained network, which are able to determine what parameters are most important for each classification. In this case, the temperature of the first drop and of the 10 and 50% distillate showed the greatest importance.

Figure 4 shows the weight map of the parameter corresponding to the temperature of the first drop in the distillation test. In this map, most northern samples were located in the blue shade areas corresponding to temperature values ranging from 35 to 40°C, while the majority of eastern samples were in the yellow and red areas, corresponding to temperatures between 42 and 46 °C.

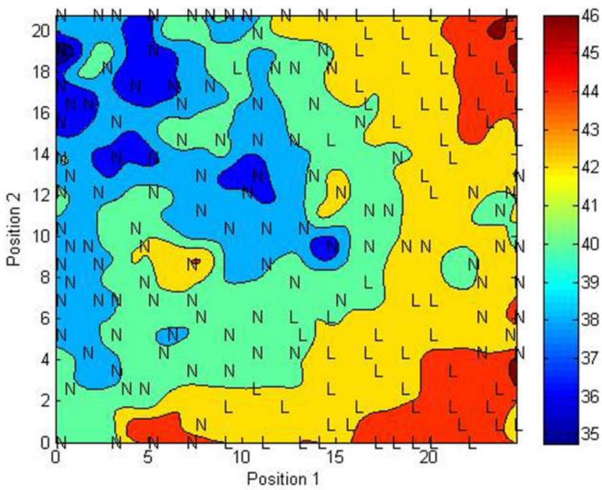

Figure 4. Weight map relative to the temperature of the first drop parameter.

The map for the temperature of the 10% distillate (Figure 5) shows the northern samples arranged in the blue and green region, corresponding to temperature values ranging between 50 and 54 °C, while the eastern samples are arranged in the yellow and red regions, with values between 56 and 60 °C.

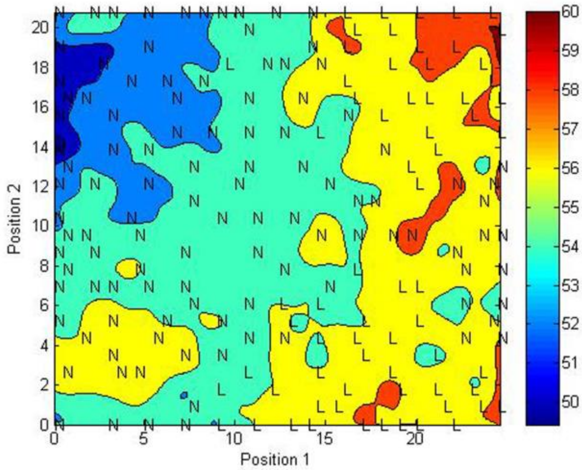

Figure 5. Weight map corresponding to the temperature of the 10% distillate parameter.

Figure 6 presents the weight map regarding the temperature of the 50% distillate, and shows that this parameter was important in discriminating eastern region samples, because they were all arranged in the yellow and red region of the map, which corresponds to temperatures between 73 and 75 °C. Due the fact thatthe samples from the northern region did not show similarity to this parameter, the area where they are located is heterogeneous. Therefore, this information is not enough to group the gasoline from the northern region, although it can differentiate them from eastern samples.

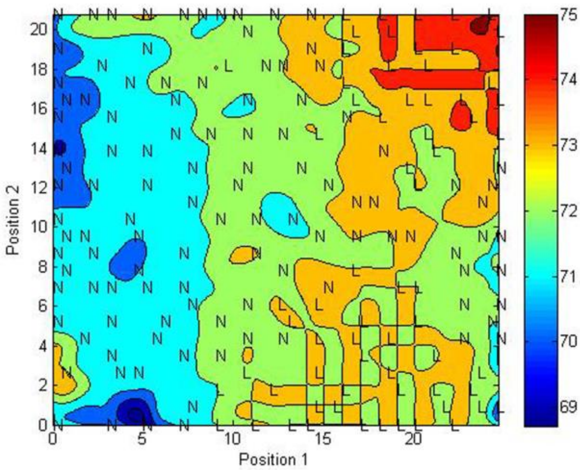

Figure 6. Weight map related to the temperature of the 50% distillate parameter.

