# Feedstock and inoculum characteristics and process parameters as predictors for methane yield in mesophilic solid-state anaerobic digestion

**GABRIELA R. NIQUINI, SUZIMARA R. SILVA, ESLY F. COSTA JUNIOR** and **ANDRÉA O.S. COSTA**

Departamento de Engenharia Química, Escola de Engenharia, Universidade Federal de Minas
Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brazil

**How to cite:** NIQUINI GR, SILVA SR, COSTA JUNIOR EF AND COSTA AOS. 2019. Feedstock and inoculum characteristics and process parameters as predictors for methane yield in mesophilic solid-state anaerobic digestion. An Acad Bras Cienc 91: e20181181. DOI 10.1590/0001-3765201920181181.

**Abstract:** In this study, several linear regression models were tested to predict the cumulative 30-day methane yield produced in mesophilic solid-state anaerobic digestion, employing diverse lignocellulosic biomass sources. Data collected from 13 studies were utilized, totalizing 86 experimental points, divided into regression and validation. Models containing higher order terms, the inverse of variables and interactions among all eleven input variables were tested. Simple linear models utilizing a single variable were unable to describe the methane production, giving an $R^2$ lower than 0.37. However, combinations of multiple variables and its inverses as only independent variable permitted an increase in simple linear models predictive capacity up to 63% of experimental variability. Higher order models presented an improvement in predictive quality: for a fourth-order multiple linear model, a validation $R^2$ of 0.8329 was achieved. In view of the obtained results, the proposed linear regression models consist in an attractive tool to propose experimental routines and to investigate new biomass sources for methane production using solid-state anaerobic digestion, significantly reducing time and cost requirements to experiments' execution.

**Key words:** lignocellulosic biomass, linear regression, methane, polynomial models, solid-state anaerobic digestion.

## INTRODUCTION

Anaerobic digestion (AD) is a natural process, in which organic matter is converted by microbes into gases which present energetic potential, such as methane ($CH_4$) and hydrogen ($H_2$), under oxygen-free conditions (Baêta et al. 2016). AD is currently explored to produce biogas using various lignocellulosic feedstock sources, such as fruit and vegetable waste, grape pomace, rice straw, wheat straw, yard waste, tree trimming and yard trimming (Brown et al. 2012, Liew et al. 2012, Zhao et al. 2014, Sheets et al. 2015, Baêta et al. 2016, El Achkar et al. 2017, 2018, Pezzolla et al. 2017, Edwiges et al. 2018).

Traditionally, biogas is produced by means of liquid anaerobic digestion (L-AD). However, recent studies have highlighted the alternative use of solid-state anaerobic digestion (SS-AD), which occurs for solid concentration above 15% (Ge et al. 2016, Xu et al. 2016, Pezzolla et al. 2017). In comparison

Correspondence to: Gabriela Rodrigues Niquini
E-mail: gabrielaniquini@gmail.com
ORCid: https://orcid.org/0000-0001-9075-0814

to L-AD, SS-AD requires smaller reactor volume for the same total solids content and reduced energy demand for heating and mixture (Xu et al. 2014, Ge et al. 2016). Nevertheless, mecanism reactions in SS-AD, as well as the influence of total solids amount (TS), feedstock composition, heat and mass transfer and temperature, are not yet clear (Xu et al. 2015).

Studies have demonstrated the influence of variables related to biomass in SS-AD process, such as cellulose, hemicellulose and lignin amounts. Inoculum characteristics, including total ammonia nitrogen (TAN) and alkalinity, also affect biogas production. Besides, SS-AD performance exhibits dependence from process variables, such as total solids content, carbon-to-nitrogen ratio (C/N), feedstock-to-effluent ratio (F/E) and particle size (Li et al. 2011b, Xu et al. 2013, 2016, Ge et al. 2016).

The performance of SS-AD processes is commonly evaluated by biochemical methane potential (BMP) tests. BMP is defined as the maximum methane amount produced by gram of volatile solid (VS) in a laboratorial experiment (Angelidaki et al. 2009, Brown et al. 2012, Liew et al. 2012). However, in literature it is brought attention to the use of alternative methods to determine methane yield, especially due to the long time required to perform a BMP test (at least 30 days) and intensive labor (Lesteur et al. 2010, Thomsen et al. 2014, Xu et al. 2014).

Mathematical models are an attractive alternative to BMP test, since these models allow a rapid estimation of methane production, reducing the requirement for laboral analyses (Thomsen et al. 2014, Xu et al. 2015, Kafle and Chen 2016). Linear regression models are relevant tools to propose experimental routines of little understood systems, since these models do not require a deep knowledge of the involved phenomena in the processes, which applies for solid-state anaerobic digestion employing lignocellulosic biomass.

In studies performed by Gunaseelan (2007), methane yield produced by L-AD from different biomass sources (fruit and vegetable solid wastes, sorghum and napiergrass) is predicted using linear regression models with basis on feedstock chemical composition. Simple linear regression models were inefficient to correlate methane yield and biomass individual components (carbohydrates, proteins, fibers, lignin and cellulose), highlighting the contribution of multiple factors to the process. On the other hand, regression models containing more than one independent variable – multiple linear regression (MLR) – were more effective to predict the methane yield.

Xu et al. (2014) considered for the first time the combination of feedstock characteristics and operational parameters in the development of MLR models for SS-AD from lignocellulosic biomass. The developed models considered solely variables linearly correlated to methane yield, which were determined by principal component analysis (PCA) method. Besides, the influence of terms containing interactions between the explanatory variables, as well as quadratic and cubic terms, was investigated only for terms containing feedstock-to-effluent ratio (F/E). This approach might have disconsidered relevant information, having in sight the probable non-linear relation between the input variables and methane yield, in addition to interactions among variables other than F/E ratio.

The present work proposes the development of linear regression models, in order to predict the 30-day cumulative methane production during SS-AD, utilizing varied lignocellulosic biomasses. Models containing all eleven explanatory variables, its inverses and combinations were constructed. Combinations of all explanatory variables were tested, including those which presented a low linear correlation to methane yield. Besides, a data set containing more experimental points, comparing to previous publications, was used for models' creation.

