Entwicklung von datengetriebenen Auswerteverfahren zur Analyse und Schätzung der Reaktorleistung von Biogasanlagen

Dissertation zur Erlangung des Doktorgrades

der Naturwissenschaften (Dr. rer. nat.)

Fakultät Naturwissenschaften

Universität Hohenheim

Institut für Lebensmittelwissenschaften und Biotechnologie

vorgelegt von

Tanja Beltramo

aus *Nishyn*

2020

Dekan: Prof. Dr. Uwe Beifuß

1. berichtende Person: Prof. Dr. Bernd Hitzmann
2. berichtende Person: Prof. Dr.-Ing. Jörg Hinrichs

Eingereicht am: 23.01.2020

Mündliche Prüfung am: 28.05.2020

Die vorliegende Arbeit wurde am 22. Januar 2013 von der Fakultät Naturwissenschaften der Universität Hohenheim als „Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften“ angenommen.

Danksagung

Diese Doktorarbeit wurde im Zeitraum von 2013 bis 2019 geschrieben. An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich auf diesem Wege unterstützt und motiviert haben.

Zuerst gebührt mein Dank Prof. Dr. Bernd Hitzmann und Prof. Dr.-Ing. Jörg Hinrichs, die meine Dissertation betreut und begutachtet haben. Einen großen Dank für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit!

Ich bedanke mich bei Dr. Cassiano Ranzan und Marius Nache für die zahlreichen interessanten Debatten und Ideen, die maßgeblich dazu beigetragen haben, dass diese Dissertation in dieser Form vorliegt. Außerdem bedanke ich mich bei Frau Dr. Monika Heiermann und Herrn Dr. Michael Klocke für die zur Verfügung gestellten Datensätzen, und bei Dr. Ulf Jeppsson für die Unterstützung bei der Implementierung der Simulationsmodelle ADM1.

Ein besonderer Dank gilt allen Kollegen bei dem Fachgebiet Prozessanalytik und Getreidetechnologie an der Universität Hohenheim für eine schöne Zeit während der Promotion und vielen neuen Freundschaften. Dr. Bianca Grote, Annika Hitzemann und Saskia Faassen - für angenehme Kaffeepausen und spannende Feierabende, Dr. Viktoria Zettel für die Freundschaft, Dr. Marc Stanke, Dr. Olivier Paquet-Durand für tiefgreifende Gespräche und tolle Filmabende, und Herbert Götz für die tolle Unterstützung bei den Versuchen.

Außerdem möchte ich mich bei Matthias Ulrich für das Korrekturlesen meiner Dissertation und für seine Geduld und Unterstützung in den Zeiten, in den ich an mich gezweifelt habe, bedanken.

Abschließend möchte ich mich bei meiner Familie bedanken, die mir immer zur Seite stand und meinen Hunden, Amadeus und Jack, für ihr Vertrauen.

Inhalt

Koautoren und ihr Beitrag

Einzelne Teile dieser Dissertation wurden bereits mit der Zustimmung des Doktorvaters Professor Dr. Bernd Hitzmann veröffentlicht. Die wissenschaftliche Arbeit in dieser Dissertation wurde teilweise in Kooperation mit den Koautoren aus der Universität Hohenheim, aus dem Leibnitz Institut für Agrartechnik und Bioökonomie und der Federal University of Rio Grande do Sul durchgeführt.

Publikation1:

"Artificial neural network prediction of biogas flow rate optimized with an ant colony algorithm"

Veröffentlicht im März 2016 in *Biosystems Engineering*, online verfügbar unter doi.org/10.1016/j.biosystemseng.2016.01.006

Mitautor 1: Dr. Cassiano Ranzan, Chemical Engineering Department of Federal University of Rio Grande do Sul, 90040-040, Porto Alegre, RS, Brasil

Aufgaben: Unterstützung bei der Implementierung des Ameisenkolonieoptimierungsalgorithmus.

Mitautor 2: Prof. Dr. Jörg Hinrichs, Universität Hohenheim, Stuttgart, Deutschland

Aufgaben: Methodische Unterstützung, Korrekturlesen der Publikation.

Aufgaben zum Eigenanteil: Implementierung der Modelle, Simulation der Daten, Durchführung der Vorhersage der Biogasausbeute, Schreiben des Artikels.

Publikation 2:

"Prediction of the biogas production using GA and ACO input features selection method for ANN model"

veröffentlicht im Januar 2019 in *Information Processing in Agriculture*, online verfügbar unter *doi.org/10.1016/j.inpa2019.01.002*

Mitautor: Dr. Michael Klocke, Leibnitz Institut für Agrartechnik und Bioökonomie, Potsdam 14469, Deutschland

Aufgaben: Koordinator des Forschungsprojekts, Erstellen der Datenbank mit Prozessdaten (das Verbundprojekt: „Prozessmikrobiologie in landwirtschaftlichen Biogasanlagen: Ermittlung der mikrobiellen Diversität sowie von hauptsächlichen verfahrenstechnischen Einflussfaktoren auf die Mikroflora: Schlussbericht zum Forschungsverbund BIOGAS-BIOCOENOSIS, mit den Förderkennzeichen BMEL 22010711 - 22028711 - 22028811 – 22028911 und Verbund-Nr. 01104265“, Korrektur des Artikels.

Aufgaben zum Eigenanteil: Implementierung der metaheuristischen Algorithmen zur Optimierung der Vorhersagemodelle, Berechnung der Modelle, Schreiben des Artikels.

Veröffentlichungen

Peer-reviewed Publikationen

Publikation 1:

“Artificial neural network prediction of biogas flow rate optimized with an ant colony algorithm”

Veröffentlicht im März 2016 in *Biosystems Engineering*

online verfügbar unter doi.org/10.1016/j.biosystemseng.2016.01.006

Autoren: Tetyana Beltramo, Cassiano Ranzan, Joerg Hinrichs, Bernd Hitzmann

Publikation 2:

“Prediction of the biogas production using GA and ACO input features selection method for ANN model”

veröffentlicht im Januar 2019 in *Information Processing in Agriculture*

online verfügbar unter *doi.org/10.1016/j.inpa2019.01.002*

Autoren: Tanja Beltramo, Michael Klocke, Bernd Hitzmann

Publikation 3:

“Evaluation of the linear and non-linear prediction models optimized with metaheuristics: application to anaerobic digestion processes”

Veröffentlicht im Juni 2019 in *Engineering in Agriculture, Environment and Food*

Online verfügbar unter doi.org/10.1016/j.eaef.2019.06.001

Autoren: Tanja Beltramo, Bernd Hitzmann

Präsentationen und Poster

Poster 1:

„Multivariate Datenanalyse zur Vorhersage der Biogasausbeute anhand chemischer und verfahrenstechnischer Prozessgrößen“

Infotag: „Von der Datenflut zum Erkenntnisgewinn – der richtige Umgang mit (Bio)-Prozessdaten“

am 29. Oktober, 2013, DACHEMA-Haus, Frankfurt am Main, Deutschland

Autoren: Tetyana Beltramo, Susanne Theuerl, Michael Klocke, Bernd Hitzmann

Poster 2:

“Estimation of Biogas Production Rate, a chemometrical approach”

9th German Conference on Chemoinformatics, International conference

10 – 12 November, 2013, Fulda, Deutschland

Autoren: Tetayna Beltramo, Susanne Theuerl, Michael Klocke, Bernd Hitzmann

Poster 3:

“Prediction of biogas production rate by means of multivariate data analysis”

Progress in Biogas III, International conference

10 – 11 September, 2014, Stuttgart, Deutschland

Autoren: Tetyana Beltramo, Susanne Theuerl, Cassiano Ranzan, Michael Klocke, Jorge O. Trierweiler, Bernd Hitzmann

Weitere Publikationen

Publikation 1:

"Estimation of the biogas production rate, a chemometrical approach"

Veröffentlicht am 11. März 2014, in *Journal of Cheminformatics*

Online verfügbar unter doi.org/10.1186/1758-2946-6-S1-P26

Authoren: Tetyana Beltramo, Susanne Theuerl, Michael Klocke, Bernd Hitzmann

Publikation 2:

„Prozessmikrobiologie in landwirtschaftlichen Biogasanlagen: Ermittlung der mikrobiellen Diversität sowie von hauptsächlichen verfahrenstechnischen Einflussfaktoren auf die Mikroflora: Schlussbericht zum Forschungsverbund BIOGAS-BIOCOENOSIS"

Veröffentlicht am 29. August, 2014, als Teilbericht des Abschlussberichtes mit den Förderkennzeichen BMEL 22010711 - 22028711 - 22028811 – 22028911 und Verbund-Nr. 01104265 in *Bornimer Agrartechnische Berichte, Heft 84*

Online verfügbar unter doi.org/10.2314/GBV: 882560409

Authoren: Michael Klocke, Susanne Theuerl, Irena Maus, Andreas Schlüter, Dirk Benndorf, Fabian Kohrs, Udo Reichl, Monika Heiermann, Robert Kausmann, Tetyana Beltramo, Bernd Hitzmann.

Publikation 3:

"Supervision of food manufacturing processes using optical process analyzers – an overview"

Veröffentlicht am 26. September 2016, in *ChemBioEng*

Online verfügbar unter doi.org/10.1002/cben.201600013

Authoren: Viktoria Zettel, Muhammad Haseeb Ahmad, Tetyana Beltramo, Bernhard Hermannseder, Annika Hitzemann, Marius Nache, Oliver Paquet-Durand, Thomas Schöck, Florian Hecker, Bernd Hitzmann

Publikation 4:

„*Mathematisch aufgeklärt*“

Veröffentlicht im September 2016, in *Labor Praxis*

Authoren: Tetyana Beltramo, Bernd Hitzmann

1 Zusammenfassung

1.1 Deutsche Version

Die Biogasherstellung ist ein sehr komplexer Prozess, der durch das Zusammenspiel unterschiedlicher Mikroorganismen und einem mehrstufigen Prozessablauf gekennzeichnet ist. Variationen in der Mikroflora können durch das verwendete Substrat, die Beschickung sowie die Temperatur gegeben sein, wobei ein Großteil der beteiligten Mikroorganismen sogar unbekannt sein kann. Die Überwachung und Analyse solcher Prozesse sind zumeist sehr zeit- und kostenintensiv.

Die VDI-Richtlinie 4630 beschreibt analytische Methoden zur Untersuchung der Vergärung organischer Stoffe, Substratcharakterisierung, Probenahme, Stoffdatenerhebung und Gärversuche, die zur Optimierung von Biogasanalgen dienen. Die in der Richtlinie beschriebenen Methoden erfordern jedoch eine entsprechende Laborausstattung, dafür ausgebildetes Personal und sind sehr aufwändig. Deshalb ist eine moderne Alternative zur Prozessanalyse und -optimierung dringend erforderlich, um gezielt wesentliche Prozessgrößen schnell, effizient und kostengünstig zu identifizieren.

Das Ziel der vorliegenden Dissertation ist es eine Methode zu entwickeln, welche eine schnelle und effiziente Analyse der Biogasproduktionsprozesse ermöglicht. In dieser Arbeit soll mithilfe von computerbasierten Verfahren untersucht werden, ob wesentliche Prozessgrößen identifiziert werden können, deren Messung für die Führung der Biogasanlagen ausschlaggebende Informationen zur Verfügung stellt. Neben mathematischen Modellen werden auch Optimierungsalgorithmen für die Identifikation der wesentlichen Einflussgrößen verwendet.

Die für die Modellierung verwendeten Datensätze wurden experimentell im Rahmen des Projektes ,,Biogas Biocoenosis‘‘ (FNR 22010711, Dr. Michael Klocke, Leibniz-Institut für Agrartechnik und Bioökonomie e. V., Potsdam) und „Biogas-Enzyme“ (FKZ 22027707, Dr. Monika Heiermann, Leibniz-Institut für Agrartechnik und Bioökonomie e. V., Potsdam) erhoben. Zusätzlich wurden Prozessdaten mithilfe des *Anaerobic Digestion Modell No.1(ADM1)* simuliert.

Die chemischen Prozessgrößen stellen die unabhängigen Prozessvariablen dar. Die Biogasausbeute wird als die abhängige Zielprozessgröße verwendet. Die Vorhersage der Biogasausbeute wird mithilfe

linearer und nicht linearer mathematischer Modelle durchgeführt. Dafür werden die *Partial-Least-Squares-Regression (PLSR)*, die *Locally-Weighted-Regression (LWR)* und *künstliche neuronale Netzte (ANN, Artificial Neural Networks)* implementiert. Um die wesentlichen unabhängigen Prozessvariablen zu identifizieren, werden folgende Optimierungsalgorithmen verwendet: ein *Ameisen-Kolonie-Optimierungsalgorithmus (ACO)* und der *genetische Algorithmus (GA)*.

Die Qualität der Vorhersage wird anhand des Vorhersagefehlers (*Root Mean Square Error, RMSE)* und des *Bestimmtheitsmaßes (R^2)* bewertet.

Anhand der entwickelten Methode konnte die gewünschte Zielgröße, Biogasausbeute, vorhergesagt werden. Die durchgeführte Variablenselektion mittels metaheuristischer Optimierungsverfahren verbesserte die Vorhersagekraft der Modelle und reduzierte gleichzeitig die Anzahl der unabhängigen Prozessvariablen. Als wesentliche Prozessvariablen wurden folgende identifiziert: hydraulische Verweilzeit, Trockensubstanz, Gehalt an Neutral-Detergentien-Faser, Säure-Detergentien-Lignin-Gehalt und n-Buttersäure. Die besten Vorhersageergebnisse wurden mit Hilfe der künstlichen neuronalen Netze erzielt. Die Modelle hatten den kleinsten Fehler der Vorhersage und das höchste Bestimmtheitsmaß.

Die erfolgreiche Umsetzung der mathematischen Verfahren zeigt, dass moderne mathematische Methoden eine gute Alternative für die Analyse und Optimierung komplexer biologischer Prozesse darstellen. Für eine abschließende Bewertung sind jedoch auch weitere experimentelle Untersuchungen notwendig, in denen die Prozessführung basierend auf den identifizierten Prozessvariablen durchgeführt wird.

1.2 Englische Version

The production of biogas is very complex process, which runs in some stages involving different microorganisms. Microbiological diversity of the process depends mainly on the composition of substrate and ambient conditions, such as process temperature. The fact is, the development and composition of the microbiological communities of the process are difficult to predict. Thus, the control and evaluation of such complex biological processes are very time consuming and expensive. In Germany the evaluation of the biogas plants can be performed according to the VDI-Norm 4630, which describes the methods for the evaluation of fermentation of organic materials including characterization of the substrate, sampling, collection of material data and fermentation tests. For that specially equipment and skilled personnel are required. Moreover, the evaluation procedure is very time consuming. That is why a new state-of-the-art alternative for the evaluation purposes is necessary to simplify and to speed up the assessment of the biogas production processes.

The aim of this doctoral thesis is the development of a fast and reliable method for the evaluation of the biogas production processes. Therefore the mathematical modelling should identify significant process variables able to evaluate the whole process. For the optimization of mathematical models metaheuristic tools were used. In this doctoral thesis two different data sets were used – experimental data and simulated data. The experimental data were collected in projects “Biogas-Biocoenosis” (FKZ 22010711, Dr. Michael Klocke, Leibnitz-Institute für Agrartechnik und Bioökonomie e.V., Potsdam) and “Biogas-Enzyme” (FKZ 22027707, Dr. Monika Heiermann, Leibnitz-Institute für Agrartechnik und Bioökonomie e.V., Potsdam). The simulated data set was generated using the Anaerobic Digestion Model No.1 (ADM1). The chemical process variables were used as the independent process variable set, while the biogas production output represented the dependent process variable. Prediction of the biogas production was done using linear and nonlinear mathematic models. Here, Partial-Least-Square-Regression (PLSR), Locally-Weighted-Regression (LWR) and Artificial Neural Networks (ANN) were implemented. In order to identify the most significant undependable process variables optimization algorithms were used, Ant Colony Optimization (ACO) and Genetic Algorithm (GA). Prediction capacity was evaluated using two model evaluation variables, Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). Figure 1 in Supplementary represents the flow chart of the developed

methodology applied for ADM1 generated data set. In Figure 2 (Supplementary) there is a flow chart of the developed methodology applied for the experimentally collected data.

The developed approaches could be successfully used for the prediction of the desired process variable, biogas production rate. The variable selection done with the help of metaheuristic optimization algorithms improved the prediction results and reduced number of the independent process variables. Hydraulic retention time, dry matter, neutral detergent fibre, acid detergent fibre and n-butyric acid were identified as the most significant ones. The best prediction was obtained using ANN models. Here, the error of prediction was low and the coefficient of determination high.

The successful implementation of the developed methodology proved mathematical models to be an effective alternative method capable to evaluate and to optimize complicated biological processes. Furthermore, it would be mandatory further experimental evaluation of the developed strategy, using the model-based process information.

2 Einführung

2.1 Einleitung

2.1.1 Energieversorgung

Die Nachhaltigkeit der Energieversorgung gewinnt heutzutage immer mehr an Bedeutung. Die wachsende Weltbevölkerung, steigende Lebensqualität der Industrieländer, der Klimawandel und die fast erschöpften Ressourcen an fossilen Brennstoffen haben in den letzten Jahrzehnten zu einer starken Entwicklung alternativer Verfahren der Energieerzeugung beigetragen (Holm-Nielsen et al. 2009). Energie aus nachwachsenden Rohstoffen, wie zum Beispiel Solarenergie, Wasserenergie, Windenergie, Bioenergie und Geothermie, steht im Fokus der zukünftigen Energiegewinnung.

2.1.2 Das Erneuerbare-Energien-Gesetz (EEG)

In Deutschland ist das Erneuerbare-Energien-Gesetz der zentrale Baustein zur Umsetzung der Nutzung erneuerbarer Energiequellen. Es trat am 4. April 2000 in Kraft (BMU, 2000) und wurde unter EEG 2004, EEG 2009, EEG 2012, PV-Novelle, EEG 2014 und EEG 2017 weiterentwickelt (EEG). Im Wortlaut besagt das Gesetzt: „ Ziel dieses Gesetzes ist es, im Interesse des Klima- und Umweltschutzes eine nachhaltige Entwicklung der Energieversorgung zu ermöglichen und den Beitrag Erneuerbarer Energien an der Stromversorgung deutlich zu erhöhen…‘‘. Das aktuelle Ziel des Gesetzes ist es den Anteil der Stromversorgung aus erneuerbaren Energiequellen bis 2050 auf 80 % auszubauen. Dafür sollte eine technologische Weiterentwicklung in diesem Bereich vorangetrieben werden.

2.1.3 Prozessphasen der Biogaserzeugung

Biogas gehört zu den Klima und Umwelt schonenden Methoden der Energiegewinnung (Nallathambi Gunaseelan 1997). Dabei entstehendes Biogas wird für die Erzeugung von Kraft, Wärme und Strom verwendet (Weiland 2010).

Biogas ist ein Gasgemisch, welches als ein Produkt bei einem anaeroben mikrobiellen Abbau der Biomasse entsteht. In dem Prozess sind eine Vielzahl unterschiedlicher Mikroorganismen beteiligt mit einer noch höheren Anzahl biochemischer Reaktionen und physikalisch-chemischer Einflussgrößen (Nielsen und Angelidaki 2008). Die Hauptkomponenten des Gasgemisches sind Methan (50 -75 %) und

Kohlendioxid (25 – 50 %). Aber auch andere Gase, wie zum Beispiel Wasserstoff, Schwefelwasserstoff und Ammoniak, entstehen während des Prozesses. Die Bildung von Methan aus der Biomasse kann durch die folgende Gleichung beschrieben werden (Gleichung 1).

$$C_cH_hO_oN_nS_s + \gamma H_2O \rightarrow xCH_4 + nNH_3 + sH_2S + (c - x)CO_2 \qquad Gleichung\ 1$$

Die Koeffizienten sind definiert als (Deublein und Steinhauser 2011):

$$x = 1/8(4c + h - 2o - 3n - 2s) \qquad Gleichung\ 2$$

$$\gamma = 1/4(4c - h - 2o + 3n + 2s) \qquad Gleichung\ 3$$

Der Biogasbildungsprozess läuft in vier Phasen ab und wird durch zahlreichen mikrobiologische, biochemische und physikalische Prozessgrößen bestimmt. Die dabei beteiligten Mikroorganismen sind oft strikt anaerob, wie zum Beispiel *Clostridia* und *Bifidobacteriaceae*, aber fakultativ anaerobe Mikroorganismen sind teilweise auch in dem Prozess aktiv, wie zum Beispiel *Streptokokken* und *Enterobakterien*. In Abbildung 1 sind Prozessphasen und die beteiligten Mikroorganismengruppen dargestellt.