Other parameters of the distillation test, such as the 90% distillate, residue and final boiling point, were not important to the segmentation of samples, as can be seen by the weight maps shown in Figure 7.

The temperatures of the 90% distillate (Figure 7a) and the final boiling point (Figure 7b) were not important because, despite having heterogeneous areas on the map, the northern and eastern samples were very similar.

The only information that can be taken from the parameter of the 90% distillate is that it differentiates the eastern samples from each other, justifying the formation of two groups.

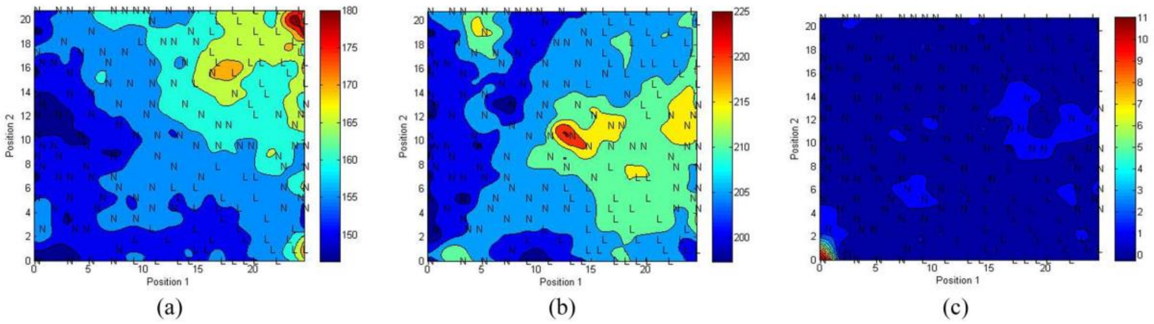

Figure 7. Weight maps relating to the parameters temperature of the 90% distillate (a), final boiling point (b) and percentage of residue (c).

The weight map of the residue (Figure 7c) is homogeneous, and it shows great similarity between the samples from the northern and eastern regions, failing to discriminate them.

Besides the distillation parameters, the parameters density and alcohol content (Figure 8) were also not important for the classification of samples. For the alcohol content, there were no large differences in the values obtained. The map of density, similarly to the temperature of the 90% distillate, only aided in the separation of the eastern samples into two groups.

4. CONCLUSION

The artificial neural network Self-organizing Map type was effective in separating gasoline C samples from different marketing regions.

The 25x25 topology and 5000 training epochs showed lower quantization error and better separation of the samples that allowed for the visualization of the groups in a topographic map.

The most significant parameters for segmentation of the samples were the temperature of the first drop and the temperature of the 10% and 50% distillates, which were mainly responsible for the classification of samples by region.

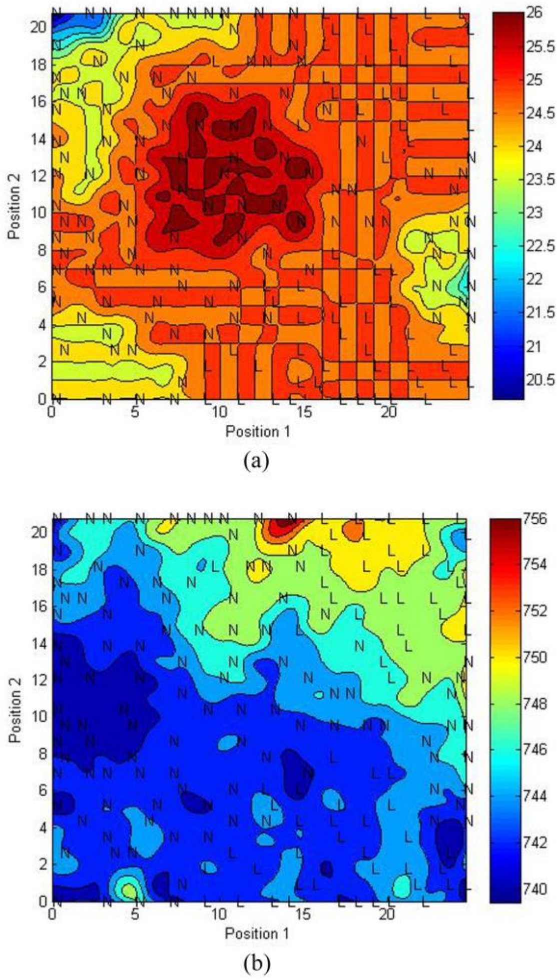

Figure 8. Weight maps relating to the parameters alcohol content (a) and density (b).