## MATERIALS AND METHODS

DATA COLLECTION

Data utilized in this study were collected from thirteen publications, which provided 86 points. Explanatory variables were divided into feedstock characteristics, inoculum characteristics and SS-AD process parameters. The feedstock characteristics applied to this study are volatile solids (VS), extractives, lignin, cellulose and hemicellulose. These variables are presented in Table I as ranges, grouped into each feedstock source. Inoculum characteristics are alkalinity and total ammonia nitrogen (TAN). SS-AD process parameters consist in feedstock-to-effluent ratio (F/E), particle size, carbon-to-nitrogen ratio (C/N) and total solids (TS). The response variable is defined as the 30-day methane yield. Some experiments were performed in a longer period than 30 days, however, only yields until the 30th day were considered at this publication. Value ranges for inoculum characteristics, process parameters and yields are presented in Table II.

The data set presents a significant variability, covering a wide range in the variables (Tables I and II), permitting the application of proposed models for several experimental conditions. Due to reduced data availability in literature for a statistical test, factors such as inoculum source, digester type, operation temperature, co-digestion and biomass pretreatment were not considered in this study. Therefore, experimental data contain solely information of mesophilic SS-AD systems, with sewage sludge as inoculum source, non-treated lignocellulosic biomass as feedstock and a TS content around 25%.

LINEAR REGRESSIONS AND STATISTICAL ANALYSIS

A statistical analysis was developed employing a correlation matrix (software R, version 3.3.1), constituted by several correlation coefficients ($r$) values, which calculate the linear dependence of each two variables, generally called $x$ and $y$, as described by Equation (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

In this equation, $x_i$ and $y_i$ represent the $i$-th measured value of variables $x$ and $y$, respectively; $\bar{x}$ and $\bar{y}$ are the mean values for the variables, and

**TABLE I**
**Ranges of feedstock characteristics.**

| Feedstock | VS | Extractives | Lignin | Cellulose | Hemicellulose |
|---|---|---|---|---|---|
| | g/100g dry biomass | | | | |
| Corn stover[a] | 89.00 - 95.97 | 6.50 - 9.90 | 15.20 - 21.70 | 33.70 - 42.23 | 15.30 - 22.87 |
| Wheat straw[b] | 83.40 - 94.68 | 13.40 - 17.00 | 15.20 - 22.10 | 32.30 - 37.90 | 15.20 - 21.80 |
| Switchgrass[c] | 89.90 - 96.90 | 11.90 - 12.10 | 17.80 - 19.30 | 31.00 - 32.30 | 16.70 - 19.50 |
| Leaves[d] | 86.90 - 93.01 | 34.70 - 36.02 | 22.70 - 23.10 | 11.10 - 12.20 | 4.20 - 11.50 |
| Maple[e] | 92.20 | 5.10 | 22.00 | 36.50 | 9.20 |
| Pine[f] | 90.50 | 14.20 | 28.30 | 26.00 | 4.40 |
| Yard trimming[g] | 91.70 - 96.53 | 14.70 - 23.40 | 21.00 - 26.02 | 20.70 - 27.40 | 9.00 - 14.20 |
| Tree trimming[h] | 98.90 - 99.60 | 9.60 - 17.00 | 27.10 - 32.90 | 23.30 - 30.80 | 11.50 - 15.90 |

[a](Zhu et al. 2010, Li et al. 2011b, Brown et al. 2012, Liew et al. 2012, Xu and Li 2012, Shi et al. 2013, Xu et al. 2013); [b](Cui et al. 2011, Brown et al. 2012, Liew et al. 2012); [c](Brown et al. 2012, Sheets et al. 2015); [d](Liew et al. 2011, 2012, Brown et al. 2012); [e](Brown et al. 2012); [f](Brown et al. 2012); [g](Brown et al. 2012, Liew et al. 2012, Xu et al. 2016); [h](Cherosky 2012, Zhao et al. 2014).

**TABLE II**
**Ranges of inoculum, process characteristics and yields.**

| Feedstock | Alkalinity (gCaCO$_3$/kg) | TAN (gN/kg) | F/E (VS$_{feed}$/ VS$_{effluent}$) | Particle size (mm) | C/N ratio | TS (%) | CH$_4$ Yield (L/ VS$_{feed}$) |
|---|---|---|---|---|---|---|---|
| Corn stover[a] | 13.20 - 19.90 | 2.00 - 4.20 | 2.00 - 7.41 | 5 - 15 | 15.10 - 39.00 | 18.50 - 27.00 | 4.21 - 237.12 |
| Wheat straw[b] | 13.20 - 18.30 | 3.30 - 4.20 | 2.00 - 6.00 | 5 - 10 | 18.79 - 36.00 | 18.50 - 22.00 | 6.14 - 123.90 |
| Switchgrass[c] | 13.20 - 14.23 | 3.30 - 4.47 | 3.00 | 5 | 15.00 - 43.00 | 18.50 - 30.00 | 46.76 - 116.90 |
| Leaves[d] | 8.90 - 18.30 | 3.30 - 4.20 | 2.00 - 8.20 | 5 - 9 | 14.25 - 25.00 | 18.50 - 26.00 | 0.61 - 75.30 |
| Maple[e] | 13.20 | 3.30 | 3.00 | 5 | 50.30 | 18.50 | 46.90 |
| Pine[f] | 13.20 | 3.30 | 3.00 | 5 | 46.20 | 18.50 | 17.00 |
| Yard trimming[g] | 13.20 - 18.30 | 3.30 - 4.30 | 0.26 - 5.00 | 5 - 9 | 11.85 - 31.92 | 18.50 - 35.00 | 0.00 - 227.81 |
| Tree trimming[h] | 14.50 - 16.50 | 2.00 - 2.90 | 2.00 - 4.00 | 6.35 - 12.70 | 17.80 - 34.90 | 13.90 - 20.00 | 9.50 - 19.06 |

[a](Zhu et al. 2010, Li et al. 2011b, Brown et al. 2012, Liew et al. 2012, Xu and Li 2012, Shi et al. 2013, Xu et al. 2013); [b](Cui et al. 2011, Brown et al. 2012, Liew et al. 2012); [c](Brown et al. 2012, Sheets et al. 2015); [d](Liew et al. 2011, 2012b, Brown et al. 2012); [e](Brown et al. 2012); [f](Brown et al. 2012); [g](Brown et al. 2012, Liew et al. 2012, Xu et al. 2016); [h](Cherosky 2012, Zhao et al. 2014).

