# Determinants of efficiency in anaerobic bio-waste co-digestion facilities: a two-stage DEA and stochastic gradient boosting classifier approach

Djavan De Clercq[a,b], Zongguo Wen[a,b,*]

a. State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, 100084 China
b. Key Laboratory for Solid Waste Management and Environment Safety (Tsinghua University), Ministry of Education of China, Tsinghua University, Beijing 100084, China

* Corresponding author: wenzg@tsinghua.edu.cn

## Abstract

The objective of this study is to explore the determinants of efficiency in industrial scale co-digestion facilities. The methodology applied to finding these determinants involved a two-stage process. First, Data Envelopment Analysis was applied to input-output data from 386 operational anaerobic digestion facilities across Germany and the USA to compute project efficiency scores. These efficiency scores were converted into a binary target variable based on the 75$^{th}$ percentile threshold of efficiency scores. Secondly, a tuned and cross-validated stochastic gradient boosted classification model was built to determine the effect of technical parameters on a project meeting this threshold or not. The results were as follows. Firstly, 75$^{th}$ percentile technical efficiencies for German and US projects were 0.82 and 0.46 respectively. Secondly, the stochastic gradient boosting model had classification accuracies of 0.77 and 0.75 for German and US projects respectively, indicating high suitability for isolating determinants of efficiency. Thirdly, the most influential features included pre-treatment stages and pre-treatment types, digester technology type, and the presence/absence of co-digestion, indicating that decision makers should focus on these technologies. This research has important implications for industrial-scale biogas facilities seeking to enhance operational efficiency and policymakers engaged in technology promotion related to anaerobic digestion.

## Highlights

- DEA used to compute efficiency scores in anaerobic co-digestion facilities in the US and Germany.
- Stochastic gradient boosting used to model determinants of efficiency in these facilities.
- The hyperparameters of the model were tuned and the model was cross validated for high accuracy.
- Pre-treatment, digester technology, and presence/absence of co-digestion were important features.

## Keywords

- Data Envelopment Analysis; Stochastic Gradient Boosting; Biogas; Anaerobic Co-Digestion; Bio-Energy.

## 1. Introduction

### *1.1. It is difficult to infer determinants of efficiency of industrial-scale co-digestion based on lab-scale, life cycle assessment, and qualitative technology-specific studies*

There are multiple factors which affect the optimality of the anaerobic co-digestion process. These include chemical composition of substrates (carbohydrates, proteins, fats, etc.), temperature, pH value, biodegradability, and more (Hagos et al., 2017). Various lab-scale studies have investigated the effect of these parameters on the anaerobic digestion (AD) process.

For instance, Tasnim et al. (2017) investigated the effect of co-digesting cow manure, kitchen waste, and water hyacinth on biogas production. Saelor et al. (2017) studied the effect of co-digesting palm oil mill effluent and empty fruit bunches on biogas production. Rodríguez-Abalde et al. (2017) optimized the co-digestion of pig slurry, pasteurized slaughterhouse waste, and glycerine for maximum methane production. Li et al. (2018) co-digested animal manure with corn stover and apple pulp to determine the optimal ratio for biogas production. Dozens of similar studies investigating the effect of numerous AD parameters on process outcomes have been conducted in recent years (Algapani et al., 2019; Cheng et al., 2018; Deepanraj et al., 2017; Ivanovs et al., 2018; Kuczman et al., 2018; Liu et al., 2016; Maragkaki et al., 2018; Mehariya et al., 2018; Menon et al., 2017; Nguyen et al., 2017; Nie et al., 2017; Pan-in and Sukasem, 2017; Qin et al., 2018; Rajagopal et al., 2017; Tonanzi et al., 2018; Ye et al., 2018; Zhang et al., 2017).

In addition to lab-scale studies, life-cycle assessment has also been used to determine eco-friendly ways to treat food waste. For instance, Lijó et al. (2017) applied a combination of LCA and DEA to analyzing the eco-efficiency in 15 agricultural biogas plants was investigated. However, the sample size was limited, and vital technical parameters such as pre-treatment steps or digester type were not investigated for their effect on eco-efficiency. Moretti et al. (2018) conducted LCA on a co-digestion facility in the Netherlands, and computed environmental indicators such as eco-toxicity, acidification, and terrestrial eutrophication. Similar studies have also applied LCA-based methods to biogas generation processes (Cristóbal et al., 2016; Edwards et al., 2017; Gao et al., 2017; Jin et al., 2015; Laso et al., 2018; Thyberg and Tonjes, 2017; Tong et al., 2018; Woon et al., 2016; Xu et al., 2015)

Other studies have focused on qualitative reviews of specific technologies in the biogas value chain. For instance, Angelidaki et al. (2018) reviewed the current status and perspectives of biogas upgrading technologies, including qualitative appraisal of the efficiencies of physicochemical and biological upgrading techniques. Hosseinipour and Mehrpooya (2019) also investigated specific upgrading technologies such as amine scrubbing and cyrogenic separation. Oreggioni et al. (2017) conducted a techno-economic comparison of pressure swing adsorption (PSA) and solvent-based processes for biomethane production, finding that PSA has lower capital and lifetime costs.

However, findings from lab-scale and LCA-based studies do not always generalize to industrial-scale anaerobic digestion facilities. For instance, the comprehensive review by Hagos et al. (2017) confirmed that “there is a large discrepancy of biogas production performance between lab scale and industrial scale biogas plants”, and that industrial applications require “further investigation”. Moreover, parameters such as waste substrate mixture deemed optimal for biogas production in experiments might even be “improper for industrial applications” (Matuszewska et al., 2016). As a result, Tyagi et al. (2018) suggested a “need for further research” into elements such as “the pretreatment processes and quality of the feed, to better realize transition from laboratory studies to full-scale applications”.

### *1.2. Past industrial-scale studies have not managed to infer determinants of efficiency from a large AD project sample size across different regions*

Several studies have applied methods such as DEA or other methods to evaluating the efficiency of anaerobic digestion processes. However, these studies have not combined a unique blend of DEA with techniques in machine learning on data from projects around the world or on sufficiently representative sample sizes. This combination is useful in identifying specific determinants of efficiency.

For example, Ahlberg-Eliasson et al. (2017) investigated 27 biogas facilities in Sweden to determine the factors affecting biogas production and nutrient content in digestate. This was based on traditional statistical methods such as Pearson's partial correlation coefficient. Bolzonella et al. (2006) compared dry AD reactors with differently sorted municipal organic solid waste inputs in a single case in Italy, but specific determinants of efficiency were not investigated. Madlener et al. (2009) performed an assessment of 41 agricultural biogas plants in Austria based on DEA and the IRIS/ELECTRE TRI multi-criteria decision analysis methods. However, that study also did not attempt to infer the determinants of efficiency with any advanced statistical learning models. Silva et al. (2018) computed the energy efficiency of a single micro-generation unit producing biogas-derived electricity from digesting swine manure on a pig farming property.

Other examples include De Clercq et al. (2017a, 2017b), which applied two-stage data envelopment analysis and multiple linear regression to identifying some of the factors which affected operational performance in specific biogas facilities. However, their studies only time-series data from two operational case studies, while the present study uses data from 386 operational facilities. Moreover, recent advances in machine learning were not applied in that research.

### *1.3. Research contributions*

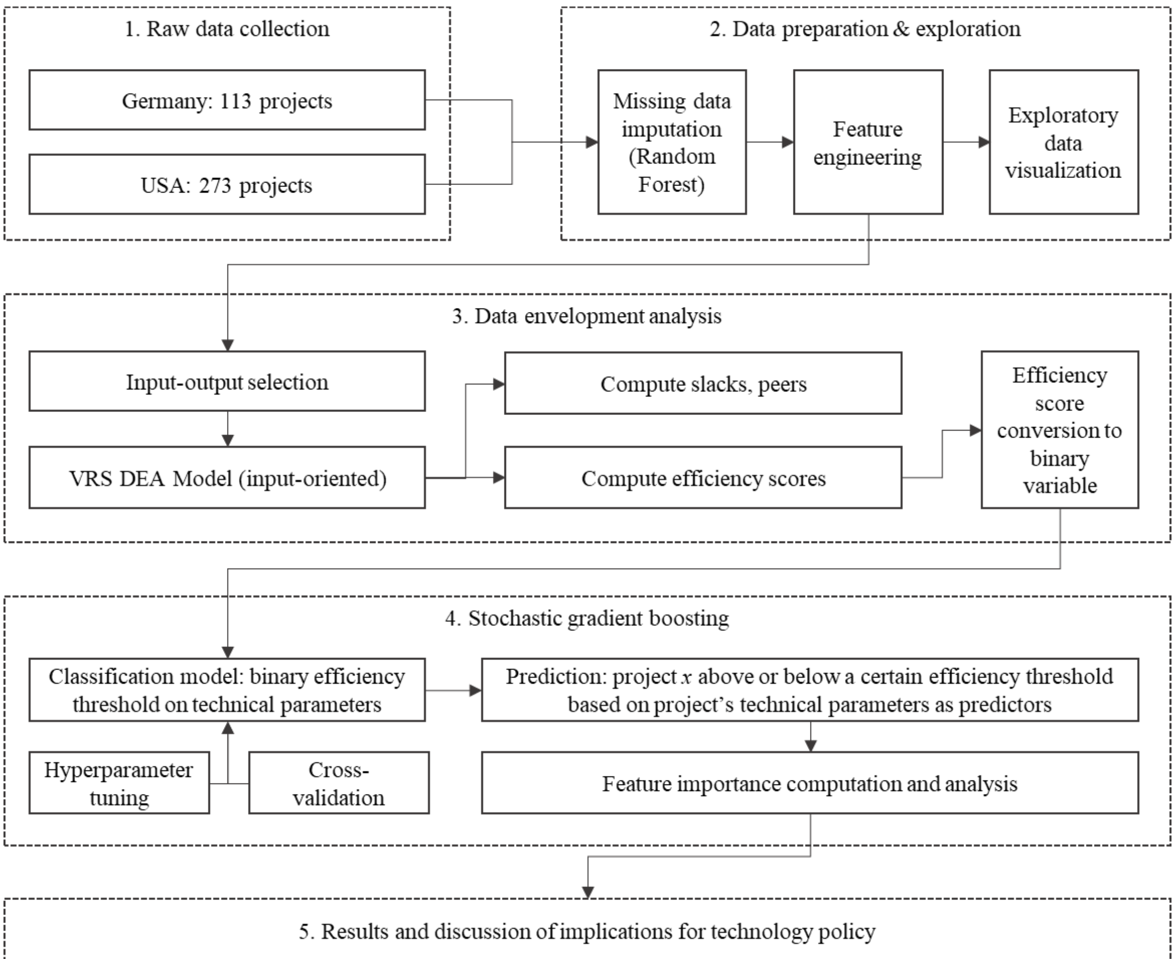
Based on the identified gaps in past research, this study applies a novel combination of DEA and stochastic gradient boosting (SGB) to identifying the determinants of efficiency in industrial-scale anaerobic co-digestion facilities for the first time. In addition, this study used a larger, cross-country dataset comprising 386 operational facilities treating a diverse set of waste types. This data scope is more representative of anaerobic digestion facilities and ensures that the results can be reasonably generalized. Moreover, the entire code and data used in this analysis can be found in the online supplement to ensure reproducibility and encourage further work in this area.

