Original Research

Parametric and Nonparametric Approaches for Detecting the most Important Factors in Biogas Production

Matjaž Ošlaj[1], Tadeja Kraner Šumenjak[2], Miran Lakota[2], Peter Vindiš[2]*

[1]Panvita d.d., Rakičan, Murska Sobota, Slovenia
[2]Faculty of Agriculture and Life Sciences, Hoče, Slovenia

Abstract

The aim of this paper is to compare results obtained via the well-known regression method ordinary least squares (OLS) and the alternative regression method called multiple model regression estimation (MM-estimation). This is motivated by the fact that exceptional crop yield observations (outliers and leverage points) can cause misleading results if least squares regression is applied. The paper demonstrates that in this case, robust regression is a more appropriate approach, with higher adjusted R-squared value. With both methods, several models have been proposed for predicting the production of biogas where various explanatory variables have been considered, such as the parameters of Weende analysis, C/N ratio, pH value, and the value of volatile fatty acids. Anaerobic digestion was carried out with a basic substrate of pig slurry and with different combinations of co-substrates, where co-substrate maize (main crop), maize (stubble crop), triticale (main crop), sorghum (main crop), a mixture of plants for biomass production (main crop), and grain maize (grain at the wax ripeness stage) were used. To optimize the anaerobic process of fermentation of substrate with co-substrate, the experimental reactor of the Nemščak biogas plant was applied. The average yield of biogas ranged from 384 Nl/kg VS to 635 Nl/kg. The resulting models revealed that crude protein (XP), starch (XS), nitrogen-free extracts (NFE), C/N ratio, volatile fatty acids (VFA), and pH value were the most important predictors affecting biogas production from different substrates. These models are helpful tools in optimising and predicting biogas production from energy crops.

Introduction

One of the greatest problems of the 21st century is increasing energy consumption and decreasing supplies of fossil fuels. The deficiency of fossil fuels has led to research on the use of renewable energy sources and thus the development of new technological processes for the production of energy [1]. Fossil fuels are limited resources, concentrated in a few geographical areas of earth. Therefore, most European countries are strongly dependent on fossil fuels imported from regions rich with them. The development of renewable energy sources such

*e-mail: peter.vindis@um.si

as biogas, obtained from national or regional biomass, increases the security of the energy supply and reduces the dependency on imported fuels [2]. Renewable energy sources play an important role in increasing the share of utilized domestic supply of energy and reducing environmental pollution, and they are also an important measure in reducing greenhouse gas emissions. The relatively new technology of energy production from biogas as a renewable energy carrier can make many valuable contributions, including reductions in emissions, in water pollution, and in soil degradation. Biogas can also be used as a vehicle fuel and has several advantages when examined from an environmental and resource-efficiency perspective when compared to other biomass-based vehicle fuels available so far. This provides the motivation for further technological development aimed at reducing costs and thereby increasing the economic competitiveness of biogas. Biogas can be converted into bio-compressed natural gas (bio-CNG) as an alternative vehicular fuel to regular compressed natural gas (CNG). Bio-CNG could play an important role in minimizing environmental pollution contributed by other carbon-based fossil fuels. A specific advantage of biogas technology is its utilisation of organic wastes and other organic by-products for energy production, as opposed to disposal via landfills, which inevitably leads to further emissions of greenhouse gases via the process of slow decomposition [3].

The energy value of alternative crops as potential energy crops compared with corn in terms of environmental performance [4] must be studied. In order to achieve a stable and effective process for the production of biogas, an appropriate substrate composition must be provided [5]. Also, some studies have shown that co-digestion of food waste with yard waste at specific ratios can improve digester operating characteristics and end performance metrics over solid state anaerobic digestion of yard waste alone [6]. The energy crops must be ensilaged to extract the optimum content of nutrients with the optimal levels of C/N ratio and the right proportion of (XP) and crude fat (XL) [7]. The composition and ensiling process of the substrate and its biodegradability in the process of anaerobic fermentation are key factors in the production process of biogas and biomethane. The content of XP, XL, and carbohydrates in the substrate affects biogas yield. Ensiling could be beneficial in terms of reducing the duration of biogas production and improving biogas yield [8]. The quality of energy crops used for the production of biogas depends on the degree of ripeness of the substrate at the time of ensilage, the proper selection of crops, and the correct technology for growing energy crops [9-12]. In southern Sweden six crops (hemp, sugar beet, maize, triticale, grass/clover ley, winter wheat) were analysed for biogas production regarding methane yield per hectare, energy input, and costs in the production and supply of crops as biogas feedstock. The highest biomass and biogas yield was observed for sugar beets. Crops with lower risk of negative environmental impact in cultivation, such as ley and hemp, produced less than half the methane energy yield per hectare. Triticale, also having less risk of negative environmental impact, gave an energy yield similar to that of winter wheat grain and maize [13].

Dioha et al. [14] studied the effect of various parameters (pH, temperature, slurry concentration, total solids, and C/N ratio) on the production of biogas. Anaerobic digestion was performed with a variety of substrates such as cow, sheep, chicken, horse, and pig manure; sugarcane; rice hulls; and silage. The results of this study found that the C/N ratio is the most important parameter affecting the production of biogas. Zhang et al. [15] in the review paper found out that pH value, the content of VFA, and the C/N ratio are key factors affecting the yield of biogas production. The anaerobic fermentation process is largely dependent on the content of VFA produced as an intermediate product of the anaerobic decomposition of complex organic units. VFA are products that have a negative impact on the process of anaerobic digestion; if their content is too high, the result may be a drop in pH in the digestate. A range of things can affect the course of the process of anaerobic fermentation of VFA. The same content of VFA can be optimal for one particular digester and inappropriate for another.

In the process of anaerobic fermentation, Oleszek et al. [16] compared the silage of wild and cultivated varieties of reed canary grass (*Phalaris arundinacea* L.). Their study showed that reed canary grass varieties are both physiologically and chemically different. They claim that a large proportion of XF and high dry matter content have an influence on the reduced production of biogas. It is estimated that the cultivated varieties of reed canary grass, in contrast to the wild varieties, are a more suitable alternative co-substrate for biogas production.