$n$ is the number of data points used to calculate the correlation coefficient.

The value of $r$ varies between -1 and 1, being -1 when correlation is total and negative, and 1 when correlation is total and positive. When the variables have no correlation, $r$ equals 0. In order to evaluate whether linear correlation between two variables is significant, hypothesis tests, such as t-test, shall be performed. The test statistic ($t^*$) in t-test is given by Equation (2):

$$t^* = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \qquad (2)$$

This value has t-Student distribution with $n-2$ degrees of freedom. $p$-value of $t^*$ is then compared to the $p$-value established for the test – which was adopted as 0.05 in this study – and, if it is smaller than the latter, the variables' correlation is statistically significant.

The coefficient of determination, denoted by $R^2$, is calculated as described in Equation (3):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \qquad (3)$$

In this equation, $\widehat{y}_i$ represents the predicted value for $i$-th experiment. $R^2$ is a statistical measure of the dependent variable behavior which can be explained by the independent variables. $R^2$ approaches 1 as the output variable variability is more accurately described by the regression.

$F$ statistic describes the significance of a regression model. For models which have intercept different from 0, it is calculated as described by Equation (4):

$$F_0 = \frac{\sum_{i=1}^{n}\left(\widehat{y}_i - \overline{y}\right)^2 \Big/ p}{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \Big/ (n-p-1)} \qquad (4)$$

$F_0$ has an F-distribution with $p$ degrees of freedom 1 and $n - p - 1$ degrees of freedom 2.

When the intercept is defined as 0, $F$ is given by Equation (5):

$$F_0 = \frac{\sum_{i=1}^{n}\left(\widehat{y}_i - \overline{y}\right)^2 \Big/ p}{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \Big/ (n-p)} \qquad (5)$$

In this case, $F_0$ has an F-distribution with $p$ degrees of freedom 1 and $n - p$ degrees of freedom 2.

Standard error of prediction ($SEP$) is a measurement of the predicted data deviation, comparing to the experimental data. It is calculated as described by Equation (6).

$$SEP = \sqrt{\frac{\left(\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2\right)}{n}} \qquad (6)$$

Among the 86 data points obtained for this study, 68 – which corresponds to approximately 79% of the data set – were used to construct the correlation matrix, as well as the statistical models, and the remaining 18 were used for models' validation. The correlation matrix and the statistical models were built using commercial softwares. Validation points were randomly selected, in such a manner that values for each variable were comprised within regression interval for every variable.

Initially, first to seventh order single linear regression models were developed. Multiple linear regression models were also evaluated, from first to fourth order. All cases were tested by evaluating effects of the variables and its inverses. In some situations, predicted methane yield was found to be lower than 0. As it happened, the prediction was set to 0, since a negative methane yield is not experimentally possible. All terms were evaluated in relation to statistic significance, including the linear coefficient, and only statistically significant terms were included, adopting the criterion of $p$-value smaller than 0.05.

**RESULTS AND DISCUSSION**

CORRELATION MATRIX

Correlation coefficients obtained for the data employed in the models' construction – regression data – are presented in Table III. Highlighted values

are those which present a $p$-value smaller than 0.05, meaning that pair of variables have a statistically significant correlation one with the other.

Results suggest that two feedstock variables (VS and lignin) and two process parameters (F/E and C/N) are strongly correlated to methane yield. Also, all these variables have a negative correlation with the yield, indicating a decrement in methane yield as these variables are linearly increased. Xu et al. (2014) also observed a strong negative correlation between feedstock-to-effluent ratio and methane production, using the Principal Component Analysis (PCA) method. This correlation is probably justified by the reduction of microbial population and alcalinity as F/E is increased (Shi et al. 2014).

C/N ratio is also a relevant factor for the anaerobic digestion process, being its maintenance within an optimum range (typically 20-30) crucial to avoid excess of total ammonia nitrogen and volatile fatty acids (VFAs) in the digestor (Li et al. 2011a). Elevated TAN and VFAs concentration might diminish methanogenic activity, causing damages to AD process (Li et al. 2011a). In Xu et al. (2014), no significative correlation between C/N ratio and cumulative methane yield was observed. The authors justify such a result as a function of data collection, since most of collected C/N values are comprised between 18 and 30. According to Xu et al. (2014), C/N ratio plays little influence in methane yield, since it presents itself within the optimum range. In the present work, a significant experimental points quantity with C/N ratio below 20 was employed, which might have been a cause for the negative correlation observed.

In literature, a consensus about the relationship between biomass constituents and methane production through AD can be observed (Gunaseelan 2007, 2009, Lesteur et al. 2010, Xu et al. 2014). Among the constituents, lignin content has been considered as one of the essential factors which restrain anaerobial biodegradation (Liew