The remainder of this study is structured as follows. Section 2 introduces the methodology used in this study. Section 3 presents the results and provides discussion surrounding the key findings. Concluding remarks are made in section 4, and the supplementary information in section 5 provides the entire dataset and code behind the analysis to ensure full reproducibility of this research.

## 2. Methodology

The methodology in this study is based on the flowchart in figure 1. Firstly, data on operational industrial-scale facilities in Germany and the USA was collected. Secondly, the data underwent imputation for missing values and feature engineering prior to efficiency computations. Thirdly, DEA was applied to projects in Germany and the USA (separately) to compute efficiency scores for projects in each country. Fourthly, stochastic gradient boosting was used to classify projects based on project technical parameters, and feature importance charts were generated to evaluate the influence of these technical parameters on efficiency. Lastly, the results and their implications were discussed from a technology selection perspective. Each sub-section below corresponds to the numbered box in figure 1.

**Figure 1: Overview of methodology**




### *2.1. Raw data collection*

Data on 113 biowaste projects in Germany was procured from the Witzenhausen Institute for Waste, Environment and Energy, based on that institute's comprehensive survey of active biowaste co-digestion projects in Germany in 2015. This comprehensive survey collected operational data on project inputs, outputs, technology types, biogas/compost end use, and more. Data on 273 projects in the United States was gathered from the US Environmental Protection Agency's AgSTAR project, which contains a database of operational agricultural and co-digestion biogas facilities. This database, which is freely available on the

AgSTAR website, provides basic input-output data and some key technical parameters of projects involved in both anaerobic co-digestion and mono-digestion.

*2.2. Data preparation and exploration*

Missing values were imputed based on the MissForest method, which is a non-parametric missing value imputation method for mixed-type data based on random forest (Stekhoven and Bühlmann, 2012). This method has been shown to outperform other common methods of imputation (such as kNN, MissPALasso, and MICE), especially where complex, non-linear interactions are suspected in the data (Feng et al., 2014). It also computationally efficient, can deal with high-dimensional data, and can impute both continuous and categorial data. Missing value plots can be found in S.1.

In addition, feature engineering was conducted in order to (1) put the data in a format acceptable for the DEA and machine learning model and (2) to provide additional information for the stochastic gradient boosting model. A full description of the data can be found in table 1 and table 2.

**Table 1: variables related to the 113 projects in Germany**

| Variable name | Notation | Type | Use | Description |
|---|---|---|---|---|
| ActualInput_tpy | $G_{IN1}$ | Numeric | DEA Input | Amount of waste input into project (t/y) |
| SFP_m3perTonWasteIn | $G_{OUT1}$ | Integer | DEA Output | Biogas produced per ton of waste ($m^3$/t) |
| MarketedCompost_tpy | $G_{OUT2}$ | Integer | DEA Output | Amount of compost sold to market (t/y) |
| MarketedLiqDig_tpy | $G_{OUT3}$ | Integer | DEA Output | Amount of liquid digestate sold to market (t/y) |
| CoDigested | $G_1$ | Integer | SGB Feature | Whether the project co-digests different waste types |
| Present_Other | $G_2$ | Factor | SGB Feature | Whether the project digests "other" waste |
| Present_Manure | $G_3$ | Factor | SGB Feature | Whether the project digests manure |
| Present_FW | $G_4$ | Factor | SGB Feature | Whether the project digests food waste |
| Present_IBW | $G_5$ | Factor | SGB Feature | Whether the project digests industrial biowaste |
| Present_GW | $G_6$ | Factor | SGB Feature | Whether the project digests green waste |
| Present_BW | $G_7$ | Factor | SGB Feature | Whether the project digests household biowaste |
| Capacity_tpy | $G_8$ | Numeric | SGB Feature | Project capacity (t/y) |
| Digester_Box | $G_9$ | Factor | SGB Feature | Whether box technology is present |
| Digester_PlugFlow | $G_{10}$ | Factor | SGB Feature | Whether plug flow technology is present |
| Digester_WetFerm | $G_{11}$ | Factor | SGB Feature | Whether wet fermentation technology is present |
| Digester_Other | $G_{12}$ | Factor | SGB Feature | Whether other technology is present |
| Digester_SingleStage | $G_{13}$ | Factor | SGB Feature | Whether single-stage technology is present |
| Temp_Meso | $G_{14}$ | Factor | SGB Feature | Whether process temperature is mesophilic |
| Pre_Crushing | $G_{15}$ | Factor | SGB Feature | Whether crushing pretreatment is present |
| Pre_Screening | $G_{16}$ | Factor | SGB Feature | Whether screening pretreatment is present |
| Pre_ManualSorting | $G_{17}$ | Factor | SGB Feature | Whether manual sorting pretreatment is present |
| Pre_MagneticSorting | $G_{18}$ | Factor | SGB Feature | Whether magnetic sorting pretreatment is present |
| Pre_SubstanceSeparation | $G_{19}$ | Factor | SGB Feature | Whether substance separation pretreatment present |
| Pre_SandRemoval | $G_{20}$ | Factor | SGB Feature | Whether sand removal pretreatment is present |
| Pre_Slurrying | $G_{21}$ | Factor | SGB Feature | Whether slurrying pretreatment is present |
| Pre_Other | $G_{22}$ | Factor | SGB Feature | Whether other pretreatment is present |
| Pre_TotalSteps | $G_{23}$ | Integer | SGB Feature | Total number of pretreatment stages in process |
| Pasteurization | $G_{24}$ | Factor | SGB Feature | Whether pasteurization is present |
| Dewatering | $G_{25}$ | Factor | SGB Feature | Whether dewatering is present |
| Efficiency Score | $E_G$ | Numeric | SGB Target | Efficiency score generated by DEA (see 2.3) |

Table 1 shows data related to the projects in Germany. The dataset included information about waste input ($G_{IN1}$) and valuable outputs such as biogas, compost, and liquid digestate ($G_{OUT1}$, $G_{OUT2}$, $G_{OUT3}$). In addition, the data included a of technical parameters related to project operation such as digester technology type, pre-treatment stages, and co-digested waste types ($G_1$, … $G_{25}$). These technical parameters were features of the stochastic gradient boosting model (see section 2.4) which was used to classify projects based on whether their efficiency score $E_G$ (see section 2.3) was above or below a certain threshold.

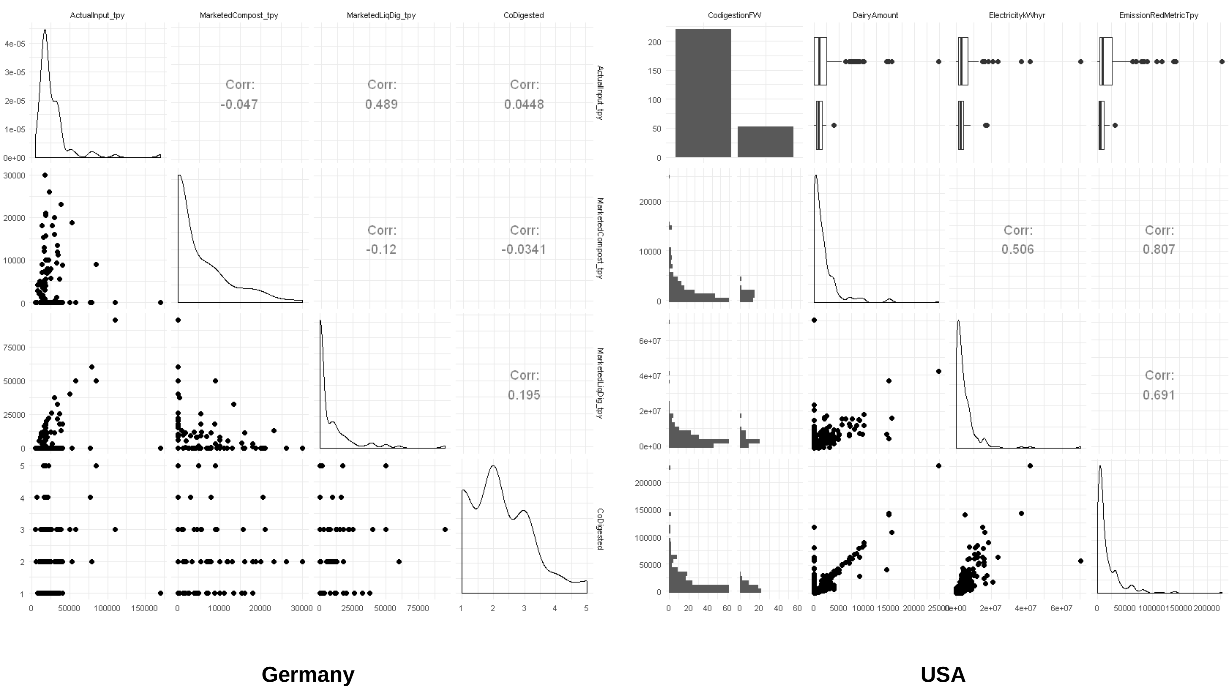
**Table 2: variables related to the 273 projects in the USA**