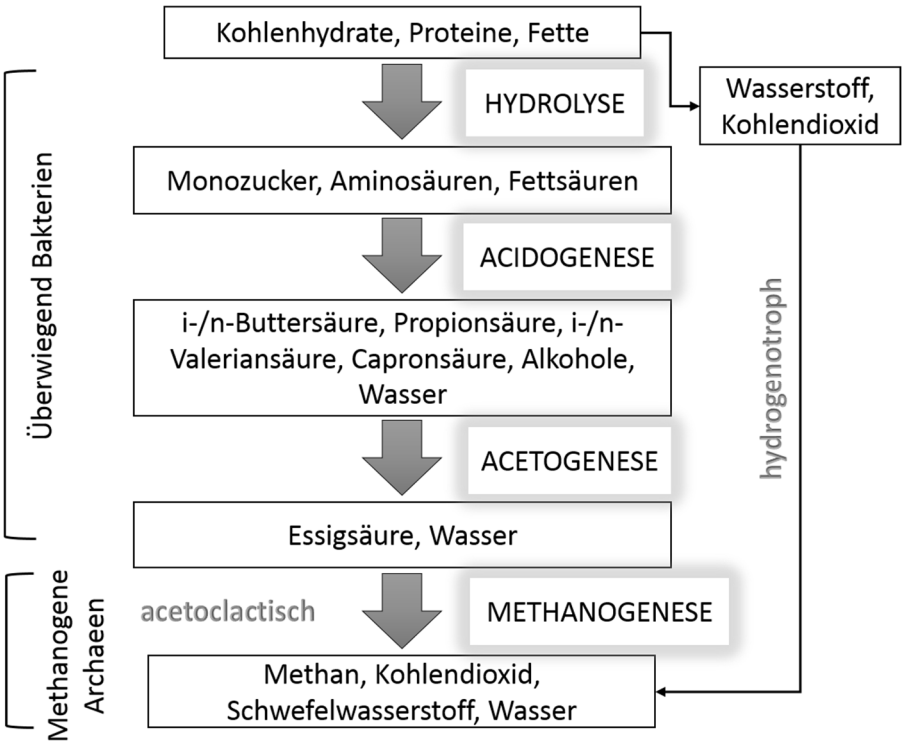

Abbildung 1: Stufen des anaeroben Biomassenabbaus nach Batstone at al., (2002).

Die erste Phase des Prozesses ist die Hydrolyse, bei der die komplexen Verbindungen des Substrates in ihre einfachen Bausteine umgewandelt werden (Wang et al. 2015). Die langkettigen Kohlenhydrate werden durch Hydrolasen, wie zum Beispiel Cellulase, Cellobiase und Xylanase, in die kurzen Bausteine zerlegt. Die Proteine werden durch die Proteasen zu Aminosäuren, während die Fette durch die Lipasen zu Fettsäuren abgebaut werden (Gleichungen 4, 5, 6).

$$C_6H_{12}O_6 \rightarrow 3CO_2 + 3CH_4 \qquad \textit{Gleichung 4}$$

$$C_{12}H_{24}O_6 + 3H_2O \rightarrow 4.5\ CO_2 + 7.5\ CH_4 \qquad \textit{Gleichung 5}$$

$$C_{13}H_{25}O_7N_3S + 6H_2O \rightarrow 6.5\ CO_2 + 6.5\ CH_4 + 3NH_3 + H_2S \qquad \textit{Gleichung 6}$$

Die Geschwindigkeit der Hydrolyse ist von der Substratzusammensetzung, Molekülgröße, Mikroorganismenaktivität und den technologischen Prozessbedingungen abhängig. Zum Beispiel braucht der Abbau der Kohlenhydrate nur wenige Stunden, bis auf die Ausnahmen von Cellulose und Hemicellulose, welche sehr langsam und nicht vollständig abgebaut werden können. Die Hydrolyse von Fetten und Proteine dagegen dauert mehrere Tage.

Die zweite Phase der Biogasherstellung ist die Acidogenese. Hierbei werden die Hydrolyse Abbauprodukte zu organischen Säuren mit der Kettenlänge C_1-C_5 umgewandelt. Dabei entstehen i-/n-Buttersäure, Propionsäure, i-/n-Valeriansäure und Essigsäure. Die einfachen Zucker werden als erstes zum Pyruvat und dann durch Milchsäurebakterien zur Milchsäure oder durch Hefe zum Ethanol umgewandelt. Dabei entstandene Fettsäuren werden schrittweise zu Essigsäure abgebaut. Aminosäuren werden zu Essigsäure, Ammonium und Kohlendioxid abgebaut.

Die dritte Phase ist die Acetogenese, bei der die essigbildenden Mikroorganismen aktiviert werden. Sie wandeln die Zwischenprodukte aus der Acidogenese in Essigsäure um. Alle acetogene Bakterien produzieren neben der Essigsäure auch Wasserstoff. Ein spezieller Fall stellt die Bildung der Essigsäure aus Kohlendioxid und Wasserstoff dar. Dies geschieht durch homo-acetogenen Bakterien, wie zum Beispiel *Acetobacterium woodii*, (Cheng 2018). In Tabelle 1 sind chemische Reaktionen der Essigsäurebildung dargestellt.

Tabelle 1: Chemische Reaktionen der Acetogenese.

Chemische Name	**Chemische Reaktion**
Kohlendioxid/ Wasserstoff	$2CO_2 + 4H_2 \rightarrow CH_3COOH + 2H_2O$
Propionsäure	$CH_3(CH_2)COOH + 2H_2O \rightarrow CH_3COOH + CO_2 + 3H_2$
Buttersäure	$CH_3(CH_2)_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2$
Valeriansäure	$CH_3(CH_2)_3COOH + 2H_2O \rightarrow CH_3COOH + CH_3CH_2COOH + 2H_2$
Isovaleriansäure	$(CH_3)_2CHCH_2COOH + HCO_3H + H_2O \rightarrow 3CH_3COOH + 5H_2$
Capronsäure	$CH_3(CH_2)_4COOH + 4H_2O \rightarrow 3CH_3COOH + 5H_2$
Glycerin	$C_8H_8O_3 + H_2O \rightarrow CH_3COOH + 3H_2 + CO_2$
Milchsäure	$CH_3CHOHCOOH + 2H_2O \rightarrow CH_3COOH + HCO_3^- + H^+ + 2H_2$
Ethanol	$CH_3(CH_2)OH + H_2O \rightarrow CH_3COOH + 2H_2$

Die letzte Phase des Biogasprozesses ist die Methanogenese. In dieser Phase sind ausschließlich strikt anaerobe Mikroorganismen aktiv, welche die gebildeten Zwischenprodukte in Methan umsetzen. Dieser Prozessschritt kann auf verschiedene Arten umgesetzt werden. Zum Beispiel, produzieren hydrogenotrophe Methanbildner Methan aus Wasserstoff und Kohlendioxid, während acetoclactische Methanbildner Essigsäure spalten. Die chemischen Reaktionen der Methanogenese sind in Tabelle 2 dargestellt.

Tabelle 2: Chemische Reaktionen der Methanogenese.

Substrat-Typ	**Chemische Reaktion**
Kohlendioxid und Wasserstoff	$4H_2 + HCO_3^- + H^+ \rightarrow CH_4 + 3H_2O$
	$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$
	$4HCOO^- + H_2O + H^+ \rightarrow CH_4 + 3HCO_3^-$
Essigsäure	$CH_3COO^- + H_2O \rightarrow CH_4 + HCO_3^-$
Methanol	$4CH_3OH \rightarrow 3CH_4 + HCO_3^- + H^+ + H_2O$
	$CH_3OH + H_2 \rightarrow CH_4 + H_2O$
	$2CH_3CH_2OH + CO_2 \rightarrow CH_4 + 2CH_3COOH$

Bei den typischen landwirtschaftlichen Biogasanlagen erfolgt die Methanbildung aufgrund der hohen Raumbelastung hauptsächlich aus Wasserstoff. Die Raumbelastung gibt an, wie viel von organischer Trockensubstanz dem Fermenter pro Zeiteinheit und Arbeitsvolumen zugeführt werden kann (Gleichung 7).

$$B_R = \frac{\dot{m} \cdot c}{V_R \cdot 100} \; [kg\ oTS\ m^{-3} d^{-1}] \qquad \textit{Gleichung 7}$$

B_R ist die Raumbelastung, c die Konzentration der organischen Substanz [% oTS], $\dot{m}$ ist zugeführte Substratmasse pro Zeiteinheit $[kgd^{-1}]$ und V_R ist das Reaktorvolumen $[m^3]$. Bei den Hochlastfermentern mit kurzen Verweilzeiten (Gleichung 8) entsteht Methan zu 70 % durch die Spaltung der Essigsäure und zu 30 % aus Wasserstoff.

$$HTR = \frac{V_R}{\dot{V}} \; [d] \qquad \textit{Gleichung 8}$$

HTR ist die hydraulische Verweilzeit, V_R das Reaktorvolumen $[m^3]$ und $\dot{V}$ ist das täglich zugeführte Substratvolumenstrom $[m^3 d^{-1}]$.

2.1.4 Prozessgrößen der Biogasherstellung

Typischerweise finden die vier Stufen des anaeroben Biomassenabbaus bei einer einstufigen Fermentation parallel statt und laufen mit unterschiedlichen Geschwindigkeiten ab, wodurch das

dynamische Verhalten des Prozesses von der Substratzusammensetzung abhängt. Die dabei beteiligten Mikroorganismen vereinen sich in komplexe Lebensgemeinschaften (Sekiguchi et al. 2001). Dies macht die Prozessdarstellung sehr komplex, da eine Vielzahl an verschiedenen Prozessgrößen berücksichtigt werden müssen. Praktische Erfahrungen zeigen, dass eine zuverlässige Prozessbeschreibung nur erreicht werden kann, wenn die komplette Prozesskette von der Substratzugabe bis hin zu den in allen Prozessphasen gebildeten Zwischenprodukten erfasst wird. Das macht die Überwachung und Steuerung des Prozesses komplex und sehr zeitintensiv. Die Prozessgrößen können in zwei Gruppen aufgeteilt werden: die biologisch-chemischen Prozessgrößen und die verfahrenstechnischen Prozessgrößen. Die Prozessgrößen sind voneinander abhängig und bilden ein sehr komplexes Netzwerk. Zu den biologisch-chemischen Prozessgrößen gehören Substratzusammensetzung, Reaktionsgeschwindigkeiten, pH-Wert, Nährstoffversorgung der Mikroorganismen, Konzentration der Zwischenprodukte, wie zum Beispiel organische Säuren und Ammoniak. Eine zentrale Rolle bei dem Prozess spielen die gebildeten Zwischenprodukte. Diese können das Wachstum und die Aktivität bestimmter Mikroorganismenstämme während der Prozessphasen hemmen (Francisci et al. 2015), was die Bildung unerwünschter Produkte oder sogar den Prozessabbruch verursachen kann. Die verfahrenstechnischen Prozessgrößen umfassen Raumbelastung, Verweilzeit, die Art der Durchmischung, Prozesstemperatur und Sauerstoffkonzentration im Fermenter und stellen eine technologische Voraussetzung für den Prozess dar. Diese haben ebenfalls einen direkten Einfluss auf das Wachstum und die Aktivität der Mikroorganismen (Karakashev et al. 2005)Eine besondere Gruppe verfahrenstechnischer Prozessgrößen bilden die hydraulische Verweilzeit und Raumbelastung, welche zu den so genannten variablen Prozessgrößen gehören und diese können variabel eingestellt werden, solange die Substratzusammensetzung konstant angehalten wird. Unter diesen Bedingungen sind die beiden Prozessgrößen miteinander gekoppelt. Eine Entkopplung kann durch die Rückführung des Gärstoffes erreicht werden.

2.2 Materialien und Methoden

2.2.1 Herkunft experimenteller Daten

Die Datengrundlage der vorliegenden Dissertation stammt aus dem Verbundprojekt „*Prozessmikrobiologie in landwirtschaftlichen Biogasanlagen*‘‘. Das Projekt wurde von dem Bundesministerium für Ernährung und Verbraucherschutz sowie dem Projektträger Fachagentur Nachwachsende Rohstoffe e.V. (FNR) unterstützt, die Förderkennzeichen der Teilprojekte sind FKZ 22028911 und FKZ 22010711 (Dr. Michael Klocke, Leibnitz Institut für Agrartechnik und Bioökonomie, Potsdam). Das Ziel des Projektes war es, den Einfluss der Prozessgrößen auf das Wachstum und die Aktivität der mikrobiologischen Lebensgemeinschaften zu untersuchen. Das Projekt „Biogas-Enzyme“, mit dem Förderkennzeichen FKZ 22027707 (Dr. Monika Heiermann, Leibnitz Institut für Agrartechnik und Bioökonomie, Potsdam) befasste sich mit der Erfassung der experimentellen Daten im Rahmen des Gesamtprojektes. Die Analyse der untersuchten landwirtschaftlichen Biogasanlagen wurde nach dem Deutschen Standard VDI 4630 durchgeführt (VDI-Fachbereich Energietechnik, 2016). Die Biogasanlagen wurden im Laufe von 18 Monaten wöchentlich beprobt. Für diese Dissertation wurden die Offlinemessdaten verwendet, gemessen bei einer Biogasanlage mit mesophilen Prozessbedingungen. Das eingesetzte Substrat setzte sich aus Mais, Grassilage, Rindergülle und Schweinemist zusammen. Der zur Verfügung gestellte Datensatz umfasste 15 unabhängigen Prozessgrößen und eine abhängige Zielgröße, die Biogasausbeute. Alle Prozessgrößen bis auf die Ausnahme der hydraulischen Verweilzeit (Hydraulic retention time, HRT) wurden offline experimentell bestimmt. HRT wurde zusätzlich berechnet. Die verwendeten Prozessmessgrößen sind in Tabelle 3 dargestellt.

Table 3: Die verwendeten Prozessmessgrößen.

Prozessgrößen **Deutsch**	**Prozessgrößen (Abkürzung)** **Englisch**	**Einheit**
Essigsäure	Acetic acid (AcA)	$g\ l^{-1}$
Saure Detergentien Faser	Acid detergent fibre (ADF)	$g\ kg^{-1}\ VS^{-1}$
Saures Detergentien Lignin	Acid detergent lignin (ADL)	$g\ kg^{-1}\ VS^{-1}$

Alkalität (Flüchtige organische Säuren/ Hydrogencarbonatpuffer)	Alkalinity ration (FOS/TAC)	-
Ammoniak	Ammonium (NH_4^+-N)	g l^{-1}
n-Buttersäure	n-butyric acid (nBA)	g l^{-1}
Iso-Buttersäure	iso-butyric acid (iBA)	g l^{-1}
Gesamttrockensubstanz	Total solids (TS)	% Fresh matter (FM)
Hydraulische Verweilzeit	Hydraulic retention time (HRT)	day
Neutrale Detergentien Faser	Neutral detergent fibre (NDF)	g kg^{-1} VS^{-1}
Flüchtige Substanzen	Volatile solids (VS)	% TS
Organische Raumbelastung	Organic loading rate (OLR)	kg VS $m^{-3}d^{-1}$
Propionsäure	Propionic acid (PA)	g l^{-1}
n-Valeriansäure	n-valeric acid (nVA)	g l^{-1}
Iso-Valeriansäure	Iso-valeric acid (iVA)	g l^{-1}

Im Rahmen des Projektes (,,Biogas Biocoenosis‘‘ (FNR 22010711, Dr. Michael Klocke, Leibniz-Institut für Agrartechnik und Bioökonomie e. V., Potsdam) und „Biogas-Enzyme“ (FKZ 22027707, Dr. Monika Heiermann, Leibniz-Institut für Agrartechnik und Bioökonomie e. V., Potsdam)) wurden mehrere Datensätze bei den untersuchten Biogasanlagen erhoben und ausgewertet.

2.2.2 Datensimulation mit Hilfe des ADM1 Modells

In der Literatur gibt es eine Reihe verschiedener Modelle, welche für die Simulation physikalischer, chemischer und biologischer Prozesse verwendet wurden (Lauwers et al. 2013). Im Bereich der Biogasherstellung wurde das erste Modell, *Activated Sludge Model 1 (ASM1),* bereits im Jahr 1987 von Henze et al. vorgestellt. Das Prinzip des *ASM1* Modells ist die Grundlage der erweiterten und später vorgestellten Modelle, wie zum Beispiel *Activated Sludge Model No. 2* (*ASM2*) (Henze 2007), *Benchmark Simulation Model No.2* (*BSM2*) (Rosen et al. 2006) und *Anaerobic Digestion Model No.1(ADM1) (Batstone 2002).* Das bekannteste Modell in Bereich der Biogasherstellung ist *Anaerobic Digestion Model No.1* (ADM1). Dieses Modell wurde von den Wissenschaftlern aus verschiedenen Fachbereichen im Auftrag der *IWA Task Group for Mathematical Modelling of Anaerobic Digestion*

Processes entwickelt und im Jahr 2002 veröffentlicht (Batstone 2002). ADM1 beschreibt die vier Phasen des anaeroben Abbaus der Biomasse bis zur Bildung vom Biogas. Die dazu gehörigen hemmenden Reaktionen werden für jeden Prozessschritt separat beschrieben. Die Berechnungen beziehen sich auf den chemischen Sauerstoffbedarf. Diese Größe beschreibt die Abbaubarkeit der Biomasse unter den angegebenen Prozessbedingungen. Das Modell ist universell und kann an einen konkreten Prozess unter Berücksichtigung der Substratzusammensetzung und der Temperatur (thermophiler oder mesophiler Prozessführung) angepasst werden (Derbal et al. 2009; Ramirez et al. 2009). Der allgemeine Prozessablauf des anaeroben Abbaus ist in Abbildung 2 dargestellt.

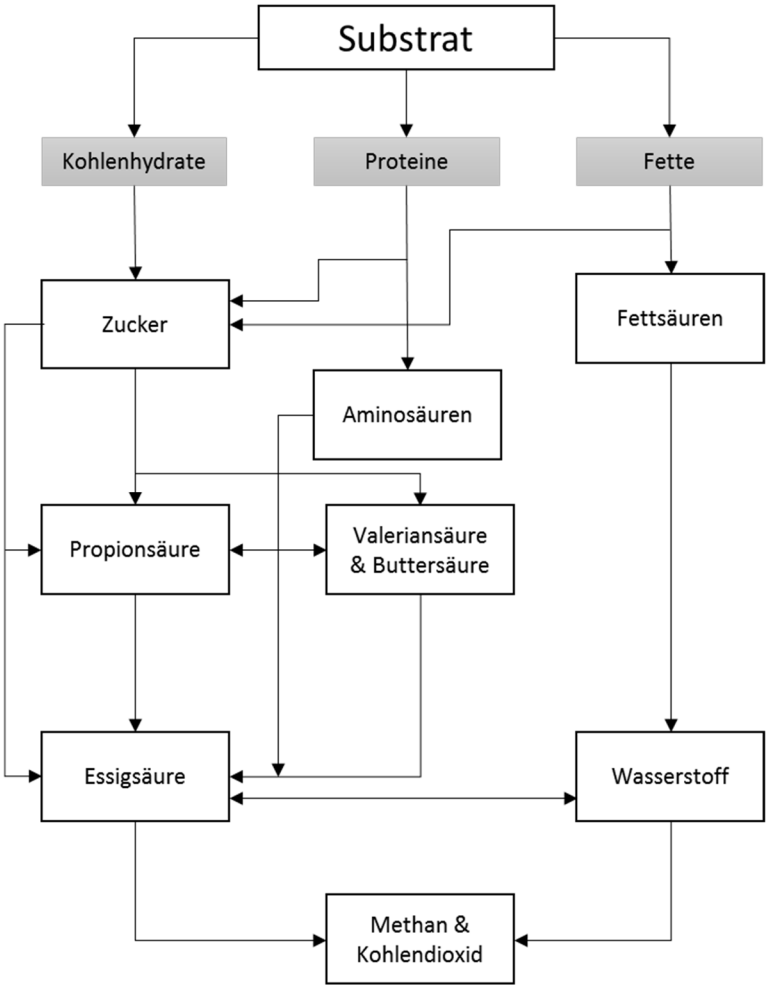

Abbildung 2: Prozess des anaeroben Abbaus in ADM1 nach Batstone et al., 2002.