In exploring the biomethane yield of various maize hybrids, Amon et al. [17] found that the quality of biogas for maize co-substrates depends on the proportions of XP and XL. Mahmood et al. [18] studied the production of biogas from several sorghum hybrids. They also studied the influence of the chemical composition of the substrate on the quality and quantity of the resulting biogas. The results of the research demonstrated that the percentage of dry matter at the time of ensiling, and the proportion of XP in the substrate, both affect the quality and quantity of the resulting biogas. Sathish et al. [19] studied the effect of certain variables on the production of biogas from rice straw substrate. The results of the research showed that the optimum temperature during the process of anaerobic fermentation, the pH value, and the substrate concentration significantly improve the production of biogas. Rath et al. [20] also studied the influence of the chemical composition of corn substrate on the production of biogas. By using results from the multiple regression model, it was found that the proportion of XL and crude fibre (XF) affects the production of biogas, while a high proportion of lignin in the composition of the substrate adversely affects the process of anaerobic fermentation.

Table 1. Review of models predicting the production of biogas and biomethane potential based on previous research.

Variable	R^2	p	Model	Author
Hemicellulose (HC), Acid detergent lignin (ADL)	0.70	<0.05	y = 0.13HC-2.00ADL+371 y = Yield of Biogas (Nl/kg VS)	Dandikas et al. [21]
Hemicellulose (HC), Acid detergent lignin (ADL)	0.83	<0.05	y = 0.25HC-3.93ADL+727 y = Yield of Methane (Nl/kg VS)	Dandikas et al. [21]
Crude protein (XP), Hemicellulose (HC), Acid detergent lignin (ADL)	0.75	<0.05	y = 0.44XP+0.16HC-3.02ADL+670 y = Yield of Biogas (Nl/kg VS)	Dandikas et al. [22]
Crude protein (XP), Hemicellulose (HC), Acid detergent lignin (ADL)	0.70	<0.05	y= 0.21XP+0.05HC-1.61ADL+370 y = Yield of Methane (Nl/kg VS)	Dandikas et al. [22]
Lignin (L)	0.698	<0.001	y= −12.804L+410.4 y = Biochemical Methane Potential (Nl/kg VS)	Triolo et al. [23]
Volatile fatty acids (VFA)	0.701	<0.001	y= 4.972VFA+167.6 y = Biochemical Methane Potential (Nl/kg VS)	Triolo et al. [23]
Cellulose (C)	0.249	<0.05	y= −3.574C +336.4 y = Biochemical Methane Potential (Nl/kg VS)	Triolo et al. [23]
Lignin (L) and Volatile fatty acids (VFA)	0.766	<0.001	y= −7.807L+3.057VFA+295.5 y = Biochemical Methane Potential (Nl/kg VS)	Triolo et al. [23]

Based on the VDI 4630 standard, Dandikas et al. [21] studied the correlation between the production of biogas and biomethane in connection with the chemical composition of energy crops: corn, barley, triticale, rye, potatoes, a mixture of grasses, alfalfa, millet, sunflower, and cup plant. Based on the statistical method of multiple regression, they created a statistical model to predict the production of biogas and another to predict the production of biomethane (Table 1). The standard Verein Deutscher Ingenieure (VDI) 4630 provides rules and specifications for tests in order to determine the biogas output of organic materials that serve the interpretation and operational optimisation of biogas plants. It thus gives comprehensive, practical, and relevant support for determining the biogas potential of organic materials. Concerning the VDI 4630 standard, Dandikas et al. [22] also studied the production of biogas and biomethane in correlation with the chemical composition of meadow plant species: resistant ryegrass, orchard grass, meadow bluegrass, meadow fescue, red clover, and white clover. From their results, and based on the multiple regression method, they created a statistical model to predict biogas production and another model to predict biomethane production (Table 1). Yet another study using the VDI 4630 standard is by Triolo et al. [23], where biomethane potential from 20 samples of various animal manures was observed. Using linear correlation, we found that the concentration of dry matter in the substrate affects the biomethane potential (y = 0.1914x + 1.0945, R^2 = 0.896). By using the multiple regression method, they also made statistical models to predict biomethane potential (Table 1).

Herrman et al. [24] studied methane production characteristics and chemical composition of 405 silages from 43 different crop species using uniform laboratory methods, with the aim to characterise a wide range of crop feedstocks from energy crop rotations and to identify the main parameters that influence biomass quality for biogas production. The study confirmed that a wide range of crops is suitable for anaerobic digestion. The most important biomass constituent that determines specific methane yield is lignin. Silage fermentation characteristics further affect methane production significantly.

The specific objective of this study is to find a correlation between the quality and quantity of biogas according to substrate composition (XP, XL, XF, C/N ratio, pH value, VFA) using multiple regression. With multiple regression, the combined effect of nutrients on biogas and biomethane production are presented and a model was developed, which will enable the prediction of biogas and biomethane production from all substrates.

When the prediction accuracy of models with one independent variable are not sufficiently high, multiple regression models are needed. An additional advantage of this statistical technique is that it enables us to study the effects, joint and individual, of all independent variables on the dependent variable. There are various statistical methods for computing the parameter estimates of multiple regression models. The best known is the OLS method, which is not efficient in the presence of outliers. In contrast, there are the robust regression methods (M-estimation, S-estimation, MM-estimation, etc.) that have not been widely used when dealing with biogas production parameters, despite the fact that they are better able to handle outliers. Different multivariate regression tools have found applications in many practical settings. The aim of this paper is to establish the most important predictors affecting biogas production from a particular

substrate and different combinations of co-substrates using multiple regression.

Materials and Methods

Design of Experiment

The experiment was carried in Slovenia, in the Prekmurje region (46°38' northern latitude, 16°5' eastern longitude, altitude 195 m) on slightly sandy soil. In the experiments, several energy crop types were sown: maize as the main crop, sorghum as the main crop, a mixture of crops for the production of biomass as the main crop, triticale as the main crop, and stubble maize after the triticale silage – in three repetitions. Treatments of maize and sorghum as the main crop came from plots of size of 224 m² (length 40 m, width 5.6 m, 8 rows with row width 0.7 m); treatments of a mixture of crops for the production of biomass as the main crop, triticale as the main crop, and stubble maize after the triticale silage came from plots of 240 m² (length 40 m, width 6 m). The plots varied in capacity depending on crop row spacing.

Analyzing Energy Crops

The composition of the substrate based on Weende analysis was observed. According to the standard method ISO 6496: 2000 [25], we analyzed the percentage of dry matter (DM) in the substrate of silage, ISO 5983: 2009 [26] for the percentage of XP in silage, ISO 6865: 2001 [27] for the percentage of XF in the silage, ISO 5984: 2003 [28] for the percentage of crude ash (XA) in the silage, and standard EC 98/64 method: 1998 [29] for the percentage of XL in the silage. For a mathematical calculation of the C/N ratio, the Kjeldahl method [30] was used to determine the value of nitrogen (N) and the standard method ISO 14235: 1998 [31] for the value of carbon (C) in the dry crop. For each tested sample, a titration for determining VFA, according to AAT proceedings was carried out [32].