**TABLE III**
**Correlation matrix for regression data.**

| Variable | VS | Ext. | Lig. | Cell. | Hcell. | Alk. | TAN | F/E | Size | C/N | TS | Yield |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VS | 1.00 | **-0.31**[*] | **0.24**[*] | 0.15 | **0.28**[*] | 0.19 | -0.22 | 0.18 | **0.35**[*] | -0.04 | -0.11 | **-0.26**[*] |
| Extractives | **-0.31**[*] | 1.00 | **0.43**[*] | **-0.94**[*] | **-0.69**[*] | **-0.35**[*] | **0.28**[*] | -0.10 | 0.14 | **-0.37**[*] | 0.13 | -0.10 |
| Lignin | **0.24**[*] | **0.43**[*] | 1.00 | **-0.57**[*] | **-0.64**[*] | -0.18 | -0.17 | -0.06 | 0.18 | 0.06 | -0.17 | **-0.30**[*] |
| Cellulose | 0.15 | **-0.94**[*] | **-0.57**[*] | 1.00 | **0.81**[*] | **0.26**[*] | **-0.31**[*] | 0.18 | -0.13 | **0.33**[*] | -0.10 | 0.18 |
| Hemicellulose | **0.28**[*] | **-0.69**[*] | **-0.64**[*] | **0.81**[*] | 1.00 | **0.27**[*] | -0.07 | 0.17 | 0.09 | -0.08 | 0.14 | 0.23 |
| Alkalinity | 0.19 | **-0.35**[*] | -0.18 | **0.26**[*] | **0.27**[*] | 1.00 | **0.37**[*] | -0.22 | 0.10 | -0.12 | 0.11 | 0.08 |
| TAN | -0.22 | **0.28**[*] | -0.17 | **-0.31**[*] | -0.07 | **0.37**[*] | 1.00 | **-0.27**[*] | -0.05 | **-0.37**[*] | **0.40**[*] | 0.02 |
| F/E | 0.18 | -0.10 | -0.06 | 0.18 | 0.17 | -0.22 | **-0.27**[*] | 1.00 | -0.01 | **0.56**[*] | -0.08 | **-0.59**[*] |
| Particle size | **0.35**[*] | 0.14 | 0.18 | -0.13 | 0.09 | 0.10 | -0.05 | -0.01 | 1.00 | -0.20 | 0.08 | -0.03 |
| C/N | -0.04 | **-0.37**[*] | 0.06 | **0.33**[*] | -0.08 | -0.12 | **-0.37**[*] | **0.56**[*] | -0.20 | 1.00 | **-0.27**[*] | **-0.38**[*] |
| TS | -0.11 | 0.13 | -0.17 | -0.10 | 0.14 | 0.11 | 0.40 | -0.08 | 0.08 | **-0.27**[*] | 1.00 | -0.14 |
| Yield | **-0.26**[*] | -0.10 | **-0.30**[*] | 0.18 | 0.23 | 0.08 | 0.02 | **-0.59**[*] | -0.03 | **-0.38**[*] | -0.14 | 1.00 |

*Significant correlations ($p$-value <0.05).

et al. 2012, Xu et al. 2014). In the present study, cellulose content did not present a strong linear correlation with the response variable, differently from the observations in Xu et al. (2014), in which a strong linear correlation between cellulose and methane yield was observed.

Fig. 1 illustrates the relationship between methane yield and each of the four variables with significant linear correlation, as presented in the correlation matrix (Table III). Plots in Fig. 1 confirm the yield diminution for largest values observed in each explanatory variable, however, the intermediate points do not exhibit a clear tendency. These results indicate that the mathematical phenomenon description might be complex, and, therefore, correlation between exploratory variables might be significant. The results presented in Table III corroborate with this hypothesis, since several significant correlations ($p$-value < 0.05) between exploratory variables are exhibited. For example, correlations between cellulose and extractives and between cellulose and hemicellulose were evaluated as -0.94 and 0.81 respectively. Such results reinforce the necessity to investigate linear regression models containing higher order terms and interactions among input variables. Having in

sight feedstock sources diversity and operational parameters considered in this story, as well as SS-AD process complexity, the variables with little correlation to methane yield were not disconsidered for the models' testing.

LINEAR REGRESSIONS

Several linear regression models were tested in order to predict the methane yield, namely: (i) simple linear regression; (ii) first-order multiple linear regression and (iii) higher-order multiple linear regression.

Table IV exhibits results for simple linear regression models, going from first to seventh order, including the exploratory variables and its inverses. With only one explanatory variable, the simple linear regression models were unable to describe methane yield as a function of any exploratory variable. Similar conclusions were verified in previous studies about methane yield prediction based on feedstock characteristics (Gunaseelan 2007, Xu et al. 2014) and operational parameters (Xu et al. 2014).

The best result for validation involved the inverse of feedstock-to-effluent ratio, which gave an $R^2$ of 0.3601 and an *SEP* of 47.80 for validation
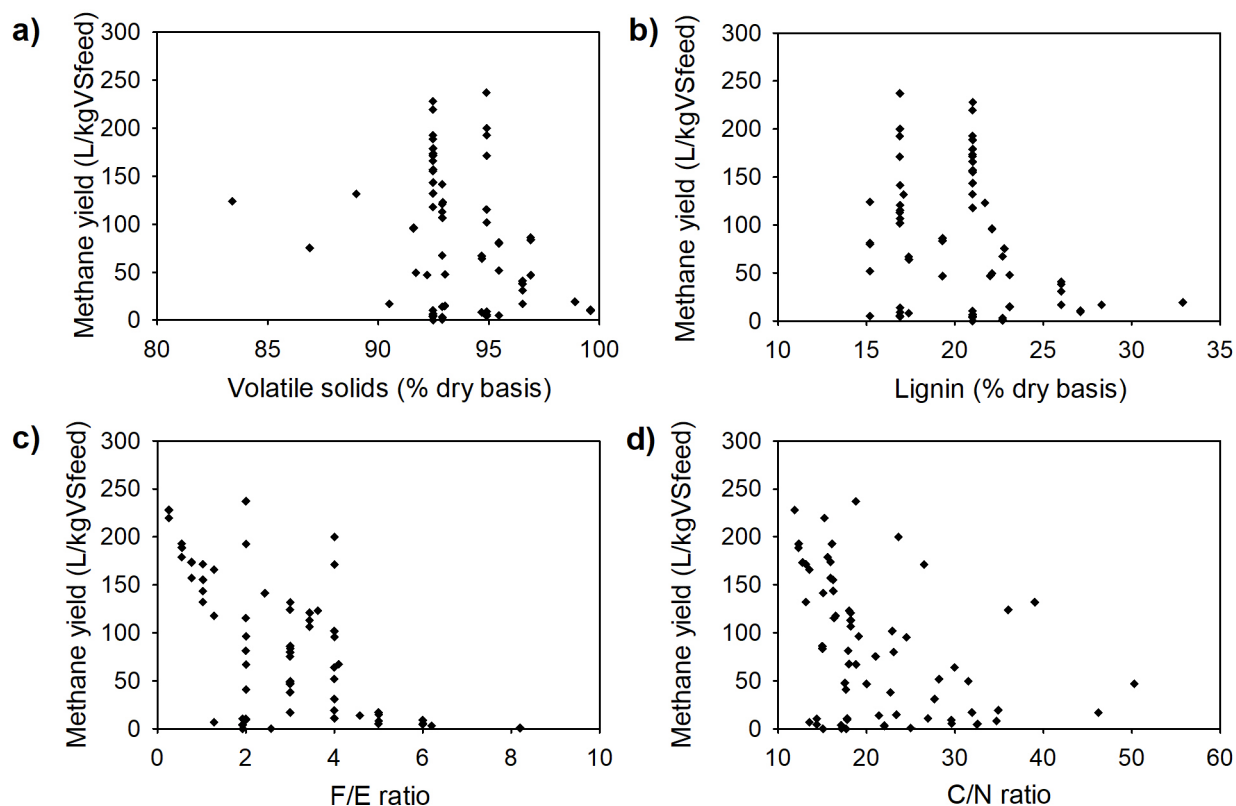
**Figure 1** - Methane yield as a function of (**a**) volatile solids, (**b**) lignin content, (**c**) feedstock-to-effluent ratio and (**d**) carbon-to-nitrogen ratio.