Für die Simulation der Prozessdaten in dieser Dissertation wurde das in Benchmark Simulation Model No 2 (BSM2) integrierte ADM1 Modells verwendet (Rosen et al. 2006). Mit diesem Modell kann der Prozessverlauf für 100 - 300 Tage mit einer Frequenz von 20 Simulationsmesswerten pro Tag in einem

Fermenter mit 3400 m^3 simuliert werden. Die Substratzusammensetzung wurde kontinuierlich variiert. Es wurde zwischen der Gülle und Grassilage gewechselt. Die Daten für die Substartcharakterisierung wurden aus der Literatur genommen (Wichern et al. 2009; Zhou et al. 2011). Mit Hilfe des Models wurden die Konzentrationen an gebildeten Zwischenprodukten, Biogasausbeute, Kohlenstoffdioxid und Wasserstoff simuliert.

2.2.3 Vorhersage der Biogasausbeute mit Hilfe mathematischer Modelle

Für die Vorhersage der Biogasausbeute wurden lineare und nicht lineare mathematische Modelle verwendet, wie zum Beispiel Regressionsmodelle und künstliche neuronale Netze. Wobei die Regression der partiellen kleinsten Quadrate (Partial Least Square Regression, PLSR) zu den linearen mathematischen Modellierungsverfahren zählt, während die lokal gewichtete Regression (Locally Weighted Regression, LWR) und künstliche neuronale Netze (Artificial Neural Networks, ANN) nicht lineare Modelle darstellen.

PLSR ist ein multivariates mathematisches Verfahren, bei dem Prädiktoren auf eine kleinere Gruppe von unkorrelierten Komponenten minimiert werden. Dabei wird eine Regression der kleinsten Quadrate für diese Komponenten anstelle der ursprünglichen Daten berechnet. Der Vorteil dieses Models im Vergleich zu der Hauptkomponentenanalyse (Principal Component Regression, PCR) ist, dass die Korrelation zwischen den unabhängigen Prädikatoren und der abhängigen vorhergesagten Größe berücksichtigt wird. Dabei wird die Anzahl der Prädikatoren reduziert, indem nur die Prädikatoren, welche mit der vorhergesagten Größe am stärksten korrelieren, in dem Modell berücksichtigt werden.

PLSR wird oft für die Vorhersage chemischer, biologischer und pharmazeutischer Größen verwendet, um die Korrelation und die Zusammenhänge zwischen den Größen zu untersuchen.

Das Leave-One-Out Kreuzvalidierungsverfahren (LOO-KV) wurde in dieser Dissertation verwendet, um die Güte der PLSR Modelle zu quantifizieren. Das Ziel der Kreuzvalidierung ist das Überprüfen von Modellen, wenn nur wenige Daten zum Kalibrieren und Testen zur Verfügung stehen. Dabei wird die Datenmatrix in einzelne Teilmengen aufgeteilt, welche für die Kalibrierung und andere für den Test verwendet werden. LOO-KV ist eine spezielle Art, bei der die Anzahl der Proben gleich der Anzahl der Teilmengen ist. Somit wird jede Probe einmal zum Testen verwendet, was eine Überprüfung jeder

Teilmenge ermöglicht. Die Anzahl der Hauptkomponenten in den Modellen wurde variiert, um die optimale Modelldimension zu ermitteln. Die Vorhersage wurde anhand der simulierten Daten durchgeführt. Für die Bewertung der Vorhersagekraft der Modelle wurden zwei Modelparameter herangezogen, die Wurzel der mittleren Fehlerquadratsumme (Root Mean Square Error, RMSE) und das Bestimmtheitsmaß (Coefficient of Determination, R^2). Die Berechnung beider Größen ist in den Gleichungen 9 und 10 dargestellt.

$$RMSE = \sqrt{\left(\sum_{k=1}^{n}\left(\widehat{y_k} - y_{s,k}\right)^2\right)/n} \qquad \text{Gleichung 9}$$

$$R^2 = 1 - \left(\sum_{k=1}^{n}\left(\widehat{y_k} - y_{s,k}\right)^2\right) / \left(\sum_{k=1}^{n}\left(y_{s,k} - \bar{y}\right)^2\right) \qquad \text{Gleichung 10}$$

$y_{s,k}$ ist ein simulierter oder gemessener Messwert, $\bar{y}$ ist der Mittelwert der $y_{s,k}$ Werte und $\widehat{y_k}$ ist der vorhergesagte Wert.

Die lokal gewichtete Regression (LWR) ist ein spezielles statistisches Modellierungsverfahren, bei dem die Regression auf der so genannten „lokalen Ebene“ durchgeführt wird. Dieses Verfahren wird für die Daten verwenden, bei denen aufgrund der vorhandenen nicht linearen Zusammenhänge eine lineare Regression der gesamten Datenmatrix nicht möglich ist. Dabei wird für jede Probe x_i in der Datenmatrix eine bestimmte Anzahl der nah liegenden Punkte (lokale Punkte) ausgewählt. Jeder Punkt wird gewichtet in Anhängigkeit der Distanz zu x_i. In dieser Dissertation wurde der Abstand zwischen den einzelnen Punkten anhand der „Mahalanobis Distanz“ berechnet (Gleichung 11).

$$D(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \sum^{-1} (\vec{x} - \vec{y})} \qquad \text{Gleichung 11}$$

Dabei bedeutet *D* der Abstand zwischen $\vec{x}$ und $\vec{y}$ und $\sum^{-1}$ die Wichtungsmatrix.

Die gewichteten Werte werden bei der Regression berücksichtigt – daher der Name die „gewichtete Regression“. In dieser Dissertation wurde die lokale Regression anhand der Hauptkomponenten durchgeführt. Die Anzahl der lokalen Punkte und die Anzahl der Hauptkomponenten wurden in den Modellen variiert. Die Vorhersagekraft der berechneten Modelle wurden mit Hilfe der oben beschriebenen Größen, RMSE und R^2, evaluiert.

Weitere in dieser Dissertation verwendete nicht lineare mathematische Modelle sind die künstlichen neuronalen Netze (Artificial Neural Networks, ANN). ANN wird für die Modellierung komplexer mathematischer Probleme verwendet, bei denen die anderen Verfahren versagen. ANN sind dem menschlichen Gehirn nachempfinden. Ein neuronales Netz ist aus einzelnen Neuronen (auch Knoten genannt) aufgebaut. Ein schematischer Aufbau eines künstlichen Neurons ist in Abbildung 3 dargestellt.

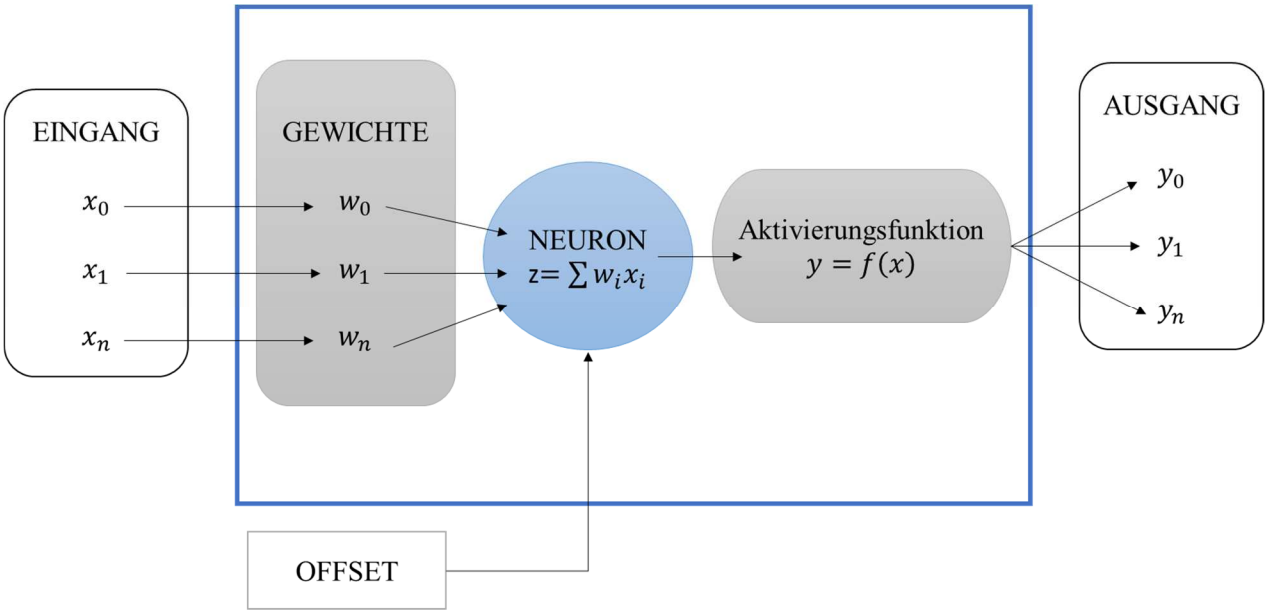

Abbildung 3: Aufbau eines Neurons.

Alle Neuronen sind miteinander über die Kanten verbunden und können somit den Informationsfluss gewährleisten. Jede Kante hat eine bestimmte Wichtung, abhängig von der Verbindungsstärke. Die Neuronen werden in Schichten angeordnet. Man unterscheidet zwischen Eingangsneuronen (input), versteckten (hidden) Neuronen und den Ausgangsneuronen (output). Die Input-Neuronen stellen die gesamte Input-Schicht dar. Sie nehmen Information von außen auf und leiten sie modifiziert weiter. Die versteckten Neuronen befinden sich zwischen den Input- und Output-Schichten und bilden eine Art internes Muster ab. In der Output-Schicht werden die Ausgabewerte des ANN zur Verfügung gestellt. Neuronale Netze können verschiedene Strukturen aufweisen. In der vorgelegten Dissertation wurde ein zweischichtiges Feed-Forward neuronales Netz mit einer Sigmoid-Funktion in der versteckten Schicht und einer linearen Funktion in der Ausgangsschicht verwendet. Aufgrund der Sigmoid-Funktionen können nicht lineare Zusammenhänge beschrieben werden. Die Sigmoid-Funktion ist in Gleichung 12 dargestellt.

$$f(x) = 1/(1 + e^{-x})$$ Gleichung 12

Für die Modellberechnung wurde die Datenmatrix in drei unabhängige Datensätze aufgeteilt: Trainieren mit ca. 70 % aller Daten, Validierung mit 15 % der Daten und Testen mit 15 % der Daten. Die Validierung diente zur Verhinderung des Übertrainierens der Modelle. Im Training wurden die Wichtungsfaktoren der Netze berechnet. Mit dem Test-Datensatz wurde die Güte der Modelle berechnet.

2.2.4 ACO und GA zur Evaluierung der Prozessgrößen

Zu den häufig verwendeten Methoden, um die komplexen mathematischen Optimierungsprobleme zu lösen, gehören metaheuristische Techniken, wie zum Beispiel der Ameisenkolonieoptimierungs-algorithmus (Ant Colony Optimization, ACO) (Dorigo und Blum 2005) und der genetische Algorithmus (Genetic Algorithm, GA) (Goldberg und Holland 1988). ACO und GA wurden nach dem Vorbild aus der Natur abgeleitet. ACO bildet das biologische System der Ameisenkolonie ab, welche auf der Suche nach Nahrungsquellen ist. Auf dem Weg zur Nahrung und zurück hinterlegen die Ameisen eine Pheromonspur, welche als ein Kommunikationssignal für die anderen Ameisen dient. Alle Ameisen laufen mit gleicher Geschwindigkeit und hinterlassen mit der gleichen Wahrscheinlichkeit eine Pheromonspur. Da das Pheromon verdunstet, enthalten die am häufigsten genutzten Strecken zur Nahrungsquelle am meisten Pheromon und sind entsprechend kürzer. Dieses Verfahren stellt das grundlegende Prinzip des ACO-Algorithmus dar. In der Dissertation wurde eine diskrete Version des ACO-Algorithmus als eine Methode der Prozessgrößenselektion verwendet. Die Pheromonkonzentration wurde für jede Prozessgröße einzeln berechnet und zusätzlich für jede Wiederholung aktualisiert. PLSR wurde für die Berechnung der Modelle verwendet. Der verwendete ACO Algorithmus hat vier Phasen: (0) Initialisierung, (1) Berechnung, (2) Wiederholungen und (3) Ergebnisse. Das Fließdiagram des Algorithmus ist in Abbildung 4 dargestellt. In der 0. Phase werden die Modellparameter und Eingangsgrößen definiert. Die 1. Phase umfasst die Modellberechnung und die Bestimmung der besten Ergebnisse für die ausgewählten Prozessgrößen. In der 2. Phase werden die besten Ergebnisse, d. h. die besten Ameisen werden untereinander verglichen und bestimmt. In den Wiederholungsschleifen wird die Pheromonkonzentration jeder Prozessgröße aktualisiert. Die 3. Phase stellt die besten Ergebnisse in Form der optimalen Kombinationen der Prozessgrößen dar.

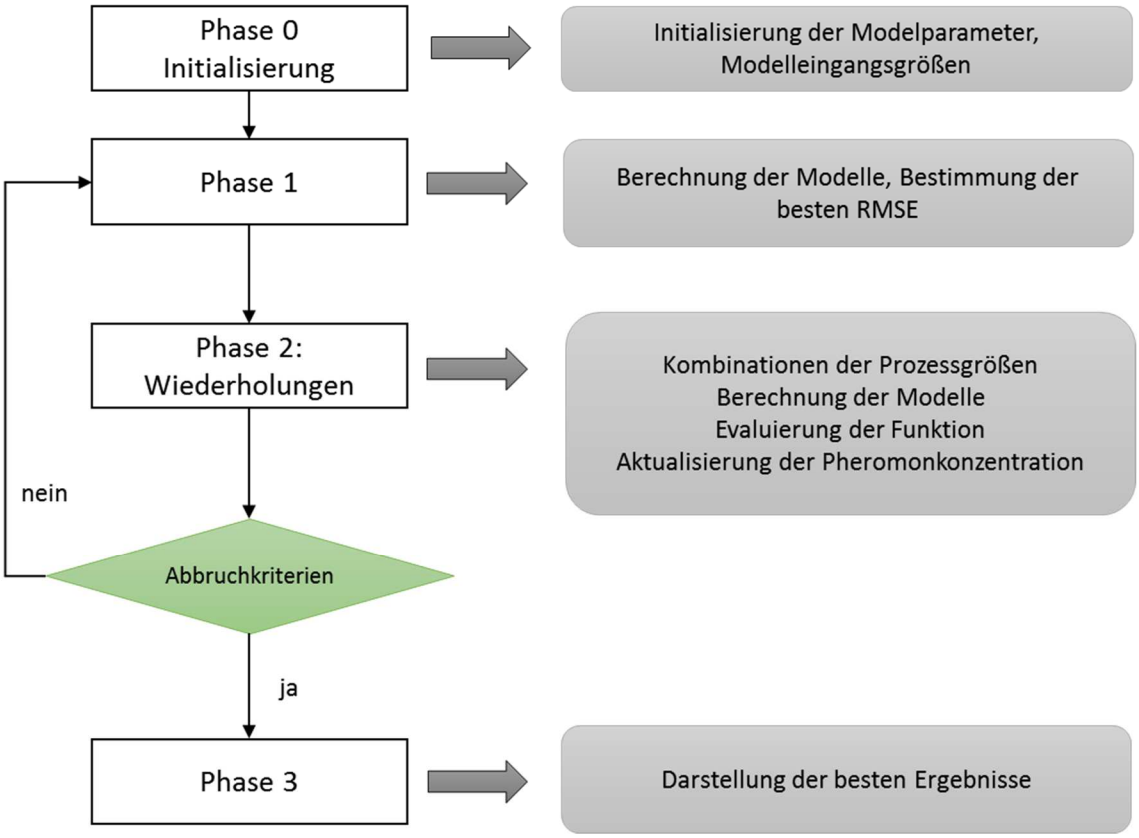

Abbildung 4: Fließschema des ACO Algorithmus.

Der genetische Algorithmus (GA) ist ein anderes mathematisches Verfahren, das oft für die Optimierung komplexer Probleme verwendet wird. Im Grunde wurde er von der Evolutionstheorie von Charles Darwin „Überleben des Stärkeren ‘‘ und den Mendelschen Gesetzen der Genetik inspiriert. Der GA wurde in den sechziger Jahren von John H. Holland vorgestellt (Goldberg und Holland 1988). Der genetische Algorithmus beschreibt eine künstliche Population mit einer Reihe evolutionärer Vorgänge, wie Mutation, Selektion und Rekombination. Das Ziel ist die besten Eigenschaften der Individuen zu extrahieren und deren Eigenschaften an die Folgegenerationen zu vererben. Der Vorteil dieses Verfahrens ist, dass kein Vorwissen oder problemspezifische Information benötigt wird, um die optimalen Lösungen zu finden. Ein schematischer Ablauf eines GA ist in Abbildung 5 dargestellt.

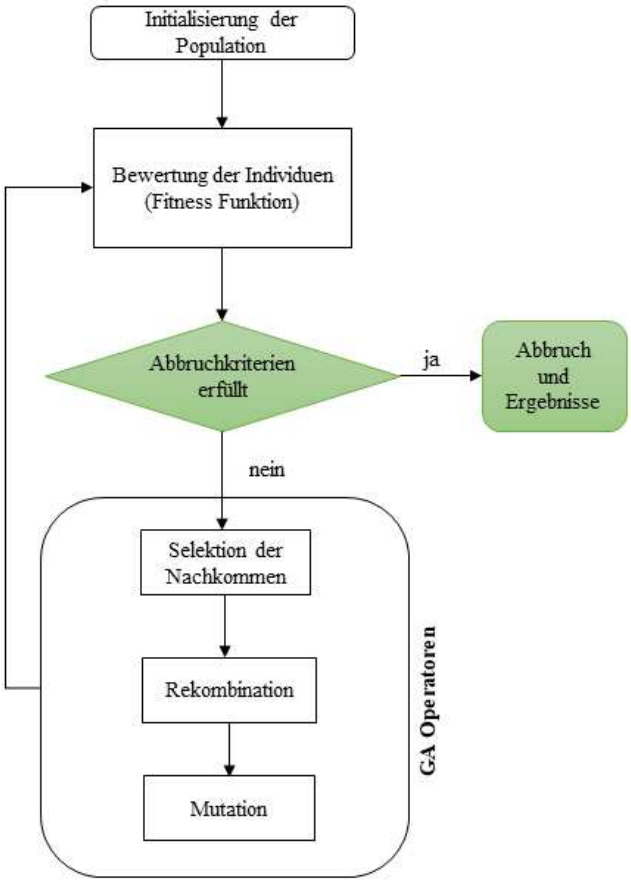

Abbildung 5: Schematischer Ablauf des genetischen Algorithmus.

In der Initialisierungsphase wird die erste Generation der Individuen zufällig erzeugt. Jedem Individuum werden entsprechend seiner Güte Werte der Fitnessfunktion zugeordnet und die Abbruchkriterien getestet. Sind die Abbruchkriterien erfüllt, werden die Ergebnisse dargestellt; anderenfalls werden die Gene rekombiniert und mutiert. Danach wird ein neuer Zyklus durchlaufen.

Der in dieser Dissertation verwendete GA enthielt in der 0. Population (Initialisierungsphase) 256 Individuen. Diese wurden mit Hilfe der Rekombinationsoperatoren und Mutation 100 Generationen optimiert. Die Auswahl der besten Individuen basierte auf der Evaluierung der RMSE Ergebnisse der berechneter PLSR Modelle.

3 Publikationen

3.1 “Artificial neural network prediction of biogas flow rate optimized with an ant colony algorithm”

Autoren: Tetyana Beltramo, Cassiano Ranzan, Joerg Hinrichs, Bernd Hitzmann

Veröffentlicht im März 2016 in *Biosystems Engineering*

online verfügbar unter doi.org/10.1016/j.biosystemseng.2016.01.006

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm

Tetyana Beltramo [a,*], Cassiano Ranzan [b], Joerg Hinrichs [a], Bernd Hitzmann [a]

[a] Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart 70599, Germany
[b] Chemical Engineering Department, Federal University of Rio Grande do Sul, 90040-040 Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 7 July 2015
Received in revised form
4 January 2016
Accepted 15 January 2016
Published online xxx

Keywords:
Neural network
Ant Colony Optimisation (ACO)
Model-driven
Modelling
Biogas flow rate

The aim of this study was to develop a fast and robust methodology to analyse the biogas production process. The Anaerobic Digestion Model No.1 was used to simulate the co-digestion of agricultural substrates. Neural network models were used to predict the biogas flow rate. With the help of the ant colony optimisation algorithm, the significant process variables were identified. Thus the model dimension was reduced and the model performance was improved. The achieved results showed that the approach gave a reliable way to analyse the biogas production process with respect to the significant process variables. This methodology could be further implemented to control the biogas production process and to manage the substrate composition.