Anaerobic Continuous Digestion of Substrate with an Energy Co-Substrate

Biogas production from energy crops and pig slurry was performed according to the German standard DIN 38414, Part 8 [33], with the experimental reactor at the Nemščak Biogas Plant. The standard is useful for determining the course of anaerobic fermentation of various organic substrates, the time course of developing gas, the composition of the gas, as well as the composition of the energy crops before and after the fermentation experiment.

Fig. 1 shows a test reactor with a 2500 litre capacity, which was filled with the inoculum and the substrate. The inoculum was taken from the Nemščak Biogas Plant, which served as the inoculum in the process of anaerobic fermentation for all experiments (its chemical composition is presented in Table 2). The digester was fed with 70% pig slurry (main substrate) and 30% co-substrates (mostly maize and triticale silage) based on volatile solids at an organic loading rate of 2.8 kg VS/(m³ d), at a hydraulic retention time of 30 days. The digester and test reactor were set to a mesophilic temperature range of 37.5ºC with a tolerance of ±1.5ºC, and the mixing in the test reactor was adjusted with a stirrer in the reactor as in commercial devices. During the experiment, the amount of VFA and the pH were also analysed. From the amount of VFA, it could be determined whether the reactor is "nourished" or "malnourished" with substrate (in our case with silage). Raising the VFA could easily lead to an acidification of the reactor content, and biogas production could stop. The optimum organic load of the reactor for the process of anaerobic fermentation was 2.8 kg VS/(m³ d). In this case the amount of VFA did not increase. Anaerobic digestion was performed with the basic substrate and with different co-substrate combinations. Altogether, 10 different mixtures of the substrate with the co-substrate (Table 3) were applied in three repetitions.

Biogas production is given in norm litres per kg of volatile solids (Nl/kg VS) and norm cubic meters per hectare (Nm³/ha); the volume of biogas production is based on norm conditions, i.e., To = 273 K and Po = 1013 mb. The ratio of organic dry matter of the substrate to organic dry matter of the inoculum was 0.4±0.1.

The amount of biogas produced was measured by using a flow meter (brand RITTER with capacity of 5 litres) for biogas. Biogas quality was measured three times a week with a gas analyser SR2-DO. Values of CO_2, O_2, and CH_4 in percentages and the value of H_2S in ppm were measured. The gas analyser has a measuring range 0-100% for CO_2, 0-25% for O_2, 0-100% for CH_4, and 0-2000 ppm for H_2S. The operation of the reactor was carried out via an automated control system that controls regulation and measurement, and monitors the yield of biogas.

Table 2. Chemical composition of the inoculum from Nemščak Farm.

DM (%)	ODM (%)	MgO (kg/m³)	P_2O_5 (kg/m³)	K_2O (kg/m³)	N (kg/m³)	NH_4-N (kg/m³)
4.8±0.2	3.5±0.1	0.9±0.1	3.0±0.2	2.9±0.1	5.0±0.3	3.2±0.1

DM – dry matter (%), ODM – organic dry matter (%), MgO – magnesium oxide (kg/m³), P_2O_5 – phosphorus pentoxide (kg/m³), K_2O – potassium oxide (kg/m³), N – nitrogen (kg/m³), NH_4-N – ammonium (kg/m³)
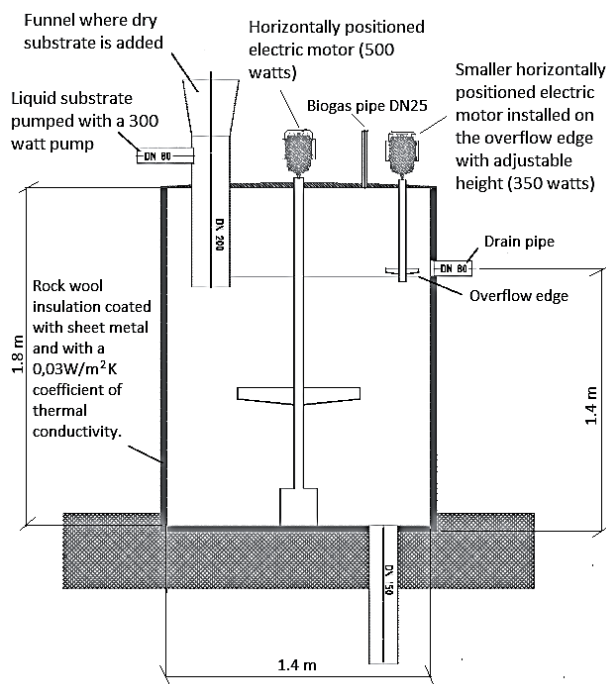
Fig. 1. Schematic diagram of a test reactor.

Statistical Data Analysis

A multiple linear regression model can be written as follows

$$Y = \beta X + \varepsilon \qquad (1)$$

…where Y is a (n×1) vector of observation of the dependent variable, X is a (n×p) matrix of independent variables, β is a (p×1) vector of unknown parameters, and ε is a (n×1) vector of random errors. The ordinary least squares (OLS) estimates for linear regression in (2) are given by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (2)$$

The assumptions of multiple regression include the assumptions of linearity, normality, independence, homoscedasticity, and multicollinearity. In regression analysis, the use of the least squares method would not be appropriate in solving the problem when the distribution of residuals is not normal, or when there are outliers. MM estimation, as introduced by Yohai [34], is a combination of two robust estimators: first the initial estimate is obtained with an S-estimator (which minimizes the dispersion of the residuals), and then it is improved with an M-estimator (which is robust in relation to outliers in the response variable, but turns out not to be resistant to leverage points).

