PAN
POLSKA AKADEMIA NAUK

JOURNAL OF WATER AND LAND DEVELOPMENT

e-ISSN 2083-4535

ITP-PIB

Polish Academy of Sciences (PAN) Institute of Technology and Life Sciences – National Research Institute (ITP – PIB)

JOURNAL OF WATER AND LAND DEVELOPMENT
DOI: 10.24425/jwld.2021.138178
2021, No. 50 (VI–IX): 229-239

# Dynamic modelling of an anaerobic reactor treating coffee wet wastewater via multiple regression model

Yans Guardia-Puebla[1)], Edilberto Llanes-Cedeño[2)], Ana Velia Domínguez-León[3)], Quirino Arias-Cedeño[1)], Víctor Sánchez-Girón[4)], Gert Morscheck[5)], Bettina Eichler-Löbermann[5)]

[1)] University of Granma, Study Center for Applied Chemistry, Cuba

[2)] Faculty of Architecture and Engineering, International SEK University, Quito, Ecuador

[3)] Language Center, University of Granma, Cuba

[4)] College of Agricultural, Food and Biosystems Engineering, Technical University of Madrid, Spain

[5)] Faculty of Agronomy and Crop Science, University of Rostock, Germany

RECEIVED 17.08.2020 REVIEWED 09.10.2020 ACCEPTED 24.05.2021

**Abstract:** A multiple regression model approach was developed to estimate buffering indices, as well as biogas and methane productions in an upflow anaerobic sludge blanket (UASB) reactor treating coffee wet wastewater. Five input variables measured (pH, alkalinity, outlet VFA concentration, and total and soluble *COD* removal) were selected to develop the best models to identify their importance on methanation. Optimal regression models were selected based on four statistical performance criteria, viz. Mallow's $C_p$ statistic ($C_p$), Akaike information criterion (*AIC*), Hannan–Quinn criterion (*HQC*), and Schwarz–Bayesian information criterion (*SBIC*). The performance of the models selected were assessed through several descriptive statistics such as measure of goodness-of-fit test (coefficient of multiple determination, $R^2$; adjusted coefficient of multiple determination, Adj-$R^2$; standard error of estimation, *SEE*; and Durbin–Watson statistic, *DWS*), and statistics on the prediction errors (mean squared error, *MSE*; mean absolute error, *MAE*; mean absolute percentage error, *MAPE*; mean error, *ME* and mean percentage error, *MPE*). The estimated model reveals that buffering indices are strongly influenced by three variables (volatile fatty acids (VFA) concentration, soluble *COD* removal, and alkalinity); while, pH, VFA concentration and total *COD* removal were the most significant independent variables in biogas and methane production. The developed equation models obtained in this study, could be a powerful tool to predict the functionality and stability for the UASB system.

**Keywords:** coffee wet wastewater, modelling, multiple regression model, upflow anaerobic sludge blanket (UASB)

## INTRODUCTION

Few crops receive as much attention as coffee in relation to the environment. The fact of growing in tropical and subtropical areas; together with being a North-South product, from the consumption and production point of view; being associated with occasions of frequent consumption; as well as being a drink related to social interaction, make coffee a product that generates interest and attention. Since high standards of environmental sustainability are continually demanded for its production, coffee processing is one of the activities that needs to adapt its production technologies to reduce environmental impact [Santos *et al.* 2009].

Only 20% of the coffee fruit is usable and the remaining 80% is waste [Houbron, Rustrian 2003]. Two types of processing exist: dry and wet. In Central America, the most widely used method of processing is wet, despite requiring the consumption of large amounts of water [Guardia-Puebla *et al.* 2013]. Coffee industry is considered one of the most polluting, with serious negative environmental impacts. This activity generates a considerable increase in organic pollution (2.4–21.9 kg $COD \cdot dm^{-3}$), and in suspended matter (1.0–10.0 kg $TSS \cdot dm^{-3}$); as well as generation of unpleasant odors, coloration and loss of visual quality, if the wastes are not treated properly [Guardia-Puebla *et al.* 2014a, b; 2016; Jung *et al.* 2012]. In addition, coffee wastewater has low pH values (<4); therefore, a severe water pollution occurs during harvest seasons, which affects the availability of water for human, industrial and recreational use [Selvamurugan *et al.* 2010].

To increase the stability and to improve the performance of the anaerobic reactors, it is necessary that the microorganisms be retained within the reactor. A technology capable of achieving these two issues is the upflow anaerobic sludge blanket (UASB) reactor. The success of the UASB system is based on the formation of an anaerobic granular sludge at the bottom of the reactor which determines the speed of the start-up stage and the efficiency of the anaerobic treatment [Chong *et al.* 2012].

The operation of the anaerobic digestion (AD) process is complex and highly dependent on the configurations of the reactors. In addition, these latter can vary significantly according to the different characteristics of the influent and operational conditions. For that reason, the system must be constantly monitored and controlled due to possible incidental instable conditions. Particularly, biogas or methane production rates, and buffering indices, can provide an indication of the overall anaerobic biomass activity in the process [Turkdogan-Aydinol, Yetilmezsoy 2010]. Since the AD process is very vulnerable to fluctuations in the input characteristics of the influent (rates of organic and hydraulic load, pH and presence of toxic organic compounds), biogas and methane productions, and buffering indices, largely depend on the conditions applied to the reactor. Therefore, the complicated interrelations that exist between the different factors involved in the AD process can be explained with statistical prediction models in which only those key variables are considered.

Regression analysis is a statistical process to estimate the relationships between variables and is widely used for prediction and forecasting [Montgomery 2013]. It includes many techniques for the modelling and analysis of various variables, when the focus is on the relationship between a dependent variable and one or more independent (or predictor) variables. In all cases, the objective is to estimate a function of the independent variables called the regression function.