### TABLE IV
### Summary of simple linear regression models.

| Model | Significance F | $R^2$ regression | SEP regression | $R^2$ validation | SEP validation |
|---|---|---|---|---|---|
| $Y = 154.65 - 24.722FE$ | $1.35 \cdot 10^{-7}$ | 0.3528 | 56.60 | 0.3310 | 48.87 |
| $Y = 45.733 + 61.211/FE$ | $6.44 \cdot 10^{-8}$ | 0.3598 | 56.29 | 0.3601 | 47.80 |
| $Y = 21.591 + 1721.4/(Ext \cdot FE)$ | $2.72 \cdot 10^{-11}$ | 0.4918 | 50.15 | 0.6109 | 37.27 |
| $Y = 136.36 \cdot Hcell/(Ext \cdot FE)$ | $1.19 \cdot 10^{-15}$ | 0.5521 | 47.09 | 0.5821 | 38.63 |
| $Y = 3.8469 \cdot Cell^2/(Lig \cdot FE)$ | $7.23 \cdot 10^{-16}$ | 0.5591 | 46.72 | 0.5300 | 40.97 |
| $Y = 19.350 + 6374.6 \cdot Cell/(Lig^2 \cdot TAN \cdot FE)$ | $3.23 \cdot 10^{-14}$ | 0.5847 | 45.34 | 0.6369 | 36.01 |
| $Y = 18.814 + 281.53 \cdot Hcell^2/(Lig \cdot TAN \cdot FE \cdot Size)$ | $1.65 \cdot 10^{-14}$ | 0.5930 | 44.89 | 0.6123 | 37.20 |
| $Y = 17.131 + 122.20 \cdot Cell \cdot Hcell \cdot Alk/(VS \cdot Lig \cdot TAN \cdot FE)$ | $8.53 \cdot 10^{-15}$ | 0.6009 | 44.45 | 0.6109 | 37.27 |

data (Table IV). For these regressions including only F/E or its inverse, the model containing the inverse presented better results for both regression and validation. Such a result highlights the gain in accuracy which can be obtained as the variables' inverses are included to regression models.

Including more variables, the model which most accurately adjusted to regression data was a seventh-order model, which presented a regression $R^2$ of 0.6009 and an *SEP* of 44.45. For validation, these values were 0.6109 and 37.27 respectively (Table IV). This model includes cellulose, hemicellulose, alkalinity, volatile solids,

lignin, total ammonia nitrogen and feedstock-to-effluent ratio. The model which most accurately adjusted to validation data was a fifth-order model, capable of predicting 63.69% of the results, with a *SEP* of 36.01 (Table IV). This model included as variables: cellulose, lignin, total ammonia nitrogen and feedstock-to-effluent ratio. Comparison plots between experimental and predicted data for both regression and validation for these two highlighted models are given in Fig. 2. Results for these mentioned models are considered to be relevant, since a simple linear model was employed. However, an improvement of the result is expected as multiple linear regressions are developed.

In Table V, results for multiple linear regression models are presented, and Fig. 3 exhibits plots comparing experimental and predicted yields for