Outliers occur very frequently in real data, and often go unnoticed. Outliers can be classified into three categories: outliers in the space of predictor variables (leverage points), outliers in the space of response variables (vertical outliers), and outliers in the space of predictor and response variables. A simple outlier diagnostic tool is a scatter plot that enables the detection of outliers in the case of linear regression or multiple linear regression with two independent variables at most. The impact of vertical outliers on the estimation of regression coefficients is usually small and mainly affects the regression intercept [35]. Besides the outliers, it is very important to detect an influential point (outliers are not necessarily influential), which is a point that greatly affects the calculated values of various estimates (slope, t-values, etc.). Cook's distance (D_i) is one of the measures for detecting high influence points; some authors have suggested that observations with $D_i > 1$ are taken as influential points [36], or a more conservative cutoff of $D_i > \frac{4}{n}$, where n is the total number of observations one could find in the literature [37]. Outliers in the space defined by the predictors can be detected by Mahalanobis distance, which computes the distance from the case

Table 3. Anaerobic digestion of substrate for various co-substrate ratios.

Substrate	Co-substrates
Pig slurry	Maize (100%) – main crop (CS1)
Pig slurry	Maize (100%) – stubble crop (CS2)
Pig slurry	Triticale (100%) – main crop (CS3)
Pig slurry	Mixture of plants for biomass production (100%) – main crop (CS4)
Pig slurry	Maize (70%) – main crop and triticale (30%) – main crop (CS5)
Pig slurry	Triticale (60%) – main crop and maize (40%) – stubble crop (CS6)
Pig slurry	Sorghum (85%) – main crop and maize (15%) – main crop (CS7)
Pig slurry	Maize (85%) – stubble crop and grain maize – grain at the wax ripeness stage (15%) (CS8)
Pig slurry	Grain maize (100%) – grain at the wax ripeness stage (CS9)
Pig slurry	Sorghum (100%) – main crop (CS10)

*CS = co-substrate

Table 4. Average values of all observed parameters of specific co-substrates.

CO SUB.	FSY (kg/ha)	DM (%)	DMY (kg/ha)	ODMY (kg/ha)	BP (Nl/kg VS)	BY (Nm³/ha)	XP (%)	XL (%)	XA (%)	NFE (%)	DXP (%)	XS (%)	XF (%)	C/N	CH₄ (%)	CO₂ (%)	O₂ (%)	H₂S (ppm)	pH	VFA (mg/l)
CS1	56696.7 ±1117	38.5 ±0.5	21381.6 ±187	20432.6 ±40.3	597 ±13.5	12209.8 ±299.6	8.3 ±0.1	8.5 ±0.3	3.8 ±0.4	64.2 ±1.1	4.2 ±0.5	71.9 ±2.3	14.9 ±0.4	45 ±3.1	66 ±0.6	33 ±0.6	0.4 ±0.1	900 ±30	7.1 ±0.6	785 ±27
CS2	28841.3 ±2265	37.2 ±0.8	10719.8 ±626	10246.3 ±554.1	513 ±5.5	5396.3 ±156.7	6.2 ±0.2	6.6 ±0.4	3.2 ±0.4	55.9 ±3.8	3.7 ±0.4	66.3 ±1.8	11.8 ±0.6	39 ±2.5	54 ±1.0	44 ±2.5	0.5 ±0.2	772 ±31	7.3 ±0.5	640 ±40
CS3	32129.7 ±111	34.9 ±1.2	11220.0 ±339	10651.6 ±282.5	345 ±6.0	3679.5 ±161.6	7.7 ±0.3	10.9 ±1.0	5.3 ±0.6	49.9 ±3.2	5.2 ±0.4	74.5 ±1.3	26.0 ±2.4	28 ±2.5	57 ±2.0	42 ±2.0	0.5 ±0.1	1050 ±190	6.1 ±0.2	360 ±13
CS4	20006.7 ±3235	26.7 ±1.0	5332.0 ±672	5075.2 ±628.7	489 ±18.5	2474.0 ±214.3	11.3 ±0.4	13.1 ±1.0	5.3 ±0.4	35.1 ±3.0	7.2 ±0.1	54.8 ±3.3	30.7 ±2.1	20 ±1.0	62 ±3.0	37 ±3.0	0.3 ±0.1	1220 ±125	8.1 ±0.3	950 ±40
CS5	48517.3 ±49	37.8 ±0.5	18339.4 ±224	17478.3 ±323.4	563 ±48.0	9850.6 ±1021.1	8.0 ±0.3	9.4 ±0.3	4.5 ±0.3	62.5 ±1.3	5.1 ±0.2	70.5 ±1.8	12.4 ±0.5	39 ±2.5	61 ±2.0	38 ±1.0	0.6 ±0.2	840 ±45	7.5 ±0.2	800 ±34
CS6	30806.3 ±534	35.8 ±0.4	11043.2 ±205	10588.7 ±53.2	384 ±30.0	4067.1 ±338.1	6.7 ±0.4	9.7 ±0.5	3.2 ±0.2	56.1 ±2.1	5.1 ±0.1	70.0 ±2.3	9.8 ±0.4	28 ±2.9	57 ±2.0	42 ±2.0	0.7 ±0.2	810 ±40	6.3 ±0.3	480 ±35
CS7	56897.3 ±1556	32.7 ±0.5	18623.3 ±261	17840.9 ±233.1	497 ±22.0	8864.1 ±302.6	7.7 ±0.5	9.4 ±0.6	4.8 ±0.4	54.8 ±2.2	3.6 ±0.1	61.4 ±1.7	14.3 ±0.4	41 ±3.1	60 ±3.0	39 ±3.0	0.2 ±0.1	800 ±130	6.9 ±0.3	750 ±100
CS8	33167.0 ±151	33.6 ±0.9	11155.2 ±316	10619.7 ±298.0	552 ±15.6	5864.1 ±298.2	7.8 ±0.4	7.2 ±0.2	3.4 ±0.3	69.5 ±1.6	5.3 ±0.2	70.7 ±1.3	13.4 ±0.4	37 ±3.6	58 ±3.0	41 ±3.0	0.5 ±0.1	790 ±20	6.5 ±0.4	829 ±20
CS9	55113.3 ±1354	34.6 ±0.7	19081.2 ±121	18242.2 ±210.8	635 ±2.0	11584.0 ±167.4	8.9 ±0.3	8.7 ±0.4	2.9 ±0.2	73.1 ±0.4	5.6 ±0.3	79.2 ±0.3	11.5 ±0.6	42 ±1.5	62 ±4.0	37 ±4.0	0.4 ±0.2	1050 ±80	7.4 ±0.3	770 ±40
CS10	57123.0 ±907	32.9 ±0.2	18821.3 ±408	17975.1 ±461.0	461 ±47.0	8272.6 ±639.6	7.6 ±0.9	11.0 ±2.6	5.6 ±0.2	55.8 ±2.8	3.0 ±0.2	62.4 ±1.7	14.0 ±0.3	41 ±1.2	59 ±2.0	40 ±2.0	0.5 ±0.1	880 ±35	6.4 ±0.3	650 ±30