Many techniques have been developed to carry out regression analysis (linear and nonlinear approaches). Therefore, several estimation models have been developed to describe biogas or methane production from UASB reactors treating organic wastes. For example, Barampouti *et al.* [2005] performed a dynamic mathematical model for the prediction of biogas production from a potato wastewater treatment plant in an UASB reactor. The technique used included regression analysis by residuals. For the model construction, the authors used seventeen parameters including the following: wastewater flow rate, reactor temperature, pH, total and soluble influent chemical oxygen demand (*COD*), volatile fatty acids (VFA), and alkalinity. Yetilmezsoy and Sapci-Zengin [2009] used a three-layer artificial neural network (ANN) model to predict *COD* removal efficiency of UASB reactors treating real cotton textile wastewater diluted with domestic wastewater. In this study, nine input parameters, such as hydraulic retention time (*HRT*), pH, *COD* influent concentration, operating temperature, alkalinity, VFA concentration, dilution ratio, organic loading rate (*OLR*), and total suspended solids (*TSS*) concentration, were used as dependent variables. Singh *et al.* [2010] assessed the performance of an UASB reactor of a wastewater treatment plant with linear and nonlinear models. In their research, partial least squares regression, multivariate polynomial regression and artificial neural networks modelling methods were applied to predict the levels of biochemical oxygen demand (*BOD*) and *COD* in the effluent, while using four input variables (*BOD*, *COD*, ammoniacal nitrogen ($NH_4$–N) and total Kjeldahl nitrogen (TKN)) measured weekly in the influent (untreated) and effluent (treated) wastewater. Turkdogan-Aydinol and Yetilmezsoy [2010] used a multiple inputs and multiple outputs fuzzy-logic-based model to predict biogas and methane production rates in a pilot scale 90 L mesophilic UASB reactor treating molasses wastewater. Five input variables such as *OLR*, total *COD* removal rate, influent alkalinity, influent pH and effluent pH were fuzzified by the use of an artificial intelligence-based approach. Yetilmezsoy [2012] used an integrated multi-objective optimization approach, within the framework of nonlinear regression based on kinetic modelling and desirability function, to optimize an UASB reactor treating poultry manure wastewater. The author developed a regression analysis based on an estimation model for biogas generated using several independent parameters, such as pH, electrical conductivity, total dissolved solids, chemical oxygen demand, alkalinity, chloride, total Kjeldahl nitrogen, ammonia, and total phosphorus. In order to develop the best model, taking into account the highest estimation performance, eight model equations including different input parameter combinations were analysed. Ramesh *et al.* [2015] developed a multiple linear model in which *COD* removal was the dependent variable, and different parameters, such as *HRT*, *OLR*, sludge loading rate, influent pH, effluent pH, inlet and outlet VFA concentration, inlet and outlet volatile suspended solids and total solids (*VSS*/*TS)* ratio, and influent and effluent *COD*, were considered as independent variables. The results of the step-wise regression method applied revealed that only four parameters (influent *COD*, effluent *COD*, volatile solids and total solids (*VS*/*TS)* ratio and influent pH) were significant on *COD* removal. Finally, Antwi *et al.* [2017] worked with artificial neural networks and multiple nonlinear regression models to estimate biogas and methane yield in an UASB reactor processing potato starch wastewater. In this research, the coefficient of multiple determination ($R^2$) of the artificial neural networks reached 98.72% and 97.93%, while the one of the multiple nonlinear regression models attained values of 93.9% and 91.08%, for both biogas and methane yield, respectively.

In most of the already mentioned research work, the selection of the best prediction model was done following some statistical performance criterion: $R^2$, adjusted coefficient of multiple determination (Adj-$R^2$), residual average (*RA*), sum of squared residuals (*SSR*), standard error of the estimate (), and *p*-value. However, to ensure that the best regression model be obtained, other criteria must also be taken into account.

Consequently, different model evaluation criteria, like Akaike information criterion (*AIC*), Hannan–Quinn criterion (*HQC*), and Schwarz-Bayesian information criterion (*SBIC*), are increasingly being used to address model selection problems. However, very little is understood about the relative efficiency of these information theoretic criteria when modelling UASB systems. Another important example is the Mallow $C_p$ statistic, which is one of the most used methods to compare all possible regressions and select the best parameter estimate. A researcher can use Mallow $C_p$ statistic to obtain a measure of bias in a reduced model. Such parameter can be extremely useful for evaluating the robustness of the reduced model obtained by the different stepwise regression procedures.

In this work, an attempt to model the performance of an UASB system for the treatment of wet coffee wastewater has been made, using different multiple regression models in which the values measured of different reactor variables have been included.

Considering the above mentioned facts, the specific objectives of this study were: (1) to develop a rapid and efficient methodology able to define operational parameters that influence the performance of an UASB reactor; (2) to identify essential process variables capable of making reliable predictions by means of various descriptive statistics; and (3) to verify the validity of the multiple regression model by several additional testing data sets to make reliable simulations and predictions. Five independent variables (pH, alkalinity, outlet FVA concentration, and total and soluble *COD* removal) obtained from the UASB process treating coffee wet wastewater were selected based on multiple linear regression analysis approach. The anaerobic process parameters were identified to optimize the performance of the UASB system.

## MATERIAL AND METHODS

### SUMMARY OF PREVIOUS STUDIES

The experimental methodology, which includes obtaining the wet coffee wastewater and the seed sludge, as well as its characteristics, the configuration and the start-up stage of the UASB reactor, the chemical reagents used, the description of the equipment, the chemical analyses performed, the gas collection system and other technical details of operation, have been documented in previous papers [Guardia-Puebla *et al.* 2013; 2014a, b].

### ALKALINITY INDICES MEASUREMENT

Obtaining alkalinity indices (alpha index (*AI*), buffer index (*BI*) and *BI-AI* ratio) was based on the determination of alkalinity due to VFA compounds (V2), the alkalinity of bicarbonates (V1) and total alkalinity (V1 + V2). A 25 cm$^3$ sample was taken and titrated with 0.02 N $H_2SO_4$ to a pH value of 5.75. The volume of acid consumed was considered as V1. Then, the titration continued until a pH value of 4.3 was obtained. This other volume of acid consumed was considered as V2. Total alkalinity was determined as the sum of V1 and V2.

The *AI* index was considered as the relationship between bicarbonate alkalinity and total alkalinity according to Equation (1).

$$AI = V_1/(V_1 + V_2) \tag{1}$$

The *BI* index expressed the relationship between the alkalinity of VFA compounds and total alkalinity (Eq. 2).

$$BI = V_2/(V_1 + V_2) \tag{2}$$

Likewise, the *BI-AI* ratio was considered as the relation between VFA compounds alkalinity and bicarbonate alkalinity (Eq. 3).

$$BI - AI \text{ ratio} = V_2/V_1 \tag{3}$$

### RESEARCH METHODOLOGY

**Regression model selection.** A multiple regression approach was used to fit a linear model for each of the dependent variables: biogas production, methane production and alkalinity indices (*AI*, *BI*, and *BI-AI*) based on the independent variables: pH, alkalinity, VFA concentration, and total and soluble COD removal. One of the assumptions of the classical regression analysis is that the model used has to be correctly specified. To determine the quality of the prediction model it is necessary to take into account some general guidelines: i) moderation or simplicity; ii) identifiability; iii) goodness of fit; iv) theoretical consistency; and v) predictive power. The specification error was assessed assuming that one or more of the following mistakes were not committed: i) omit a relevant variable; ii) include an unnecessary variable; iii) adopt a wrong functional form; iv) incorrect specification of the stochastic disturbance term; and v) measurement errors [Ramesh *et al.* 2015]. The consequences of including irrelevant variables in a model are, fortunately, not serious. However, when a legitimate variable of the model is omitted the consequences are very serious: the coefficients of the variables are inconsistent and violate the usual hypothesis testing procedures. Therefore, the selection of the optimal regression model was based on information about the goodness of the adjustments provided by four statistical parameters.

The mean squared error (*MSE*) characterizes the estimate of the variance of the deviations from the fitted model (Eq. 4), given by:

$$MSE_{\text{model}} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p - 1} \tag{4}$$

where: is the observed value; is the predicted value by the fitted model, *n* is the number of observations and *p* is the number of independent variables included in the model.

The Mallow's $C_p$ statistic was calculated according to Equation (5):

$$C_p = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{MSE(\text{full})} - n + 2p \tag{5}$$

where: *MSE*(full) is the mean squared error of the model when all independent variables are included on it.

If a fitted model has little bias, $C_p$ should be close to *p*-value. It is desirable to have a small $C_p$ as long as the value is not much greater than *p*.

The *AIC* (Eq. 6) was calculated from:

$$AIC = 2\ln(RMSE) + \frac{2p}{n} \tag{6}$$

where: *RMSE* is the root mean squared error during the estimation period, *p* is the number of estimated coefficients in the fitted model, and *n* is the sample size used to fit the model.