| Variable | Notation | Type | Use | Description |
|---|---|---|---|---|
| CattleAmount | $U_{IN1}$ | Numeric | DEA Input | Number of cattle in the project |
| DairyAmount | $U_{IN2}$ | Numeric | DEA Input | Number of dairy animals in the project |
| PoultryAmount | $U_{IN3}$ | Numeric | DEA Input | Number of poultry in the project |
| SwineAmount | $U_{IN4}$ | Numeric | DEA Input | Number of swine in the project |
| ElectricitykWhyr | $U_{OUT1}$ | Numeric | DEA Output | Electricity produced (kWh/y) |
| EmissionRedMetricTpy | $U_{OUT2}$ | Numeric | DEA Output | Emissions reduced per year (t/y) |
| Digester_ASBR | $U_1$ | Factor | SGB Feature | Whether anaerobic batch reactor technology present |
| Digester_CompleteMix | $U_2$ | Factor | SGB Feature | Whether complete mix reactor technology present |
| Digester_CoveredLagoon | $U_3$ | Factor | SGB Feature | Whether covered lagoon technology present |
| Digester_FixedFilm | $U_4$ | Factor | SGB Feature | Whether fixed film technology present |
| Digester_HPF | $U_5$ | Factor | SGB Feature | Whether horizontal plug flow technology present |
| Digester_IBR | $U_6$ | Factor | SGB Feature | Whether induced blanket reactor technology present |
| Digester_MPF | $U_7$ | Factor | SGB Feature | Whether mixed plug flow reactor technology present |
| Digester_ModPF | $U_8$ | Factor | SGB Feature | Whether modular plug flow technology present |
| Digester_PlugFlow | $U_9$ | Factor | SGB Feature | Whether plug flow reactor technology present |
| Digester_VPF | $U_{10}$ | Factor | SGB Feature | Whether vertical plug flow technology present |
| Digester_Other | $U_{11}$ | Factor | SGB Feature | Whether other reactor technology present |
| Digester_Unknown | $U_{12}$ | Factor | SGB Feature | Whether reactor technology is unknown |
| Project_Centralized | $U_{13}$ | Factor | SGB Feature | Whether project scale is centralized |
| Project_FarmScale | $U_{14}$ | Factor | SGB Feature | Whether project scale is farm scale |
| Project_MutiFarm | $U_{15}$ | Factor | SGB Feature | Whether project scale is multiple farm scale |
| Project_Research | $U_{16}$ | Factor | SGB Feature | Whether project scale is research scale |
| Cattle | $U_{17}$ | Factor | N/A | Whether cattle manure is present in the project |
| Dairy | $U_{18}$ | Factor | N/A | Whether dairy manure is present in the project |
| Poultry | $U_{19}$ | Factor | N/A | Whether poultry manure is present in the project |
| Swine | $U_{20}$ | Factor | N/A | Whether swine manure is present in the project |
| Codigestion | $U_{21}$ | Factor | SGB Feature | Whether the project co-digests different waste types |
| CodigestionFW | $U_{22}$ | Factor | SGB Feature | Whether food waste is co-digested in the project |
| Efficiency Score | $E_U$ | Numeric | SGB Target | Efficiency score generated by DEA (see 2.3) |

Table 1 shows data related to the projects in the USA. The dataset included information about the number of livestock generating manure input in each project ($U_{IN1}$, $U_{IN2}$, $U_{IN3}$, $U_{IN4}$) and valuable outputs such as electricity production and reduced emissions ($U_{OUT1}$, $U_{OUT2}$). In addition, the data included a of technical parameters related to project operation such as digester technology type and whether co-digested was present in the project ($U_1$, … $U_{22}$). These technical parameters were features of the stochastic gradient boosting model (see section 2.4) which was used to classify projects based on whether their efficiency score $E_U$ (see section 2.3) was above or below a certain threshold.

Figure 2 presents a pairplot with distributions and correlations of some selected variables in each project. For the German projects, the mean values for waste input, marketed compost, and marketed liquid digestate were 25,774 t/y, 4,749 t/y. and 7,368 t/y respectively. German projects co-digested between one and five different waste types. For US projects, mean values for dairy livestock, electricity output, and emission reductions per years were 1,919 units, 4,706,731 kWh/y, and 17,773 t/y respectively. In addition, co-digestion of various waste types occurred in 108 projects, while co-digestion with food waste in particular occurred in 53 projects.

**Figure 2: Pairplot of selected variables across projects in both countries**




*2.3. Data Envelopment Analysis*

Data Envelopment Analysis (DEA) is a linear programming method which constructs a non-parametric, piecewise frontier over given production data, to calculate efficiencies relative to this frontier. DEA measures the relative efficiency of decision making units (DMUs) – for example, a firm, hospital, or engineering project –which have the same inputs and outputs. This study applied DEA to evaluating the relative efficiency of co-digestion projects in the USA and Germany. In other words, DEA was used to determine whether there exist a certain AD projects that can produce more output with a similar usage of inputs (Joro and Korhonen, 2015). If such a project exists, then the implication is that the other AD projects are inefficient.

DEA has seen applications in numerous fields outside of biogas. These include wastewater treatment (Castellet and Molinos-Senante, 2016), agricultural eco-efficiency (Angulo-Meza et al., 2018), supply chain sustainability (Izadikhah and Saen, 2018), museum performance (Basso et al., 2018), toll road performance (Regalado López and Campos Cacheda, 2018), and many more, as described in the comprehensive survey by Sueyoshi et al. (2017).

In this study, we deal with $n$ biogas projects (DMUs) with input and output matrices $X = (x_{ij}) \in R^{m\times n}$ and $Y = (y_{rj}) \in R^{s\times n}$ respectively. The production possibility set, which is assumed to be based on non-negative data (i.e. $X \geq 0$ and $Y \geq 0$) is defined as

$$T = \{(x, y) : x \geq X\lambda, y \leq Y\lambda, \lambda \geq 0\} \tag{1}$$

where $\lambda$ is a non-negative vector in $R^n$. An input-oriented DEA model can be defined as:

$$\theta^* = \min \theta, \text{ s.t.}$$

$$\begin{aligned}
&\textstyle\sum_{j=1}^{n} \lambda_i x_{ij} \leq \theta x_{io} && i = 1,2, \dots, m; \\
&\textstyle\sum_{j=1}^{n} \lambda_i y_{rj} \leq y_{ro} && r = 1,2, \dots, s; \\
&\textstyle\sum_{j=1}^{n} \lambda_i = 1 && \\
&\lambda_i \geq 0 && j = 1,2, \dots, n.
\end{aligned} \tag{2}$$

where $y_{ro}$ and $x_{io}$ and *r*th and *i*th output and input respectively. Additional mathematical details have been covered extensively in the literature (Aparicio et al., 2017; Chen et al., 2017; Cooper et al., 2004; Hatami-Marbini et al., 2017; Joro and Korhonen, 2015; Olesen and Petersen, 2016; Zhu, 2014).

For the German projects, a one-input, three-output model was constructed, using variables $G_{IN1}$ as input and $G_{OUT1}$, $G_{OUT2}$, and $G_{OUT3}$ as outputs. For the USA projects, a four-input, two-output model was built, using variables $U_{IN1}$, $U_{IN2}$, $U_{IN3}$, $U_{IN4}$ as inputs and $U_{OUT1}$, $U_{OUT2}$ as outputs (see table 2 for descriptions).

After DEA scores were computed for each group of projects in Germany and the USA, the efficiency scores were converted to a binary number, in order to create a target variable for the gradient boosting classification model. A project was allocated a score of "1", indicating high efficiency, if its efficiency was at the 75th percentile (i.e. equal to the value below which 75% of other observations may be found). A project was allocated a score of "0" otherwise, indicating low efficiency. Defining efficiency categories based on percentiles is consistent with previous approaches in the DEA literature (Hsu and Lin, 2007).

*2.4. Stochastic gradient boosting for classification of efficiency scores based on technical process features*

Supervised learning methods attempt to obtain a mapping $f: X \rightarrow Y$ based on training data, where $X \subset R^D$ and $Y \subset R$ denote the input and output space respectively (James et al., 2013). The training data can be characterized as input-output pairs, $(x_i, y_i)_{i=1}^{N} \subset X \times Y$. The function $f(x)$ is learned based on empirical loss minimization, which can be characterized as

$$\hat{f}(x) = \operatorname{argmin}_{f(x)} \sum_{i=1}^{N} \mathcal{L}(y_i, f(x_i)) \tag{3}$$

where $y_i$ is the true label (i.e. efficiency score of "1" or "0") of input $x_i$. The loss function $\mathcal{L}(\cdot,\cdot)$ evaluates the difference between a prediction $f(x_i)$ and the truth $y_i$.