FSY – fresh silage yields (kg/ha), DM – dry matter (%), DMY – dry matter yields (kg/ha), ODMY - organic dry matter yields (kg/ha), BP – biogas production (Nl/kg VS), BY – biogas yields (Nm³/ha), XP – crude protein (%), XL – crude fat (%), XA – crude ash (%), NFE – nitrogen free extracts (%), DXP - digestible crude protein (%), XS – starch (%), XF – crude fibre (%), C/N – C/N ratio, CH₄ – methane (%), CO₂ – carbon dioxide (%), O₂ – oxygen (%), H₂S – hydrogen sulphide (%), pH – pH value, VFA – volatile fatty acids (mg/l)

to the centroid of all cases for the predictor variables. Any distance larger than the critical value depending on sample size and number of predictors indicates an observation that is an outlier in the space defined by the predictors [36]. The centered leverage value is an index of the influence of each data point on the regression slope. Larger values indicate a greater degree of influence on the slope. The critical value for a small sample size is $\frac{3k}{n}$, where k is the number of predictors and n is the number of observations [38].

The multiple regression-based modeling study was performed on an Intel Core i5-3470T processor 3.6 GHz, 4 GB DDR3 computer system. The software packages SPSS (IBM SPSS Statistics 21) and R (RX 64 3.2.5) were used for statistical analysis. Multiple linear regression (MLR) analysis (stepwise) was performed in order to develop models for predicting biogas production in Nl/kg VS and biogas yield in Nm³/ha, where the explanatory variables were XP, NFE, XF, DM, XS, XA, VFA, and pH, and the C/N ratio. Because the assumptions were violated in most cases, the robust MM-method that addresses outliers in both the dependent and independent variables was also used [34]. In addition, the value of adjusted R-squared was calculated for each model, which measured the percentage of variance explained by the independent variables. Correlations between observed variables were estimated using Pearson's and Spearman's correlation.

Results and Discussion

The average yield of fresh silage per hectare (FSY) of energy co-substrates (Table 4) ranged from 20006 kg in the co-substrate mixture of plants for biomass production (CS4) to 57123 kg of fresh silage mass for sorghum (CS10). The lowest DM content at 26.7% was found in the mixture of plants for biomass production (CS4), and the highest in the maize main crop (CS1) at 38.5%. The highest average DMY was found in co-substrate maize main crop (CS1), 21381.2 DM/ha, while the lowest DMY was found in the co-substrate mixture of plants for biomass production (CS4), which was 5332 kg of DM/ha. The average yield of biogas (BP) ranged from 384 Nl/kg VS in co-substrate maize (70%) main crop and triticale (30%) main crop (CS5), to 635 Nl/kg VS in grain maize (100%) grain at the wax ripeness stage (CS9). Comparable results for BP investigated for co-substrates are also confirmed by other studies [10, 12-13, 17-18, 39-41]. The values of the C/N ratios ranged from 20 to 45/1. The optimum yield of biogas falls into the C/N ratio range from 20/1 to 40/1 [7, 14]. The methane content in biogas of the co-substrate ranged from 54% in maize stubble crop (CS2), up to 66% in maize main crop (CS1). The lowest content of hydrogen sulphide was detected in the co-substrate - maize stubble crop (CS2), with 772 ppm and the highest in the mixture of plants for biomass production (CS4), i.e., 1220 ppm. Hydrogen sulphide (H_2S) is always present in biogas, and the
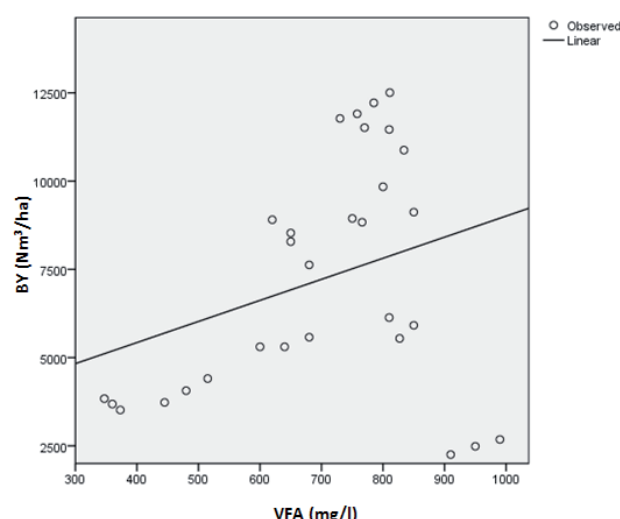

Fig. 2. Linear correlation (OLS estimation) between VFA and BY.

concentrations of H_2S should normally be below 200-1500 ppm [42]. The pH of the substrates (CS1-CS10) ranged from 6.1 to 8.1. Similar studies in the field of anaerobic fermentation in the same pH range of the substrates were also carried out by other researchers [6, 43-45]. The ideal pH range for anaerobic digestion in the digester has been reported to be between 6.8 to 7.4 [46]. The average value of VFA in the process of anaerobic fermentation ranged from 360 to 950 mg/l. It is well known that high VFA concentration causes a decrease in pH values and results in toxic reactor conditions. Digester stability also depends on the VFA:alkalinity ratio. Callaghan et al. [47] reported that there are three critical values: if the ratio is under 0.4 the digester should be stable, if the ratio is 0.4-0.8 some instability will occur, and if the ratio is over 0.8 the digester is significantly unstable. In our case, the VFA:alkalinity ratio during the experiment was between 0.03 and 0.07, which indicates that the digester was very stable.

Pearson's and Spearman's correlation coefficients between the parameters under investigation are presented in Tables 5 and 6. The highest linear correlation coefficient between BP was estimated for NFE, with a value of 0.57, while the lowest linear correlation was estimated between BP and XA, with a value of -0.44. For the variable BY, the highest positive linear association is again indicated between NFE, with a value of 0.69, and negative between XF, with a Pearson's coefficient value of -0.52. These results are in agreement with the report by Dandikas et al. [21], who established that for different crops, biogas yield had a positive correlation with NFE. Rath et al. [20] found that in different maize genotypes there was a negative effect of XF on biogas yield. The content of methane has a statistically significant positive correlation only with XP among the parameters of the Weende analysis. The highest Pearson's correlation coefficient between BP and these free single variables (C/N, pH, VFA) was estimated for

Table 5. Pearson's (top diagonal) and Spearman's (bottom diagonal) correlation coefficients of variables of Weende analysis and biogas production.