Notice that *AIC* is a function of the variance of the model residuals, penalized by the number of estimated parameters. In general, the model that minimizes the mean square error without using too many coefficients in relation to the amount of data available will be selected.

The Hannan–Quinn criterion (*HQC*) (Eq. 7) was calculated from:

$$HQC = 2\ln(RMSE) + \frac{2p\ln(\ln(n))}{n} \tag{7}$$

This criterion uses a different penalty for the number of estimated parameters. The Schwarz-Bayesian information criterion (*SBIC*) (Eq. 8) was calculated from:

$$SBIC = 2\ln(RMSE) + \frac{p\ln(n)}{n} \tag{8}$$

Again, the penalty for the number of estimated parameters is different from that of the other criteria.

**Multiple regression approach.** The general form of the multiple regression approach used in this study was developed from measurements recorded at equally spaced time intervals. The dependent variables (*AI*, *BI*, *BI-AI* index, biogas production and methane production) were denoted by *y*, the input variables by $x_1$, $x_2$, …, $x_k$ (pH, alkalinity, outlet FVA concentration, and total and soluble *COD* removal), and a random error term was added (Eq. 9). Coefficients $\beta_0$, $\beta_1$, $\beta_k$, which were usually unknown, were subsequently estimated by the regression analysis.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \hat{\varepsilon} \tag{9}$$

where: $x_1$, $x_2$, and $x_k$ are the represented terms for the quantitative predictors, and *k* is the number of independent regressors excluding the constant term.

For the purpose of modelling, several assumptions were considered: linearity of the models, constant variance and homoscedasticity, non-autocorrelation, explicative variables are stochastic, non-multicollinearity, normal distribution of the errors, and specification bias does not exist.

**Statistics for the fitted model and residual analysis.** To evaluate the performance of the model and the goodness of the fit, several descriptive statistical parameters were selected: $R^2$ (Eq. 10), Adj-$R^2$ (Eq. 11), standard error of estimation (Eq. 12), and Durbin–Watson statistic (*DWS*) (Eq. 13).

$$R^2 = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{10}$$

$$\text{Adj-}R^2 = \frac{\left(\frac{n-1}{n-p-1}\right)\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2}$$

$$SEE = \sqrt{MSE} \tag{12}$$

$$DWS = \frac{\sum_{i=1}^{n-1}\left(e_{i+1} - e_i\right)^2}{\sum_{i=1}^{n} e^2} \tag{13}$$

where: $y_i$ denotes the observed value, is the arithmetic mean of the observed data, and $e_i = y_i - \hat{y}_i$ is the residual of the response variable for an individual *i*.

On the other hand, the goodness of the adjustments obtained with the prediction models were evaluated with four statistical criteria that were applied to the prediction errors (Eq. 14). These criteria are defined in the Eqs. (15)–(19): mean squared error (*MSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), mean error (*ME*), and mean percentage error (*MPE*).

$$e_i = y_i - \hat{y}_i \tag{14}$$

$$MSE = \frac{\sum_{i=1}^{n} {e_i}^2}{n-1} \tag{15}$$

$$MAE = \frac{\sum_{i=1}^{n}|e_i|}{n} \tag{16}$$

$$MAPE = \frac{\frac{100\sum_{i=1}^{n}|e_i|}{y_i}}{n} \tag{17}$$

$$ME = \frac{\sum_{i=1}^{n} e_i}{n} \tag{18}$$

$$MPE = \frac{\frac{100\sum_{i=1}^{n} e_i}{y_i}}{n} \tag{19}$$

## RESULTS AND DISCUSSION

### UASB PROCESS

The UASB was operated for a period of about seventy days at different hydraulic retention time (*HRT* is 21.5 h, 18.5 h, and 15.5 h) after the start-up stage of the anaerobic system was completed. The operation began with an *HRT* of 21.5 h and *OLR* of 3.6 kg *COD*·m$^{-3}$·d$^{-1}$; subsequently, *HRT*s were step wisely shortened to 18,5 h and 15.5 h with increases *OLR*s of 3.8 kg *COD*·m$^{-3}$·d$^{-1}$ and 4.1 kg COD·m$^{-3}$·d$^{-1}$, respectively. Despite the different organic and hydraulic loading conditions, the UASB system successfully treated the coffee wet wastewater [Guardia-Puebla *et al.* 2014b].

Table 1 shows the summary of the descriptive statistics of all the variables taken in the study, which were classified into independent variables and dependent variables. The mean and standard deviation for each variable was calculated for 45 observations. Total and soluble *COD* removal remained in the range of 53.5–81.2% and 62.5–85.6% and standard deviations of 8.0% and 6.5%, and coefficient of variation between 8.35 and 11.52%, respectively; alkalinity had a mean of 1685.87 mg CaCO$_3$·dm$^{-3}$ and standard deviation of 305.72 mg CaCO$_3$·dm$^{-3}$; the pH was in the range of 6.53–8.38; and the total VFA concentration was 222.28 ± 16.64 mg·dm$^{-3}$. Meanwhile, the alkalinity indices showed the lowest coefficients of variation (between 3.5 and 6.68%), and the biogas production and methane concentration had values of 0.254 ± 0.012 dm$^3$·d$^{-3}$ and 51.3 ± 7.27% and coefficients of variation of 4.84% and 14.18%, respectively.

### REGRESSION MODELS SELECTION FOR THE EXPERIMENTAL DATA

Mallow's $C_p$ statistic is a powerful technique for model selection in regression. Mallow's $C_p$ is an objective measure of the degree of bias in a reduced model and it is extremely useful in measuring the level of bias of the parameter estimates, $\beta_k$. Essentially, researchers should select the reduced model with the highest Adj-$R^2$, Lowest *MSE*, and lowest Mallow's $C_p$ value. These selection criteria are highly appropriate when researchers are interested in data description and parameters estimation [Zuccaro 1992].

The *AIC* handle a trade-off between the goodness-of-fit and the complexity of the model, i.e. it provides a relative appraisal of

**Table 1.** Descriptive statistics

| Variables | Average | Standard deviation | Coefficient of variation (%) | Minimum | Maximum | Range |
|---|---|---|---|---|---|---|
| **Independent variables** | | | | | | |
| pH | 7.55 | 0.63 | 8.35 | 6.53 | 8.38 | 1.85 |
| Alkalinity | 1685.87 | 305.72 | 18.13 | 1204.0 | 2208.0 | 1004.0 |
| VFA concentration | 222.28 | 16.64 | 7.49 | 185.08 | 260.28 | 75.20 |
| Total *COD* removal | 69.52 | 8.0 | 11.52 | 53.5 | 81.2 | 27.7 |
| Soluble *COD* removal | 77.31 | 6.5 | 8.35 | 62.5 | 85.6 | 23.14 |
| **Dependent variables** | | | | | | |
| *AI* | 0.49 | 0.017 | 3.5 | 0.45 | 0.53 | 0.08 |
| *BI* | 0.50 | 0.020 | 4.08 | 0.47 | 0.54 | 0.07 |
| *BI-AI* ratio | 1.00 | 0.067 | 6.68 | 0.89 | 1.13 | 0.24 |
| Biogas production | 0.254 | 0.012 | 4.84 | 0.237 | 0.285 | 0.048 |
| Methane concentration | 51.33 | 7.27 | 14.18 | 40.0 | 61.0 | 21.0 |

Source: own study.

the information loss when a certain model is used to estimate data. However, the main disadvantage of this method is that it does not provide a numerical value that allows determining the quality of the model. The model that best fits a series of data is the one that results with the lowest *AIC* value. This parameter, therefore, not only provides information on the goodness of the fit, but also avoids an over-adjustment when choosing the model that minimizes the loss of information [Wang, Liu 2006].