5. REFERENCES AND NOTES

[1] Borsato, D.; Galão, O. F., Moreira, I. Combustíveis Fosseis: Carvão e Petróleo. Londrina: Eduel; 2009, pp. 132-136.

[2] Silva, F. L. N.; Santos Jr, J. R.; Moita Neto, J. M.; Da Silva, R. L. G. N. P.; Flumignan, D. L.; Oliveira, J. E. *Quim. Nova*. **2009**, 32, 56. [CrossRef]

[3] Deisingh, A. K.; Stone, D. C.; Thompson, M. *Int. J. Food Sci. Technol.* **2004**, 39, 587. [CrossRef]

[4] Liao, S. *Expert Syst Appl*. **2005**, *1*, 93. [CrossRef]

[5] Borsato, D; Pina, M. V. R.; Spacino, K. R.; Scholz, M. B. S.; Androcioli Filho, A. *Eur. Food Res. Technol*. **2001**, *233*, 533. [CrossRef]

[6] Kovács, Z. L. Redes Neurais Artificiais: Fundamentos e Aplicações. São Paulo: Editora Acadêmica, 1996, pp.163.

[7] Haykin, S. Neural Networks: Principles and Practices. Porto Alegre: Bookman; 2001, pp. 483-500.

[8] Huang, D.; Gentili, R. J.; Reggia, J. A. *Neural Net*works **2015,** *63*, 208. [CrossRef]

[9] Palomo, E. J.; North, J.; Elizondo, D.; Luque, R. M.; Watson, T. *Neural Networks* **2012**, *32*, 275. [CrossRef]

[10] Kohonen, T. Self Organizing Maps. Series in Information Sciences. Heidelberg: Springer-Verlag, 1997.

[11] Nobrega, M. M.; Bona, E.; Yamashita, F. *Materials Sci. Eng. C. Mater. Bio. Apply*. **2013,** *33*, 4331. [CrossRef]

[12] Link, J. V.; Lemes, A. L. G.; Marquetti, I.; Scholz, M. B. S.; Bona, E. *Food. Res. Inter.* **2014**, *59*, 1. [CrossRef]

[13] Debska, B.; Guzowska-Swider, B. *Analytica Chimica Acta*. **2011**, *705*, 283. [CrossRef]

[14] Vuković, N.; Miljković, Z. *Neural Networks* **2015**, *63*, 31. [CrossRef]

[15] Kosic, D. *Neural Networks* **2015**, *63*, 79. [CrossRef]

[16] Read, S. J.; Monroe, B. M.; Brownstein, A. L., Yang, Y.; Chopra, G.; Miller, L. C. *Psychological Review* **2010**, *117*, 61. [CrossRef]

[17] Karelina, K.; Liu, Y.; Alzate-Correa, D.; Wheaton, K. L.; Hoyt, K. R.; Arthur, J. S. C.; Obrietan, K. *Neuroscience* **2015**, *285*, 292. [CrossRef]

[18] ASTM-American Society for Testing and Materials. ASTM D86: Distillation of Petroleum Products at Atmospheric Pressure. 10th. ed. West Conshohocken: ASTM; 2001.

[19] ASTM-American Society for Testing and Materials. ASTM D-1298: Standard Test Method for Density, Relative Density, or API Gravity of Crude Petroleum and Liquid Petroleum Products by Hydrometer Method. 10th. ed. West Con-shohocken: ASTM; 2001.

[20] Bona, E.; Silva, R. S. S. F.; Borsato, D.; Bassoli, D. G. *Acta Sci. Technol*. **2012**, *34*, 11. [CrossRef]