Gradient boosting fits an additive model and using sequentially grown decision trees as base learners. Decision trees contain internal and terminal nodes; the internal nodes perform binary splitting and the terminal nodes provide predictions. A decision tree with $L$ terminal nodes can be illustrated as

(4)

$$T(x; C, R) = \sum_{l=1}^{L} C_l I(x \in R_l)$$

where $R = \{R_l, l = 1, \dots L\}$ where $L$ indicate the disjoint regions of the input space separated by the decision tree. $C = \{C_l, l = 1, \dots L\}$ refers to the response of each terminal node. The indicator function $I(condition)$ returns 1 or 0 depending on whether a condition is met. Unlike fitting a single large decision tree to the data, the boosting approach learns slowly by fitting small trees with just a few terminal nodes that depend on trees that have already been grown. This simultaneously prevent overfitting, while also improving $\hat{f}$ in areas where it does not perform well (James et al., 2013). When decision trees are chosen as base learners in this way, the additive model is the sum of the trees such that

$$f(x) = \sum_{m=0}^{M} T(x; C_m, R_m) \tag{5}$$

where $T(x; C_m, R_m)$ refers to the $m^{\text{th}}$ tree, and where $C_m$ and $R_m$ are the tree's parameters. A greedy forward additive model subsequently solves Eq. (5) through fitting a single tree $T(x; C_m, R_m)$ at each step based on existing model $f_{m-1}(x)$,

$$(C_m, R_m) = \text{argmin}_{(C_m, R_m)} \sum_{i=1}^{N} \mathcal{L}(y_i, f_{m-1}(x_i) + T(x_i; C_m, R_m)) \tag{6}$$

According to Friedman (2001), at stage $m$, the model computer current residuals $\tilde{y}_{im}$, $y = 1, \dots N$ based on the prior tree model $f_{m-1}$. It then trains the $m^{\text{th}}$ tree $T(x; C_m, R_m)$ based on dataset $(x_i, \tilde{y}_{im})_{i=1}^{N}$ using

$$(C_m, R_m) = \text{argmin}_{(C_m, R_m)} \sum_{i=1}^{N} \mathcal{L}(\tilde{y}_{im}, T(x; C_m, R_m)) \tag{7}$$

In Eq. (7), $\tilde{y}_{im}$ is computed via the negative gradient of the loss function,

$$\tilde{y}_{im} = -\left[\frac{\partial L\big(y_i, f(x_i)\big)}{\partial f(x_i)}\right]_{f(x) = f_{m-1}}, i = 1, \dots, N \tag{8}$$

Generalizability and computational efficiency are enhanced via shrinkage and subsampling during the boosting process. Shrinkage scales the contribution of each tree:

$$f_m(x) = f_{m-1}(x) + vT(x; C_m, R_m) \tag{9}$$

where the parameter $v$ $(0 < v < 1)$ dictates the learning rate during gradient boosting. The stochastic element of gradient boosting entails random subsampling of the training data, which tends to reduce correlation between the trained decision trees at successive iterations (Friedman, 2002; Xiong et al., 2018).

Hyperparameter tuning coupled with 10-fold cross validation repeated 10 times were used to choose the parameters of the model. The parameters tuned included (1) the number of boosting iterations; (2) max tree depth; (3) shrinkage; and (4) minimum terminal node size. Cross-validation is suitable method for overcoming data scarcity (in this case, only 386 observations), where a "holdout data" approach would be unreasonable. Finding the best combination of parameters is crucial to optimize and obtain the best model

(Climent et al., 2018). Additional details on parameter tuning and cross-validation can be found in Hastie et al. (2009) and Touzani et al. (2018).

The predictive power of the stochastic gradient boosting algorithm was evaluated based on the accuracy metric, which computes the percentage of instances where predicted labels are equal to the actual labels.

### *2.5. Computation*

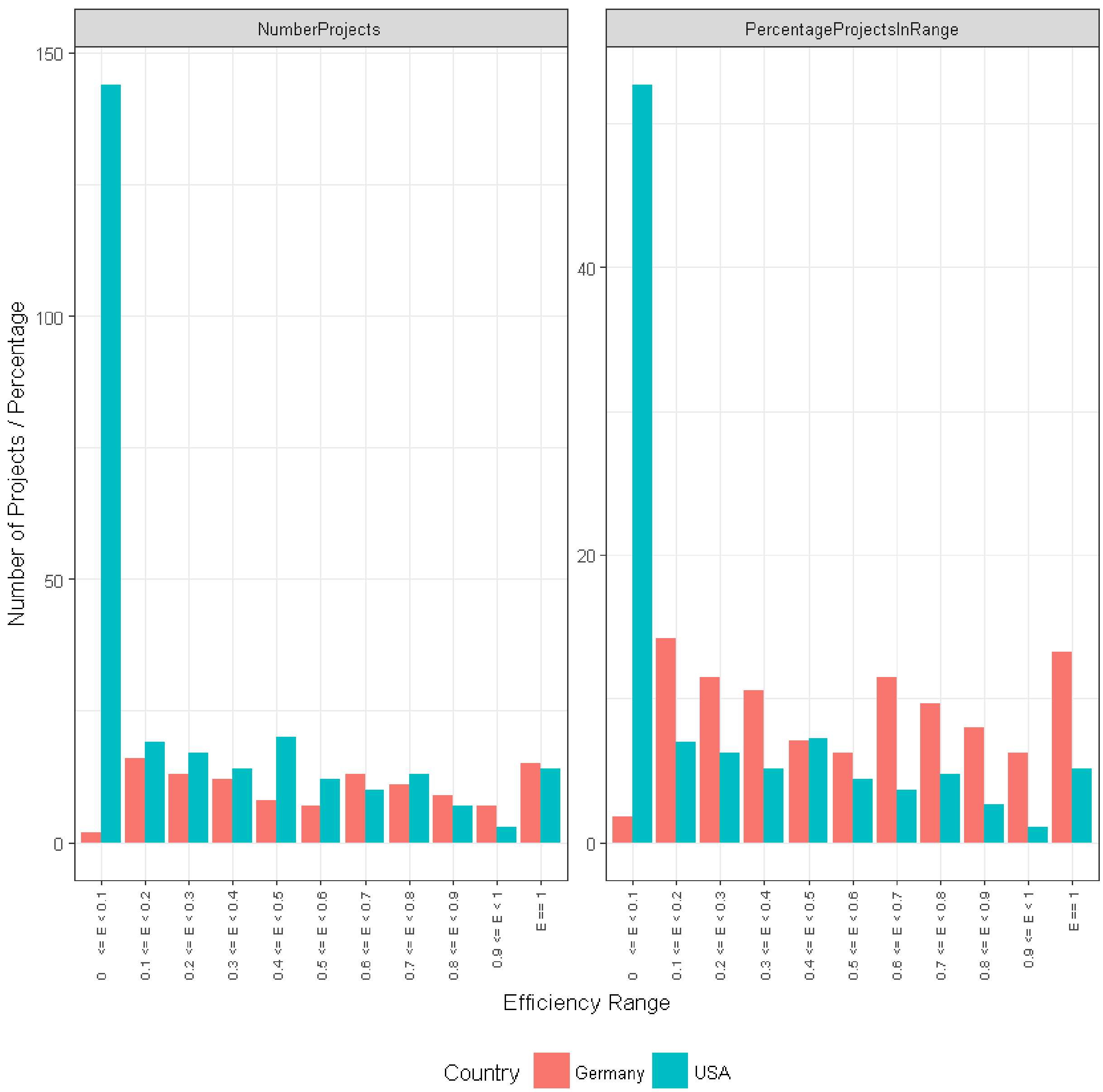
All computation was conducted in R, with Jupyter Notebook as the IDE. For data imputation, the “missForest” package was used. For DEA, the “Benchmarking” package was used. For stochastic gradient boosting, parameter tuning and cross validation, the “gbm” and “caret” packages were used. Additional packages for data cleaning and visualization included “tidyverse” and “ggplot2”. All data and code can be found here: https://github.com/djavandeclercq/BiogasDEA_SGB.

## 3. Results and discussion

### *3.1. DEA results*

A summary of the efficiency scores is presented in figure 3, from which it is evident that there was significant variation across the projects in the two countries. Regarding the 113 German projects, the minimum, 1st quartile, mean, 3rd quartile, and maximum efficiency scores were 0.03, 0.27, 0.56, 0.82, and 1.00 respectively. Based on these values, the efficiency threshold for German projects was set as 0.82 for the subsequent binary classification problem. For the 273 projects in the USA, the minimum, 1st quartile, mean, 3rd quartile, and maximum efficiency scores were 0.00, 0.00, 0.24, 0.46, and 1.00 respectively. Based on these values, the efficiency threshold for projects in the USA was set as 0.46 for the subsequent binary classification problem.

**Figure 3: summary of efficiency scores**



Comparing the efficiency values between the two countries is infeasible, since the input-output values used in the DEA models for each country's projects are different. We are therefore unable to comment on whether projects in one country were on average more efficient than projects in the other country.

Nevertheless, comparison between projects within each country is possible. Figure 3 demonstrates that most projects in the USA were skewed towards low efficiency. For example, 52% of projects fell within the 0.00 to 0.10 efficiency range, while only 1.8% of projects in Germany fell in this range. In addition, based on the third quartile values, we can infer that the majority of projects in the USA fell below a relatively low efficiency threshold: 75% of projects in the USA fell below 0.46 efficiency, while in Germany 75% of projects fell below 0.82 efficiency. In addition, only 5.1% of projects in the USA were fully efficient, while 13.3% of projects in Germany were fully efficient. Additional information can be found in S.2.

The variations in efficiency values between the two countries might be a result of several factors. Firstly, the input-output combinations chosen in the DEA model set up were different, primarily due to data availability. Another potential factor in the average efficiency being higher in Germany may be related to the fact that anaerobic digestion technology is more mature in that country compared to the USA (De Clercq et al., 2017b).

*3.2. Classification results*

Figure 4 visualizes cross-validated accuracy values across two of the tuned hyperparameters: boosting iterations and maximum tree depth. For German projects, the best-performing model for predicting whether a certain project met an efficiency threshold or not had an accuracy of 0.77 (left panel of figure 4). This accuracy was achieved with parameters values of 0.1 (shrinkage), 3 (max tree depth), 10 (minimum terminal node size), and 100 (boosting iterations). For USA projects, the best-performing model had an accuracy of 0.75 (right panel of figure 4). The parameter values for this accuracy were 0.1 (shrinkage), 3 (max tree depth), 10 (minimum terminal node size), and 50 (boosting iterations).