Similarly, the *SBIC* also enables the selection of models between a finite set of models, due to that probability function is very narrowly related with the information-theoretic criteria *AIC*. Both *SBIC* and *AIC* methods introduce a penalization term for the number of parameters in the model; thereby, the lowest value implies a minor number of explanatory variables, a better adjustment, or both. Nevertheless, the penalization term of the *SBIC* is greater than the *AIC* method. Essentially, the two penalized criteria are based on two different model selection approaches: *AIC* is aimed to find the best adjustment model to the data, meanwhile *SBIC* is designed to identify the true model [Acquah 2010].

The *HQC* is another criterion for model selection and is an alternative to the other *AIC* and *SBIC* criteria. The method allows obtaining the measure of goodness of fit from a statistical model. In addition, it is a model selection criterion among a finite set of models. When the numerical values of the dependent variable are identical, in order to compare all the estimates, it is used to compare the estimated models [Shittu, Asemota 2009].

In this study of regression, twenty-two mathematical models were solved and automatically sorted according to the four information-theoretic criteria considered. Nevertheless, with representative motives only the more promissory eight models will be showed. Table 2 shows the comparison among the eight models selected according to $C_p$, *AIC*, *HQC*, and *SBIC*. In general, the models selected have a single structure; the number of variables is between 2 and 3 that guarantees the simplicity of the relations. The prediction models for buffering indices were defined as a function of four operating independent variables (pH, alkalinity, VFA concentration, and soluble *COD* removal); while, the models for the prediction of biogas production and methane yield were established as a function of three process variables (pH, VFA concentration, and total *COD* removal). For the selection of the best prediction models the lowest values of $C_p$ were considered; on the contrary, the highest values of *AIC*, *HQC* and *SBIC* were considered as the best criteria.

## GOODNESS-OF-FIT TEST AND RESIDUAL ANALYSIS

The goodness-of-fit of a statistical model describes the way that a data set is adjusted. In general, the measures of goodness-of-fit summarize the discrepancy between the observed values and the expected values in a model. The best adjustment models obtained according to the dependent variables studied, defined as a function of the independent variables, are shown in Table 3. Also, the performance criteria parameters ($R^2$, Adj-$R^2$, and *DWS*) calculated, obtained from that models, are summarized in Table 3. The performances of the buffering indices are given as follow: $AI = f(\text{VFA}_{\text{conc}}, SCOD_{rem})$, $BI = f(\text{pH}, SCOD_{\text{rem}})$, and *BI-AI* ratio $= f(\text{alk}, \text{VFA}_{\text{conc}}, SCOD_{\text{rem}})$; meanwhile, both parameters that characterize the quality of gas are a function of two variables: methane prod. $= f(\text{VFA}_{\text{conc}}, TCOD_{\text{rem}})$, and biogas prod. $= f(\text{pH}, TCOD_{\text{rem}})$. High values of $R^2$ and Adj-$R^2$ were obtained by all models (interval of variation for $R^2$ and Adj-$R^2$ and were between 0.854–0.928 and 0.848–0.925, respectively), except for biogas production where the values were more modest (0.658 and 0.651, respectively).

Commonly, the goodness-of-fit is measured through the $R^2$, which shows the variation proportion of the dependent variables explained by the independent variables. Nevertheless, a better practice is to use Adj-$R^2$ due to, sometimes, $R^2$ would be able to provide a too much optimist result of the regression. For both parameters, values will be between 0 and 1; closer to 1, the adjustment will be better.

High degrees of precision and a good deal of the reliability of the models were indicated by low values of *SEE* (Tab. 3). Standard error of estimation indicates the standard deviation of *y* values regarding to the estimated regression line, which is

**Table 2.** Comparison of eight mathematical models according to the selection of the statistical performance criterion considered

| Depended variables | Statistical criterion | Independent variables included in the model | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ACD | ACE | AE | BCE | C | CD | CE | E |
| *AI* index | $C_p$ | 56.72 | 4.25 | 128.50 | 4.20 | 518.24 | 56.39 | 2.30* | 183.53 |
| | *AIC* | −9.65 | −10.48 | −9.06 | −10.48 | −8.01 | −9.71 | −10.55* | −8.92 |
| | *HQC* | −9.59 | −10.42 | −9.01 | −10.42 | −7.98 | −9.67 | −10.50* | −8.89 |
| | *SBIC* | −9.49 | −10.32 | −8.94 | −10.32 | −7.93 | −9.59 | −10.43* | −8.84 |
| *BI* index | $C_p$ | 4.75 | 2.05* | 2.33 | 3.12 | 47.95 | 8.45 | 7.02 | 5.12 |
| | *AIC* | −8.68 | −8.75 | −8.76* | −8.72 | −8.06 | −8.62 | −8.65 | −8.72 |
| | *HQC* | −8.62 | −8.69 | −8.71* | −8.66 | −8.03 | −8.57 | −8.61 | −8.69 |
| | *SBIC* | −8.52 | −8.59 | −8.64* | −8.56 | −7.98 | −8.50 | −8.53 | −8.64 |
| *BI-AI* ratio | $C_p$ | 6.73 | 2.35 | 4.19 | 2.05* | 51.98 | 8.67 | 5.19 | 4.32 |
| | *AIC* | −6.23 | −6.33 | −6.31 | −6.34* | −5.61 | −6.21 | −6.29 | −6.33 |
| | *HQC* | −6.17 | −6.27 | −6.26 | −6.28 | −5.58 | −6.16 | −6.24 | −6.30* |
| | *SBIC* | −6.07 | −6.17 | −6.19 | −6.18 | −5.53 | −6.09 | −6.17 | −6.25* |
| Biogas production | *Cp* | 3.93 | 5.46 | 47.02 | 5.30 | 4.68 | 2.01* | 3.46 | 56.71 |
| | *AIC* | −9.66 | −9.62 | −8.99 | −9.63 | −9.69 | −9.73* | −9.69 | −8.93 |
| | *HQC* | −9.60 | −9.56 | −8.94 | −9.57 | −9.66 | −9.68* | −9.65 | −8.90 |
| | *SBIC* | −9.50 | −9.46 | −8.87 | −9.47 | −9.61* | −9.61 | −9.57 | −8.85 |
| Methane production | *Cp* | 2.15* | 17.02 | 29.18 | 19.48 | 186.92 | 7.02 | 28.62 | 32.00 |
| | *AIC* | −10.12* | −9.79 | −9.63 | −9.75 | −8.49 | −10.02 | −9.64 | −9.63 |
| | *HQC* | −10.06* | −9.73 | −9.58 | −9.69 | −8.46 | −9.98 | −9.59 | −9.60 |
| | *SBIC* | −9.95* | −9.63 | −9.51 | −9.59 | −8.41 | −9.90 | −9.52 | −9.55 |

Explanations: A = pH, B is alkalinity, C = VFA concentration, D = total *COD* removal, and E = soluble *COD* removal; the asterisk indicates the best value selected for each statistical criterion.
Source: own study.