1. Introduction

1.1. Anaerobic digestion

The production of biogas is an important part of the renewable energy supply system (Weiland, 2010). Furthermore, it has proved to be a feasible energy supply technology with development capacity and environmental advantages (Fehrenbach et al., 2008). Biogas production is based on anaerobic digestion (AD) of biomass and is characterised through a complex microbiological structure. The product of the process is a gas-mixture consisting mainly of methane (CH_4, 55–70%) and carbon dioxide (CO_2, 30–45%). The main compound, methane, has a high energy potential and can be used for heating and electricity production purposes. The biogas production efficiency can be estimated with respect to the activity of the methane-producing microorganisms. Their growth and activity depend on many ambient conditions, such as substrate composites, temperature, pH, and the interdependences between the microbial populations. One of the most used parameters to evaluate AD processes is the biogas production rate. This process parameter is directly correlated with the development of the methane-producing microbial populations and characterises the effectiveness of their activity. There are research studies describing the

* Corresponding author.
E-mail address: t.beltramo@uni-hohenheim.de (T. Beltramo).
http://dx.doi.org/10.1016/j.biosystemseng.2016.01.006

Nomenclature

Abbreviation	*Description*
ACO	Ant colony optimisation
AD	Anaerobic digestion
ADM1	Anaerobic digestion model No.1
ANN	Artificial neural networks
BSM2	Benchmark simulation model No.2
CH_4	Methane
COD	Chemical oxygen demand
CO_2	Carbon dioxide
MLP	Multi-layered perceptron
MLR	Multiple linear regression
ODE	Ordinary differential equations
PCR	Principal component regression
PLS	Partial least square regression
R^2	Coefficient of determination
RMSE (c/p)	Root mean square error (calibration/prediction)
S_{aa}	Amino acids
S_{ac}	Acetic acid
S_{fa}	Long chain fatty acids (LCFA)
S_i	Inert solutes
S_{in}	Inorganic nitrogen
SNV	Standard normal variate
S_{su}	Monosaccharides
UASB	Upflow anaerobic sludge blanket digestion
X_c	Composites
X_{ch}	Carbohydrates
X_i	Inert particulates
X_{li}	Lipids
X_{pr}	Proteins

dependence of the biogas production rate on the ambient conditions, with respect to the microorganisms which breakdown biomass to CH_4 and CO_2 (Amon et al., 2007; Biernacki, Steinigeweg, Borchert, & Uhlenhut, 2013). Therefore the measurement of the biogas production rate represents a proper approach to evaluate the entire AD process.

1.2. *Simulation modelling*

In recent years, mathematical modelling has become a widespread technique to optimise and to control diverse AD processes. There are many theoretical models describing AD systems by defining its biochemical, biological and physicochemical processes. The most used model is the Anaerobic digestion model No. 1 (ADM1), developed by the IWA Task Group for 'Mathematical Modelling of Anaerobic Digestion Processes' (Batstone et al., 2002). This model was basically created to simulate AD of sludge from waste water treatment plants. Later modifications of the model have developed ADM1 into a major AD modelling approach successfully applied to different digestion systems (Lauwers et al., 2013). ADM1 describes the most important process steps from substrate degradation up to biogas accumulation, using dynamic balances, kinetics and acid-base equilibria equations. Such reactions as decomposition of organic acids, ammonia and bicarbonate as well as stripping of methane and carbon dioxide are described in the model. The ADM1 model is universal and can be adjusted for specific processes. There are many applications of ADM1 in the literature, e.g. a modified ADM1 model for AD of the grass, maize, green weed silage and agro-waste application (Biernacki et al., 2013; Gali, Benabdallah, Astals, & Mata-Alvarez, 2009), ADM1 for AD of cyanide-containing substrate (Zaher et al., 2006) as well as applications of ADM1 for mesophilic and thermophilic conditions (Wichern et al., 2010). There is also an application of ADM1 used as an optimisation tool to improve the biogas production rate with respect to the identification of optimal rates for the different solid waste streams and the corresponding hydraulic retention times (Zaher, Jeppson, Steyer, & Chen, 2009). In this regard ADM1 has proved to be a feasible technique to simulate different types of AD systems with respect to various substrates and different ambient conditions.

1.3. *Artificial neural networks*

Artificial neural networks (ANN) represent a popular method to model the relationships within complex structures, such as biological systems. ANN makes it possible to reveal the interdependence patterns within the biological system without previous knowledge about metabolic and kinetic processes of the system (Haider, Pakshirajan, Singh, & Chaudhry, 2008). The ANN models are data-driven, approximating the non-linear relations between the independent input process variables and the dependent output variable. The development of the neural logic was inspired by the central nervous system of animals, in particular the brain. ANN describes a structured system with a number of interconnected neurons ordered in layers. The frequently used structure of the neural networks is the multi-layered perceptron (MLP), which consists of input, hidden and output layers. In comparison to other regression methods (Kessler, 2007) such as the multilinear regression (MLR), the principal component regression (PCR) and the partial least squares regression (PLS), ANN methodology enables a more reliable approximation of the relations between the input variables, such as substrate composites, feed rate and the predicted output variable (Gueguim Kana, Oloke, Lateef, & Adesiyan, 2012). Its advantage is the non-linear sigmoid function at the hidden layer, which provides a computational flexibility in comparison to the linear regression methods and provides an accurate prediction performance of the desired variable (Hitzmann & Kullick, 1994; Hitzmann, Ritzka, Ulber, Schöngarth, & Broxtermann, 1998).

This powerful tool has also proved to be useful for the evaluation and optimisation of anaerobic digestion processes (Ward, Hobbs, Holliman, & Jones, 2007). ANN has also been used to optimise the biogas production from a waste digester (Abu Qdais, Bani Hani, &, Shatnawi, 2010; Holubar et al., 2000) and to control the anaerobic digestion process (Holubar et al., 2002). Kana et al. employed ANN to evaluate the biogas production on sawdust and other co-substrates (Gueguim Kana et al., 2012). Also the prediction of trace compounds in biogas (Strik, Domnanovich, Zani, Braun, & Holubar, 2005) and prediction of the methane fraction from field scale landfill bioreactors (Özkaya, Demir, & Bilgili, 2007) has been

performed with the help of ANN. Neural logic has also been used to simulate the biological hydrogen production in an upflow anaerobic sludge blanket digestion (UASB) (Mu & Yu, 2007). In view of the described implementations, neural logic has proved to be an effective methodology to evaluate and to optimise AD processes.

1.4. Ant colony optimisation

Ant colony optimisation (ACO) is a recently represented optimisation algorithm (Blum, 2005) and describes one of the most popular swarm intelligence techniques. The development of this method was inspired by insect colonies which possess an outstanding social structure. The basis of the algorithm is the structure of the ant's natural behaviour looking for food. On their way to the food source and back to the nest they leave a pheromone trail. The pheromone trail serves as an indirect means of communication between the ants, identifying the pathways to the food source. All the ants move with the same speed and spread pheromone at the same rate. The pheromone evaporates at a constant rate as well. Hence the shortest pathways will be mostly used and contain accordingly the highest concentration of pheromone (Allegrini & Olivieri, 2011). This principle of the "shortest path" is used by the ACO algorithm. The described methodology has already been applied successfully in many scientific fields. It was used to solve sequential problems like the travelling salesman problem (Dorigo & Gambardella, 1997; Dorigo & Schutze, 2004). In the medical field ACO has been implemented for solving problems in protein folding (Shmygelska & Hoss, 2005) and for prediction of major compatibility complex (MHC) class II binders (Karpenko, Shi, & Dai, 2005). In bioinformatics the ACO approach has been employed to optimise multiple sequence alignment (Moss & Johnson, 2003). It has also been used as an optimisation solution in food science for characterisation of wheat flour (Ranzan et al., 2014) as well as a non-invasive control of pH and lactate in porcine meat (Nache, Scheier, Schmidt, & Hitzmann, 2015). The interdisciplinary application of ACO has shown it to be a widely suitable methodology, with potential to be applied to the optimisation of AD processes.

The aim of this study was to develop a fast and reliable approach to analyse the biogas production process with respect to the biogas flow rate. The developed methodology should represent a simple tool to be applied to real processes. In this work the ADM1 model was used to generate data so that its complexity could be compared with the developed methodology. The optimisation technique is used to identify the significant process variables, which can simplify the model dimension and improve prediction performance. In practice, it will reduce the analytical time and costs and can be used to manage the substrate composition.

2. Material and methods

2.1. Simulation of the co-digestion of multiple substrates

For data generation, a modified version of ADM1 (Rosen & Jeppsson, 2006) was used. The implemented model is based on the ADM1 structure (Batstone et al., 2002), which is embedded in the benchmark simulation model No.2 framework (BSM2). ADM1 is consistent with BSM2. The calculations are based on the units of kg COD (chemical oxygen demand) m^{-3}. The implementation in Matlab® Simulink platform was performed using ordinary differential equations (ODE). The simulation model includes 19 process rate equations, 6 acid-base rate equations, 3 gas transfer rate equations and corresponding inhibition balances. The water phase equations are provided for both soluble and particulate matter, including 32 equations. The gas phase equations are presented for hydrogen, carbon dioxide and methane gases. The simulation was performed for 100 days with a frequency of 20 simulation points per simulation day. The co-digestion of cow-manure and grass-silage was modelled. The feed charge of the substrates was varied in a defined sequence. The volumetric flow rate was 170 $m^3\ d^{-1}$ and the volume of the liquid phase was 3400 m^3. The dilution rate was not varied during simulation. Further information about the steady state parameters have been given by Bornhöft, Hanke-Rauschenbach, and Sundmacher (2013). Cow-manure substrate was loaded into the fermenter on days 0, 30, 60 and 85. Grass silage was loaded on days 15, 45, 70 and 90. The input variables describing the substrates used were taken from the literature (Wichern et al., 2010; Zhou, Löffler, & Kranert, 2011) and are presented in Table 1.

2.2. The signal noise

The measurements in practice always contain an undesirable part of the signal, the noise. Noise can appear through an inappropriate analytical method, through the interference of the equipment, or human factors can influence the measurements. In contrast to this, the simulation models represent an ideal reproduction of the real system containing the known interactions of system components and predictable development steps. In order to enable a theoretical model to reflect a real system, a certain noise factor should be necessarily integrated into the simulation model. In this work Gaussian noise of 5% was included. The probability density function p(z) is represented as a normal distribution Eq. (1).

$$p(z) = {}^{1}/_{\sigma\sqrt{2\pi}} e^{-(z-\mu)^2/2\sigma^2} \quad (1)$$

where μ describes the mean value, z is the random variable and σ the standard deviation. For the generation of noise in the simulated data, the standard Matlab® function random was used.

2.3. Data pre-processing

Data pre-processing represents a prior step performed before any data-based analysis (Wold, Sjöström, & Eriksson, 2001). For our purposes, a normalisation method was used. Equations (2) and (3) represent the mathematical computation employed in the normalisation procedure, which is called the standard normal variate method (SNV).

$$X_{i,norm} = (X_i - \overline{X})/\sigma(X_i) \quad (2)$$

Table 1 – The ADM1 input variables.

Variable	Abbreviation	Unit	Cow-manure	Grass-silage
Composites	X_c	kg COD m^{-3}	10.1	10.1
Carbohydrates	X_{ch}	kg COD m^{-3}	29.6	47.9
Proteins	X_{pr}	kg COD m^{-3}	7.1	14.7
Lipids	X_{li}	kg COD m^{-3}	5.9	1.5
Inert particulates	X_i	kg COD m^{-3}	46.5	30.9
Monosaccharides	S_{su}	kg COD m^{-3}	13.4	0
Amino acids	S_{aa}	kg COD m^{-3}	3.4	0
LCFA	S_{fa}	kg COD m^{-3}	0.9	0
Acetic acid	S_{ac}	kg COD m^{-3}	0	2.6
Inorganic nitrogen	S_{in}	kmol N m^{-3}	0.2	0
Inert solutes	S_i	kg COD m^{-3}	7.1	2.4

$$\sigma_i = \sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \Big/ (n-1)} \quad (3)$$

where X_i is a simulated value including the addition of noise, and $\overline{X}$ is the mean value of X_i. The computed σ represents the standard deviation of X_i.

2.4. Computational platform

The computing of models and further treatment of the simulation data was performed on Matlab® Version 7.10 (2010a) (The MathWorks Inc., Natick, USA) on a processor AMD Phenom ™ II X2 B57 with 3.2 GHz platform.

2.5. Artificial neural networks

An artificial neural network consists of neurons ordered in layers, an input layer, a hidden layer and an output layer. The input neurons represent the independent process variables. The output neuron is the dependent predicted process variable, in this case the biogas flow rate. The hidden layer transforms the input information. The number of hidden layers as well as the number of hidden neurons can be varied in order to get the best model structure and to improve its prediction performance. The ANN model used was a two-layer feedforward network with a sigmoid transfer function in the hidden-layer and a linear transfer function at the output-layer. The training was performed using the Levenberg–Marquardt backpropagation algorithm imbedded in the Matlab® Neural Networks Toolbox. The sigmoid transfer function at the hidden layer has an advantage, that also the non-linear behaviour of the process can be predicted. It takes the input (any values between plus and minus infinity) and squashes the output into the range 0–1 (Hagan & Demuth, 2002) as it is shown in equation Eq. (4).

$$f(x) = 1/(1+\exp^{-x}) \quad (4)$$

The data-set was split into three independent data-blocks, the training data (70% of the data, the first 70 days), the validation data (15% of the data, days 71–86) and the test data (15% of the data, days 87–101). The training data-set was used to adjust the models according to the error. The validation data was used to measure the network generalisation and to stop the training before overfitting. The test-set was used for the prediction. The calculated models were evaluated according to two statistical model parameters, the root mean square error (RMSE), which represents the accuracy of the prediction models, and the coefficient of determination (R^2), which estimates the model robustness. Equations (5) and (6) illustrate the computation methodology of the model parameters.

$$RMSE = \sqrt{\left(\sum_{i=1}^{n} \left(\hat{y} - y_{s,i}\right)^2\right) \Big/ n} \quad (5)$$

$$R^2 = 1 - \left(\sum_{i=1}^{n} \left(\hat{y}_i - y_{s,i}\right)^2\right) \Big/ \left(\sum_{i=1}^{n} \left(y_{s,i} - \overline{y}\right)^2\right) \quad (6)$$

where $\hat{y}$ refers to the predicted value, $y_{s,i}$ represents simulated value and $\overline{y}$ is the mean value of y_{si}. The element n gives the number of samples. RMSE and R^2 were computed for training, validation and test models.

2.6. Ant colony optimisation as a variable selection tool

The implemented ACO algorithm is a discrete version (Ranzan et al., 2014). The ACO algorithm was used to perform the variable selection. The basis of the ACO approach is a random factor, which correlates directly with the function of the pheromone trail concentration. Thus the pheromone concentration represents an independent source for the variable evaluation. It is calculated and updated for each variable after each iteration step. Thus only the variables (s) with a high pheromone trail concentration (vector p) were selected out of available variable (N) (Allegrini & Olivieri, 2011). The selection probability of a variable prob(n) is described in Equation (7).

$$prob(n) = p(n) \Big/ \sum_{n=1}^{N} p(n) \quad (7)$$

For the calculation of the models the partially least square regression (PLSR) method was used. The ACO algorithm is performed in four phases, the initiation phase (0), the calculation phase (1), the cycles (2) and the results (3). In the 0th phase all specifications are defined, such as the default variables, the model type, initial pheromone concentration

Table 2 – ACO model specifications.

Used regression model	Partial least square regression
Number of principal components	5
Number of input variables	11
Number of ants	100
Number of iterations	50
Initial pheromone trail concentration	10^{-6}
Evaporation rate per iteration	0.5

and pheromone evaporation rate, the input variables, the number of ants, the number of iterations and the principal components number. In the 1st phase, the objective function is solved and the initialisation of best global output variables is performed. The 2nd phase represents a number of iterations, where the best ants with the best used variable combinations are selected and compared with the best global stored. In the 3rd phase, the results are presented. The ACO model specifications are shown in Table 2.

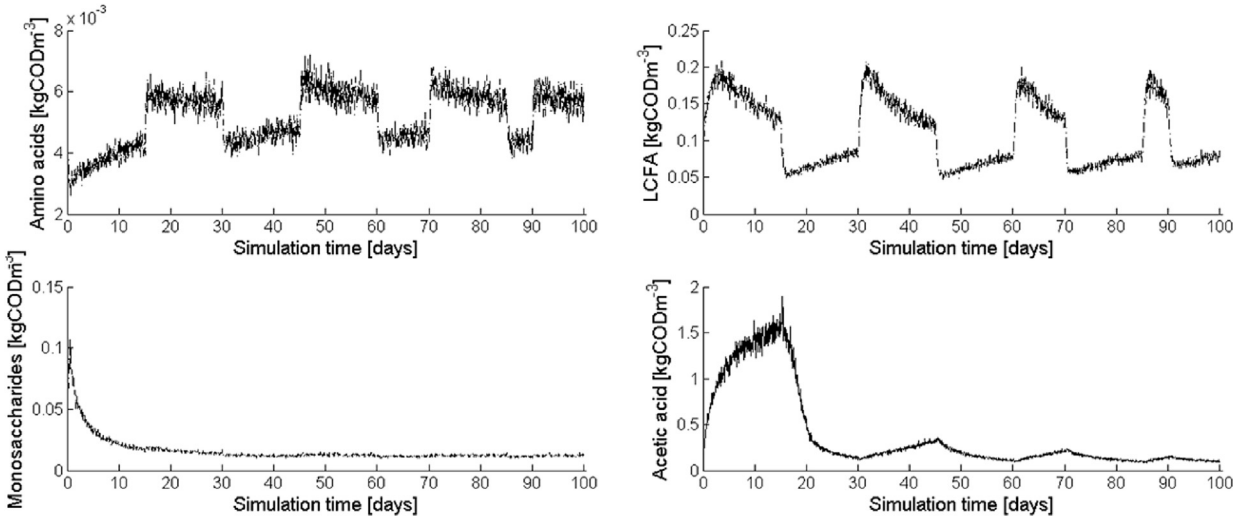

Fig. 1 – Evolution of amino acids, LCFA, monosaccharides and acetic acid. The simulation was performed for 100 days using ODE implemented in ADM1 model with a frequency of 20 simulation points per simulation day with respect to 5% noise.