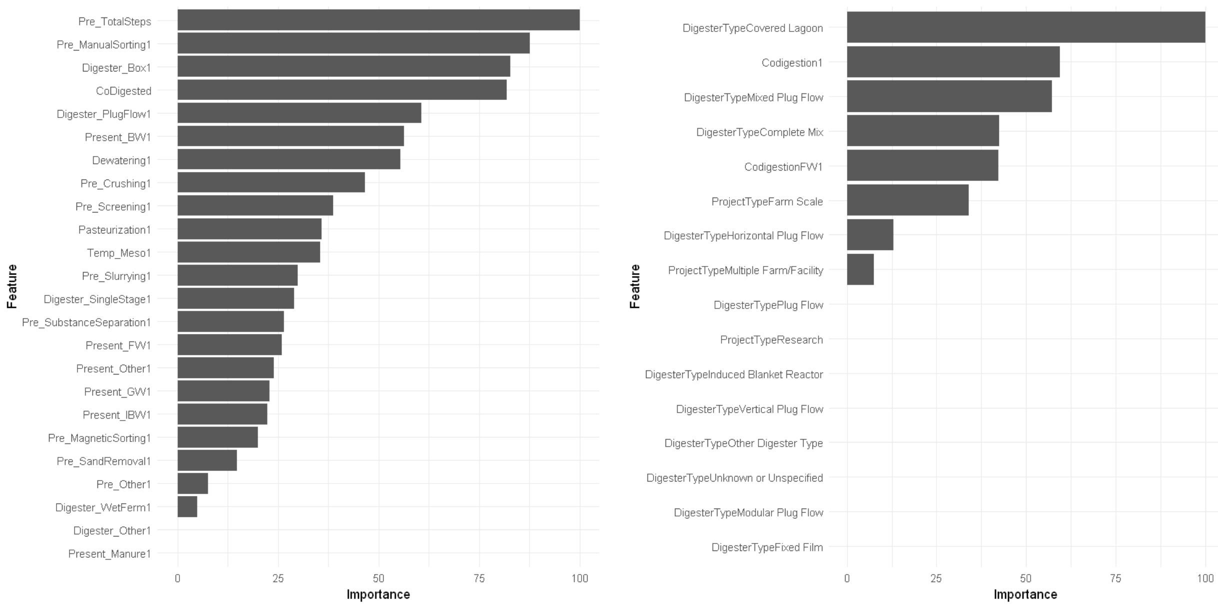
**Figure 4: gradient boosted classifier model cross-validated accuracy as a function of tuning parameters**


**Germany** **USA**

It is useful to interpret the derived approximation of $f(x)$ in this study to better understand the impact of particular input variables that are most influential in contribution to its variance (Friedman, 2001). The relative feature importance values illustrated in figure 5 demonstrate the technical parameters of projects in both countries which had the highest impact on efficiency. Are measures of importance are scaled to have a maximum value of 100.

For German projects (left panel of figure 5) the top ten features in terms of relative importance were (1) the total number of pre-treatment stages in the project; (2) the presence/absence of manual sorting as a pre-treatment step; (3) whether or not the digester incorporated box-type technology; (4) the presence/absence of co-digestion of different waste types in the project; (5) whether or not the digester incorporated plug-flow technology; (6) the presence/absence of bio-waste in project waste input; (7) the presence/absence of digestate dewatering; (8) the presence/absence of crushing pre-treatment; (9) the presence/absence of screening pre-treatment; and (10) the presence/absence of digestate pasteurization.

**Figure 5: Ranked relative feature importance associated with efficiency threshold classification**




**Germany** **USA**

For the German projects, four out of the ten most importance features for the efficiency classification model were related to pre-treatment. This is intuitive and confirms past experimental research, which has demonstrated the importance of pre-treatment for optimal anaerobic digestion conditions. For instance, Choi et al. (2018) found that thermal hydrolysis pretreatment of sewage sludge enhanced anaerobic digestion. Rafieenia et al. (2018) demonstrated the effectiveness of various pre-treatment methods applied to digestion of food waste, including heat shock, alkaline treatment, and aeration. Surra et al. (2018) showed that hydrogen peroxide pre-treatment increased the digestibility of maize cob waste. Kim et al. (2018) showed the effectiveness of hot water pre-treatment as a chemical-free method to improve methane production from the digestion of rice straw. Other recent studies on the benefits of pre-treatment are numerous (Carlini et al., 2018; Kang et al., 2018; Neumann et al., 2018; Rodriguez et al., 2015; Serrano et al., 2016; Yang et al., 2017; Zeynali et al., 2017).

For projects in the USA (right panel of figure 5), the top eight most influential features were (1) whether the project was of the covered lagoon digester type; (2) the presence/absence of co-digestion of different waste types in the project; (3) whether the project was of the mixed plug flow digester type; (4) whether the project was of the complete mixture digester type; (5) whether or not food waste was co-digested; (6) whether the project was farm-scale; (7) whether the project was of the horizontal plug flow digester type; (8) whether the project was of multiple farm scale.

One of the top eight features was related to co-digestion, which confirms the results of experimental studies. For instance, (Barua et al., 2018) showed that co-digestion of water hyacinth and cooked food waste resulted in favorable synergies during the AD process. (Velásquez Piñas et al., 2018) demonstrated that co-digestion of maize, grass silage, and cattle manure had better had better performance than mono-digestion of these substrates. (Andriamanohiarisoamanana et al., 2018) showed that co-digesting meat and bone meal, crude glycerol, and dairy manure enhanced methane yield and increased ammoniacal nitrogen in the digestate.

(Cong et al., 2018) investigated co-digestion of grass and forbs, and found that co-digestion led to enhanced bio-methane yield and reduced lag-phase. (Barua et al., 2019) showed that co-digestion of water hyacinth and banana peels enhanced biogas generation. Other studies have made similar findings on the benefits of co-digestion, including coffee husks with microalgal biomass (Passos et al., 2018), rice straw with *hydrilla verticillata* (Kainthola et al., 2019), and durian shell and various livestock manure (Shen et al., 2018).

Food waste co-digestion was an influential feature, and experimental studies have also confirmed the beneficial effects of food waste co-digestion. For example, (F. M. S. Silva et al., 2018) found that energy production was maximized via co-digestion of food waste with sewage sludge and glycerol. (Jiang et al., 2018) found that co-digestion of food waste and pig manure helped to inactivate pathogenic microorganisms such as *Salmonella*. (Z. Li et al., 2018) showed that anaerobic co-digestion of sewage sludge and food waste enhanced hydrogen gas content by 62.4%.

## 4. Conclusion

This study applied data envelopment analysis and a stochastic gradient boosting classifier to uncovering the determinants of efficiency in 113 anaerobic digestion facilities in Germany and 273 anaerobic digestion facilities in the USA. The results revealed the impacts of various technology parameters on process efficiency.

For German projects, the top ten most influential technology parameters included (1) the total number of pre-treatment stages in the project; (2) the presence/absence of manual sorting as a pre-treatment step; (3) whether or not the digester incorporated box-type technology; (4) the presence/absence of co-digestion of different waste types in the project; (5) whether or not the digester incorporated plug-flow technology; (6) the presence/absence of bio-waste in project waste input; (7) the presence/absence of digestate dewatering; (8) the presence/absence of crushing pre-treatment; (9) the presence/absence of screening pre-treatment; and (10) the presence/absence of digestate pasteurization.

For projects in the USA, the top eight most influential features were (1) whether the project was of the covered lagoon digester type; (2) the presence/absence of co-digestion of different waste types in the project; (3) whether the project was of the mixed plug flow digester type; (4) whether the project was of the complete mixture digester type; (5) whether or not food waste was co-digested; (6) whether the project was farm-scale; (7) whether the project was of the horizontal plug flow digester type; (8) whether the project was of multiple farm scale.

These results may be of interest to decision-makers involved in anaerobic digestion technical selection who wish to maximize efficiency in industrial-scale facilities.

**Acknowledgements**

The authors gratefully acknowledge the financial support from the "Thirteenth Five-Year" National Key R&D Program of China (2018YFC190080), National Science Fund for Distinguished Young Scholars of China (71825006) and General Programs (71774099). The responsibility for any error rests solely with the authors. The contents of this paper reflect the views of the authors and do not necessarily indicate acceptance by the sponsors.

## References


Ahlberg-Eliasson, K., Nadeau, E., Levén, L., Schnürer, A., 2017. Production efficiency of Swedish farm-scale biogas plants. Biomass and Bioenergy 97, 27–37. https://doi.org/https://doi.org/10.1016/j.biombioe.2016.12.002

Algapani, D.E., Qiao, W., Ricci, M., Bianchi, D., M. Wandera, S., Adani, F., Dong, R., 2019. Bio-hydrogen and bio-methane production from food waste in a two-stage anaerobic digestion process with digestate recirculation. Renew. Energy 130, 1108–1115. https://doi.org/https://doi.org/10.1016/j.renene.2018.08.079

Amirteimoori, A., 2007. DEA efficiency analysis: Efficient and anti-efficient frontier. Appl. Math. Comput. 186, 10–16. https://doi.org/https://doi.org/10.1016/j.amc.2006.07.006

Andriamanohiarisoamanana, F.J., Saikawa, A., Kan, T., Qi, G., Pan, Z., Yamashiro, T., Iwasaki, M., Ihara, I., Nishida, T., Umetsu, K., 2018. Semi-continuous anaerobic co-digestion of dairy manure, meat and bone meal and crude glycerol: Process performance and digestate valorization. Renew. Energy 128, 1–8. https://doi.org/https://doi.org/10.1016/j.renene.2018.05.056

Angelidaki, I., Treu, L., Tsapekos, P., Luo, G., Campanaro, S., Wenzel, H., Kougias, P.G., 2018. Biogas upgrading and utilization: Current status and perspectives. Biotechnol. Adv. 36, 452–466. https://doi.org/https://doi.org/10.1016/j.biotechadv.2018.01.011

Angulo-Meza, L., González-Araya, M., Iriarte, A., Rebolledo-Leiva, R., Soares de Mello, J.C., 2018. A multiobjective DEA model to assess the eco-efficiency of agricultural practices within the CF + DEA method. Comput. Electron. Agric. https://doi.org/https://doi.org/10.1016/j.compag.2018.05.037

Aparicio, J., Cordero, J.M., Pastor, J.T., 2017. The determination of the least distance to the strongly efficient frontier in Data Envelopment Analysis oriented models: Modelling and computational aspects. Omega 71, 1–10. https://doi.org/https://doi.org/10.1016/j.omega.2016.09.008

Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Manage. Sci. 30, 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Barua, V.B., Rathore, V., Kalamdhad, A.S., 2019. Anaerobic co-digestion of water hyacinth and banana peels with and without thermal pretreatment. Renew. Energy 134, 103–112. https://doi.org/https://doi.org/10.1016/j.renene.2018.11.018

Barua, V.B., Rathore, V., Kalamdhad, A.S., 2018. Comparative evaluation of anaerobic co-digestion of water hyacinth and cooked food waste with and without pretreatment. Bioresour. Technol. Reports 4, 202–208. https://doi.org/https://doi.org/10.1016/j.biteb.2018.11.002