regression and validation of these models. The first-order MLR model is capable of describing 71% of experimental variability – with a *SEP* of 32.05 – and includes as explanatory variables: feedstock extractives, inoculum alkalinity and feedstock-to-effluent ratio. This result is already more accurate than the results obtained for simple linear regression. In order to improve the accuracy, higher-order models were tested. A fourth-order model presents an $R^2$ value of 0.8329 and an *SEP* of 24.42, and this result is considered as quite satisfactory, due to experimental methane yield broad range (0 – 237.12). The smallest *SEP* found by Xu et al. (2014) by multiple linear regression was 10.07, however, a range of 1-157.3 for yield was used. Also, the study developed by Xu et al. (2014) employed 40 data points for regression
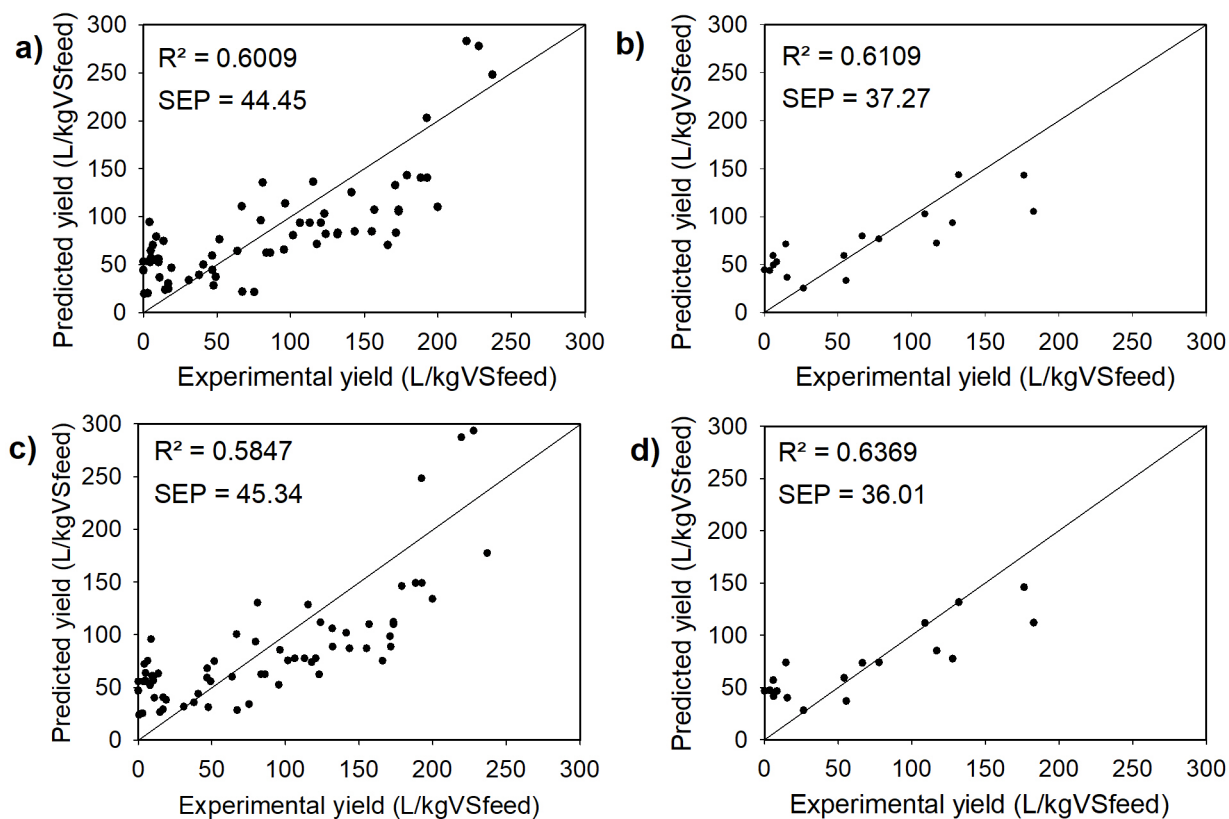


**Figure 2** - Comparison between experimental and predicted yield for (**a**) regression and (**b**) validation of the simple linear model with the highest regression $R^2$, and for (**c**) regression and (**d**) validation for the model with the highest validation $R^2$. *SEP*: Standard error of prediction.

**TABLE V**
**Summary of models obtained by multiple linear regression.**

| *First order* |
|---|
| $Y = -778.07 - 3.27107\,Ext + 31.541\,Alk - 26.869\,FE + 7174.9/Alk + 42.551/FE$ |
| Significance F: $5.38 \cdot 10^{-11}$ |
| $R^2$ model: 0.6056 |
| SEP model: 44.19 |
| $R^2$ validation: 0.7124 |
| SEP validation: 32.05 |
| *Second order* |
| $Y = -402.72 + 10.280\,Cell/FE - 36.402/FE^2 + 12.523\,Hcell/TAN + 8931.8/(Cell \cdot Size) - 611.61\,Alk/TS + 1.0404\,Alk^2 + 9547.7/TS$ |
| Significance F: $2.87 \cdot 10^{-15}$ |
| $R^2$ model: 0.7395 |
| SEP model: 35.91 |
| $R^2$ validation: 0.8100 |
| SEP validation: 26.04 |
| *Third order* |
| $Y = -92.922 - 3452.2/FE - 34.099/FE^2 - 16.785*Ext + 6825.6/Cell + 22.999 \cdot Hcell + 337580/(VS \cdot FE) - 313.31\,Hcell \cdot TAN/VS - 806.84\,FE/(Ext \cdot TAN)$ |
| Significance F: $1.65 \cdot 10^{-19}$ |
| $R^2$ model: 0.8265 |
| SEP model: 29.31 |
| $R^2$ validation: 0.8303 |
| SEP validation: 24.62 |
| *Fourth order* |
| $Y = -3699.0/FE + 0.0080732\,Ext^2 \cdot Lig - 2.0551\,Ext^2 \cdot Cell/VS - 36.940/FE^2 + 21.454\,Hcell + 361590/(VS \cdot FE) - 289.43\,Hcell \cdot TAN/VS - 764.84\,FE/(Ext \cdot TAN)$ |
| Significance F: $1.52 \cdot 10^{-20}$ |
| $R^2$ model: 0.8310 |
| SEP model: 28.93 |
| $R^2$ validation: 0.8329 |
| SEP validation: 24.42 |

and 9 for validation, and the present work used 68 and 18 data points for regression and validation, respectively. Therefore, Xu et al. (2014) obtained lower *SEP* values, however, this present study employed more data points and a wider methane yield range, being more generalist.

Inclusion of terms involving interactions among input variables – cross terms – to linear regression models was necessary to improve prediction performance, possibly due to the complexity of SS-AD and the significant correlation among the process' exploratory variables. Feedstock-to-effluent ratio was found to be essential in methane prediction, since it composes all regression models presented in this study. Besides, cellulose, hemicellulose and extractives were recurrent, in particular for models which presented a significant prediction quality, i. e. validation $R^2$ greater than 0.80.