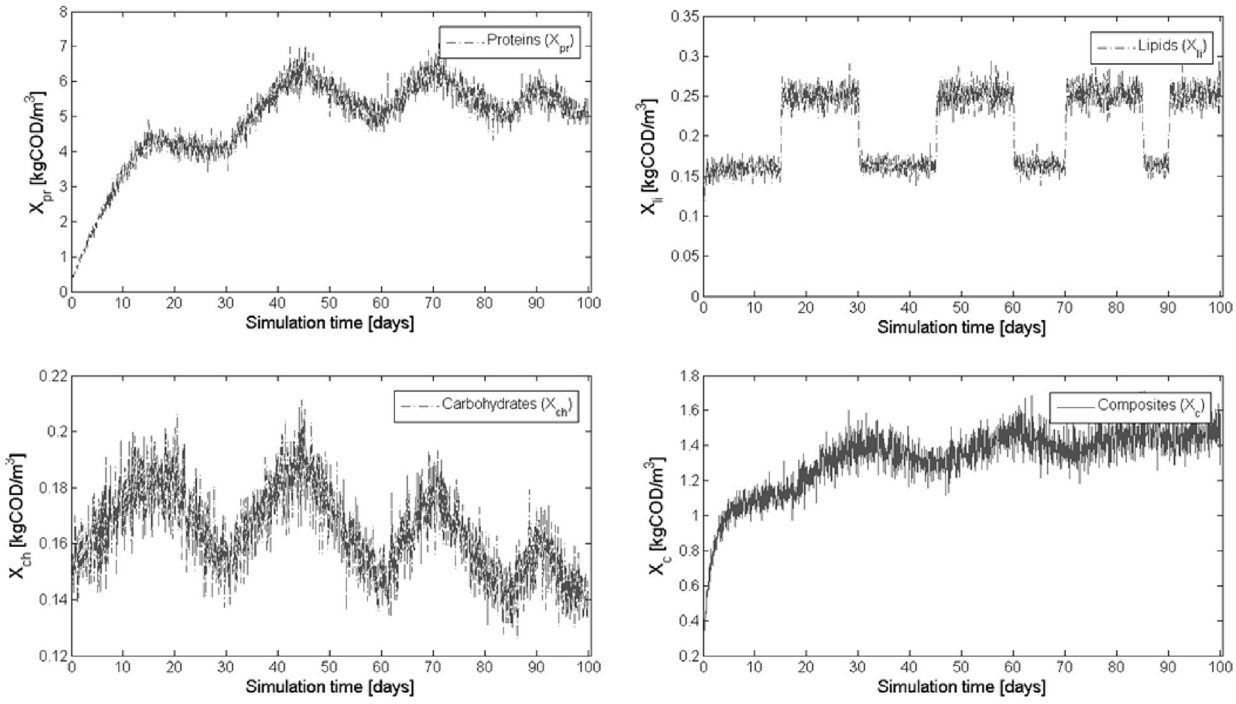

Fig. 2 – Evolution of proteins, lipids, carbohydrates and composites. The simulation was performed for 100 days using ODE implemented in ADM1 model with a frequency of 20 simulation points per simulation day with respect to 5% noise.

Table 3 – ANN prediction of biogas flow rate using 11 input neurons.

Number of hidden neurons	10	3	1
Iterations	14	48	46
RMSE training [%]	5.08	5.52	5.8
RMSE validation [%]	5.02	4.97	4.9
RMSE test [%]	4.91	5.10	5.8
R^2 training [–]	0.93	0.92	0.9
R^2 validation [–]	0.91	0.91	0.91
R^2 test [–]	0.92	0.91	0.89

3. Results and discussion

3.1. Simulation results

The co-digestion of cow-manure and grass-silage substrates with a definite feed sequence during 100 days was performed using the modified version of ADM1. The dynamic evolution of the variables used for the further biogas flow prediction as well as time evolution of the biogas flow are presented in Figs.

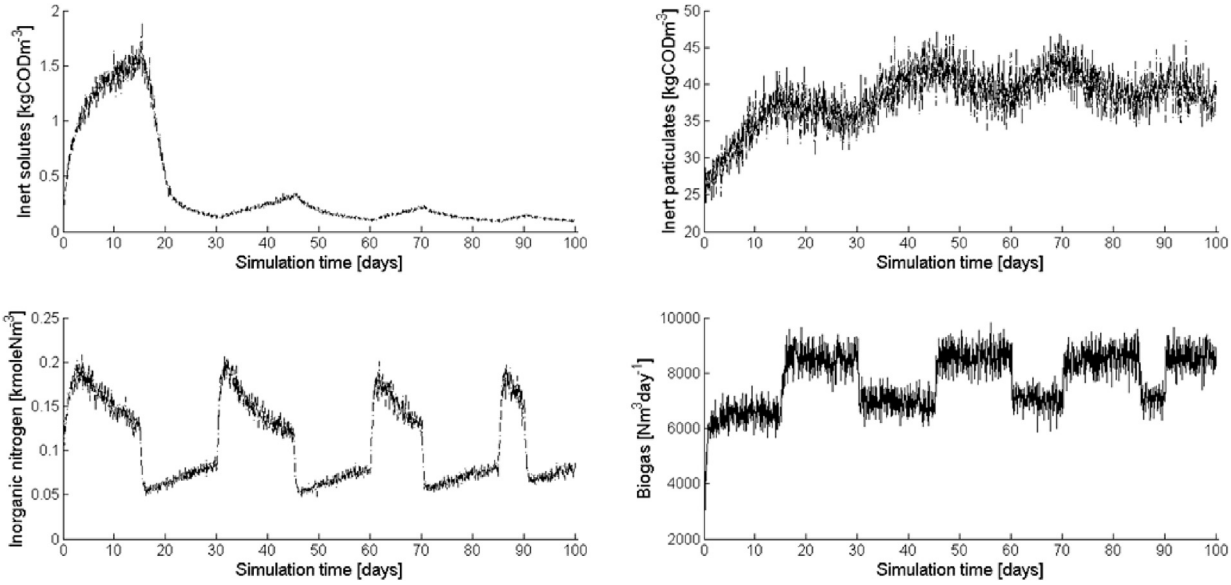

Fig. 3 – Evolution of inert solutes, inert particulates, inorganic nitrogen and biogas flow rate. The simulation was performed for 100 days using ODE implemented in ADM1 model with a frequency of 20 simulation points per simulation day with respect to 5% noise.

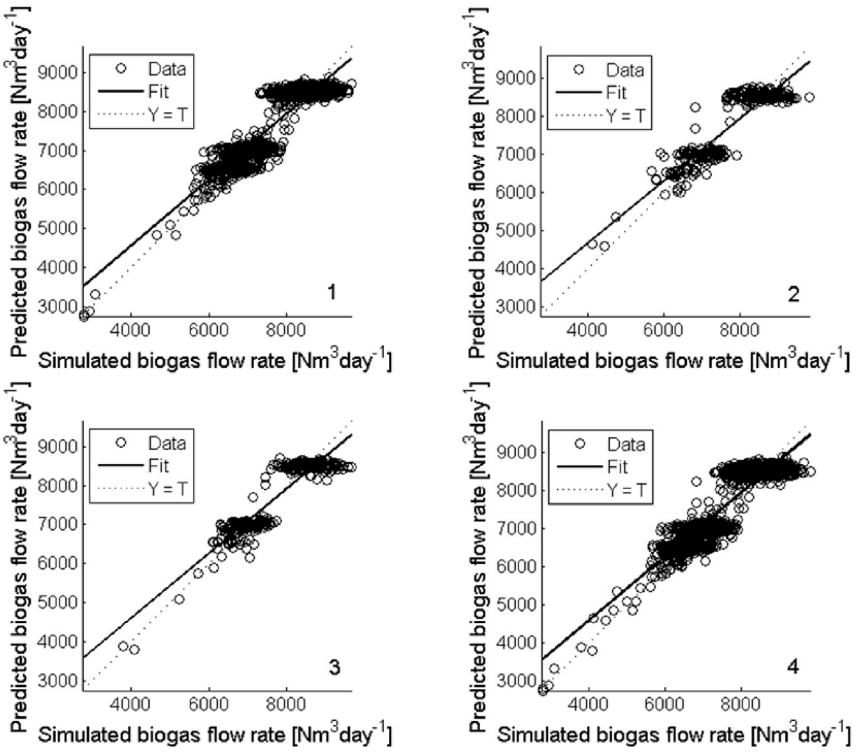

Fig. 4 – ANN regression results of biogas flow rate prediction using 11 input neurons and 10 hidden neurons. Here the regression results of training [1], validation [2], testing [3] and all data [4] are represented.

Table 4 – Input variables with the corresponding pheromone concentration.

Variable name	Abbreviation	Pheromone concentration [units]
Inert solutes	S_i	0.69
Inert particulates	X_i	0.71
Acetic acid	S_{ac}	0.74
Inorganic nitrogen	S_{in}	0.74
Sugars	S_{su}	0.76
Composites	X_c	0.78
Lipids	X_{li}	0.81
LCFA	S_{fa}	0.81
Amino acids	S_{aa}	0.86
Proteins	X_{pr}	0.87
Carbohydrates	X_{ch}	0.95

Table 5 – Prediction of biogas flow rate using 5 input neurons, whose pheromone concentration was higher than 0.8.

Number of hidden neurons	10	3	1
Number of iterations	14	13	19
RMSE training [%]	5.4	5.5	6.05
RMSE validation [%]	4.91	5	5.32
RMSE test [%]	5.35	5.1	5.6
R^2 training	0.91	0.91	0.89
R^2 validation	0.91	0.89	0.89
R^2 test	0.91	0.89	0.88

1–3. The dynamic evolution of amino acids, sugars, acetic acid and total LCFA are presented in Fig. 1.

In Fig. 2 the dynamic evolution of proteins, lipids, carbohydrates and particulate composites is shown.

The dynamic evolution of inert solutes, inert particulates, inorganic nitrogen and the biogas flow rate are presented in Fig. 3.

3.2. *The ANN prediction of the biogas flow rate using eleven process variables*

The prediction of the biogas flow rate was performed with the help of multilayer ANN models using 11 input neurons. The number of hidden neurons was varied to identify the best model structure. The prediction results are shown in Table 3.

The models with 11 input neurons and 10 hidden neurons performed the best prediction (Table 3). Here the RMSE was 5%, and the R^2 reached 0.92. The regression performance of the ANN models using 11 input variables and 10 hidden neurons are shown in Fig. 4.

3.3. *ACO-optimised ANN prediction of the biogas flow rate*

The accumulation of the pheromone trail was used as a quality feature to identify the significant process variables. The computed pheromone concentration of the input variables was distributed in range of 0.69–0.95 units. The process variables acetic acid, inorganic nitrogen, inert solutes as well as inert particulates had the pheromone concentration in

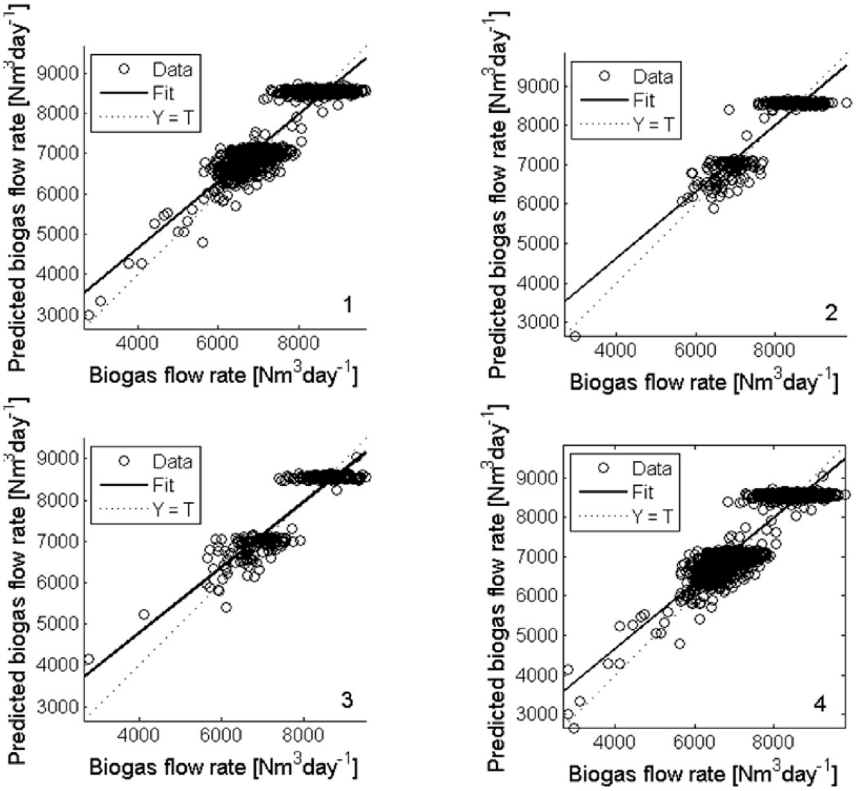

Fig. 5 – ANN regression results of biogas flow prediction optimised with ACO using 5 input neurons and 3 hidden neurons. Here the regression of training [1], validation [2], testing [3] and all data [4] are presented.

Table 6 – Prediction of biogas flow rate using 3 input neurons, whose pheromone concentration was higher than 0.85.

Number of hidden neurons	10	3	1
Number of iterations	11	18	37
RMSE training [%]	5.72	5.81	6.18
RMSE validation [%]	5.34	5.25	6.18
RMSE test [%]	5.52	5.22	6.1
R^2 training	0.9	0.89	0.89
R^2 validation	0.91	0.91	0.89
R^2 test	0.9	0.88	0.88

Table 7 – Prediction of biogas flow rate using 3 input neurons, whose pheromone concentration was less than 0.74.

Number of hidden neurons	10	3	1
Number of iterations	28	42	30
RMSE training [%]	9.3	11	11.04
RMSE validation [%]	8.84	9.66	10.01
RMSE test [%]	9.45	10.35	10.05
R^2 training	0.71	0.66	0.57
R^2 validation	0.72	0.61	0.51
R^2 test	0.61	0.61	0.56

range of 0.69–0.74 units. The calculated pheromone concentration of sugars and composites was between 0.75 and 0.8 units. The process variables amino acids, LCFA, carbohydrates, proteins and lipids had pheromone concentration higher than 0.81 units. The selected process variables with the corresponding pheromone concentration are presented in Table 4.

Concerning the results of the ACO variable analysis the first step of the variable reduction was performed, selecting the input variables, whose pheromone concentration was higher than 0.8 units. Here the amino acids, LCFA, carbohydrates, proteins and lipids were used to predict the biogas flow rate. Despite reduction of the number of input variables, a good prediction of the biogas flow rate could be achieved (Table 5). Nevertheless the RMSE was generally higher and the R^2 lower in comparison to the models with 11 input variables. The best prediction of the biogas flow rate was done using 3 hidden neurons. Here the calculated error of the test models was 5.1%, while the R^2 reached 0.89 (Fig. 5).

The second step of the variable reduction was performed using the variables with a calculated pheromone concentration greater than 0.85 units using amino acids, carbohydrates and proteins as inputs. The further reduction to 3 input neurons was successful and the best prediction was done using 3 hidden neurons (Table 6). Here the computed prediction error was 5.22%, while the R^2 reached 0.88 (Fig. 6).

For the evaluation purpose the prediction of the biogas flow rate was done using the process variables with the lowest pheromone concentration. Here inert solutes (0.69), inert particulates (0.71) and acetic acid (0.74) were selected. The prediction performance of the models using these process variables, whose pheromone concentration was lower, was less successful (Table 7). The RMSE of regression models was greater than 10% and the R^2 lower in comparison to the ANN models with the variables having a high pheromone concentration.

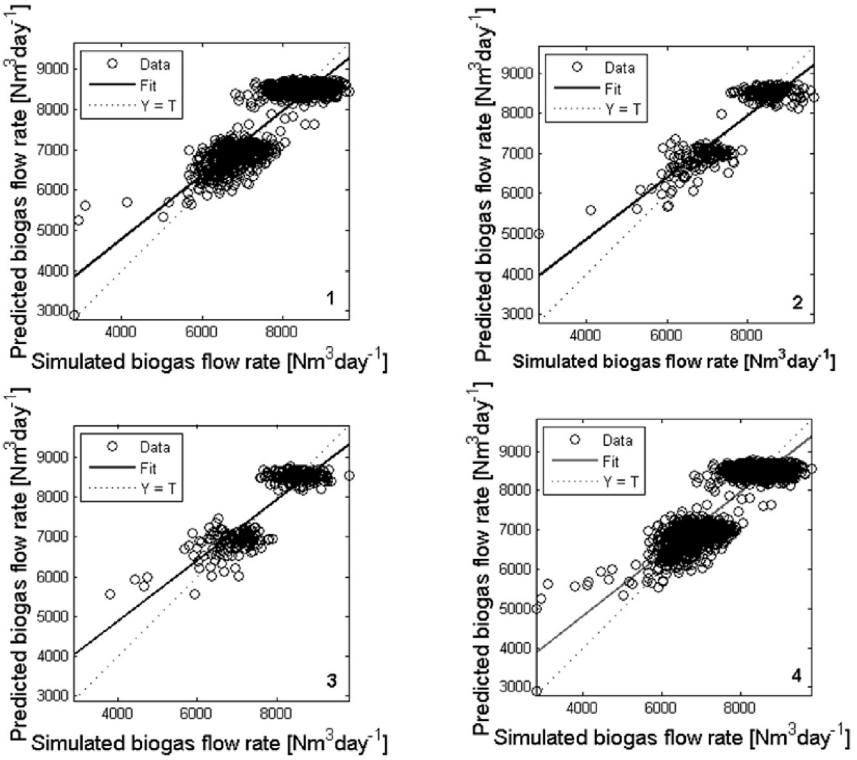

Fig. 6 – ANN regression results of biogas flow rate prediction optimised with ACO using 3 input neurons and 3 hidden neurons. Here the regression of training [1], validation [2], testing [3] and all data [4] are presented.

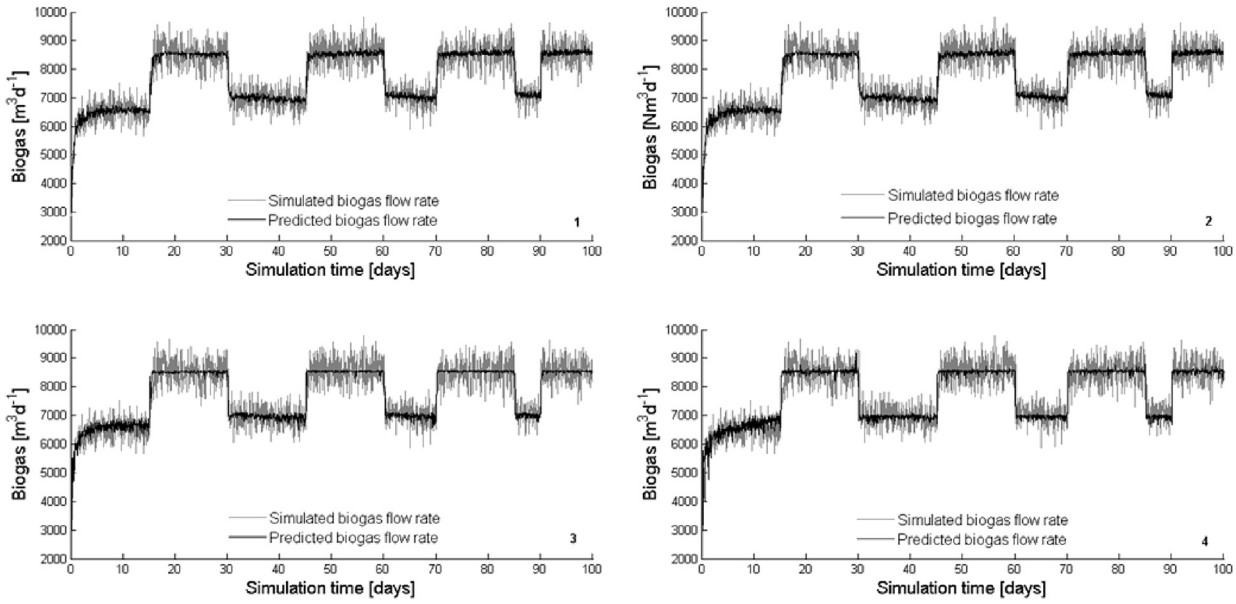

Fig. 7 – The dynamic evolution of the biogas flow rate performed with the ANN using 11 input neurons and 10 hidden neurons [1]; using 5 input neurons and 3 hidden neurons [2]; using 3 input neurons and 10 hidden neurons [3]; using 3 input neurons with the lowest pheromone trail concentration and 10 hidden neurons [4].

The metaheuristic method used has proved to be a feasible tool to identify the significant process variables, based on the correlation with the output variable. The models using the less significant process variables showed less successful prediction performance in comparison to the models using the significant process variables. The dynamic evolution of the biogas flow rate of the calculated models is presented in Fig. 7.

For evaluation purposes, a prediction of the biogas flow rate with respect to different substrate compositions for 50 days was done. The substrates are represented in Table 8. The used ANN model included three input neurons, here amino acids, carbohydrates and proteins and 3 hidden neurons. The biogas flow rate was used as the output neuron. The input data were split into three blocks, here training (70%, first 35 days), validation (15%, days 36–42) and testing (15%, days 43–50). The best and more or less constant biogas output was achieved using the substrate composition S2. The substrate composition S1 provided a strongly deviating biogas output, while the substrate composition S3 showed a periodic variation of the biogas flow rate (Fig. 8).

