Fuel 156 (2015) 20–25

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel

Ethanol fuel adulteration with methanol assessed by cyclic voltammetry and multivariate calibration

Romário E.L. Abreu, José E.M. Paz, Adenilton C. Silva, Márcio J.C. Pontes, Sherlan G. Lemos *

Departamento de Química, Universidade Federal da Paraíba, CP5093, 58051-970 João Pessoa, PB, Brazil

HIGHLIGHTS

- Simple and fast assessment method of ethanol fuel adulteration.
- Simultaneous multivariate electroanalytical determination of methanol and ethanol.
- Variable selection was crucial for accurate quantification.
- GA-MLR shows the best prediction performance.
- Technology: voltammetry.

ARTICLE INFO

Article history:
Received 14 December 2014
Received in revised form 11 March 2015
Accepted 14 April 2015
Available online 23 April 2015

Keywords:
Biofuels
Chemometrics
Variable selection
Signal processing
Quality control

ABSTRACT

In this work, it is proposed the one-voltammogram simultaneous determination of methanol and ethanol in hydrated ethyl alcohol fuel (HEAF) samples by cyclic voltammetry and multivariate calibration. To perform the determination, different multivariate calibration approaches were tested namely partial least-squares regression (PLS) and multivariate linear regression (MLR). To minimize collinearity problems in MLR, three variable selection algorithms were evaluated: the successive projections algorithm (SPA), the stepwise (SW) formulation, and the genetic algorithm (GA). Variable selection could also provide an improvement in prediction results when compared to full-voltammogram PLS. In this sense, genetic algorithm PLS (GA-PLS), interval PLS (iPLS), and synergy interval PLS (siPLS) were evaluated. An excellent analytical performance was obtained for both alcohols (RMSEP < 0.8% w.w^{-1}) with GA-MLR performed on data transformed by standard normal variate (SNV). The developed method presented a wide linear working range for both alcohols and a LOD = 0.9% w.w^{-1} for methanol sufficient to assess the adulteration of HEAF samples in accordance to Brazilian regulatory agency. The proposed method is simple (with no sample pre-treatment), accurate, rapid, presents low-cost, and can be conducted at distribution sites using a portable potentiostat.

1. Introduction

In Brazil, ethanol is widely employed as an automotive vehicle fuel. It can be found in the anhydrous form (blended with gasoline) and as hydrated ethyl alcohol fuel (HEAF). The production and consumption of this fuel are increasingly larger, making ethanol more susceptible to adulterations. Specifically, a common non-conformity refers to the addition of methanol to HEAF. In Brazil, the acceptable maximum level of methanol in ethanol fuel is 1.0 % w.w^{-1} [1].

There is a predominance of spectroanalytical methods for the determination of methanol in HEAF samples [2–4]. In Brazil, the official method uses gas chromatography [5]. Some papers report the use of electroanalytical methods to monitor the quality of biofuels [6–8]. Bueno and Paixão [9] proposed a method to evaluate the adulteration process of ethanol fuel with water based on capacitance measurements of a copper interdigitated electrode and non-supervised pattern recognition methods. Pereira and coworkers [10] developed an interesting method using cyclic voltammetry to the simultaneous determination of ethanol and methanol in fuel ethanol at a bare gold electrode. In that method, ethanol has been selectively detected at +0.19 V and both compounds at +1.20 V, with methanol being determined by subtracting the currents at those potentials after applying a correction factor to estimate the contribution of ethanol at +1.20 V [10]. However, some drawbacks could be observed such as the narrow linear working ranges for both alcohols (0.00–0.35% w.w^{-1} for ethanol, and 0.00–0.15%

* Corresponding author. Tel./fax: +55 83 32167438.
E-mail address: sglemos@pq.cnpq.br (S.G. Lemos).

http://dx.doi.org/10.1016/j.fuel.2015.04.024
0016-2361/

w.w^{-1} for methanol), and the correction factor that must be determined for each calibration procedure.

Silva and coworkers [11] proposed a screening analysis method to identify HEAF adulteration by methanol based on the voltammetric analysis at copper electrodes aided by supervised pattern recognition techniques, classifying samples as "adulterated" or "unadulterated", with excellent classification rates. Screening analysis is an interesting approach, since it could provide a reduction in the number of samples submitted to official review by the chromatographic method, reducing the cost of official inspection programs. However, in the absence of conditions to perform the official method, it is important to have alternative methods with analytical features that meet the resolution of the analytical problem and, if possible, at a lower cost.

In this sense, the present work proposes the use of the same voltammetric procedure carried out by Silva and coworkers [11] allied to multivariate calibration techniques to develop a one-voltammogram simultaneous determination of methanol and ethanol in HEAF samples. The association of electroanalytical techniques with chemometric tools are shown to be most suitable to evaluate the quality of fuels accurately, rapidly, and at low-cost [9,11,12]. As in [11], this method also employed a copper electrode, which presents electrocatalytic activity toward the anodic oxidation of simple alcohols and small organic molecules in highly alkaline medium [13–15], contrary to gold electrode that requires the addition of methanol to the electrolyte to perform the simultaneous determination of methanol and ethanol [10].

