Industrial Crops & Products 120 (2018) 330–342

Contents lists available at ScienceDirect

Industrial Crops & Products

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

Optimization of specific methane yield prediction models for biogas crops based on lignocellulosic components using non-linear and crop-specific configurations

Moritz von Cossel[a,*,1], Jens Möhring[b,1], Andreas Kiesel[a], Iris Lewandowski[a]

[a] University of Hohenheim, Institute of Crop Science, Biobased Products and Energy Crops, Fruwirthstr. 23, 70599, Stuttgart, Germany
[b] University of Hohenheim, Institute of Crop Science, Biostatistics, Fruwirthstr. 23, 70599, Stuttgart, Germany

ARTICLE INFO

Keywords:
Anaerobic digestion
Biogas yield
Crop biomass
Lignocellulose
Meta-analysis
Prediction models

ABSTRACT

Basing the prediction of specific methane yield (SMY) of crop biomass on lignocellulosic components has become a promising tool in biogas plant management and bioenergy policies. Most studies on SMY prediction provide linear or non-linear models across crops with lignin content as major regressor variable. To determine the effect of crop-specific regressions, a meta-analysis was conducted using data from 14 published studies (518 observations) and three of the authors' own experiments (160 observations). In total, 678 observations of biomass components and SMY from 13 potential biogas crops were included. The data were used to validate seven published models and both develop and cross-validate new linear and non-linear models with and without crop-specific regressions. Available models showed correlations between r = 0.12 and 0.51. New models reached correlations of up to r = 0.66. Both crop-specific intercepts and slopes as well as non-linear regressions significantly increased model predictability. Of these, crop-specific intercepts brought about the largest improvement but still allowed easy use and interpretability. Therefore, it was shown that the use of biomass source information can help optimize the precision of SMY prediction.

1. Introduction

In a growing bioeconomyoptimizing cascading approaches for the conversion of lignocellulosic plant resources for bioenergy (Belgacem and Gandini, 2008) such as lignocellulosic biomass fractionation depolymerization and valorization (Bjelić et al., 2018; Gallezot, 2012; Grilc and Likozar, 2017; Grilc et al., 2014, 2016), will become essential to meet future demand for both biobased products and biomass-based renewable energies (Zörb et al., 2018). In this context, biogas production from anaerobic co-digestion of plant biomass (Weiland, 2010) is an important strategy both to provide renewable energy and to induce multiple benefits for farms under aspects of sustainable intensification (Blumenstein et al., 2018).

A recent focus of research in this field has turned towards the investigation of quick and reliable evaluation methods (including prediction models) to determine the substrate quality of plant biomass for biogas processing (Bekiaris et al., 2015; Godin et al., 2015; Raju et al., 2011; Triolo et al., 2014; Ward, 2016). These methods can support the selection of both new biogas crops and their combination within novel crop rotation systems, which are required for a more environmental benign biogas production in the long term (Dauber et al., 2010; Glemnitz and Brauckmann, 2016). Furthermore, rapid quality estimation enables biogas plant owners to optimize the substrate production and purchase, since the substrate quality of plant biomass for biogas processing widely varies with both plant species and management practices (Amon et al., 2007a; Weiland, 2010). This biogas substrate quality can be represented by either the substrate-specific methane yield (SMY) or the substrate-specific biogas yield (SBY), both expressed in norm litre per kg of volatile solids (l_N kg^{-1}VS) (Dandikas et al., 2014). SBY includes yield-relevant components like methane and yield-irrelevant components like carbon dioxide. There is a strong correlation between SMY and SBY (r > 0.98; Dandikas et al., 2014; Rath et al., 2013), but SMY is the more important parameter as methane represents the yield-relevant proportion of biogas. Optimized models for SMY

Abbreviations: ADF, Acid detergent fibre; ADL, Acid detergent lignin; CL, Cellulose; CV, Cross validation; HC, Hemicellulose; MLR, Multiple linear regression; NDF, Neutral detergent fibre; RES, Non-lignocellulosic biomass fraction; rMSE, square root of the mean square error; SBY, specific biogas yield; SMY, specific methane yield; VS, Volatile solids; WCCS, Whole crop cereal silage

* Corresponding author.
E-mail addresses: mvcossel@gmx.de (M. von Cossel), moehring@uni-hohenheim.de (J. Möhring).
[1] First authors.

https://doi.org/10.1016/j.indcrop.2018.04.042
Received 9 September 2017; Received in revised form 12 April 2018; Accepted 15 April 2018

prediction could improve and support (i) the estimation of the bioenergy potential of crops or residues for different regions or climatic scenarios (Niu et al., 2016), and (ii) the breeding of new plant varieties for biogas purposes (Herrmann and Rath, 2012). For these purposes, three main criteria have to be met. Models need to be (i) accurate enough to ensure a satisfactory relevance of the predictions, (ii) suitable for a large range of different types of plant biomasses and (iii) simple enough to enable its implementation in practice. In this respect, an SMY prediction model not only needs to be as precise as possible but also time- and cost-efficient. Therefore, variables used for the prediction need to be measured in an easy, fast and cheap way (Godin et al., 2015). These restrictions limit the number of available variables and consequently the potential maximum accuracy of prediction models since both the total costs and duration of measurements increase with the number of variables.