The achieved results showed the developed approach to be a fast and feasible method to analyse bioreactor performance with respect to the biogas flow rate. The implemented ANN models could successfully predict the dynamic evolution of the process using a simple model structure in comparison to the ADM1 models. The ADM1 models include 19 process rate equations, 6 acid-base rate equations, 3 gas transfer rate equations, inhibition balances and 32 water phase equations for soluble and particulate matter. The ADM1 models require 11 kinetic parameters and rates to be additionally estimated for each metabolic process. The ANN models however are data-driven and can predict the variable evolution without any knowledge of the metabolic and kinetic processes of the system. In comparison to other implementations of the ANN models, the used ANN models have a low number of input neurons and hidden neurons. For example Özkaya used 8 input neurons and 15 hidden neurons to predict the methane fraction (Özkaya et al., 2007), Sahinkaya used 6 input neurons and 20 hidden neurons to predict sulphate, acetate and sulphide concentration (Sahinkaya, Özkaya, Kaksonen, & Puhakka, 2007). The implemented ACO algorithm provides a novel approach in the context of variable analysis and has not been previously implemented with AD systems. Other authors have used genetic algorithms to optimise the ANN models (Abu Qdais et al., 2010; Gueguim Kana et al., 2012; Sahinkaya et al., 2007). In this work the ACO approach was implemented to select the significant process variables with respect to the correlation with the biogas flow rate. Thus a systematic variable reduction could be performed. An effective variable selection enabled the ANN models to predict the biogas flow rate using three input neurons and three hidden neurons. Thus the developed approach represents a fast and robust evaluation method, which can be used to predict biogas flow rate and to

Table 8 – Substrate compositions.

Days	S1	S2	S3
0	Corn silage	Cow manure	Cow manure
5	Corn silage	Grass silage	Corn silage
10	Cow manure	Grass silage	Grass silage
15	Cow manure	Cow manure	Cow manure
20	Grass silage	Grass silage	Corn silage
25	Grass silage	Grass silage	Grass silage
30	Corn silage	Cow manure	Cow manure
35	Cow manure	Grass silage	Corn silage
40	Grass silage	Grass silage	Grass silage
45	Corn silage	Cow manure	Cow manure
50	Grass silage	Grass silage	Corn silage

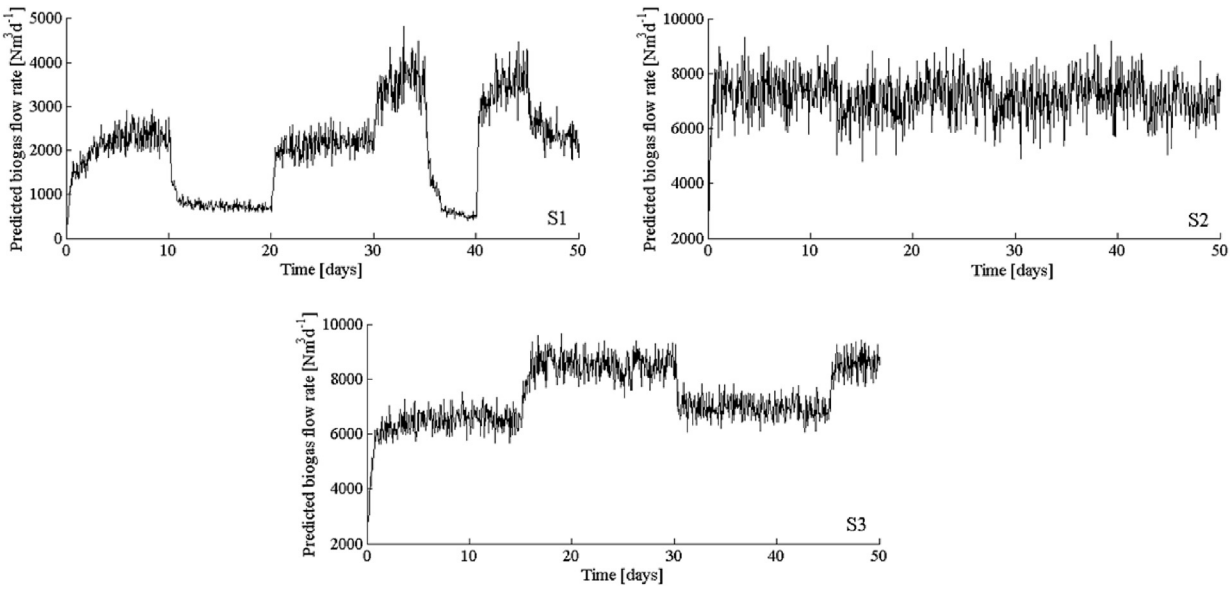

Fig. 8 – Prediction of biogas flow rate with respect to different substrate compositions. [S1] and [S3] co-digestion of corn silage, cow manure and grass silage with different feed sequences; [S2] co-digestion of cow-manure and grass silage.

identify the significant process variables. The developed approach could be used as a tool to assist the management of substrate composition, for example to identify the best possible substrate components and the feed sequence. The process variables identified as important can be directly used for the prediction of the process performance with respect to the biogas output enabling a fast and easy way to analyse the biogas process, potentially saving time and costs.

4. Conclusion

The aim of the study was to develop an evaluation tool able to analyse the biogas production process with respect to the biogas flow rate. The ADM1 model was applied for data generation and for comparison with the simpler ANN approach used here. The advantage of the ANN models is that they are data-driven and don't need any prior knowledge about the process kinetics. The implemented optimisation tool was used to evaluate the process variables and to identify the significant ones. It helped to minimise the model dimension and to improve its prediction performance. The developed methodology can be used for the evaluation of the real biogas production processes, to identify the significant process variables and potentially reduce the analytical costs and time. Thus the operating engineer could use simple models to evaluate the feed composition and improve the process output.

REFERENCES

Abu Qdais, H., Bani Hani, K., & Shatnawi, N. (2010). Modeling and optimization of biogas production from waste digester using artificial neural network and genetic algorithm. *Resources, Conservation and Recycling*, 54, 359–363.

Allegrini, F., & Olivieri, A. C. (2011). A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least – squares analysis. *Analytica Chimica Acta*, 699, 18–25.

Amon, T., Amon, B., Kryvoruchko, V., Zollitsch, W., Mayer, K., & Gruber, L. (2007). Biogas production from maize and dairy cattle manure – influence of biomass composition on the methane yield. *Agriculture, Ecosystems and Environment*, 118, 173–182.

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., et al. (2002). *Anaerobic digestion Model No. 1 (ADM)*. IWA Publishing (Intl Water Assoc).

Biernacki, P., Steinigeweg, S., Borchert, A., & Uhlenhut, F. (2013). Application of Anaerobic Digestion Model No. 1 for describing anaerobic digestion of grass, maize, green weed silage and industrial glycerine. *Bioresource Technology*, 127, 188–194.

Blum, C. (2005). Ant colony optimization: introduction and recent trends. *Physics of Life reviews*, 2, 353–373.

Bornhöft, A., Hanke-Rauschenbach, R., & Sundmacher, K. (2013). Steady-state analysis of the Anaerobic Digestion Model No.1 (ADM1). *Nonlinear dymanics*, 73, 535–549.

Dorigo, M., & Gambardella, L. M. (1997). Ant colonies for the travelling salesman problem. *Bio Systems*, 43, 73–81.

Dorigo, M., & Schutze, T. (2004). *Ant colony optimization*. Cambridge, MA, USA: The MIT Press.

Fehrenbach, H., Giegrich, J., Reinhardt, G., Sayer, U., Gretz, M., Lanje, K., et al. (2008). Kriterien einer nachhaltigen Bioenergienutzung im globalen Maßstab. *UBA-Forschungsbericht*, 206, 41–112.

Gali, A., Benabdallah, T., Astals, T., & Mata-Alvarez, J. (2009). Modified version of ADM1 model for agro-waste application. *Bioresource Technology*, 100, 2783–2790.

Gueguim Kana, E. B., Oloke, J. K., Lateef, A., & Adesiyan, M. O. (2012). Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm. *Renewable Energy*, 46, 276–281.

Hagan, M. T., & Demuth, H. B. (2002). An introduction to the use of neural networks in control systems. *International Journal of Robust and Nonlinear Control*, 12(11), 959–985.

Haider, M. A., Pakshirajan, K., Singh, A., & Chaudhry, S. (2008). Artificial neural network – genetic algorithm approach to optimize media constituents for enhancing lipase production by a soil microorganism. *Applied Biochemistry Biotechnology, 144*, 225–235.

Hitzmann, B., & Kullick, T. (1994). Evaluation of pH field effect transistor measurement signals by neural networks. *Analytica Chimica Acta, 294*, 243–249.

Hitzmann, B., Ritzka, A., Ulber, R., Schöngarth, K., & Broxtermann, O. (1998). Neural networks as a modeling tool for the evaluation and analysis of FIA signals. *Journal of Biotechnology, 65*(1), 15–22.

Holubar, P., Zani, L., Hager, M., Fröschl, W., Radak, Z., & Braun, R. (2000). Modelling of anaerobic digestion using self-organizing maps and artificial neural networks. *Water science and Technology, 41*(12), 149–156.

Holubar, P., Zani, L., Hager, M., Fröschl, W., Radak, Z., & Braun, R. (2002). Advanced controlling of anaerobic digestion by means of hierarchical neural networks. *Water research., 36*, 2582–2588.

Karpenko, O., Shi, J., & Dai, Y. (2005). Prediction of MHC class II binders using the ant colony search strategy. *Artificial Intelligence in Medicine, 35*(1–2), 147–156.

Kessler, W. (2007). *Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik*. Weinheim: WILEY-VCH GmbH und Co. KGaA.

Lauwers, J., Appels, L., Thompson, I. P., Degrève, J., Impe, J. F., & Dewil, R. (2013). Mathematical modelling of anaerobic digestion of biomass and waste: power and limitations. *Progress in Energy and Combustion Science, 39*, 383–402.

Moss, J., & Johnson, C. G. (2003). An ant colony algorithm for multiple sequence alignment in bioinformatics. *Artificial Neural Nets and Genetic Algorithm*, 182–186.

Mu, Y., & Yu, H.-Q. (2007). Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. *International Journal of Hydrogen Energy, 32*, 3308–3314.

Nache, M., Scheier, R., Schmidt, H., & Hitzmann, B. (2015). Non-invasive lactate and pH-monitoring in porcine meat using Raman spectroscopy and chemomerics. *Chemometrics and Intelligent Laboratory Systems, 142*, 197–205.

Özkaya, B., Demir, A., & Bilgili, M. S. (2007). Neural network prediction for the methane fraction in biogas from field-scale landfill bioreactors. *Environmental Modelling & Software, 22*, 815–822.

Ranzan, C., Strohm, A., Ranzan, L., Trierweiler, L. F., Hitzmann, B., & Trierweiler, J. O. (2014). Wheat flour characterization using NIR and spectral filter based on ant colony optimization. *Chemometrics and Intelligent Laboratory Systems, 132*, 133–140.

Rosen, C., & Jeppsson, U. (2006). *Aspects on ADM1 implementation within the BSM2 framework*. http://www.iea.lth.se/publications/Reports/LTH-IEA-7224.pdf.

Sahinkaya, E., Özkaya, B., Kaksonen, A. H., & Puhakka, J. A. (2007). Neural network prediction of thermophilic (65 °C) sulfidogenic fluidized-bed reactor performance for the treatment of metal-containing wastewater. *Biotechnology and Bioengineering, 97*(4), 780–787.

Shmygelska, A., & Hoss, H. H. (2005). An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem. *BMC bioinformatics, 6*(30), 1–22.

Strik, D. P. B. T. B., Domnanovich, A. M., Zani, L., Braun, R., & Holubar, P. (2005). Prediction of trace compounds in biogas from anaerobic digestion using the Matlab Neural Network Toolbox. *Environmental Modelling & Software, 20*, 803–810.

Ward, A. J., Hobbs, P. J., Holliman, P. J., & Jones, D. L. (2007). Optimization of the anaerobic digestion of agricultural resources. *Bioresource Technology, 99*, 7928–7940.

Weiland, P. (2010). Biogas production: current state and perspectives. *Applied Microbial Biotechnology, 85*, 849–860.

Wichern, M., Gehring, T., Fischer, K., Andrade, D., Lübken, M., Koch, K., et al. (2010). Monofermentation of grass silage under mesophilic conditions: measurements and mathematical modeling with ADM1. *Bioresource Technology, 100*, 1675–1681.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS – regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*, 109–130.

Zaher, U., Jeppson, U., Steyer, J. P., & Chen, S. (2009). GISCOD: general integrated solid waste co-digestion model. *Water resources, 43*(10), 2717–2727.

Zaher, U., Moussa, M., Widyatmika, I., van Der Stehen, P., Gijzen, H., & Vanrolleghem, P. (2006). Modelling anaerobic digestion acclimatization to a biogradable toxicant: application to cyanide. *Water Science Technology, 54*, 129–137.

Zhou, H., Löffler, D., & Kranert, M. (2011). Model-based prediction of anaerobic digestion of agricultural substrates for biogas production. *Bioresource Technology, 102*, 10819–10828.

3.2 "Prediction of the biogas production using GA and ACO input features selection method for ANN model"

Autoren: Tanja Beltramo, Michael Klocke, Bernd Hitzmann

Veröffentlicht im Januar 2019 in *Information Processing in Agriculture*

Online verfügbar unter doi.org/10.1016/j.inpa2019.01.002

Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 6 (2019) 349–356

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

Prediction of the biogas production using GA and ACO input features selection method for ANN model

Tanja Beltramo [a,*], *Michael Klocke* [b], *Bernd Hitzmann* [a]

[a] *Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart 70599, Germany*
[b] *Leibniz Institute for Agricultural Engineering and Bioeconomy, Potsdam 14469, Germany*

ARTICLE INFO

Article history:
Received 6 June 2018
Received in revised form
19 December 2018
Accepted 7 January 2019
Available online 11 January 2019

Keywords:
Ant colony optimization
Artificial neural networks
Biogas
Genetic algorithm

ABSTRACT

This paper presents a fast and reliable approach to analyze the biogas production process with respect to the biogas production rate. The experimental data used for the developed models included 15 process variables measured at an agricultural biogas plant in Germany. In this context, the concentration of volatile fatty acids, total solids, volatile solids acid detergent fibre, acid detergent lignin, neutral detergent fibre, ammonium nitrogen, hydraulic retention time, and organic loading rate were used. Artificial neural networks (ANN) were established to predict the biogas production rate. An ant colony optimization and genetic algorithms were implemented to perform the variable selection. They identified the significant process variables, reduced the model dimension and improved the prediction capacity of the ANN models. The best prediction of the biogas production rate was obtained with an error of prediction of 6.24% and a coefficient of determination of $R^2 = 0.9$.

1. Introduction

Biogas belongs to the renewable energy sources, which underwent a worldwide expanding technological development over the last decades. The basis of the biogas production is anaerobic digestion (AD) and subsequent biomethanation of biomass and organic wastes, e.g. from husbandry or industrial production processes. AD is a complex process involving complex microbial consortia with numerous metabolic processes and kinetic reactions [1,2]. While the biogas production is a complex and long lasting biological process, the use of the conventional analytical methods, e.g. according to German VDI norm 4630 [3], is time- and equipment-consuming and, hence, expensive. Numerous process variables must be taken into consideration and be controlled to evaluate the process.

Mathematical modelling represents a quick and cheap alternative to the conventional analytics, among others especially chemometric methods and metaheuristics. These powerful tools are helpful to identify the process structure and to analyze the correlations between the process components without any prior knowledge about the process correlations [4]. One of the most popular mathematic methods applied for the optimization of biological systems are the artificial neural networks (ANN) [5,6]. In agricultural sciences it was used for visual identification of orange varieties [7], to detect plant diseases [8], to improve milk service platform [9] and to estimate the biophysical variables [10]. This approach was successfully implemented in field of AD systems in order to predict the process intermediates, to optimize the bioreactor performance and to improve the process conditions. As an example, Strik et al. used ANN to predict the trace compounds in biogas from anaerobic digestion [11]. Here the ANN models were successfully used to predict hydrogen

* Corresponding author.
E-mail address: tanja.beltramo@gmail.com (T. Beltramo).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.01.002

sulfide and ammonia concentrations in biogas under dynamical conditions. The developed approach can be used to control and to reduce the toxic trace compounds in fuel cells. In terms of the prediction of the bioreactor performance, Sahinkaya et al. implemented the neural networks to analyze the thermophilic sulfidogenic fluidized-bed reactor [12], which help to control the operational conditions improving the process cycles. Also based on ANN, Ozkaya et al. predicted the methane fraction in biogas [13]. Here the anaerobic digestion process was evaluated using the leachate characteristics measured in two different time periods used to predict the biogas output. Kanat et al. evaluated the biogas production rate [14] using the ANN models based on the data measured by steady-state and abnormal process conditions to control the thermophilic bioreactors. Bernari et al. used an automated medium scale prototype for anaerobic co-digestion of olive mill wastewater [15]. The described implementations of ANN used for the modelling of biogas digestion process based on different process variables used for control and evaluation either of the separate process parameters or of the entire process, proved to be a reliable technique to optimize and to control AD processes. In comparison to other existing models, such as the anaerobic digestion model No.1 (ADM1) [16], neural networks have a simple structure. The ADM1 model includes 19 process rate equations, six acid-base reactions, three gas transfer reactions, a number inhibition balances and over 30 water phase equations for soluble and particulate matter. Moreover, ADM1 requires a number of kinetic parameters and rates need to be estimated for each process. Thus the complexity of the microbiological degradation processes limits the application to a certain course of the process. In comparison to ADM1, ANN models are data-driven and do not require any prior knowledge about the kinetics and the microbiological digestion processes [17]. ANN models are represented by the multilayered perceptron, that enables to evaluate the correlation between the independent process input variables and the dependent output variables without any prior knowledge about their interrelations.

Metaheuristic techniques, such as genetic algorithms (GA) [18] and ant colony optimization (ACO) [19] have been applied to solve complex problems within biological systems. For example, ACO was implemented in the medical sciences for solving problems in protein folding. Here ACO was used to predict the protein conformation based on the amino acids sequences [20]. The used ACO represents a successful tool to solve the problems in bioinformatics, where other state-of-the-art methods failed. In field of bioinformatics, GA was used to detect recombination problems [21] and to analyze an enzyme kinetic process [22]. In food science ACO and GA were implemented for the flour characterization based on the NIR spectral filter data [23] and to predict pH and lactate in porcine meat [24]. In agricultural sciences Silalahi et al. used GA based on infrared spectral data for the identification of ripeness of oil palm [25], while Mehdizadeh et al. for simulation of greenhouse processes [26]. The interdisciplinary use of metaheuristics proved it to be reliable optimization tools and opens the perspective to be also applied for the optimization of AD systems.

The main objective of this study was to develop a reliable tool able to predict the bioreactor performance with respect to the biogas production rate. It should improve the biogas production process and could select the significant process features in order to simplify the analytical evaluation procedure. The developed methodology represents a fast and adaptable method to be used for different AD processes under real conditions based on the real anaerobic digestion process data. In detail, it should be able to identify the correlations within the process and predict accurately the process evolution. For this approach, ANN was used to predict the biogas production rate. As variable selection tools, GA and ACO were applied, which were compared according to their feasibility and usefulness in field of AD optimization. In that sense, the optimization step is unique due to the algorithm combination method and the application of ACO for the real data of an anaerobic digestion process.

2. Materials and methods

2.1. *Origin of experimental data*

The experimental data used in this study are off-line measurements collected at a biogas plant in Germany over a time-period of ten months with a frequency of one sampling per week. The used data originated from the joint project BIOGAS-ENZYME supported by the German Federal Ministry of Food and Agriculture (BMEL), grant no. 22027707 [27]. The measurements were done in July 2011 and from February till October 2012. In this biogas plant four feedstocks, i.e. maize and grass silages together with pig and cattle manure, were co-digested at mesophilic temperature (42 °C). From digestate samples, the concentrations of acetic acid, propionic acid, n- and iso-butyric acid, n- and iso-valeric acid and ammonium (NH_4^+-N) were determined as well as the contents of total solids (TS), volatile solids (VS), neutral detergent fibre (NDF), acid detergent fibre (ADF), the organic loading rate, alkalinity ratio and acid detergent lignin (ADL). Hydraulic retention time (HRT) was additionally calculated. The measured variables are represented in Table 1. TS and VS serve generally as

Table 1 – The measured process variables (measured with a frequency one measurement per week).