Multivariate calibration models were built based on partial least squares (PLS) regression or multiple linear regression (MLR), which generally requires the use of a tool for selecting non-redundant variables in order to minimize multicollinearity problems [16]. Different variable selection techniques associated to MLR were evaluated: genetic algorithm (GA) [17], stepwise (SW) formulation and successive projections algorithm (SPA) [18].

GA, SW, and SPA were chosen because they are based on very different selection principles. Genetic algorithm mathematically simulates the evolutionary process of a living species [19], and could be divided into three phases. In the first phase, the voltammograms are codified by using a binary code assigned randomly, where each potential represents a gene population. Then, a regression model is calculated for each gene (second phase), estimating the fitness function. Genes that show good fitness are selected for third phase, the reproduction. The operations selection, crossover and mutation of chromosomes are performed in this phase. The second and third phases are repeated until the designated number of generations is achieved. In SW, the algorithm checks the importance of each variable, including or eliminating variables based on a decision rule. At each iteration, a new model is obtained and the effect of the included variable is evaluated by an *F*-test. A variable with an *F*-value larger than a critical *F* is included in the model. On the other hand, SPA aims to select variables that have minimum collinearity among them. Therefore, the algorithm performs three steps: the first step consists of projection operations carried out on the calibration matrix to produce several chains of variables. The second step produces and evaluates MLR models for each chain of variable and selects the best one. Finally, the third step eliminates variables in the selected chain without collinearity with the concentration vector.

In case of PLS regression, the selection of fewer predictor variables previously to modeling is a practice that also could provide significant improvement in prediction results, once variables not related to the response and noise incorporated into regression models could be discarded [20]. Thus, we also intend to evaluate the feasibility of different variable selection techniques associated to PLS in the simultaneous determination of methanol and ethanol. The following variable selection techniques were evaluated: genetic algorithm (GA-PLS) [21]; interval partial least squares (iPLS) [22] and synergy interval partial least squares (siPLS) [23]. In iPLS algorithm, the voltammogram is split into smaller equal-length intervals and a PLS regression model is developed for each interval. In siPLS, a PLS model is developed for all possible combinations of two, three or four intervals [20].

2. Material and methods

2.1. Data set

The data set for the study consists of 120 voltammograms of HEAF samples obtained in a previous work [11]. In summary, 120 HEAF samples were randomly collected from different fuel distributors and ethanol content was assessed with standard chromatographic method [5]. From this set, 60 samples were adulterated with methanol (Merck), from 0.9% up to 17.0% w.w^{-1}. Ethanol concentration varied from 77.0% w.w^{-1} up to 94.5% w.w^{-1}.

Voltammograms were obtained in a conventional three-electrode electrochemical setup using an Ag/AgCl (KCl sat.) electrode and a platinum wire as reference and auxiliary electrodes, respectively. A copper wire (d = 2.0 mm) embedded in a glass rod was the working electrode, which was previously activated in the supporting electrolyte (1.0 mol L^{-1} NaOH) by applying a potential of −0.1 V for 5 s. All voltammograms were recorded from −0.9 V to 0.8 V by applying a potential step of 5 mV and a scan rate of 50 mV s^{-1}, after addition of 100 μL of ethanol fuel in 10 mL of supporting electrolyte at room temperature. Each sample was analyzed in triplicate and the average voltammograms were employed in the calculations.

2.2. Data analysis

The total number of variables in the original voltammograms was initially reduced from 656 to 200 by applying a cubic spline interpolation function. This reduction is necessary once GA performance decreases when >200 input variables are used due to the increasing risk of overfitting and to the size of the search domain, which becomes too great at higher variables/objects ratios [21]. Consequently, dataset comprised 120 samples (voltammograms) with 200 variables (potentials).

Multivariate regression models were obtained from samples divided into calibration (96 samples) and prediction (24 samples) subsets by applying the classic Kennard–Stone (KS) uniform sampling algorithm [24] to the voltammograms. Calibration samples were used in the model-building process. Leave-one-out cross-validation was carried out to select the number of factors (for PLS models) or variables (for MLR). Prediction samples were employed only in the final evaluation by comparing regression models in terms of the lowest root mean squares error of prediction (RMSEP), the relative error percentage (REP), and number of latent or true variables employed. RMSEP is calculated as:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_i^n (\hat{y} - y_i)^2}{n}}, \quad (1)$$

where $\hat{y}$ are the predicted concentration values, y are the real concentration values and n is the number of samples in prediction set. On the other hand, the relative error percentage (REP) denotes an average estimate of the relative error, and is calculated as:

$$\mathrm{REP} = \frac{\mathrm{RMSEP}}{\bar{y}} \times 100,$$

where $\bar{y}$ is the average real concentration value for each analyte. All voltammetric data were mean-centered before modeling procedures. Additionally, models were built on non-transformed and