**Table 3.** Summary of multiple regression results for the best-fit models

| Model | Descriptive statistics | | | | |
|---|---|---|---|---|---|
| | $R^2$ | Adj-$R^2$ | *SEE* | *DWS* | *p*-value |
| $AI$ = 0.000895VFA$_{\text{conc}}$ + 0.003821$SCOD_{\text{rem}}$ | 0.928 | 0.925 | 0.004 | 2.099 | 0.0000 |
| $BI$ = 0.037804pH + 0.002709$SCOD_{\text{rem}}$ | 0.920 | 0.863 | 0.057 | 2.314 | 0.0000 |
| $BI$-$AI$ ratio = 0.0000705alk + 0.001148VFA$_{\text{conc}}$+ 0.008103$SCOD_{\text{rem}}$ | 0.881 | 0.843 | 0.038 | 2.511 | 0.0000 |
| Biogas prod. = 0.000907VFA$_{\text{conc}}$+ 0.000758$TCOD_{\text{rem}}$ | 0.658 | 0.651 | 0.007 | 2.116 | 0.0000 |
| Methane prod. = 0.001512pH + 0.001709$TCOD_{\text{rem}}$ | 0.854 | 0.848 | 0.006 | 2.074 | 0.0000 |

Explanations: $R^2$ = coefficient of multiple determination, Adj-$R^2$ = adjusted coefficient of multiple determination, *SEE* = standard error of estimation, *DWS* = Durbin–Watson statistic, and *p*-value <0.05 was considered as significant.
Source: own elaboration.

frequently used as a measure that summarizes the goodness-of-fit of a model.

The best-known test for detecting serial correlation is the *DWS*. Currently, it is common to include reports of the *DWS* in the regression analysis, together with the $R^2$, Adj-$R^2$, *t*-statistic, among others. As a rule, there is no first order correlation when the *DWS* value is close to a value 2, either positive or negative. However, if the value closes to 0, the presence of a perfect positive correlation in the residuals is indicated. On the contrary, being a negative correlation is evidenced when a value close to 4 is obtained.

In all models, the *DWS* was found in the non-autocorrelation region among the residues; so, multiple linear regression models are appropriate. On the other hand, statistical models with the lowest *p*-value were considered significant. Probability values below 0.05 specify that the model is significant at 95% of probability. Based on the *p*-values given in Table 3, all mathematical models were properly selected according to the statistical performance criteria considered. The regression

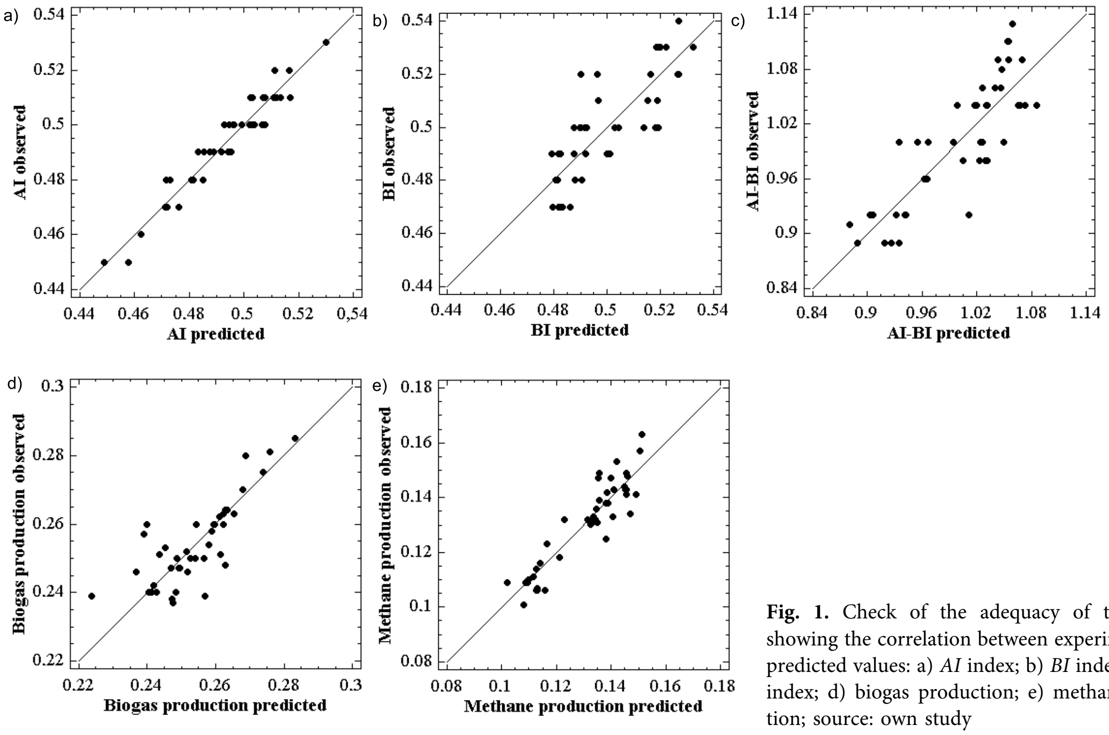

**Fig. 1.** Check of the adequacy of the models showing the correlation between experimental and predicted values: a) *AI* index; b) *BI* index; c) *BI-AI* index; d) biogas production; e) methane production; source: own study

analysis showed a small deviation in the prediction of the models obtained. The check of the adequacy of the models showed a good correlation between the observed and predicted values, shown in Figure 1. The cluster point around the diagonal line indicates a good fit of models.

It is a common practice to use the residues to check the assumptions of the model, since the residues will have a normal distribution with zero mean and constant variance if the assumptions are met. To evaluate the model performances, descriptive statistic and residual analysis are given in Tables 4 and 5, respectively. When the fit is better, the residuals will be smaller and, consequently, statistics on the prediction errors will be small.

Good models are those that meet more adjustment quality criteria. However, in circumstances where a criterion is not met, the model obtained will not necessarily be unfeasible from a practical point of view. If the normality in the residuals is not satisfied, that assumption will not be decisive, that is, the methodology is more or less robust to the lack of normality. Another aspect to consider is that under conditions of similar quality when adjusting models, the simplest model should always be preferred.

## BUFFERING INDEX MODELLING

In this research, pH, alkalinity, VFA concentration, and *SCOD* removal values were selected as variable inputs to model the buffering indices performances (*AI*, *BI*, and *BI-AI* ratio). The correlation for buffering indices between testing outputs and the experimental data is depicted in Figure 2.

**Table 4.** Summary of descriptive statistic

| Statistic | Calculation | Results | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *AI* | | *BI* | | *BI-AI* ratio | | biogas production | | methane production | |
| | | observed | predicted | observed | predicted | observed | predicted | observed | predicted | observed | predicted |
| $s^2$ | $\sum_{i=1}^{n}(x_i-\bar{x})^2/_{n-1}$ | 0.0003 | 0.0003 | 0.0004 | 0.0005 | 0.0039 | 0.0070 | 0.0002 | 0.0001 | 0.0002 | 0.0002 |
| $SD$ | $\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/_{n-1}}$ | 0.0175 | 0.0174 | 0.0192 | 0.0212 | 0.0621 | 0.0834 | 0.0123 | 0.0118 | 0.0157 | 0.0158 |
| $CV$ | $^{s}/_{\bar{x}}100$ | 3.54% | 3.52% | 3.83% | 4.24% | 6.24% | 8.40% | 4.85% | 4.66% | 12.09% | 12.12% |
| $s_{\bar{x}}$ | $^{s}/\sqrt{n}$ | 0.0026 | 0.0026 | 0.0029 | 0.0032 | 0.0093 | 0.0124 | 0.0018 | 0.0018 | 0.0023 | 0.0024 |

Explanations: $s^2$ = variance, *SD* = standard deviation, *CV* = coefficient of variation,  = standard error.
Source: own study.