Basso, A., Casarin, F., Funari, S., 2018. How well is the museum performing? A joint use of DEA and BSC to measure the performance of museums. Omega 81, 67–84. https://doi.org/https://doi.org/10.1016/j.omega.2017.09.010

Bolzonella, D., Pavan, P., Mace, S., Cecchi, F., 2006. Dry anaerobic digestion of differently sorted organic municipal solid waste: A full-scale experience. Water Sci. Technol. https://doi.org/10.2166/wst.2006.232

Carlini, M., Castellucci, S., Mennuni, A., 2018. Water hyacinth biomass: chemical and thermal pre-treatment for energetic utilization in anaerobic digestion process. Energy Procedia 148, 431–438. https://doi.org/https://doi.org/10.1016/j.egypro.2018.08.106

Castellet, L., Molinos-Senante, M., 2016. Efficiency assessment of wastewater treatment plants: A data envelopment analysis approach integrating technical, economic, and environmental issues. J. Environ. Manage. 167, 160–166. https://doi.org/https://doi.org/10.1016/j.jenvman.2015.11.037

Chen, L., Wang, Y.-M., Lai, F., 2017. Semi-disposability of undesirable outputs in data envelopment analysis for environmental assessments. Eur. J. Oper. Res. 260, 655–664. https://doi.org/https://doi.org/10.1016/j.ejor.2016.12.042

Cheng, H., Hiro, Y., Hojo, T., Li, Y.-Y., 2018. Upgrading methane fermentation of food waste by using a hollow fiber type anaerobic membrane bioreactor. Bioresour. Technol. 267, 386–394. https://doi.org/https://doi.org/10.1016/j.biortech.2018.07.045

Choi, J.-M., Han, S.-K., Lee, C.-Y., 2018. Enhancement of methane production in anaerobic digestion of sewage sludge by thermal hydrolysis pretreatment. Bioresour. Technol. 259, 207–213. https://doi.org/https://doi.org/10.1016/j.biortech.2018.02.123

Climent, F., Momparler, A., Carmona, P., 2018. Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. J. Bus. Res. https://doi.org/https://doi.org/10.1016/j.jbusres.2018.11.015

Cong, W.-F., Moset, V., Feng, L., Møller, H.B., Eriksen, J., 2018. Anaerobic co-digestion of grass and forbs – Influence of cattle manure or grass based inoculum. Biomass and Bioenergy 119, 90–96. https://doi.org/https://doi.org/10.1016/j.biombioe.2018.09.009

Cooper, W.W., Seiford, L.M., Zhu, J. (Eds.), 2004. Handbook on Data Envelopment Analysis, International Series in Operations Research & Management Science. Kluwer Academic Publishers, Boston. https://doi.org/10.1007/b105307

Cristóbal, J., Limleamthong, P., Manfredi, S., Guillén-Gosálbez, G., 2016. Methodology for combined use of data envelopment analysis and life cycle assessment applied to food waste management. J. Clean. Prod. 135, 158–168. https://doi.org/https://doi.org/10.1016/j.jclepro.2016.06.085

De Clercq, D., Wen, Z., Caicedo, L., Cao, X., Fan, F., Xu, R., 2017a. Application of DEA and statistical inference to model the determinants of biomethane production efficiency: A case study in south China. Appl. Energy 205, 1231–1243. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.08.111

De Clercq, D., Wen, Z., Gottfried, O., Schmidt, F., Fei, F., 2017b. A review of global strategies promoting the conversion of food waste to bioenergy via anaerobic digestion. Renew. Sustain. Energy Rev. 79, 204–221. https://doi.org/https://doi.org/10.1016/j.rser.2017.05.047

De Clercq, D., Wen, Z., Lu, X., Caicedo, L., Cao, X., Fan, F., 2017c. Determinants of efficiency in an industrial-scale anaerobic digestion food waste-to-biogas project in an Asian megacity based on data envelopment analysis and exploratory multivariate statistics. J. Clean. Prod. 168, 983–996. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.09.062

Deepanraj, B., Sivasubramanian, V., Jayaraj, S., 2017. Multi-response optimization of process parameters in biogas production from food waste using Taguchi – Grey relational analysis. Energy Convers. Manag. 141, 429–438. https://doi.org/https://doi.org/10.1016/j.enconman.2016.12.013

Edwards, J., Othman, M., Crossin, E., Burn, S., 2017. Anaerobic co-digestion of municipal food waste and sewage sludge: A comparative life cycle assessment in the context of a waste service provision. Bioresour. Technol. 223, 237–249. https://doi.org/https://doi.org/10.1016/j.biortech.2016.10.044

Feng, L., Nowak, G., O'Neill, T.J., Welsh, A.H., 2014. CUTOFF: A spatio-temporal imputation method. J. Hydrol. 519, 3591–3605. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.11.012

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378. https://doi.org/https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232.

Gao, A., Tian, Z., Wang, Z., Wennersten, R., Sun, Q., 2017. Comparison between the Technologies for Food Waste Treatment. Energy Procedia 105, 3915–3921. https://doi.org/https://doi.org/10.1016/j.egypro.2017.03.811

Hagos, K., Zong, J., Li, D., Liu, C., Lu, X., 2017. Anaerobic co-digestion process for biogas production: Progress, challenges and perspectives. Renew. Sustain. Energy Rev. 76, 1485–1496. https://doi.org/https://doi.org/10.1016/j.rser.2016.11.184

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed. Springer-Verlag, New York, NY, USA. https://doi.org/10.1007/978-0-387-84858-7

Hatami-Marbini, A., Ebrahimnejad, A., Lozano, S., 2017. Fuzzy efficiency measures in data envelopment analysis using lexicographic multiobjective approach. Comput. Ind. Eng. 105, 362–376. https://doi.org/https://doi.org/10.1016/j.cie.2017.01.009

Henriques, I.C., Sobreiro, V.A., Kimura, H., Mariano, E.B., 2018. Efficiency in the Brazilian banking system using data envelopment analysis. Futur. Bus. J. 4, 157–178. https://doi.org/https://doi.org/10.1016/j.fbj.2018.05.001

Hosseinipour, S.A., Mehrpooya, M., 2019. Comparison of the biogas upgrading methods as a transportation fuel. Renew. Energy 130, 641–655. https://doi.org/https://doi.org/10.1016/j.renene.2018.06.089

Hsu, C.-S., Lin, J.-R., 2007. Mutual Fund Performance and Persistence in Taiwan: A Non-Parametric Approach. Serv. Ind. J. 27, 509–523. https://doi.org/10.1080/02642060701411658

Ivanovs, K., Spalvins, K., Blumberga, D., 2018. Approach for modelling anaerobic digestion processes of fish waste. Energy Procedia 147, 390–396. https://doi.org/https://doi.org/10.1016/j.egypro.2018.07.108

Izadikhah, M., Saen, R.F., 2018. Assessing sustainability of supply chains by chance-constrained two-stage DEA model in the presence of undesirable factors. Comput. Oper. Res. 100, 343–367. https://doi.org/https://doi.org/10.1016/j.cor.2017.10.002

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, 1st ed. Springer-Verlag, New York, NY, USA. https://doi.org/10.1007/978-1-4614-7138-7

Jiang, Y., Dennehy, C., Lawlor, P.G., Hu, Z., Yang, Q., McCarthy, G., Tan, S.P., Zhan, X., Gardiner, G.E., 2018. Inactivation of Salmonella during dry co-digestion of food waste and pig manure. Waste Manag. 82, 231–240. https://doi.org/https://doi.org/10.1016/j.wasman.2018.10.037

Jin, Y., Chen, T., Chen, X., Yu, Z., 2015. Life-cycle assessment of energy consumption and environmental impact of an integrated food waste-based biogas plant. Appl. Energy 151, 227–236. https://doi.org/https://doi.org/10.1016/j.apenergy.2015.04.058

Joro, T., Korhonen, P., 2015. Extension of Data Envelopment Analysis with Preference Information. Springer Science+Business Media, New York, NY, USA.

Kainthola, J., Kalamdhad, A.S., Goud, V. V, 2019. Optimization of methane production during anaerobic co-digestion of rice straw and hydrilla verticillata using response surface methodology. Fuel 235, 92–99. https://doi.org/https://doi.org/10.1016/j.fuel.2018.07.094

Kang, X., Sun, Y., Li, L., Kong, X., Yuan, Z., 2018. Improving methane production from anaerobic digestion of Pennisetum Hybrid by alkaline pretreatment. Bioresour. Technol. 255, 205–212. https://doi.org/https://doi.org/10.1016/j.biortech.2017.12.001

Kim, M., Kim, B.-C., Nam, K., Choi, Y., 2018. Effect of pretreatment solutions and conditions on decomposition and anaerobic digestion of lignocellulosic biomass in rice straw. Biochem. Eng. J. 140, 108–114. https://doi.org/https://doi.org/10.1016/j.bej.2018.09.012

Kuczman, O., Gueri, M.V.D., De Souza, S.N.M., Schirmer, W.N., Alves, H.J., Secco, D., Buratto, W.G., Ribeiro, C.B., Hernandes, F.B., 2018. Food waste anaerobic digestion of a popular restaurant in Southern Brazil. J. Clean. Prod. 196, 382–389. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.05.282

Laso, J., Margallo, M., García-Herrero, I., Fullana, P., Bala, A., Gazulla, C., Polettini, A., Kahhat, R., Vázquez-Rowe, I., Irabien, A., Aldaco, R., 2018. Combined application of Life Cycle Assessment and linear programming to evaluate food waste-to-food strategies: Seeking for answers in the nexus approach. Waste Manag. 80, 186–197. https://doi.org/https://doi.org/10.1016/j.wasman.2018.09.009

Li, K., Liu, R., Cui, S., Yu, Q., Ma, R., 2018. Anaerobic co-digestion of animal manures with corn stover or apple pulp for enhanced biogas production. Renew. Energy 118, 335–342. https://doi.org/https://doi.org/10.1016/j.renene.2017.11.023