	BP	BY	CH_4	XP	XF	XL	NFE	XA	DM	XS
BP		0.74**	0.49**	0.22	-0.36*	-0.43*	0.57**	-0.44*	0.20	0.21
BY	0.77**		0.53**	-0.07	-0.52**	-0.35	0.69**	-0.23	0.49**	0.40*
CH_4	0.48**	0.49**		0.54**	0.17	0.19	0.10	0.09	-0.06	-0.04
XP	0.36*	0.20	0.67**		0.65**	0.59**	-0.35	0.34	-0.68**	-0.32
XF	-0.25	-0.16	0.34	0.42*		0.69**	-0.79**	0.61**	-0.63**	-0.42*
XL	-0.66**	-0.46*	0.13	0.28	0.53**		-0.67**	0.68**	-0.64**	-0.43*
NFE	0.76**	0.70**	0.24	0.18	-0.57**	-0.62**		-0.60**	0.62**	0.73**
XA	-0.49**	-0.27	0.10	0.20	0.68**	0.64**	-0.60**		-0.48**	-0.53*
DM	0.33	0.41*	0.02	-0.27	-0.27	-0.48**	0.39*	-0.45**		0.64**
XS	0.31	0.33	-0.01	0.06	-0.28	-0.28	0.57**	-0.46*	0.50**	

See List of Abbreviations for abbreviation meanings.
*The correlation values are significant at the 0.05 level. **The correlation values are significant at the 0.01 level.

Table 6. Pearson's (top diagonal) and Spearman's (bottom diagonal) correlation coefficients of variables measured in a chemical laboratory and in biogas production.

	BP	BY	C/N	CH_4	CO_2	O_2	H_2S	pH	VFA
BP		0.74**	0.58**	0.49**	-0.51*	-0.20	-0.07	0.53**	0.73**
BY	0.77**		0.84**	0.53**	-0.52**	-0.17	-0.20	0.13	0.31
C/N	0.58**	0.85**		0.18	-0.21	-0.20	-0.47**	-0.07	0.13
CH_4	0.48**	0.49**	0.31		-0.98**	-0.40*	0.23	0.30	0.50**
CO_2	-0.51*	-0.48**	-0.33	-0.98**		0.46*	-0.20	-0.33	-0.51**
O_2	-0.17	-0.17	-0.31	-0.42*	0.48**		-0.20	-0.30	-0.33
H_2S	-0.03	-0.05	-0.19	0.34	-0.29	-0.15		0.33	0.13
pH	0.50*	0.17	0.09	0.27	-0.32	-0.33	0.24		0.63**
VFA	0.59**	0.23	0.05	0.55**	-0.55**	-0.31	0.11	0.55**	

See List of Abbreviations for abbreviation meanings.
*The correlation values are significant at the 0.05 level. **The correlation values are significant at the 0.01 level.

VFA, with a value of 0.73 (Table 6); moreover, all these variables have a statistically significant positive linear correlation with BP. This is not the case for BY and CH_4, where only one significant correlation is obtained; C/N is positively correlated with BY (0.84) and VFA with CH_4 (0.50). Dioha et al. [14] reported that among the main parameters affecting biogas production are C/N ratio, pH, and VFA. To compare the parametric and nonparametric approaches, one can observe that for the pairs (BP, XF), (BY, XS), and (BY, XF), only Pearson's coefficients are significant; moreover, in the last pair values for both, coefficients differ widely (Pearson's coefficient is -0.52 and Spearman's coefficient is -0.16). De Winter et al. argued [48] that normally distributed data have similar expected values for both coefficients; thus, the parametric procedure in our data analysis for some pairs could lead to incorrect conclusions. In Fig. 2 the scatter plot shows the correlation between VFA and BY, where one can observe the presence of outliers that can cause misleading results if the parametric approach is applied.

The results of robust MM-estimation as compared to OLS regression results are interesting to observe. Ryan [49] argued that robust regression estimation techniques are almost as good as the OLS method in cases where the assumptions of OLS regression are met, and that they perform much better than OLS in other cases. It is also well known that the OLS estimator produced unstable prediction estimates in the presence of a non-normal distribution of errors. In the models, all variables show a statistically significant influence on the dependent variable ($p \leq 0.05$). Models were constructed separately

for variables of the Weende analysis, on the one hand, and for the variable pH, content of VFA, and C/N ratio on the other. Table 7 includes only models with an adjusted R-squared value higher than 0.54.

Outliers and influential points were examined using the Mahalanobis distance, Cook's Distance, and the centered leverage value (Table 7). The Mahalanobis distance did not detect outliers in the space of the

Table 7. Coefficient estimates for OLS and MM-methods.

Variable	OLS-Estimation	MM-Estimation
MODEL 1 (dependent variable is BP (Nl/kg VS))		
Intercept	106.87 (0.70)	128.48 (0.93)
XP	29.15 (3.32)**	30.11 (4.28)***
XS	-4.86 (-2.07)*	-5.65 (-2.65)*
NFE	8.55 (5.40)***	9.04 (6.72)***
Adj. R^2	0.54	0.58
Mahal. Distance (max)	9.26	
Cook's Distance (max)	0.14	
Centered leverage value (max)	0.32	
MODEL 2 (dependent variable is BP (Nl/kg VS))		
Intercept	154.49 (1.16)	210.62 (1.01)
XP	45.54 (4.87)***	46.84 (2.14)*
XL	-24.92 (-3.10)**	-28.48 (1.02)
NFE	3.78(2.85)**	3.29 (1.52)
Adj. R^2	0.61	0.65
Mahal. Distance (max)	12.79	
Cook's Distance (max)	1.20	
Centered leverage value (max)	0.44	
MODEL 3 (dependent variable is BP (Nl/kg VS))		
Intercept	-123.96 (-1.30)	-101.73 (0.80)
pH	33.12 (2.22)*	29.49 (1.36)
C/N	5.92 (5.95)***	5.88 (6.65)***
VFA	0.26 (4.40)***	0.27 (3.81)***
Adj. R^2	0.78	0.78
Mahal. Distance (max)	8.49	
Cook's Distance (max)	0.26	
Centered leverage value (max)	0.29	
MODEL 4 (dependent variable is BY (Nm^3/ha))		
Intercept	-11872.24 (-3.16)**	-14615.24 (-4.87)***
pH	924.42 (1.92)	1238.05 (3.76)***
C/N	351.14 (8.58)***	371.91 (12.79)***
Adj. R^2	0.72	0.80
Mahal. Distance (max)	8.39	
Cook's Distance (max)	0.41	
Centered leverage value (max)	0.29	