**Table 5.** Residual analysis

| Variable | *MSE* | *MAE* | *MAPE* | *ME* | *MPE* |
|---|---|---|---|---|---|
| *AI* | $0.2 \cdot 10^{-4}$ | 0.0039 | 0.7988 | $0.1 \cdot 10^{-4}$ | –0.0044 |
| *BI* | 0.0032 | 0.0501 | 10.0 | 0.0042 | 0.4306 |
| *BI-AI* ratio | 0.0014 | 0.0313 | 3.1172 | $0.2 \cdot 10^{-4}$ | –0.1262 |
| Biogas production | $0.5 \cdot 10^{-4}$ | 0.0050 | 1.9853 | $0.8 \cdot 10^{-4}$ | –0.0193 |
| Methane production | 6.9598 | 2.0777 | 4.1861 | –0.1349 | –0.8104 |

Explanations: *MSE* = mean squared error, *MAE* = mean absolute error, MAPE = mean absolute percentage error, ME = mean error, and MPE = mean percentage error.
Source: own elaboration.

The respective close ranges of *AI*, *BI*, and *BI-AI* ratio were between 0.45 and 0.53, 0.47 and 0.54, and between 0.89 and 1.13 for observed values; and between 0.45 and 0.53, 0.44 and between 0.55 and 0.80 for predicted values, respectively, which suggest a good-fit of the selected models to the dataset. Moreover, the similar average values closer to the unity obtained indicate a satisfactory prediction for the UASB system: observed values of 0.49, 0.5, 1.0, and predicted values of 0.49, 0.50, 0.99 for *AI*, *BI*, *BI-AI* index values, respectively.

Figure 3 shows the visual agreements between experimental data and predicted values of buffering index. The highest AI values were observed at the initial reactor operation associated with the start-up stage of the system. Nevertheless, when the stability conditions were reached, the values were in the range of 0.45–0.51, although these values were obtained by evaluating the maximum *OLR*.

In addition, the highest BI-AI ratio values were obtained in the initial stage of operation; however, when stability in the system was reached, the values decreased. Although literature recommends that BI-AI index be less than 0.3, that value is associated with the treatment of domestic wastewater [Lahav, Morgan 2004]. Pérez and Torres [2008] reached an adequate range of BI-AI ratio between 0.44 and 0.55 treating wastewater from the cassava starch process in an anaerobic filter. The authors concluded that *COD* removal efficiency and biogas production were closely related to the variation of the buffering indices.

## PREDICTION OF BIOGAS AND METHANE PRODUCTION

The values measured and predicted by the biogas production and methane yield models obtained were plotted in Figure 4. A close pattern of variation among the measured and predicted values for both variables is evident suggesting a good predictive capability of the selected models. Biogas increased along with the increasing indicating a positive correlation. The variation range of biogas production observed was between 0.23–0.28 $dm^3 \cdot d^{-1}$, while a variation range of 0.22–0.28 $dm^3 \cdot d^{-1}$ was obtained by the prediction model. Furthermore, biogas production was in the range of 0.10–0.16 for both observed and predicted values. Methane fractions were maintained in the range of 45–58%.

## PREDICTION OF BIOGAS AND METHANE PRODUCTION

Barampouti *et al.* [2005] correlated biogas production using the multiple regression technique and residue analysis. The authors determined three mathematical models by correlating biogas production with several independent variables. The most strongly correlated variables were wastewater flow rate, total influent COD concentration, and soluble influent and effluent COD concentration. However, although the three models had similar capacities to estimate the biogas production rate, the ability to predict and control the values of the dependent variable was different. Yetilmezsoy and Sakar [2008] obtained a prediction model to quantify the biogas production rate. Experimental results obtained from three different operating phases were performed through nonlinear regression analysis. Nonlinear modelling study showed that *HRT* and influent *COD* concentration were found to be main operational variables, which directly affect biogas production rate and *COD* removal efficiency. On the other hand, Turkdogan-Aydinol and Yetilmezsoy [2010] defined two models that predict the performance of biogas and methane production rates as a function of five process variables. The best-fit-models for biogas production or methane production were based on five different model components (*OLR*, volumetric *TCOD* removal rate, alkalinity, inlet and outlet pH). The non-linear regression variable results showed that volumetric *TCOD* removal rate and effluent pH had more importance than other model components in prediction of both biogas and methane production rates. A multiple regression model for the estimation of biogas production from landfill leachate treatment system using leachate characteristics was developed by Akkaya *et al.* [2015]. In that research, the model obtained was based on six different independent variables (*COD*, conductivity, alkalinity, pH, total phosphorus, and total Kjeldahl nitrogen). The developed multiple regression model shows sufficient prediction performance, for that reason the authors concluded that the developed equation