Xu et al. (2014) evaluated simple and multiple linear regression models for a data basis similar to the one used in this study. The authors highlighted the relevance of F/E ratio, as well as lignin, cellulose and extractives contained in the feedstock, to
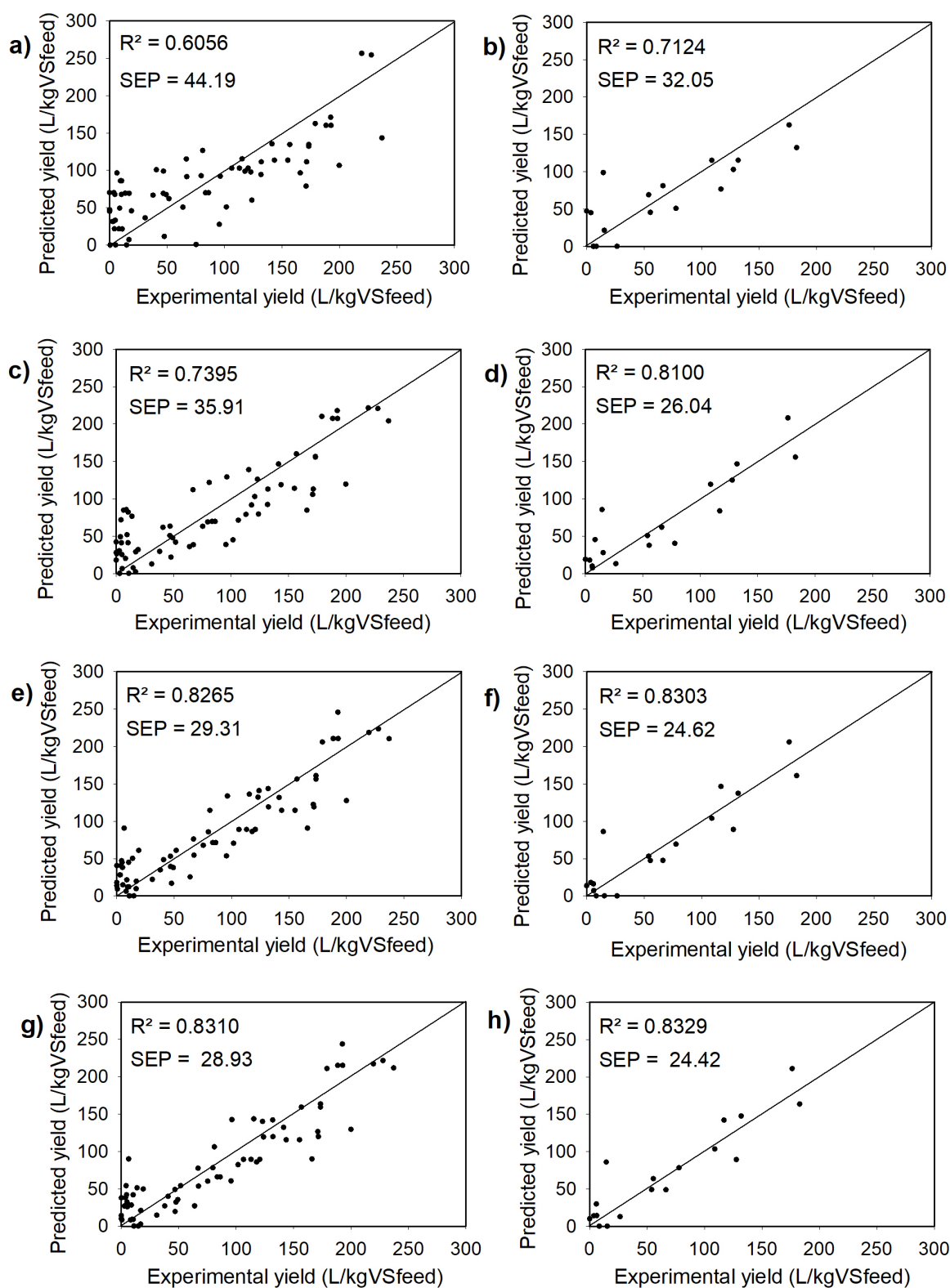
**Figure 3** - Comparison between experimental and predicted data for regression and validation, for (**a**), (**b**) first-order, (**c**), (**d**) second-order, (**e**), (**f**) third-order and (**g**), (**h**) fourth-order models. *SEP*: Standard error of prediction.

accurately estimate the cumulative 30-day methane yield by using multiple linear regression models. In addition, the interaction between lignin and F/E and the presence of quadratic and cubic terms involving F/E were crucial for prediction quality.

Although C/N ratio presented a significant correlation with methane yield, as demonstrated in the correlation matrix (Table III), this variable was not verified in any developed model with satisfactory predictive capacity.

Results presented in this study corroborate with conclusions achieved by Xu et al. (2014) about the relevance of terms which contemplate non-linear correlation between the exploratory variables and methane yield, as well as interaction among two or more variables as a single term in the linear models, due to the complex nature of the SS-AD process.

Differently from published papers in literature, variables which did not present a significant linear correlation to methane yield (Table III) were included in tested models, and a positive effect of such an inclusion was observed in the polynomials predictive quality. These results confirm the initial hypothesis, which is the nontriviality in the exploratory variables selection to describe methane production in linear regression models, due to the complexity of SS-AD phenomenon. Other highlight is the achievement of models constructed by means of simple and multiple linear regression, involving terms containing multiple combinations among variables and its inverses, improving the predictive quality as the polynomial order was increased.

## CONCLUSIONS

Simple and multiple linear models were developed to predict solid-state anaerobic digestion methane yield. Predictions models involved explanatory variables, its inverses and combinations, and the regression accuracy was improved for higher-order models. The input variables considered as

relevant were: feedstock-to-effluent ratio, amounts of cellulose, hemicellulose and extractives contained in the feedstock and total ammonia nitrogen concentration in the inoculum. Among those variables, F/E ratio was essential to methane yield prediction, being identified in all satisfactory models. The developed linear models might be used as references to propose experiments and to evaluate a biomass source potential to produce methane through SS-AD. In this manner, it is possible to reduce costs and time required to perform laboratorial studies.

## AUTHOR CONTRIBUTIONS

Gabriela Rodrigues Niquini and Suzimara Reis da Silva participated together in literature review, data collection and organization, methodology proposal, development of models, discussion of results and writing. They also worked on manuscript submission and reviewing. Esly Ferreira da Costa Junior worked actively on the development of models and review of the manuscript. Andréa Oliveira Souza da Costa made the supervision of the project.

## REFERENCES

ANGELIDAKI I, ALVES M, BOLZONELLA D, CAMPOS JL, GUWY AJ, KALYUZHNYI S, JENICEK P AND VAN LIER JB. 2009. Defining the biomethane potential (BMP) of solid organic wastes and energy crops: a proposed protocol for batch assays. Water Sci Technol 59: 927-934.

