Bioresource Technology 154 (2014) 80–86

Contents lists available at ScienceDirect

Bioresource Technology

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

Statistical prediction of biomethane potentials based on the composition of lignocellulosic biomass

Sune Tjalfe Thomsen [a], Henrik Spliid [b], Hanne Østergård [a,*]

[a] Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark DTU, DK-2800 Kgs. Lyngby, Denmark

[b] Section of Statistics and Data Analysis, Department of Applied Mathematics and Computer Science, Technical University of Denmark DTU, DK-2800 Kgs. Lyngby, Denmark

HIGHLIGHTS

- A statistical model for predicting *BMP* from lignocellulosic material is developed.
- The true effect of lignin and carbohydrates on *BMP* is described.
- The best prediction is proposed using a canonical linear mixture model.
- An expression for prediction founded on the largest dataset to date, is presented.

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 3 December 2013
Accepted 7 December 2013
Available online 13 December 2013

Keywords:
Biomethane potential (*BMP*)
Mixture model
Lignocellulose
Biogas
Anaerobic digestion (AD)

ABSTRACT

Mixture models are introduced as a new and stronger methodology for statistical prediction of biomethane potentials (BPM) from lignocellulosic biomass compared to the linear regression models previously used. A large dataset from literature combined with our own data were analysed using canonical linear and quadratic mixture models. The full model to predict *BMP* ($R^2 > 0.96$), including the four biomass components cellulose (x_C), hemicellulose (x_H), lignin (x_L) and residuals ($x_R = 1 - x_C - x_H - x_L$) had highly significant regression coefficients. It was possible to reduce the model without substantially affecting the quality of the prediction, as the regression coefficients for x_C, x_H and x_R were not significantly different based on the dataset. The model was extended with an effect of different methods of analysing the biomass constituents content (D_A) which had a significant impact. In conclusion, the best prediction of *BMP* is $pBMP = 347x_{C+H+R} - 438x_L + 63D_A$.

1. Introduction

Biomethane potential (*BMP*) measurements are very time-consuming, as up to 90 days are required as a standard incubation time (Hansen et al., 2004; Gerber et al., 2013; Angelidaki et al., 2009). Therefore, it is attractive to use faster methods when estimating how much methane gas it is possible to produce from a given biomass. This is especially the case when making theoretical studies without access to laboratory facilities, or when a fast prediction of *BMP* from a new biomass is required.

Theoretical methods of predicting *BMP* (*pBMP*) have been available since 1933 when Symons and Buswell made their theoretical and laboratory studies of anaerobic digestion of carbohydrates where they presented what later would be known as Buswell's formula (Symons and Buswell, 1933). This formula expresses the maximum output of methane gas in a complete anaerobic digestion of organic matter, and is calculated from the chemical sum formula of the organic material:

$$C_nH_aO_b + \left(n - \frac{a}{4} - \frac{b}{2}\right)H_2O \rightarrow \left(\frac{n}{2} + \frac{a}{8} - \frac{b}{4}\right)CH_4 + \left(\frac{n}{2} - \frac{a}{8} + \frac{b}{4}\right)CO_2$$

Even though Buswell's formula were designed for estimating the ultimate *BMP* from a biomass based on the sum formula, it can also be used on each of the biomass constituents. This means, that the formula can determine the theoretical *BMP* on cellulose (x_C), hemicellulose (x_H), protein, lipids, etc. of biomass, if compositional data are available. In these cases, it is also possible to exclude a contribution from non-convertible biomass constituents such as lignin (x_L) and ash.

Even though *BMP* can be predicted with Buswell's formula, one important factor is not taken into account, namely the recalcitrance of the biomass in question. When dealing with pure substrates, such as sugars or lipids recalcitrance is not important to include. However, when dealing with e.g., lignocellulosic substrates, the shielding effect of the lignocellulosic matrix will decrease the *BMP* (Azhar and Stuckey, 1994). The extent to which

* Corresponding author. Address: DTU Risø Campus, Frederiksborgvej 399, P.O. Box 49, Building 301, 4000 Roskilde, Denmark. Tel.: +45 21326955.

E-mail address: haqs@kt.dtu.dk (H. Østergård).

0960-8524/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biortech.2013.12.029

such an effect takes place tends to be correlated with the compositions of the biomass (Labatut et al., 2011).

Often chemical oxygen demand (COD) is used to estimate *BMP*, however this method suffers from some of the same inconsistencies as Buswell's formula. When measuring COD a total oxidation of organic material is made, and therefore neither biomass recalcitrance nor the contribution of non-convertible lignin is taken into account, forcing the COD method to over-estimate the *pBMP*.

Determining *pBMP*s through regression models is a relatively new methodology initiated within the last decade (Monlau et al., 2012; Triolo et al., 2011; Gunaseelan, 2007, 2009). The current study focuses only on determining *pBMP* from lignocellulosic biomasses. This has only been addressed in a few previous studies which are presented in Table 1.

As seen in Table 1, the previously proposed regression models assume that the lignin content is the single most import biomass constituent, when predicting *BMP* (Triolo et al., 2011; Monlau et al., 2012). This is paradoxical since lignin does not contribute to the formation of methane in the anaerobic digestion (AD) process, but rather is acting as the glue that ties the lignocellulosic matrix together while making a physical barrier around the carbohydrates (Albersheim et al., 2011). In that way, the regression models proposed so far have been contradictory to AD theory, since the content of degradable biomass constituents such as cellulose and hemicellulose is not accounted for in the models. This implies that a biomass with low content of lignin will give rise to a high *pBMP* regardless of the carbohydrate content. Furthermore, Triolo et al. (2011) found that when including both x_L and x_C as regression variables, x_C contributed negatively to the model (Table 1, row 3). This is also contrary to AD theory, since carbohydrates are the main substrates in the AD and, therefore, should imply a positive regression coefficient.