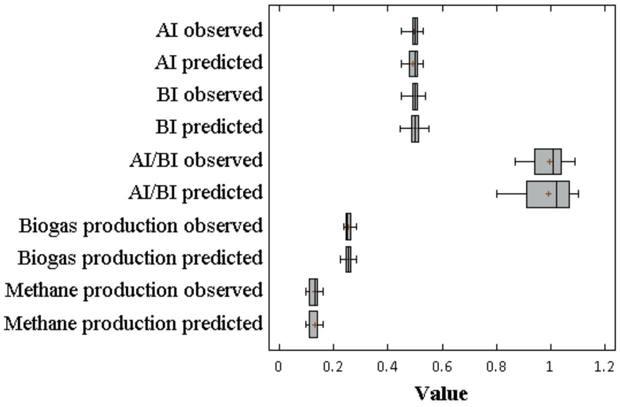

**Fig. 2.** Box-and-whiskers plot; source: own study

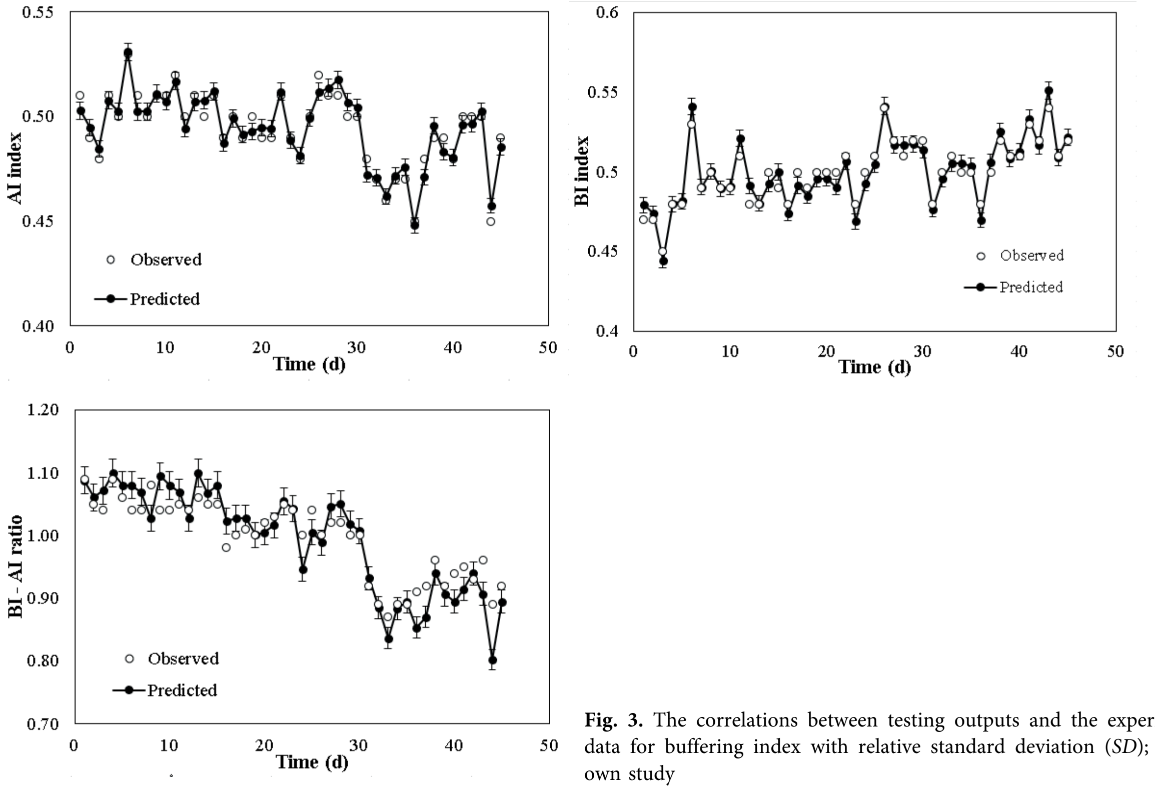

**Fig. 3.** The correlations between testing outputs and the experimental data for buffering index with relative standard deviation (*SD*); source: own study

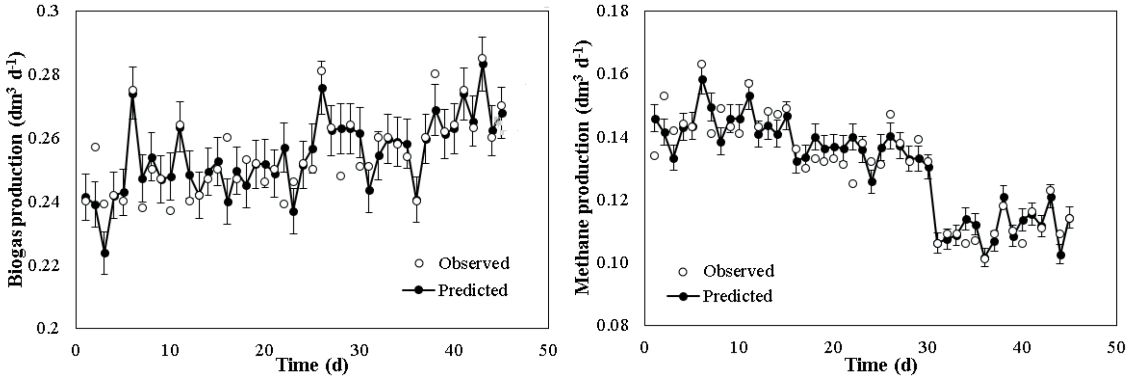

**Fig. 4.** A head to head comparison of performances for observed and predicted values of biogas production and methane production with relative standard deviation (*SD*); source: own study

model is a good biogas production predictor tool. Also, ANTWI *et al.* [2017] estimated the biogas production and methane yield from an UASB reactor with the multiple nonlinear regression approach. Statistical analysis of the regression input variables revealed that *COD*, VFA concentration, and *HRT* variables were the most significant ones in the prediction of biogas production and methane yield.

The small deviations (between 0.48% and 6.0%) in the validation of the five models achieved indicated the suitability of the proposed integrated approach and suggested that this methodology could be successfully adapted to the design and operation of a mesophilic UASB reactor treating coffee wet wastewater. The correlation between the measured and the model predicted values of the dependent variable is considered an important parameter to indicate the predictive ability of the model. In terms of prediction, all regression models in this study reach high levels of predictive capacity. The amount of variance explained exceeds 84% and the maximum expected error rate is 12%, except for biogas production. In context, the results support the validation of the models and provide levels of safety in the regression models as the basis for developing operation and control strategies for anaerobic reactors.

## CONCLUSIONS

The behaviour of the buffering indices, and the biogas and methane productions generated in an UASB reactor treating coffee wet wastewater were evaluated and modelled. Data obtained were used to predict the behaviour of the system without using biomodelation mechanisms, which involve a great degree of complexity and insecurity. Both total and soluble *COD* removal efficiencies observed were higher than 75% and 80%, respectively, under various organic and hydraulic loading conditions. Alkalinity indices showed a small variation range between 0.45 – 0.54, and the biogas production and methane production have values of 0.254 ± 0.012 $dm^3 \cdot d^{-1}$ and 51.3 ± 7.27%, respectively. The multiple regression model approach modelling methodology for the construction of dynamic models proved to be very satisfactory. The selection of the best prediction model requires the development of novel approaches in order to improve the reliability of the reactor performance. The optimal regression model selection was based on the selection of four statistical performance criteria: Mallow's $C_p$ statistic, *AIC*, *HQC*, and *SBIC*. Proper selection of regression models requires the use of low $C_p$ values and higher *AIC*, *HQC*, and *SBIC* values. The predictive abilities of the models obtained were evaluated through various goodness-of-fit tests and residual analysis. In terms of prediction, all the regression models achieve high levels of predictive accuracy, with the interval of variation for $R^2$ and Adj-$R^2$ of 0.854 – 0.928, and 0.848 – 0.925, respectively, except for biogas production which had more modest values (0.658 y 0.651, respectively). In terms of explanation, the estimated model reveals that buffering indices are strongly influenced by three variables, viz. Volatile fatty acids (VFA) concentration, soluble chemical oxygen demand (*COD*) removal, and alkalinity. Meanwhile, VFA concentration and total *COD* removal were the most significant independent variables in biogas production. On the other hand, regression variable results showed that pH and total *COD* removal were found to be more important for methane production. Choosing the most appropriate model representing the extension of the experimental data can help to recognize possible technical faults and to reduce operating costs of plants in the planning stage. As a main result from this work, the developed equation models obtained are a good predictor tool for the upflow anaerobic sludge blanket (UASB) reactor treating coffee wet wastewater.

## ACKNOWLEDGEMENT

This study was partly supported by the Federal Ministry of Education and Research (BMBF) project InFertRes (CUB17WTZ-042).

## REFERENCES

Acquah H. 2010. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship [online]. Journal of Development and Agricultural Economics. Vol. 2. Iss. 1 p. 1–6. [Access 15.07.2020]. Available at: http://www.academicjournals.org/app/webroot/article/article1379662949_Acquah.pdf

Akkaya E., Ahmet D., Gamze V. 2015. Estimation of biogas generation from a UASB reactor via multiple regression model. International Journal of Green Energy. Vol. 12 p. 185–189. DOI 10.1080/15435075.2011.651754.