Li, Z., Chen, Z., Ye, H., Wang, Y., Luo, W., Chang, J.-S., Li, Q., He, N., 2018. Anaerobic co-digestion of sewage sludge and food waste for hydrogen and VFA production with microbial community analysis. Waste Manag. 78, 789–799. https://doi.org/https://doi.org/10.1016/j.wasman.2018.06.046

Lijó, L., Lorenzo-Toja, Y., González-García, S., Bacenetti, J., Negri, M., Moreira, M.T., 2017. Eco-efficiency assessment of farm-scaled biogas plants. Bioresour. Technol. 237, 146–155. https://doi.org/https://doi.org/10.1016/j.biortech.2017.01.055

Liu, C., Li, H., Zhang, Y., Liu, C., 2016. Improve biogas production from low-organic-content sludge through high-solids anaerobic co-digestion with food waste. Bioresour. Technol. 219, 252–260. https://doi.org/https://doi.org/10.1016/j.biortech.2016.07.130

Madlener, R., Antunes, C.H., Dias, L.C., 2009. Assessing the performance of biogas plants with multi-criteria and data envelopment analysis. Eur. J. Oper. Res. 197, 1084–1094. https://doi.org/https://doi.org/10.1016/j.ejor.2007.12.051

Maragkaki, A.E., Fountoulakis, M., Kyriakou, A., Lasaridi, K., Manios, T., 2018. Boosting biogas production from sewage sludge by adding small amount of agro-industrial by-products and food waste residues. Waste Manag. 71, 605–611. https://doi.org/https://doi.org/10.1016/j.wasman.2017.04.024

Matuszewska, A., Owczuk, M., Zamojska-Jaroszewicz, A., Jakubiak-Lasocka, J., Lasocki, J., Orliński, P., 2016. Evaluation of the biological methane potential of various feedstock for the production of biogas to supply agricultural tractors. Energy Convers. Manag. 125, 309–319. https://doi.org/https://doi.org/10.1016/j.enconman.2016.02.072

Mehariya, S., Patel, A.K., Obulisamy, P.K., Punniyakotti, E., Wong, J.W.C., 2018. Co-digestion of food waste and sewage sludge for methane production: Current status and perspective. Bioresour. Technol. 265, 519–531. https://doi.org/https://doi.org/10.1016/j.biortech.2018.04.030

Menon, A., Wang, J.-Y., Giannis, A., 2017. Optimization of micronutrient supplement for enhancing biogas production from food waste in two-phase thermophilic anaerobic digestion. Waste Manag.

59, 465–475. https://doi.org/https://doi.org/10.1016/j.wasman.2016.10.017

Moretti, M., Van Dael, M., Malina, R., Van Passel, S., 2018. Environmental assessment of waste feedstock mono-dimensional and bio-refinery systems: Combining manure co-digestion and municipal waste anaerobic digestion. J. Clean. Prod. 171, 954–961. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.10.097

Neumann, P., Barriga, F., Álvarez, C., González, Z., Vidal, G., 2018. Process performance assessment of advanced anaerobic digestion of sewage sludge including sequential ultrasound–thermal (55 °C) pre-treatment. Bioresour. Technol. 262, 42–51. https://doi.org/https://doi.org/10.1016/j.biortech.2018.03.057

Nguyen, D.D., Yeop, J.S., Choi, J., Kim, S., Chang, S.W., Jeon, B.-H., Guo, W., Ngo, H.H., 2017. A new approach for concurrently improving performance of South Korean food waste valorization and renewable energy recovery via dry anaerobic digestion under mesophilic and thermophilic conditions. Waste Manag. 66, 161–168. https://doi.org/https://doi.org/10.1016/j.wasman.2017.03.049

Nie, Y., Tian, X., Zhou, Z., Li, Y.-Y., 2017. Impact of food to microorganism ratio and alcohol ethoxylate dosage on methane production in treatment of low-strength wastewater by a submerged anaerobic membrane bioreactor. Front. Environ. Sci. Eng. 11, 6. https://doi.org/10.1007/s11783-017-0947-1

Olesen, O.B., Petersen, N.C., 2016. Stochastic Data Envelopment Analysis—A review. Eur. J. Oper. Res. 251, 2–21. https://doi.org/https://doi.org/10.1016/j.ejor.2015.07.058

Oreggioni, G.D., Luberti, M., Reilly, M., Kirby, M.E., Toop, T., Theodorou, M., Tassou, S.A., 2017. Techno-economic analysis of bio-methane production from agriculture and food industry waste. Energy Procedia 123, 81–88. https://doi.org/https://doi.org/10.1016/j.egypro.2017.07.252

Pan-in, S., Sukasem, N., 2017. Methane production potential from anaerobic co-digestions of different animal dungs and sweet corn residuals. Energy Procedia 138, 943–948. https://doi.org/https://doi.org/10.1016/j.egypro.2017.10.062

Passos, F., Cordeiro, P.H.M., Baeta, B.E.L., de Aquino, S.F., Perez-Elvira, S.I., 2018. Anaerobic co-digestion of coffee husks and microalgal biomass after thermal hydrolysis. Bioresour. Technol. 253, 49–54. https://doi.org/https://doi.org/10.1016/j.biortech.2017.12.071

Qin, Y., Wu, J., Xiao, B., Hojo, T., Li, Y.-Y., 2018. Biogas recovery from two-phase anaerobic digestion of food waste and paper waste: Optimization of paper waste addition. Sci. Total Environ. 634, 1222–1230. https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.03.341

Rafieenia, R., Pivato, A., Lavagnolo, M.C., 2018. Effect of inoculum pre-treatment on mesophilic hydrogen and methane production from food waste using two-stage anaerobic digestion. Int. J. Hydrogen Energy 43, 12013–12022. https://doi.org/https://doi.org/10.1016/j.ijhydene.2018.04.170

Rajagopal, R., Bellavance, D., Rahaman, M.S., 2017. Psychrophilic anaerobic digestion of semi-dry mixed municipal food waste: For North American context. Process Saf. Environ. Prot. 105, 101–108. https://doi.org/https://doi.org/10.1016/j.psep.2016.10.014

Regalado López, F.J., Campos Cacheda, J.M., 2018. An Approximation to Technical Efficiency in Spanish Toll Roads through a DEA approach. Transp. Res. Procedia 33, 386–393. https://doi.org/https://doi.org/10.1016/j.trpro.2018.11.005

Rodríguez-Abalde, Á., Flotats, X., Fernández, B., 2017. Optimization of the anaerobic co-digestion of pasteurized slaughterhouse waste, pig slurry and glycerine. Waste Manag. 61, 521–528. https://doi.org/https://doi.org/10.1016/j.wasman.2016.12.022

Rodriguez, C., Alaswad, A., Mooney, J., Prescott, T., Olabi, A.G., 2015. Pre-treatment techniques used for anaerobic digestion of algae. Fuel Process. Technol. 138, 765–779. https://doi.org/https://doi.org/10.1016/j.fuproc.2015.06.027

Saelor, S., Kongjan, P., O-Thong, S., 2017. Biogas Production from Anaerobic Co-digestion of Palm Oil Mill Effluent and Empty Fruit Bunches. Energy Procedia 138, 717–722. https://doi.org/https://doi.org/10.1016/j.egypro.2017.10.206

Serrano, A., Siles, J.A., Martín, M.A., Chica, A.F., Estévez-Pastor, F.S., Toro-Baptista, E., 2016. Improvement of anaerobic digestion of sewage sludge through microwave pre-treatment. J. Environ. Manage. 177, 231–239. https://doi.org/https://doi.org/10.1016/j.jenvman.2016.03.048

Shen, J., Zhao, C., Liu, Y., Zhang, R., Liu, G., Chen, C., 2018. Biogas production from anaerobic co-digestion of durian shell with chicken, dairy, and pig manures. Energy Convers. Manag. https://doi.org/https://doi.org/10.1016/j.enconman.2018.06.099

Silva, F.M.S., Mahler, C.F., Oliveira, L.B., Bassin, J.P., 2018. Hydrogen and methane production in a two-stage anaerobic digestion system by co-digestion of food waste, sewage sludge and glycerol. Waste Manag. 76, 339–349. https://doi.org/https://doi.org/10.1016/j.wasman.2018.02.039

Silva, F.P., de Souza, S.N.M., Kitamura, D.S., Nogueira, C.E.C., Otto, R.B., 2018. Energy efficiency of a micro-generation unit of electricity from biogas of swine manure. Renew. Sustain. Energy Rev. 82, 3900–3906. https://doi.org/https://doi.org/10.1016/j.rser.2017.10.083

Skovsgaard, L., Jacobsen, H.K., 2017. Economies of scale in biogas production and the significance of flexible regulation. Energy Policy 101, 77–89. https://doi.org/https://doi.org/10.1016/j.enpol.2016.11.021

Stekhoven, D.J., Bühlmann, P., 2012. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28, 112–118.