Note: Statistics in parentheses are t statistics. Significant codes: ***0.001 **0.01 *0.05.

predictors (the critical value was not exceeded); thus, with two predictors, the critical value equalled 12.24 ($\alpha = 0.05$), and with three predictors the critical value is 13.67 where n = 30. On the other hand, in Model 2 at least one point was influential ($D_i > 1$); by using more sophisticated criteria ($D_i > \frac{4}{n}$), it was possible to find more influential points. Additionally, centered leverage statistics indicated that two cases exceeded the critical value $0.3(\frac{3k}{n} = \frac{9}{30})$, and in the last case (Model 4), where only two predictors were involved, the critical value $0.2(\frac{3k}{n} = \frac{6}{30})$ was also exceeded.

Model 1 predicts BP (Nl/kg VS) using three regressors: XP, XS, and NFE. Using both OLS and robust regression, all variables were significant at the 0.05 level, and the model obtained by the MM-method yielded an improved goodness of fit ($R^2 = 0.58$) compared to that obtained by the OLS method ($R^2 = 0.54$). These findings agree with those reported by Dandikas [22], where XP was the prominent variable in their model for predicting biogas yield. In contrast, Weissbach [50] reported that XP was not significant for biogas yield prediction. In Model 2, instead of XS, XL was the explanatory variable. This shows that the coefficients for XL and NFE are not significantly different from zero at the 0.05 level in the MM-estimation. Rath et al. [20] found that XL was an important predictor affecting the specific biogas yield of maize. Models that include the variables pH, C/N ratio, and VFA reach a higher adjusted R-squared value than those including parameters of the Weende analysis. Model 4, obtained using the MM-estimate, has the adjusted R-squared value of 0.80, and both variables are significant at the 0.001 level. Further analysis involving a larger number of observations is needed to confirm the fact that MM-estimation is a good choice for this type of data.

Conclusions

The purpose of the research was to examine the correlation between substrate composition and biogas production. The highest biogas production was obtained with maize, while the other investigated mixtures of energy crops yielded lower values. Based on multiple regression, several linear models for predicting the production of biogas were suggested, where our goal was to compare OLS results with an MM-estimation. The results show that the MM-estimation did a better job. It was found that XP, XS, NFE, C/N ratio, VFA, and pH value were the most important predictors affecting biogas production from a particular substrate and co-substrates.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

References

1. AL SEADI T., RUTZ D., PRASSL H., KÖTTNER M., FINSTERWALDERF F., VOLK S., JANSSEN R., GRMEK M. Handbook on biogas. Agency of restructuring energy, Ljubljana, 1, **2010**.
2. COMPARETTI A., FEBO C., GRECO C., ORLANDO S. Current state and future of biogas and digestate production. Bulg. J. Agric. Sci. **19**, 1, **2013**.
3. KHAN I.U., OTHMAN M.H.D., HASHIM H., MATSUURA T., ISMAIL A.F., REZAEI-DASHTARZHANDI M., WAN AZELEE I. Biogas as a renewable energy fuel – A review of biogas upgrading, utilisation and storage. Energ. Convers. Manage. **150**, 277, **2017**.
4. GONZÁLES-GARCIA S., BACENETTI J., NEGRI M., FIALA M., ARROJA L. Comparative environmetal performance of three different annual energy crops for biogas production in Northern Italy. J. Clean. Proud. **43**, 1, **2013**.
5. SCHATTAUER A., ADBOUN E., WEILAND P., PLÖCHL M., HEIERMANN M. Abundance of trace elements in demonstration biogas plants. Biosyst. Eng. **108**, 57, **2011**.
6. BROWN D., LEBO Y. Solid state anaerobic co-digestion of yard waste and food waste for biogas production. Bioresour. Technol. **127**, 275, **2013**.
7. WANG X., LU X., LI F., YANG G. Effects of Temperature and Carbon-Nitrogen (C/N) Ratio on the Performance of Anaerobic Co-Digestion of Dairy Manure, Chicken Manure and Rice Straw: Focusing on Ammonia Inhibition. PLoS ONE. **9** (5), 1, **2014**.
8. KAFLE G.H., KIM S.H. Effects of chemical compositions and ensiling on the biogas productivity and degradation rates of agricultural and food processing by-products. Bioresour. Technol. **142**, 553, **2013**.
9. GAO J., CHEN L., YAN Z., WANG L. Effect of ionic liquid pretreatment on the composition, structure and biogas production of water hyacinth (*Eichhornia crassipes*). Bioresour. Technol. **132**, 361, **2012**.
10. BAUER A., LEONHARTSBERGER C., BÖSCH P., AMON B., FRIEDL A., AMON T. Analysis of methane yields from energy crops and agricultural by-products and estimation of energy potential from sustainable crop rotation systems in EU-27. Clean. Technol. Envir. **12**, 153, **2010**.
11. VERVAEREN H., HOSTYN K., GHEKIERE G., WILLEMS B. Biological ensilage additives as pretreatment for maize to increase the biogas production. Renew. Energy. **35** (9), 2089, **2010**.
12. OSLAJ M., MURSEC B., VINDIS P. Biogas production from maize hybrids. Biomass Bioenergy. **34** (11), 1538, **2010**.
13. GISSÉN C., PRADE T., KREUGER E., NGES I.A., ROSENQVIST H., SVENSSON S.E., LANTZ M., MATTSSON J.E., BÖRJESSON P., BJÖRNSSON L. Comparing energy crops for biogas production - Yields, energy input and costs in cultivation using digestate and mineral fertilisation. Biomass Bioenergy. **64**, 199, **2014**.

14. DIOHA I.J., IKEME C.H., NAFI'U T., SOBA N.I. YUSUF M.B.S. Effect pf carbon to nitrogen ratio on biogas production. IRJNS. **1** (3), 1, **2013**.

15. ZHANG B., SU H., BAEYENS J., TAN T. Reviewing the anaerobic digestion of food waste for biogas production. Renew. Sust. Energ. Rev. **38**, 383, **2014**.