Most SMY prediction models already published are either completely based on lignocellulosic components or include them as important variables (Alaru et al., 2011; Amon et al., 2007b; Dandikas et al., 2014, 2015; Godin et al., 2015; Gunaseelan, 2007; Herrmann et al., 2016; Monlau et al., 2012; Rath et al., 2013; Thomsen et al.2014; Triolo et al., 2011, 2012). This is because (i) the determination of lignocellulosic components is a cheaper and faster approach than biogas batch experiments and (ii) the biogas yield is often correlated with fibre components such as lignin and cellulose (Dandikas et al., 2014; Kiesel et al., 2017; Thomsen et al.2014; Triolo et al., 2011). Therefore, this paper also focusses on lignocellulosic-component-based prediction models. In addition to chemical analysis, the lignocellulosic components can be predicted using near infrared spectroscopy (NIRS). Direct prediction of SMY using NIRS data has also recently been reported (Godin et al., 2015; Triolo et al., 2014). Both prediction models with variables from lignocellulosic components or with NIRS data are applied to either diverse biomasses from a single crop or to biomass from different crops. Most studies include several crops and show models used for prediction of SMY across crops (Alaru et al., 2011; Dandikas et al., 2014; Herrmann et al., 2016; Thomsen et al., 2014; Triolo et al., 2011). However, some studies are exclusively on maize (Amon et al., 2007b; Godin et al., 2015; Rath et al., 2013). Models developed across several crop species show promising accuracies reflected by coefficients of determination (R^2) ranging from 0.40 (Alaru et al., 2011) to 0.96 (Thomsen et al., 2014) for prediction from lignocellulosic composition and 0.8 for NIRS-based SMY prediction (Godin et al., 2015; Triolo et al., 2014). The lignocellulosic-composition-based prediction models published so far were estimated with relatively small number of observations ($N < 70$) (Alaru et al., 2011; Dandikas et al., 2014, 2015; Thomsen et al., 2014; Triolo et al., 2011) and often lack proper validation. In all these models, the content of acid detergent lignin (ADL) is the most influential component since it is not digestible and also inhibits the degradation of a certain amount of cellulose within the cell walls (Triolo et al., 2011). For this reason, ADL is used in prediction models either solely (Dandikas et al., 2014; Triolo et al., 2011) or in combination with other components including hemicellulose (HC), cellulose (CL), crude fat, raw protein and reducing sugars (Dandikas et al., 2014; Rath et al., 2013; Triolo et al., 2011). Simple or multiple linear regressions were used to find promising prediction models. Non-linear or crop-specific extensions of the multiple linear regression (MLR) approach have already been suggested (Dandikas et al., 2014; Thomsen et al., 2014), but rarely performed. The benefits of these extended models are unclear. Thomsen et al. (2014) found no significance of non-linear two-way interaction terms (e.g. ADL × HC) excluding squared terms (e.g. ADL × ADL) given that main effects are included in the model. Godin et al. (2015) however found increasing predictability by fitting models with non-linear twofold interaction terms (including squared terms) when using NIRS data. They also distinguished between observations from green-dried, silage-dried and wet plant biomass, and concluded that the physical condition of the biomass does not influence predictability (Godin et al., 2015). In contrast, Triolo et al. (2011) compared different kinds of biomasses (plant biomass and animal manure) and found that they differ in their SMY. Dandikas et al. (2015) found that a crop-specific prediction model based on two grassland species outperforms prediction from an across-crop model (Dandikas et al., 2014). Neither of these extensions to SMY prediction models (adding non-linear terms and crop-specific terms) has been systematically compared for different plant biomass crops.

The overall objective of this study is to assess whether SMY can be sufficiently well predicted based on biomass composition using a large dataset. The sub-objectives are to (i) create a dataset as large as possible combining at best all data from studies of lignocellulosic-based SMY predictions, (ii) use this dataset to validate existing SMY prediction models, (iii) select and cross-validate alternative linear and non-linear crop-specific regression models, and (iv) compare the suitability of these approaches for SMY prediction.

2. Materials and methods

This section starts with a description of the dataset used for meta-analysis including observations for a range of different countries (e.g. Belgium, Denmark, Estonia) taken from both external (literature) and internal (authors' own experiments) data. Details of the MLR approach and the model validation are then described. Finally two possible parameterizations of the same full model are compared using MLR approach.

2.1. External data: literature search and selection of relevant data

A literature search was conducted using Scopus (Elsevier, The Netherlands). First, articles were searched for using specific search terms. Then relevant articles were selected considering four criteria. The literature search conducted from August until mid-September 2016 was performed using the terms 'biogas', 'chemical composition' or 'lignin'. The search field was set to 'all fields'. About 1600 articles were found. These articles were limited as follows

- the article should provide data on both lignocellulosic components and SMY,
- the SMY should be measured using dry matter plant material (not fresh or ensilaged material),
- the fibre fractions should be analyzed according to VDLUFA, 6.5.3 (Van Soest and Wine, 1967), and
- the biomass should not be highly lignified (lignin content < 10% of dry matter).

This resulted in a set of identified articles. Further relevant articles were found in the reference lists of the identified articles. In addition to peer-reviewed literature dissertations and official reports of governmental institutions found via Google Scholar (Google California U.S.) (written in English or German) were taken into account (Eberl et al. 2014; Kaiser, 2007; Zeise and Fritz, 2012; Zeise et al., 2016). Data from articles were subdivided into studies e.g. if they were published by different authors. In total, 518 observations on 11 crops taken from 14 articles (Table 1) were included. Not all crops from all articles were included, due to the selection criteria mentioned above. Details on the volume of crop-specific data are shown in Table 1.

2.2. Internal data

Internal data were taken from three studies conducted at two locations in southwest Germany. The plant material (samples) for chemical analysis was taken from field trials conducted in Renningen (Latitude: 48.7404 DD; Longitude: 8.9219 DD; Altitude: 486 m) and Hohenheim (Latitude: 48.71504 DD; Longitude: 9.2113 DD; Altitude: 400 m), both located in southwest Germany. Both geographic and long-term climatic site conditions have been previously described by Von