Variable name (abbreviation)	Unit
Acetic acid (AcA)	$g\,l^{-1}$
Acid detergent fibre (ADF)	$g\,kg^{-1}\,VS^{-1}$
Acid detergent lignin (ADL)	$g\,kg^{-1}\,VS^{-1}$
Alkalinity ration (FOS/TAC)	–
Ammonium (NH_4^+-N)	$g\,l^{-1}$
n-Butyric acid (nBA)	$g\,l^{-1}$
Iso-butyric acid (iBA)	$g\,l^{-1}$
Total solids (TS)	% Fresh matter (FM)
Hydraulic retention time (HRT)	day
Neutral detergent fibre (NDF)	$g\,kg^{-1}\,VS^{-1}$
Volatile solids (VS)	% TS
Organic loading rate (OLR)	$kg\,VS\,m^{-3}\,d^{-1}$
Propionic acid (PA)	$g\,l^{-1}$
n-Valeric acid (nVA)	$g\,l^{-1}$
Iso-valeric acid (iVA)	$g\,l^{-1}$

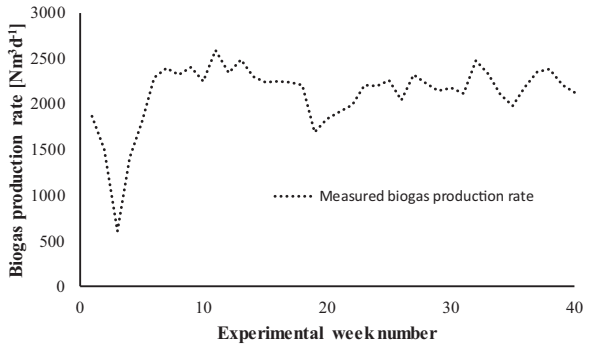

Fig. 1 – Evolution of the biogas production rate during sampling period.

reference values for other substrate compounds, such as proteins, nitrogen and trace elements. ADF, ADL and NDF describe the content of nutrients. HRT is the average period that a given quantity of biomass material remains in the digester. Organic loading rate indicates the quantity of dry solids loaded per m^3 of digester volume and unit of time. Alkalinity ratio indicates the quantity of volatile organic acids in relation to the buffer capacity of carbonate. Biogas production rate measured during the sampling period is presented in Fig. 1. The succession of organic acids, TS, VS, NDF, ADL, ADF, alkalinity ratio, NH_4^+-N, HRT, and OLR is shown in Supplementary, Figs. S1–S4.

2.2. Computational platform

For the development of the neural network models the Matlab® Neural Network Toolbox Version 7.10 2010a (The MathWorks Inc., Natick, USA) provided with a user's guide [28] was used. A Matlab script generated by the Neural Network Toolbox loaded the data, trained, validated and tested the models. The data were normalized before application in the neural network using a script written in Matlab. The implemented optimization tools, here the ant colony optimization and the genetic algorithm were written in Matlab. All Matlab scripts are available free of costs from the corresponding author. The treatment of the experimental data and the model calculations were carried out on a processor AMD Phenom ™ II X2 B57 with 3.2 GHz.

2.3. Pre-processing technique

Data pre-processing is an essential step of any data analysis [29]. In this work a normalization method, the standard normal variate (SNV) was used. Eqs. (1) and (2) represent the mathematical computation of SNV-transformation.

$$\sigma = \sqrt{\sum_{k=1}^{n} \left(x_k - \bar{x}\right)^2 / (n-1)} \quad (1)$$

$$x_{k,norm} = \left(x_k - \bar{x}\right)/\sigma \quad (2)$$

Here x_k is a measured value, while $\bar{x}$ is the mean value of x_k. The computed σ represents the standard deviation of x_k.

2.4. Artificial neural networks

For the prediction of the biogas production rate, a two layered feedforward neural network from the Matlab® ANN Toolbox was applied. The ANN network was trained using the Levenberg-Marquardt algorithm, embedded in Matlab ANN Toolbox. ANN included an input layer, a hidden layer with a sigmoid activation function and an output layer with a linear output function. A random initialization was used to provide more accurate prediction results. The sigmoid transfer function at the hidden layer enables to treat the data with non-linear features. It takes the input (any values between plus and minus infinity) and squashes the output into the range 0–1 [30]. The sigmoid activation function is presented in Eq. (3).

$$f(x) = 1/(1 + \exp^{-x}) \quad (3)$$

The data of the measured variables were used as input neurons, while the biogas production rate was calculated by the output neuron. To identify the optimal model structure the number of hidden neurons was varied from 3 to 20. The data set was split into three data sets for training (70%, 28 samples), validation (15%, 6 samples) and test (15%, 6 samples). The network is trained to find the optimal weights, which minimize the cost functional. Here the RMSE training was used as a cost functional and was minimized. The validation data served to overcome overfitting and is used to stop training, which was determined by the failure of the error decrease for six consecutive iterations. Using the test data the prediction of the biogas production rate was done. The assessment of the calculated models was done using the root mean square error (RMSE) and the coefficient of determination (R^2), which are presented in Eqs. (4) and (5). RMSE was used to evaluate the accuracy of the models, while R^2 to evaluate their robustness. Here $\widehat{y_k}$ refers to the predicted value, whereas y_k represents the measured value and $\bar{y}$ is the mean value of y_k. n is the number of samples.

$$RMSE = \sqrt{\left(\sum_{k=1}^{n} \left(\widehat{y_k} - y_k\right)^2\right)/n} \quad (4)$$

$$R^2 = 1 - \left(\sum_{k=1}^{n} \left(\widehat{y_k} - y_k\right)^2\right) / \left(\sum_{k=1}^{n} \left(y_k - \bar{y}\right)^2\right) \quad (5)$$

RMSE and R^2 were computed for training, validation and test models.

2.5. Metaheuristic methods

For the variable selection purpose two metaheuristic methods were implemented, the genetic algorithm and an ant colony optimization. GA represents a mathematical interpretation of the Darwin's evolution theory, which is based on the natural rule of survival and the Mendel's principles of inheritance [31]. Due to the selection of the best individuals and further mutation and cross-over transformation of their genome only the best qualities will be transferred to the next generation [32]. In this work, the implemented GA was used to identify the significant process variables [33]. The principle of the

GA variable selection is based on the lowest cross-validated RMSE of the calculated models. Thus, the significant process variables are those used in the partially least square regression models (PLSR) with the lowest cross-validated RMSE. For the applied GA approach the following algorithm parameters were set: PLS as regression model type, cross over points of 2, mutation rate of 0.005, 15 individuals with a population size of 256 and 100 runs. PLSR is a multivariate statistical tool, which identifies the latent variables (factors) regarding the most of variance in data and linear correlation between the process variables. The cross-over operator is related to the natural reproduction cross-over and will be set randomly. At each step GA selects randomly 15 individuals out of a set population during all runs.

The principle of ACO was inspired by natural behavior of ants seeking for food [23]. They leave on the way to food and back to the nest a pheromone trail, which serves as a natural guide for other ants. Accordingly the "shortest ways" have the highest pheromone concentration and will be frequently used by ants. Hence, the pheromone concentration can serve as an independent criterion for the variable selection [19]. In the approach used here, the measured process variables represent the possible ways to the "food source", namely the biogas production rate. The virtual ants use different combinations of measured process variables to find out the best possible solution. The pheromone concentration will be calculated for each process variable separately and depends on the quality of correlation with the predicted variable. Generally, pheromone concentration is in range between 0 and 1. The measured process variables, which correlated best with the biogas production rate and are mostly used by the virtual ants, have a higher pheromone concentration in comparison to those correlating less with the predicted variable. The measured variables which had the highest pheromone concentration are the ones most important for the prediction of the biogas production rate. For ACO the following algorithm parameters were applied: 100 ants, PLS as regression method with a principal component number von 1–15, initial pheromone concentration of 10^{-6} and pheromone evaporation rate of 0.5 per iteration, 50 iterations. Here the pheromone trail is an independent evaluation parameter, which will be adapted by ants to determine the solutions to the problem. It will be updated for each variable separately regarding the evaporation rate during the ACO implementation. The implemented ACO algorithm proceeds in three stages, the 0th stage is the initiation, where the model parameters will be set. In the 1st stage the objective function will be calculated and the best global output variables will be defined, following by a number of cycles. In the last stage the results will be presented (Table 2).

3. Results & discussion

3.1. *Ann prediction of the biogas production rate*

The input layer of the implemented ANN included 15 neurons, which were represented by the measured process variables; the number of hidden neurons was varied from 3 to 20 in order to get the optimal model structure. The biogas production rate resulted as an output neuron. In general, the results of 5–20 hidden neurons did not show significant differences. The best prediction results with a small-dimensioned model structure are shown in Table 3. The less successful prediction of the biogas production rate is not shown.

As can be seen in Table 3, the prediction of biogas production rate using 15 input neurons was successful. The best results were achieved using 5 and 10 hidden neurons, although the results obtained with more than 5 neurons are not significantly different and were not shown. The models with 10 hidden neurons were more robust and accurate in comparison to models with 5 hidden neurons. Here the training RMSE was 4.81% and the training R^2 was 0.90. The test RMSE was 9.66%, while the test R^2 reached 0.80. The evolution of the predicted biogas production rate done with ANN using 15 input neurons and 10 hidden neurons is presented in Fig. 2.

3.2. *Optimized ANN prediction of the biogas production rate*

The metaheuristic tools implemented in the model were used to identify the significant process variables and to improve model performance. The calculated pheromone concentration served as an independent evaluation criterion to select the significant process variables. It depends on the correlation with the predicted biogas production rate. The calculated value equivalent to the pheromone concentration in the ant

Table 2 – Stages of the ACO algorithm.

Stage	Routines
1. Initiation	Initial pheromone concentration Model type Pheromone evaporation rate Number of ants Number of iterations Number of principal components
2. Calculation and iterations	Objective function will be calculated The best global output variables will be defined Iterations will be performed The best ants will be selected and compared with the best stored ones Update of the pheromone trail Selection of the best variable combinations
3. Results	Display of the results

Table 3 – ANN prediction of the biogas production rate using 15 input neurons.

Number of input neurons	15	15
Number of hidden neurons	5	10
Number of iterations	12	8
RMSE training [%]	5.89	4.81
RMSE validation [%]	13.66	5.15
RMSE test [%]	**13.08**	**9.66**
R^2 training [–]	0.90	0.88
R^2 validation [–]	0.46	0.70
R^2 test [–]	**0.76**	**0.80**

The results of the prediction were highlighted in bold. Other results belong to validation and training models.

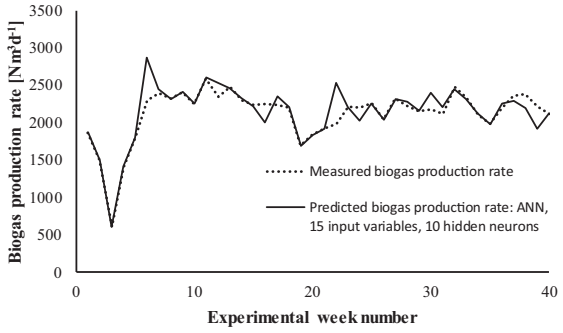

Fig. 2 – Prediction of the biogas production rate done with ANN using 15 input variables and 10 hidden neurons.

model of all variables is shown in Supplementary (Table S1). The measured process variables with a virtual pheromone concentration value higher than 0.5 were identified as the most significant ones. The ACO model identified seven significant process variables, namely TS, HRT, NDF, ADL, alkalinity ratio, n-butyric acid, and iso-valeric acid.

The GA algorithm selected nine process variables, namely HRT, TS, NDF, ADF, ADL, propionic acid, n-butyric acid, iso-butyric acid, and n-valeric acid. Five process variables, namely HRT, TS, NDF, ADL, and n-butyric acid were identified as process relevant by both optimization algorithms.

The selected variables were used for the further prediction of the biogas production rate. Therefore, nine significant process variables selected by GA, seven significant process variables selected by ACO and five significant process variables selected by both optimization algorithms were used. The prediction results are presented in Table 4.

Regarding in Table 4 presented results a good prediction performance could be achieved using a small-dimensioned model structure. The ACO- and GA-optimized ANN models showed generally similar results. For the GA-optimized ANN models more hidden neurons were required. A more accurate prediction was obtained using the ACO selected process variables and 10 hidden neurons. Here, the training RMSE was 3.53%, while the training R^2 was 0.98. The test RMSE was 10.37%, and the test R^2 was 0.83. The models with the significant process variables selected by both algorithms showed the best result. Here, the best model performance was achieved using 5 input neurons and 5 hidden neurons. The training models had RMSE of 3.49% and an R^2 of 0.98. By the test models RMSE was reduced to 6.24%, the R^2 reached 0.9. The regression performance of the GA-ACO-optimized ANN models is shown in Fig. 3.

The dynamic evolution of the predicted biogas production rate done with the optimized ANN models are shown in Figs. 4–6.

The achieved results proved the implemented approach to be a feasible methodology to analyze the biogas production process. The implemented ANN models predicted successfully the biogas production rate. In contrast to the approach used in this study, previously published studies used more complex ANN to achieve a good prediction performance in context of biomethanation processes. For example, Ozkaya et al. used eight input neurons and 15 hidden neurons to predict the methane fraction of biogas [13]. In another study published by Sahinkaya et al., six input neurons and 20 hidden neurons were required to predict sulphate, acetate and sulphide concentrations in a thermophilic sulfidogenic fluidized-bed reactor [12]. In our study we applied five input neurons and five hidden neurons to predict the biogas production rate. Thus, an effective simplification of the model dimension was done due to the implemented optimization algorithms. In addition, the used metaheuristics enabled an accurate variable selection by defining the significant measured process variables. The approach using an ACO algorithm is unique in the field of assessment of AD systems. Several publications used a genetic algorithm to optimize ANN models. Abu Qdais et al. used GA to optimize the methane output of the bioreactor and to define the best

Table 4 – Prediction results done using the optimized ANN models.

	GA optimized		ACO optimized		GA-ACO optimized
Number of input neurons	9	9	7	7	5
Number of hidden neurons	10	15	5	10	5
Number of iterations	14	12	16	13	18
RMSE training [%]	6.98	13.90	11.21	3.53	3.49
RMSE validation [%]	11.68	13.89	6.97	7.07	3.30
RMSE test [%]	**13.70**	**11.86**	**12.37**	**10.37**	**6.24**
R^2 training [–]	0.95	0.85	0.94	0.98	0.98
R^2 validation [–]	0.62	0.74	0.56	0.56	0.96
R^2 test [–]	**0.80**	**0.89**	**0.75**	**0.83**	**0.90**

The results of the prediction were highlighted in bold. Other results belong to validation and training models.

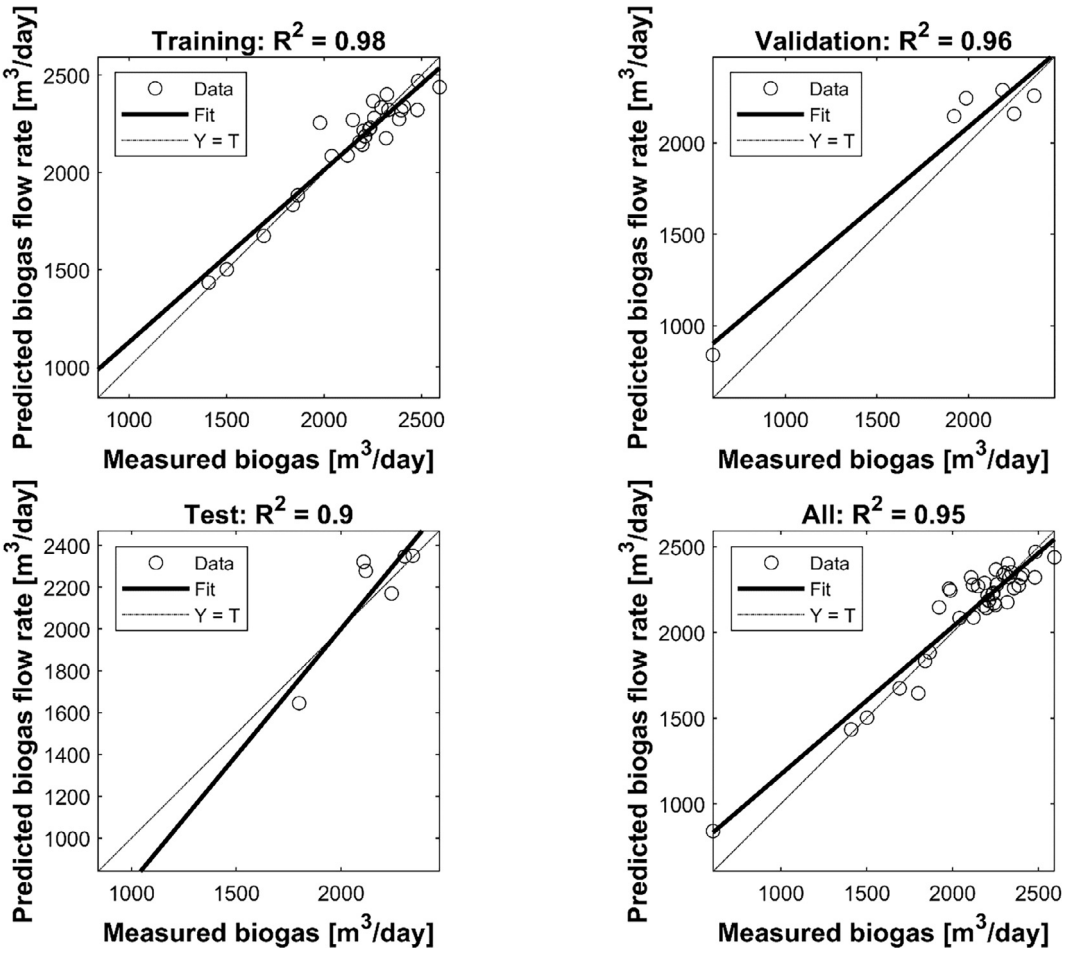

Fig. 3 – Regression performance of the GA-ACO-optimized ANN models: (1): training; (2): validation; (3): test; and (4): all data.

operational conditions while Gueguim Kana et al. (2012) applied GA to improve the biogas production on saw dust and other co-substrates [34,35]. The use of two optimization algorithms, namely ACO and GA, was aimed to assess their feasibility in field of variable selection by AD systems. Thus five significant measured process variables could be identified, which enabled an accurate prediction of the biogas production rate. Thus, the developed approach represents a fast and robust method to analyze the process evolution. In comparison to other commonly accepted models, such as ADM1, the methodology developed in this study requires only a small number of process variables to perform a successful evaluation of the process. In contrast to it, ADM1 needs a comparatively huge number of process variables and kinetic parameters as well as rates additionally determined for each kind of substrate, that makes its application complex.

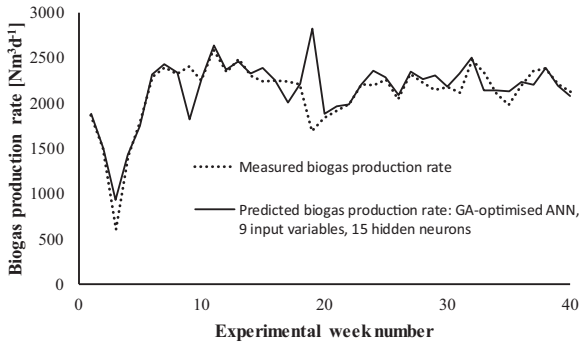

Fig. 4 – Prediction of the biogas production rate calculated using the GA-optimized ANN with nine input neurons and 15 hidden neurons.

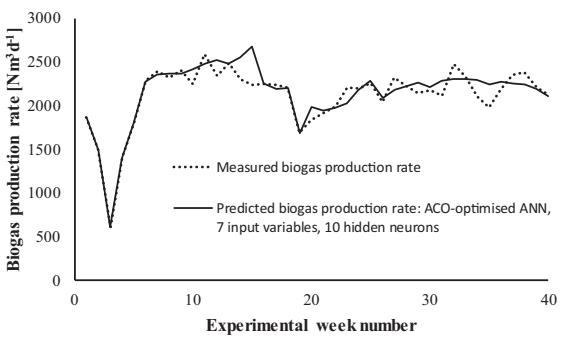

Fig. 5 – Prediction of the biogas production rate calculated using the ACO-optimized ANN with seven input variables and 10 hidden neurons.