16. OLESZEK M., KRÓL A., TYS J., MATYKA M., KULIK M. Comparison of biogas production from wild and cultivated varieties of reed canary grass. Bioresour. Technol. **156**, 303, **2014**.

17. AMON T., AMON B., KRYVORUCHKO V., ZOLLITSCH W., MAYER K., GRUBER L. Biogas production from maize and dairy cattle manure − Influence of biomass composition on the methane yield. Agric. Ecosyst. Environ. **118**, 173, **2007**.

18. MAHMOOD A., ULLAH H., IJAZ M., JAVAID M.M., SHAHZAD A.N., HONERMEIER B. Evaluation of sorghum hybrids for biomass and biogas production. Aust. J. Crop. Sci. **7** (10), 1456, **2013**.

19. SASHISH S., VIVEKANANDAN S. Optimization of Different Parameters Affecting Biogas Production from Rice Straw: An analytical Approach. J. Simulat. 78, **2011**.

20. RATH J., HEUWINKEL H., HERRMANN A. Specific Biogas Yield of Maize Can Be Predicted by the Interaction of Four Biochemical Constituents. Bioenergy. Res. **6** (3), 939, **2013**.

21. DANDIKAS V., HEUWINKEL H., LICHTI F., DREWES J.E., KOCH K. Correlation between biogas yield and chemical composition of energy crops. Bioresour. Technol. **174**, 316, **2014**.

22. DANDIKAS V., HEUWINKEL H., LICHTI F., DREWES J.E., KOCH K. Correlation between Biogas Yield and Chemical Composition of Grassland Plant Species. Energy Fuels. **29** (11), 7221, **2015**.

23. TRIOLO J.M., WARD A.J., PEDERSEN L., SOMER S.G. Characteristics of Animal Slurry as a Key Biomass for Biogas Production in Denmark. Biomass Now - Sustainable Growth and Use. InTech - Open Access Publisher. 307, **2013**.

24. HERRMAN C., IDLER C., HEIERMANN M. Biogas crops grown in energy crop rotations: Linking chemical composition and methane production characteristics. Bioresour. Technol. **206**, 23, **2016**.

25. SIST ISO 6496:2000. Animal feeding stuffs, Determination of moisture and other volatile matter content.

26. SIST ISO 5983:2009. Animal feeding stuffs, Determination of nitrogen content and calculation of crude protein content, Part 2: Block digestion and steam distillation method.

27. SIST EN ISO 6865:2001. Animal feeding stuffs, Determination of crude fibre content − Method with intermediate filtration (ISO 6865:2000).

28. SIST ISO 5984:2003. Animal feeding stuffs, Determination of crude ash.

29. EU Directive 98/64/EC of 3 September **1998**. Establishing community methods of analysis for the determination of amino acids, crude oils and fats.

30. KJELDAHL J. Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern (New method for the determination of nitrogen in organic substances), Zeitschrift für analytische Chemie. **22** (1), 366, **1998**.

31. ISO 14235:1998. Soil quality - Determination of organic carbon by sulfochromic oxidation.

32. AAT Biogas Technology, **2017**. Available online: http://www.aat-biogas.at/ (accessed on 07.10.2017).

33. DIN 38 414. Deutsche Norm (Schlamm und Sedimente – Grupe S), **1985**.

34. YOHAI V.J. High breakdown-point and high efficiency robust estimates for regression. Ann. Stat. **15** (2), 642, **1987**.

35. STURM J.E., DE HAAN J. How robust is the relationship between economic freedom and economic growth? Applied. Economics. **33** (7), 839, **2001**.

36. STEVENS J. P. Outliers and influential data points in regression analysis. Psychological Bulletin. **95** (2), 334, **1984**.

37. BOLLEN K.A., JASCKMAN R. Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Eds.), Modern methods of data analysis. Newbury Park, CA: Sage. 257, **1990**.

38. VELLEMAN P. F., WELSCH R. E. Efficient computing of regression diagnostics. The American Statistician. **35**, 234, **1981**.

39. BAUER A., MAYR H., HOPFNER-SIXT K., AMON T. Detailed monitoring of two biogas plants and mechanical solid-liquid separation of fermentation residues. J. biotechnol. res. **142** (1), 56, **2009**.

40. NEGRI M., BACENETTI J., BRAMBILLA M., MANFREDINI A., CANTORE A., BOCCHI S. Biomethane production from different crop systems of cereals in Northern Italy. Biomass Bioenergy. **63**, 321, **2014**.

41. SEPPÄLÄ M., PYYKKÖNEN V., LAINE A., RINTALA J. Methane production from maize in Finland - Screening for different maize varieties and plant parts. Biomass Bioenergy. **46**, 282, **2012**.

42. SUN Q., LI H., YAN J., LIU L., YU Z., YU X. Selection of appropriate biogas upgrading technology-a review of biogas cleaning, upgrading and utilisation. Renew. Sust. Energ. Rev. **51**, 521, 2015.

43. LI Y., PARK S.Y, ZHU J. Solid-state anaerobic digestion for methane production from organic waste. Renew. Sustain. Energy. Rev. **15** (1), 821, **2011**.

44. LIN J., ZUO J., GAN L., LI P., LIU F., WANG K., CHEN L., GAN H. Effects of mixture ratio on anaerobic co-digestion with fruit and vegetable waste and food waste of China. J. Environ. Sci. **23** (8), 1403, **2011**.

45. NGES I.A., BJÖRNSSON L. High methane yields and stable operation during anaerobic digestion of nutrient-supplemented energy crop mixtures. Biomass Bioenergy. **47**, 62, **2012**.

46. MAO C., FENG Y., WANG X., REN G. Review on research achievements of biogas from anaerobic digestion. Renew. Sustain. Energy. Rev. **45**, 540, **2015**.

47. CALLAGHAN F.J., WASE D.A.J., THAYANITHY K., FORSTER C.F. Continuous co-digestion of cattle slurry with fruit and vegetable wastes and chicken manure. Biomass. Bioenergy. **22** (1), 71, **2002**.

48. DE WINTER J.C.F, GOSLIN S.D., POTTER J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. Psychol. Methods. **21** (3), 273, **2016**.

49. RYAN T.P. Modern Regression Methods, New York, NY: John Wiley & Sons, Inc., 1, **2009**.

50. WEISSBACH F. Die Bewertung von nachwachsenden Rohstoffen für die Biogasgewinnung. Teil I: Das Gasbildungspotenzial der fermentierbaren Nährstoffe. Pflanzenbauwissenschaften, **13** (2), 72, **2009**.