BAÊTA BEL, LIMA DRS, FILHO JGB, ADARME OFH, GURGEL LVA AND AQUINO SF. 2016. Evaluation of hydrogen and methane production from sugarcane bagasse hemicellulose hydrolysates by two-stage anaerobic digestion process. Bioresour Technol 218: 436-446.

BROWN D, SHI J AND LI Y. 2012. Comparison of solid-state to liquid anaerobic digestion of lignocellulosic feedstocks for biogas production. Bioresour Technol 124: 379-386.

CHEROSKY P. 2012. Anaerobic Digestion of Yard Waste and Biogas Purification by Removal of Hydrogen Sulfide. PhD Thesis, The Ohio State University, Department of Food, Agricultural and Biological Engineering, 103 p.

CUI Z, SHI J AND LI Y. 2011. Solid-state anaerobic digestion of spent wheat straw from horse stall. Bioresour Technol 102: 9432-9437.

EDWIGES T, FRARE L, MAYER B, LINS L, MI TRIOLO J, FLOTATS X AND DE MENDONÇA COSTA MSS. 2018. Influence of chemical composition on biochemical methane potential of fruit and vegetable waste. Waste Manag 71: 618-625.

EL ACHKAR JH, LENDORMI T, HOBAIKA Z, SALAMEH D, LOUKA N, MAROUN RG AND LANOISELLÉ JL. 2017. Anaerobic digestion of nine varieties of grape pomace: Correlation between biochemical composition and methane production. Biomass Bioener 107: 335-344.

EL ACHKAR JH, LENDORMI T, SALAMEH D, LOUKA N, MAROUN RG, LANOISELLÉ JL AND HOBAIKA Z. 2018. Influence of pretreatment conditions on lignocellulosic fractions and methane production from grape pomace. Bioresour Technol 247: 881-889.

GE X, XU F AND LI Y. 2016. Solid-state anaerobic digestion of lignocellulosic biomass: Recent progress and perspectives. Bioresour Technol 205: 239-249.

GUNASEELAN VN. 2007. Regression models of ultimate methane yields of fruits and vegetable solid wastes, sorghum and napiergrass on chemical composition. Bioresour Technol 98: 1270-1277.

GUNASEELAN VN. 2009. Predicting ultimate methane yields of *Jatropha curcus* and *Morus indica* from their chemical composition. Bioresour Technol 100: 3426-3429.

KAFLE GK AND CHEN L. 2016. Comparison on batch anaerobic digestion of five different livestock manures and prediction of biochemical methane potential (BMP) using different statistical models. Waste Manag 48: 492-502.

LESTEUR M, BELLON-MAUREL V, GONZALEZ C, LATRILLE E, ROGER JM, JUNQUA G AND STEYER JP. 2010. Alternative methods for determining anaerobic biodegradability: A review. Process Biochem 45: 431-440.

LI Y, PARK SY AND ZHU J. 2011a. Solid-state anaerobic digestion for methane production from organic waste. Renew Sust Energ Rev 15: 821-826.

LI Y, ZHU J, WAN C AND PARK SY. 2011b. Solid-state anaerobic digestion of corn stover for biogas production. Trans ASABE 54: 1415-1421.

LIEW LN, SHI J AND LI Y. 2011. Enhancing the solid-state anaerobic digestion of fallen leaves through simultaneous alkaline treatment. Bioresour Technol 102: 8828-8834.

LIEW LN, SHI J AND LI Y. 2012. Methane production from solid-state anaerobic digestion of lignocellulosic biomass. Biomass Bioener 46: 125-132.

PEZZOLLA D, DI MARIA F, ZADRA C, MASSACCESI L, SORDI A AND GIGLIOTTI G. 2017. Optimization of solid-state anaerobic digestion through the percolate recirculation. Biomass Bioener 96: 112-118.

SHEETS JP, GE X AND LI Y. 2015. Effect of limited air exposure and comparative performance between thermophilic and mesophilic solid-state anaerobic digestion of switchgrass. Bioresour Technol 180: 296-303.

SHI J, WANG Z, STIVERSON JA, YU Z AND LI Y. 2013. Reactor performance and microbial community dynamics during solid-state anaerobic digestion of corn stover at mesophilic and thermophilic conditions. Bioresour Technol 136: 574-581.

SHI XS, YUAN XZ, WANG YP, ZENG SJ, QIU YL, GUO RB AND WANG LS. 2014. Modeling of the methane production and pH value during the anaerobic co-digestion of dairy manure and spent mushroom substrate. Chem Eng J 244: 258-263.

THOMSEN ST, SPLIID H AND ØSTERGÅRD H. 2014. Statistical prediction of biomethane potentials based on the composition of lignocellulosic biomass. Bioresour Technol 154: 80-86.

XU F AND LI Y. 2012. Solid-state co-digestion of expired dog food and corn stover for methane production. Bioresour Technol 118: 219-226.

XU F, LI Y AND WANG Z-W. 2015. Mathematical modeling of solid-state anaerobic digestion. Prog Energy Combust Sci 51: 49-66.

XU F, SHI J, LV W, YU Z AND LI Y. 2013. Comparison of different liquid anaerobic digestion effluents as inocula and nitrogen sources for solid-state batch anaerobic digestion of corn stover. Waste Manag 33: 26-32.

XU F, WANG F, LIN L AND LI Y. 2016. Comparison of digestate from solid anaerobic digesters and dewatered effluent from liquid anaerobic digesters as inocula for solid state anaerobic digestion of yard trimmings. Bioresour Technol 200: 753-760.

XU F, WANG Z-W AND LI Y. 2014. Predicting the methane yield of lignocellulosic biomass in mesophilic solid-state anaerobic digestion based on feedstock characteristics and process parameters. Bioresour Technol 173: 168-176.

ZHAO J, ZHENG Y AND LI Y. 2014. Fungal pretreatment of yard trimmings for enhancement of methane yield from solid-state anaerobic digestion. Bioresour Technol 156: 176-181.

ZHU J, WAN C AND LI Y. 2010. Enhanced solid-state anaerobic digestion of corn stover by alkaline pretreatment. Bioresour Technol 101: 7523-7528.