Sueyoshi, T., Yuan, Y., Goto, M., 2017. A literature study for DEA applied to energy and environment. Energy Econ. 62, 104–124. https://doi.org/https://doi.org/10.1016/j.eneco.2016.11.006

Surra, E., Bernardo, M., Lapa, N., Esteves, I., Fonseca, I., Mota, J.P., 2018. Maize cob waste pre-treatments to enhance biogas production through co-anaerobic digestion with OFMSW. Waste Manag. 72, 193–205. https://doi.org/https://doi.org/10.1016/j.wasman.2017.11.004

Tasnim, F., Iqbal, S.A., Chowdhury, A.R., 2017. Biogas production from anaerobic co-digestion of cow manure with kitchen waste and Water Hyacinth. Renew. Energy 109, 434–439. https://doi.org/https://doi.org/10.1016/j.renene.2017.03.044

Thyberg, K.L., Tonjes, D.J., 2017. The environmental impacts of alternative food waste treatment technologies in the U.S. J. Clean. Prod. 158, 101–108. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.04.169

Tonanzi, B., Gallipoli, A., Gianico, A., Montecchio, D., Pagliaccia, P., Di Carlo, M., Rossetti, S., Braguglia, C.M., 2018. Long-term anaerobic digestion of food waste at semi-pilot scale: Relationship between microbial community structure and process performances. Biomass and Bioenergy 118, 55–64. https://doi.org/https://doi.org/10.1016/j.biombioe.2018.08.001

Tong, H., Shen, Y., Zhang, J., Wang, C.-H., Ge, T.S., Tong, Y.W., 2018. A comparative life cycle assessment on four waste-to-energy scenarios for food waste generated in eateries. Appl. Energy 225, 1143–1157. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.05.062

Touzani, S., Granderson, J., Fernandes, S., 2018. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy Build. 158, 1533–1543. https://doi.org/https://doi.org/10.1016/j.enbuild.2017.11.039

Tyagi, V.K., Fdez-Güelfo, L.A., Zhou, Y., Álvarez-Gallego, C.J., Garcia, L.I.R., Ng, W.J., 2018. Anaerobic co-digestion of organic fraction of municipal solid waste (OFMSW): Progress and challenges. Renew. Sustain. Energy Rev. 93, 380–399. https://doi.org/https://doi.org/10.1016/j.rser.2018.05.051

Velásquez Piñas, J.A., Venturini, O.J., Silva Lora, E.E., Calle Roalcaba, O.D., 2018. Technical assessment of mono-digestion and co-digestion systems for the production of biogas from anaerobic digestion in Brazil. Renew. Energy 117, 447–458. https://doi.org/https://doi.org/10.1016/j.renene.2017.10.085

Woon, K.S., Lo, I.M.C., Chiu, S.L.H., Yan, D.Y.S., 2016. Environmental assessment of food waste valorization in producing biogas for various types of energy use based on LCA approach. Waste Manag. 50, 290–299. https://doi.org/https://doi.org/10.1016/j.wasman.2016.02.022

Xiong, D., Gui, Q., Hou, W., Ding, M., 2018. Gradient boosting for single image super-resolution. Inf. Sci. (Ny). 454–455, 328–343. https://doi.org/https://doi.org/10.1016/j.ins.2018.04.075

Xu, C., Shi, W., Hong, J., Zhang, F., Chen, W., 2015. Life cycle assessment of food waste-based biogas generation. Renew. Sustain. Energy Rev. 49, 169–177. https://doi.org/https://doi.org/10.1016/j.rser.2015.04.164

Yang, S., McDonald, J., Hai, F.I., Price, W.E., Khan, S.J., Nghiem, L.D., 2017. Effects of thermal pre-treatment and recuperative thickening on the fate of trace organic contaminants during anaerobic digestion of sewage sludge. Int. Biodeterior. Biodegradation 124, 146–154. https://doi.org/https://doi.org/10.1016/j.ibiod.2017.06.002

Ye, M., Liu, J., Ma, C., Li, Y.-Y., Zou, L., Qian, G., Xu, Z.P., 2018. Improving the stability and efficiency of anaerobic digestion of food waste using additives: A critical review. J. Clean. Prod. 192, 316–326. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.04.244

Zeynali, R., Khojastehpour, M., Ebrahimi-Nik, M., 2017. Effect of ultrasonic pre-treatment on biogas yield and specific energy in anaerobic digestion of fruit and vegetable wholesale market wastes. Sustain. Environ. Res. 27, 259–264. https://doi.org/https://doi.org/10.1016/j.serj.2017.07.001

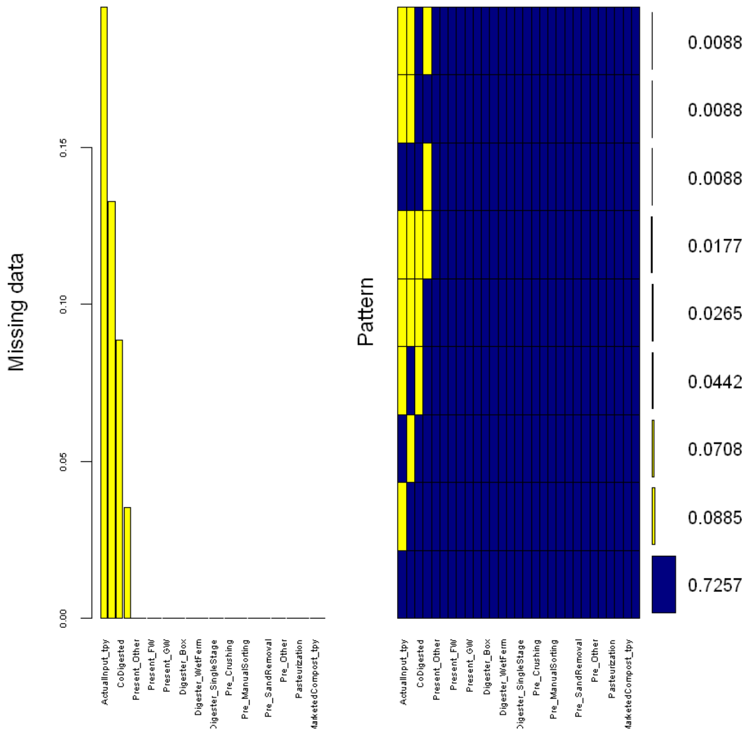
Zhang, J., Li, W., Lee, J., Loh, K.-C., Dai, Y., Tong, Y.W., 2017. Enhancement of biogas production in anaerobic co-digestion of food waste and waste activated sludge by biological co-pretreatment. Energy 137, 479–486. https://doi.org/https://doi.org/10.1016/j.energy.2017.02.163

Zhu, J., 2014. Quantitative Models for Performance Evaluation and Benchmarking, 3rd ed, International Series in Operations Research & Management Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-06647-9

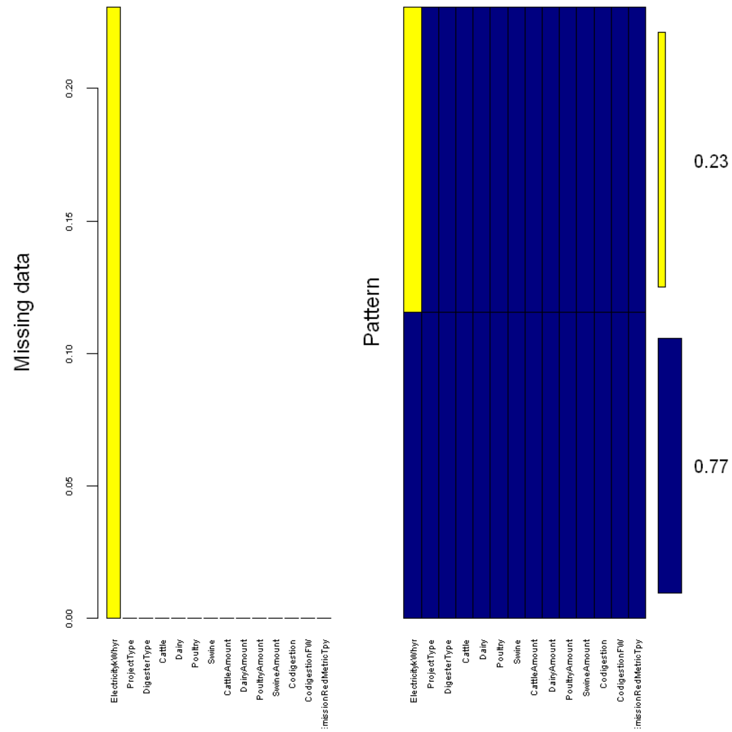
**Supplementary information**

The entire dataset and code used in this study can found online at the following GitHub URL: https://github.com/djavandeclercq/BiogasDEA_SGB. We welcome researchers exploring biogas efficiency to use the data and expand on/adapt the code for their own purposes, and hope that they in turn will also make their future analyses open-source and fully reproducible.

**S1. Missing values in each dataset**

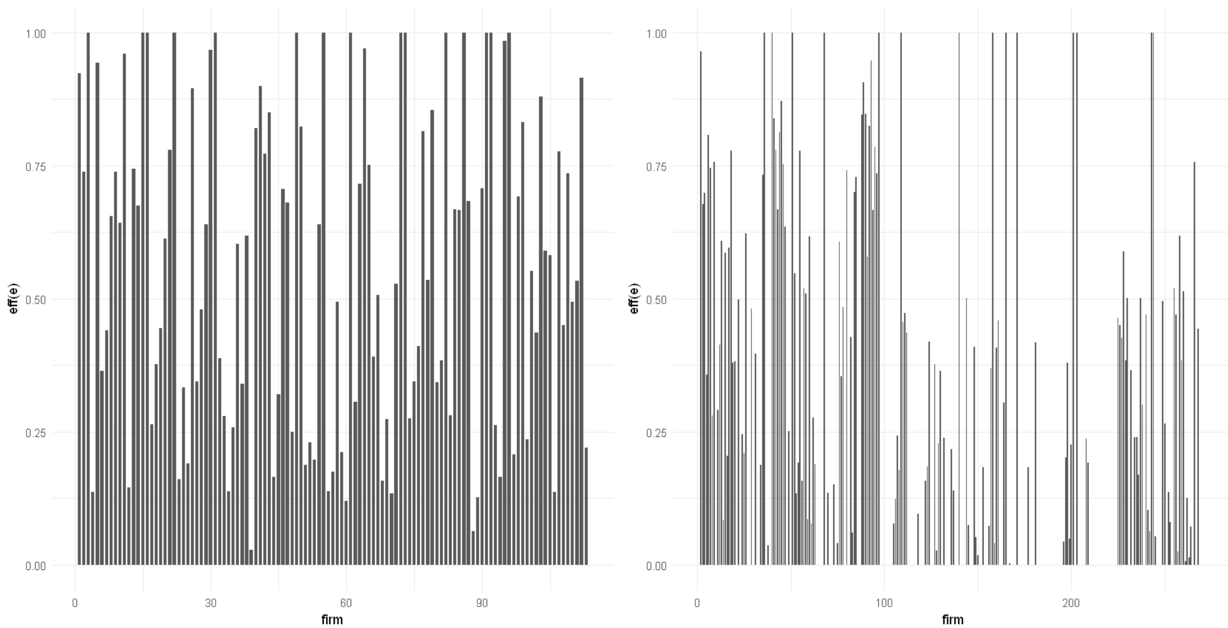



Germany




USA

## S2. Efficiency scores




Germany USA