The possible misinterpretations in the previous prediction models may reflect the relative nature of the compositional data. Biomass composition is most often presented as % of total solids (TS), % of volatile solids (VS) or w/w%. This results in a constraint on the data, since the components add up to 100%. Normally, regression coefficients are interpreted as the change in the dependent variable due to a unit change in the independent variable while keeping everything else constant, but with compositional data it is not possible to change one proportion while keeping the others constant. Due to this constraint, the space in which each component can be varied is obviously strongly restricted, which previously has not been addressed in relation to *BMP*. Similar issues have been taken into account elsewhere, especially for chemical mixtures where compositional data also are predominant. Here, a wide range of regression models, known as mixture models, have been developed (Cornell, 2011; Prakasham et al., 2009; Scheffe, 1963). It might be advantageous to view the compositional data as a chemical mixture, thus investigating the effect of the different biomass constituents on *pBMP* in a mixture model.

In mixture models, the variables are proportionate nonnegative amounts of different constituents, $0 \leqslant x_i \leqslant 1$, $i = 1,2,\ldots,q$ where $\sum_{i=1}^{q} x_i = 1$. In our case, the variables are the main biomass constituents of lignocellulosic biomass: Cellulose (x_C), hemicellulose (x_H), and lignin (x_L). Since the variables sum up to one, an additional variable (x_R) which is often called 'residuals' in relation to biomass composition, is included in the model. In this way everything which is not carbohydrates or lignin is characterised as residuals, $x_R = 1 - (x_C + x_H + x_L)$. Introducing residuals is not new to the area of determining biomass composition (Sluiter et al., 2010; Thomsen et al., 2012). However, x_R has not been considered in previous models as a regression variable (Table 1), which might be problematic, since x_R might contain methane yielding biomass constituents such as lipids, fatty acids, pectin, proteins and tannins.

In the present study, both a canonical linear mixture model, $pBMP = \sum_{i=C,H,L,R} \beta_i x_i$, as well as a canonical quadratic mixture model, $pBMP = \sum_i \beta_i x_i + \sum\sum_{i<j} \beta_{ij} x_i x_j$ (where indices i and j refer to the components C, H, L and R) will be investigated. In this way, models for predicting *BMP*, which are in accordance to AD theory, will be developed. The regression coefficients will be estimated from a large dataset from literature combined with data prepared for this study.

2. Methods

2.1. BMP test performed for this study

Biomasses tested for *BMP* for this study were cassava stalks, cocoa pods, groundnut straw, lucerne cake, maize cobs, maize stalks, oil palm empty fruit bunches (oil palm EFB), plantain leaves, plantain trunks, rice straw, vetch hay and rye straw mixed, rye straw and vetch hay. For the determination of biogas potentials prepared for this study, triplicate-samples of all biomasses were distributed in 1 l serum flasks (effective volume 1125 ml) in amounts of 1 g volatile solids (VS) per 100 ml active volume. The samples were inoculated with 150 ml of effluent from a lab-scale biogas reactor treating cattle manure and water was added to a total active volume of 300 ml. For subtraction of biogas produced by the inoculum, flasks containing only inoculum and water were also prepared. The flasks were sealed with rubber septum and metal screw plugs and the samples were incubated at 55 °C for a period of 50 days, hereafter, no more gas production was observed. The CH_4 production in the flasks was measured by collecting 0.5 ml of headspace gas using a gas tight syringe and analysing the CH_4 concentration in the sample by gas chromatography (HP 6890; Agilent). Measurements were carried out in increasing intervals ranging from 2 days in the beginning to 8 days in the end of the digestion trials. Biomass composition has been assessed previously (Thomsen et al., 2012; Carter et al., 2012).

2.2. Literature search and selection

In order to find relevant data for determining the best possible regression model, we aimed to construct as large a dataset as possible. The literature search was done with a systematic approach where all combinations of two lists of search criteria (Table 2) were applied. The search engine Scopus was used until April 5 2013 and

Table 1
Previously presented regression models for determining *pBMP* from composition of lignocellulosic biomass.

Regression model	Prediction model[a]	Reference	Biomass used (number of samples used generating the equation)	R^2
$\alpha_0 + \alpha_L x_L$	$pBMP = 461 - 258\, x_L$	Triolo et al. (2011)	Energy crops ($n = 10$)	0.76
$\alpha_0 + \alpha_L x_L$	$pBMP = 380 - 65\, x_L$	Monlau et al. (2012)	Raw and pretreated sunflower stalks ($n = 8$)	0.92
$\alpha_0 + \alpha_L x_L + \alpha_C x_C$	$pBMP = 447 - 277\, x_L - 7\, x_C$	Triolo et al. (2011)	Energy crops ($n = 10$)	0.77
Buswell on carbohydrates	$pBMP = 414\, x_C + 423\, x_H$	Symons and Buswell (1933)	Theoretical model	–

[a] The regression coefficients have been transformed to the unit of the variables used in this study which is w/w instead of w/w% used in the references. x_L is the lignin content, x_C is the cellulose content and x_H is hemicellulose.