Antwi P., Jianzheng L., Portia O.B., Jia M., En S., Kaiwen D., Francis K.B. 2017. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with backpropagation artificial neural network. Bioresource Technology. Vol. 228 p. 106–115. DOI 10.1016/j.biortech.2016.12.045.

Barampouti E.M., Mai S.T., Vlyssides A.G. 2005. Dynamic modeling of biogas production in an UASB reactor for potato processing wastewater treatment. Chemical Engineering Journal. Vol. 106. Iss. 1 p. 53–58. DOI 10.1016/j.cej.2004.06.010.

Chong S., Sen T.K., Kayaalp A., Ang H.M. 2012. The Performance Enhancements of UASB reactors for domestic sludge treatment – A state-of-the-art review. Water Research. Vol. 46 p. 3434–3470. DOI 10.1016/j.watres.2012.03.066.

Guardia-Puebla Y., Jiménez-Hernández J., Pacheco-Gamboa R., Rodríguez-Pérez S., Sánchez-Girón V. 2016. Multiple responses optimization on the anaerobic co-digestion of coffee wastewater with manures [online]. Ciencias Técnicas Agropecuarias. Vol. 25. Iss. 3 p. 54–64. [Access 15.07.2020]. Available at: http://scielo.sld.cu/pdf/rcta/v25n3/rcta06316.pdf

Guardia-Puebla Y., Rodríguez-Pérez S., Cuscó-Varona Y., Jiménez-Hernández J., Sánchez-Girón V. 2014a. Two-phase anaerobic digestion of coffee wet wastewater: Effect of recycle on anaerobic process performance [online]. Ciencias Técnicas Agropecuarias. Vol. 23. Iss. 1 p. 25–31. [Access 15.07.2020]. Available at: http://scielo.sld.cu/pdf/rcta/v23n1/rcta04114.pdf

Guardia-Puebla Y., Rodríguez-Pérez S., Jiménez-Hernández J., Sánchez-Girón V. 2013. Performance of a UASB reactor treating coffee wet wastewater [online]. Ciencias Técnicas Agropecuarias. Vol. 22. Iss. 3 p. 35–41. [Access 15.07.2020]. Available at: http://scielo.sld.cu/pdf/rcta/v23n2/rcta09214.pdf

Guardia-Puebla Y., Rodríguez-Pérez S., Jiménez-Hernández J., Sánchez-Girón V., Morgan-Sagastume J., Noyola A. 2014b. Experimental design technique is useful tool to compare anaerobic systems. Renewable Bioresources. Vol. 2. Iss. 3 p. 1–12. DOI 10.7243/2052-6237-2-3.

Houbron E., Larrinaga A., Rustrian E. 2003. Liquefaction and methanization of solid and liquid coffee wastes by two phase anaerobic digestion process [online]. Water Science & Technology. Vol. 48. Iss. 6 p. 255–262. [Access 15.07.2020]. Available at: https://pubmed.ncbi.nlm.nih.gov/14640226

Jung K.-W., Kim D.-H., Lee M.-Y., Shin H.-S. 2012. Two-stage UASB reactor converting coffee drink manufacturing wastewater to hydrogen and methane. International Journal of Hydrogen Energy. Vol. 37 p. 7473–7481. DOI 10.1016/j.ijhydene.2012.01.150.

Lahav O., Morgan B. 2004. Titration methodologies for monitoring of anaerobic digestion in developing countries – A review. Journal of Chemical Technology and Biotechnology. Vol. 79 p. 1331–1341. DOI 10.1002/jctb.1143.

Montgomery D. 2013. Design and analysis of experiments. 8th ed. Hoboken. John Wiley & Sons, Inc. ISBN 978-1118-14692-7 pp. 757.

Pérez A, Torres P. 2008. Indices de alcalinidad para el control del tratamiento anaerobio de aguas residuales fácilmente acidificables [Alkalinity indices for control of anaerobic treatment of readily acidifiable wastewaters]. Ingeniería y Competitividad. Vol. 10. Iss. 2 p. 41–52. DOI 10.25100/iyc.v10i2.2473.

Ramesh N., Vennila G., Abdul B.J., Ramesh S., Magesh K.P. 2015. Energy production through organic fraction of municipal solid waste a multiple regression modeling approach. Ecotoxicology and Environmental Safety. Vol. 134 p. 350–357. DOI 10.1016/j.ecoenv.2015.08.027.

Santos Dos J.S, Santos Dos M.L., Conti M.M., Santos Dos S.N., Oliveira De E. 2009. Evaluation of some metals in Brazilian coffees cultivated during the process of conversion from conventional to organic agriculture. Food Chemistry. Vol. 115 p. 1405–1410. DOI 10.1016/j.foodchem.2009.01.069.

Selvamurugan M., Doraisamy P., Maheswari M., Nandakumar N.B. 2010. High rate anaerobic treatment of coffee processing wastewater using upflow anaerobic hybrid reactor [online]. Iran Journal of Environmental Health and Science Engineering. Vol. 7.

Iss. 2 p. 129–136. [Access 10.07.2020]. Available at: https://www.sid.ir/FileServer/JE/102620100204.pdf

Shittu O., Asemota M. 2009. Comparison of criteria for estimating the order of autoregressive process: A Monte Carlo approach [online]. European Journal of Scientific Research. Vol. 30. Iss. 3 p. 409–416. [Access 10.07.2020]. Available at: http://www.eurojournals.com/ejsr.htm

Singh K.P., Basant N., Malik A., Jain G. 2010. Modeling the performance of "up flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches – A case study. Analytica Chimica Acta. Vol. 658 p. 1–11. DOI 10.1016/j.aca.2009.11.001.

Turkdogan-Aydinol F.I., Yetilmezsoy K. 2010. A fuzzy-logic-based model to predict biogas and methane production rates in a pilot-scale mesophilic UASB reactor treating molasses wastewater. Journal of Hazardous Materials. Vol. 182 p. 460–471. DOI 10.1016/j.jhazmat.2010.06.054.

Wang Y., Liu Q. 2006. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock – Recruitment relationships. Fisheries Research. Vol. 77 p. 220–225. DOI 10.1016/j.fishres.2005.08.011.

Yetilmezsoy K. 2012. Integration of kinetic modeling and desirability function approach for multi-objective optimization of UASB reactor treating poultry manure wastewater. Bioresource Technology. Vol. 118 p. 89–101. DOI 10.1016/j.biortech.2012.05.088.

Yetilmezsoy K., Sakar S. 2008. Development of empirical models for performance evaluation of UASB reactors treating poultry manure wastewater under different operational conditions. Journal of Hazardous Materials. Vol. 153 p. 532–543. DOI 10.1016/j.jhazmat.2007.08.087.

Yetilmezsoy K., Sapci-Zengin Z. 2009. Stochastic modeling applications for the prediction of COD removal efficiency of UASB reactors treating diluted real cotton textile wastewater. Stochastic Environmental Research and Risk Assessment. Vol. 23 p. 13–26. DOI 10.1007/s00477-007-0191-5.

Zuccaro C. 1992. Mallow's Cp statistic and model selection in multiple linear regression. International Journal of Market Research. Vol. 34. Iss. 2 p. 1–13. DOI 10.1177/147078539203400204.