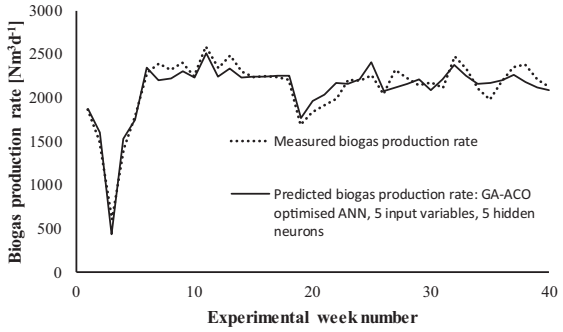

Fig. 6 – Prediction of the biogas production rate calculated using ACO-GA optimized ANN with five input variables and 5 hidden neurons.

4. Conclusion

The approach developed in this study represents a fast and reliable method to evaluate the biogas production process. Herewith the evaluation and prediction of typical process variables namely the biogas production rate could be performed. For the prediction of the biogas production rate, ANN models were implemented. The used ANN models had a simple structure [36] performing a robust and accurate prediction of the biogas production rate. The optimization tools were used to evaluate the process variables selecting the significant ones. The used variable selection tools made it possible to reduce the model dimension and to improve its performance. The best results were gained using ACO-GA optimized ANN models. Here the prediction error was reduced to 6.24% and R^2 increased to 0.90. The developed approach can be further used to develop an on-line control, which will help to improve the process conditions and to prevent possible process failures. The variable evaluation tools can support the operating engineer with information about the main process correlations. The developed approach demonstrates that ANN in combination with GA and ACO optimization tools showed reliable results in evaluation of the biogas production rate. Here the neural logic could predict the process development, while the optimization tools could improve the prediction capacity by selection of the significant process variables. Moreover, for the modelling a small number of data sets was required. This strategy can be used as a control operator to evaluate the process development based on the measured data. It can be used as an alternative approach to replace the computationally intensive and time-consuming ADM1 as well as to speed up the simulation procedure of biological processes [37]. Another way to use the developed approach is to evaluate the composition of the substrates. The intelligent model can rapidly estimate the best process conditions, accurately analyzing the process variables regarding the complex non-linear process behavior. It will help to improve the process development, to gain the highest biogas output, saving time, costs and to avoid the time-consuming and expensive analytics. The developed approach can be used for different AD processes regarding chemical and technical influence coefficients. Nevertheless, it is recommended in future studies to focus on the evaluation of the effects of the individual input variables. For that a more elaborated data is required.

Conflict of interest

None.

Acknowledgement

This work was part of the joint projects BIOGAS-ENZYME and BIOGAS-BIOCOENOSIS supported by the German Federal Ministry of Food and Agriculture (BMEL), grant nos. 22027707, 22010711 and 22028911 [27].

Our sincere gratitude to Robert Kausmann and Susanne Theuerl for the support by this research.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.inpa.2019.01.002.

REFERENCES

[1] Klang J, Theuerl S, Szewzyk U, Huth M, Tölle R, Klocke M. Dynamic variation of the microbial community structure during the long-time mono-fermentation of maize and sugar beet silage. Microb Biotechnol 2015;8:764–75.
[2] Theuerl S, Kohrs F, Benndorf D, Maus I, Wibberg D, Schlüter A, et al. Community shifts in a well-operating agricultural biogas plant: how process variations are handled by the microbiome. Appl Microbiol Biotechnol 2015;99:7791–803.
[3] Friedmann H, Görtz G, Helm M, Kaltschmitt M, Knopf U, et al. Fermentation of organic materials – characterisation of the substrate, sampling, collection of material data fermentation tests (Vdi 4630). In: Vdi-Handbuch Energietechnik, Verein Deutscher Ingenieure. Düsseldorf: Beuth; 2006. p. 1–48.
[4] Kessler W. Multivariate Datenanalyse: Für Die Pharma-, Bio- Und Prozessanalytik; Ein Lehrbuch. Weinheim: Wiley-Vch; 2007. in German.
[5] Hitzmann B, Ritzka A, Ulber R, Schöngarth K, Broxtermann O. Neural networks as a modeling tool for the evaluation and analysis of FIA signals. J Biotechnol 1998;65:15–22.
[6] Hitzmann B, Kullick T. Evaluation of ph field effect transistor measurement signals by neural networks. Anal Chim Acta 1994;294:243–9.
[7] Sabzi S, Abbaspour-Gilandeh Y, Garcia-Mateos GA. New approach for visual identification of orange varieties using neural networks and metaheuristic algorithms. Inform Process Agric 2018;5:162–72.
[8] Golhani K, Balasundram SK, Vadamalai G, Pradhan B. A review of neural networks in plant disease detection using hyperspectral data. Inform Process Agric 2018;5:354–71.
[9] Ma W, Fan J, Li Q, Tang YA. Raw milk service platform using BP neural network and fuzzy inference. Inform Process Agric 2018;5:308–19.
[10] Sehgal VK, Charkraborty D, Sahoo RN. Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. Inform Process Agric 2018;3:107–18.
[11] Strik DP, Domnanovich AM, Zani L, Braun R, Holubar P. Prediction of trace compounds in biogas from anaerobic

digestion using the matlab neural network toolbox. Environ Modell Softw 2005;20:803–10.
[12] Sahinkaya E, Ozkaya B, Kaksonen AH, Puhakka JA. Neural network prediction of thermophilic (65 °C) sulfidogenic fluidized-bed reactor performance for the treatment of metal-containing wastewater. Biotechnol Bioeng 2007;97 (4):780–7.
[13] Ozkaya B, Demir A, Bilgili M. Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environ Modell Softw 2007;22:815–22.
[14] Kanat G, Saral A. Estimation of biogas production rate in a thermophilic UASB reactor using artificial neural networks. Environ Model Assess 2009;14:607–14.
[15] Bernardi B, Zema DA, Tamburino V, Zimbalatti G. An automated medium scale prototype for anearobic co-digestion of olive mill wastewater. Inform Process Agric 2017;4:316–20.
[16] Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, et al. Anaerobic digestion model No. 1: [(Adm1)]. London: Iwa Publ; 2002.
[17] Beltramo T, Ranzan C, Hinrichs J, Hitzmann B. Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm. Biosyst Eng 2016;143:68–78.
[18] Goldberg DE, Holland JH. Genetic algorithms and machine learning. Mach Learn 1988;3:95–9.
[19] Dorigo M, Blum C. Ant colony optimization theory: a survey. Theoret Comput Sci 2005;344:243–78.
[20] Shmygelska A, Hoos HH. An ant colony optimisation algorithm for the 2d and 3d hydrophobic polar protein folding problem. BMC Bioinf 2005;6:30.
[21] Kosakovsky Pond Sl, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. Bioinformatics 2006;22:3096–8.
[22] Lindner PFO, Hitzmann B. Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the fisher information matrix. J Theor Biol 2006;238:111–23.
[23] Ranzan C, Strohm A, Ranzan L, Trierweiler LF, Hitzmann B, Trierweiler JO. Wheat flour characterization using NIR and spectral filter based on ant colony optimization. Chemometr Intell Lab 2014;132:133–40.
[24] Nache M, Scheier R, Schmidt H, Hitzmann B. Non-invasive lactate- and pH-monitoring in porcine meat using Raman spectroscopy and chemometrics. Chemometr Intell Lab 2015;142:197–205.
[25] Silalahi DD, Reano CE, Lansigan FP, Panopio RG, Bantayan NC. Using genetic algorithm neural network on near infrared spectral data for ripeness grading of oil palm (Elaeis Guineensis Jacq.) fresh fruit. Inform Process Agric 2016;4:252–61.
[26] Taki M, Mehdizadeh SA, Rahnama M, Rahmati-Joneidabad MR. Applied machine learning in greenhouse simulation; new application and analysis. Inform Process Agric 2018;2:253–68.
[27] Klocke M, editor. Prozessmikrobiologie In Landwirtschaftlichen Biogasanlagen: Ermittlung Der Mikrobiellen Diversität Sowie Von Hauptsächlichen Verfahrenstechnischen Einflussfaktoren Auf Die Mikroflora. Potsdam-Bornim: Leibniz-Institut Für Agrartechnik Potsdam-Bornim E.V.; 2014 [in German].
[28] Demuth H, Beale M. Neural network toolbox user's guide. Natick: The Mathworks; 2000.
[29] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab 2001;58:109–30.
[30] Hagan MT, Demuth HB, Jesus OD. An introduction to the use of neural networks in control systems. Int J Robust Nonlinear Control 2002;12:959–85.
[31] Mccall J. Genetic algorithms for modelling and optimisation. J Comput Appl Math 2005;184:205–22.
[32] Leardi R. Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection. J. Chemometrics 1994;8:65–79.
[33] Tsai C-F, Eberle W, Chu C-Y. Genetic algorithms in feature and instance selection. Knowl-Based Syst 2013;39:240–7.
[34] Gueguim Kana EB, Oloke JK, Lateef A, Adesiyan MO. Modeling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renew Energ 2012;46:276–81.
[35] Abu Qdais H, Bani Hani K, Shatnawi N. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resour Conserv Recy 2010;54:359–63.
[36] Li C, Li Z, Wu J, Zhu L, Yue J. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Inform Process Agric 2018;5:11–20.
[37] Samadianfard S, Ghorbani MA, Mohammadi B. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled-hybrid firefly optimizer algorithm. Inform Process Agric 2018;5:465–76.

3.3 “Evaluation of the linear and non-linear prediction models optimized with metaheuristics: application to anaerobic digestion processes”

Authoren: Tanja Beltramo, Bernd Hitzmann

Veröffentlicht im Juni 2019 in *Engineering in Agriculture, Environment and Food*

Online verfügbar unter doi.org/10.1016/j.eaef.2019.06.001

Engineering in Agriculture, Environment and Food 12 (2019) 397–403

Contents lists available at ScienceDirect

Engineering in Agriculture, Environment and Food

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

Evaluation of the linear and non-linear prediction models optimized with metaheuristics: Application to anaerobic digestion processes

Tanja Beltramo*, Bernd Hitzmann

Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, 70599, Germany

ARTICLE INFO

Keywords:
Anaerobic digestion model No.1
Ant colony optimization
Artificial neural networks
Genetic algorithm
Locally weighted regression
Partial least square regression

ABSTRACT

This research represents an evaluation study of the linear and non-linear mathematical methods applied to predict the biogas flow rate in anaerobic digestion processes. The anaerobic digestion model No.1 was used to generate the process data. For the prediction of the biogas flow rate the partially least squares regression, the locally weighted regression and the artificial neural networks were used. Two metaheuristic tools, here a genetic algorithm and an ant colony optimization algorithm were applied to improve the prediction models. They carried out the variable selection procedure. The implemented mathematical models could successfully perform the prediction of the biogas flow rate. Nevertheless, more robust and accurate prediction of the biogas flow rate was done with the help of the artificial neural networks. Here the error of prediction was about 9% while the coefficient of determination reached 0.97.

1. Introduction

The biogas production is a complex anaerobic digestion (AD) process with a high number of microbial populations and metabolites. The recent scientific knowledge about the process control is not extensive enough, even for the measurement of biogas production rate or for the analysis of the concentration of accumulated intermediates (Donoso-Bravo et al., 2011; Björnsson et al., 2000). A better process development could be achieved, if a reliable process monitoring system was developed, which could evaluate the entire process with respect to the microbial activities. The direct monitoring of the microbial activities is very complicated, but it can be indirectly measured for example by means of the biogas production rate or other process variables. In literature there are different mathematical models, which describe the AD processes, for example the well-established anaerobic digestion model No. 1 (ADM1). ADM1 was developed by the *IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes* (Batstone, 2002). Primarily ADM1 was used to simulate the AD processes of activated and primary sludge. Later it was successfully applied to simulate the AD processes with respect to different substrates and different reactor conditions (Lauwers et al., 2013). For example it was applied for the agro-waste simulation (Galí et al., 2009), for the anaerobic digestion of grass, maize, green weed silage and industrial glycerin (Biernacki et al., 2013) as well as for the fermentation of the cattle-manure and the renewable energy crops (Lübken et al., 2007). Other mathematical solutions used for the monitoring of the AD processes are represented by the chemometric methods, i. a. the regression models (Kessler, 2008). For example the multilinear regression technique (MLR) is a classical linear least squares regression method mostly used for experimental design. Another well-known linear regression technique is the principal component regression (PCR), which combines the principal component analysis (PCA) and MLR. The partial least squares regression (PLSR) is another linear regression method related to PCR, but differs with respect to the calculated variance (Wold et al., 1984). For PLSR the correlation between the predicted variables and predictors will be used, while PCR method considers only the correlation of the predictors (Grote et al., 2014). The locally weighted regression (LWR) is a non-linear regression method (Cleveland et al., 1988). The principle of LWR is based on finding a number of local samples in data, which can be linearly approximated. The above described regression tools were successfully implemented in the field of monitoring of the AD processes. Thus PCR was used to control the sewage sludge anaerobic digester (Reed et al., 2013) and to predict the effects of toxics in anaerobic granular sludge (Costa et al., 2009). PLSR was implemented to predict the total solids in bio-slurry (Ihunegbo et al., 2012), to control the methanogen density (Zhang et al., 2002), to predict the intermediates in the biogas process (Holm-Nielsen and Esbensen, 2011) and to analyze the AD process (Beltramo et al., 2014). LWR was used to identify ammonium and acetate (Dahlbacka et al., 2013) and to analyze the ammonia removal from the lagoon biogas (Lin et al., 2014).

* Corresponding author.
E-mail address: tanja.beltramo@gmail.com (T. Beltramo).

https://doi.org/10.1016/j.eaef.2019.06.001
Received 6 June 2018; Received in revised form 6 May 2019; Accepted 1 June 2019
Available online 11 June 2019

Artificial neural networks (ANN) belong to the cognitive sciences and represent a non-linear statistical learning tool (Wasserman, 1993). Neural networks belong to a black-box modelling and are used to evaluate the processes without prior knowledge about the structure or dependencies within the system (Hitzmann et al., 1997) and a modified tool in computer aided diagnostics (Wang et al., 2017). The ANN models were frequently used also to evaluate the AD processes. For example it was used to predict the biogas production rate (Abu Qdais et al., 2010; Beltramo et al., 2016), to control the anaerobic digestion processes (Holubar, 2002), to predict the methane fraction in biogas (OZKAYA et al., 2007) and to evaluate the trace compounds in biogas (Strik et al., 2005).

Metaheuristic methods represent a new approach used to solve the complex mathematical problems, where the conventional methods fail. These include swarm intelligence or genetic algorithms. The ant colony optimization (ACO) belongs to the swarm intelligence algorithms (Dorigo and Blum, 2005) based on the structure of the ant's natural behavior looking for their food. The ants leave a pheromone trace on the way to the food, which serves as a communication means for other ants. ACO was used to investigate the difficulties by protein folding (Shmygelska and Hoos, 2005) and to calculate the major compatibility complex (MHC) class II binders (Karpenko et al., 2005). It was implemented as an optimization tool to evaluate the wheat flour (Ranzan et al., 2014) and to predict the pH level and lactate in porcine meat (Nache et al., 2015). In the field of anaerobic digestion processes ACO was used as a variable selection method (Beltramo et al., 2016). An ACO-GA combined model was used in economic sciences to predict the corporate bankruptcy (Zhang et al., 2013).

Other popular metaheuristic methods are the genetic algorithms (GA). The GA approach represents a mathematical interpretation of the Darwin's evolution theory based on the “survival of the fittest” and the Mendel's inheritance laws (Whitley, 1994). In this regard the genome of the best individuals undergoes genetic changes through mutation and cross-over transformations. As a result, the next generation will be provided with the best features. In the mathematical sense of view, the fittest individuals represent the optimal solutions for a definite problem. GA was also applied to optimize the AD processes, for example to improve the biogas production on saw dust and other substrates (Gueguim Kana et al., 2012) and from a waste digester (Abu Qdais et al., 2010) as well as to simulate the biological hydrogen production (Mu and Yu, 2007).

The main objective of this research was to investigate the appropriateness of different linear and non-linear strategies used for the monitoring of the AD processes. Furthermore, the evaluation of the optimization tools according to the feature selection capability was done. The process development data were simulated by means of a reliable and well accepted model from the literature. Thereby, the substrate configuration was varied in order to get a fair comparison to the real process conditions. The flow chart of the modelling procedure is presented in Fig. 1.

2. Materials and methods

2.1. Computational platform

For the development of the neural network models the Matlab® Neural Network Toolbox Version 7.10 2010a provided with a user's guide (Demuth and Beale, 2000) was used. For the LWR and PLSR models, the normalization routine, the noise factor and the metaheuristic algorithms Matlab scripts were written (The MathWorks Inc., Natick, USA). The treatment of the generated data and the model calculations were carried out on a processor AMD Phenom ™ II X2 B57 with 3.2 GHz.

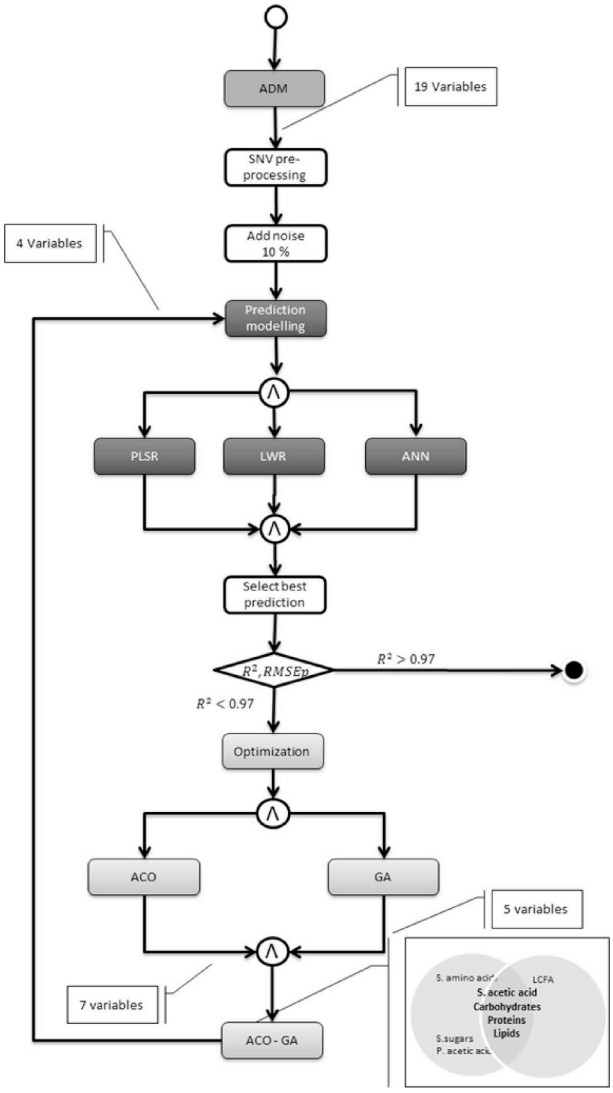

Fig. 1. Flowchart of the modelling procedure.

2.2. Data generation

The modified ADM1 (Rosen et al., 2006) was implemented to generate the process data (Rosen and Jeppsson). The simulation methodology is presented in Beltramo et al., (2016). The simulation procedure was performed for 50 days of the co-fermentation of three agricultural substrates corn-silage, cow-manure and grass-silage. The feed sequence and the substrates flow rate were varied. The input variables and parameters used to describe the substrate types were taken from the literature (Wichern et al., 2009), (Zhou et al., 2011). In Table 1 the used substrates, the sequence of the substrate feed and the

Table 1
Reactor conditions. The flow chart of the ADM1 is represented in Supplementary, Fig. S1.

Time range [day]	Substrate type	Flow rate [$m^3 d^{-1}$]
0–10	Corn-silage	100
11–20	Cow –manure	170
21–30	Grass-silage	100
31–35	Corn-silage	170
36–40	Cow –manure	100
41–45	Grass-silage	170
46–50	Corn-silage	100

substrate flow rates are represented.

2.3. Signal noise

The undesirable signal modifications will be defined as a signal noise. There are many factors, which can cause the data noise, such as the inappropriate analytical methods, defective equipment, problems by the storage or processing of the data. In order to recover the original data a noise factor will be added to the simulated data. In this work the Gaussian noise of 10% was added. Its probability density function *p(z)* is represented as a normal distribution Eq. (1).

$$p(z) = (1/\sigma\sqrt{2\pi})e^{-(z-\mu)/2\sigma^2} \quad \text{Eq. (1)}$$

Here μ describes the mean value, while z is the random variable and σ the standard deviation.

2.4. Data pre-processing

Data pre-processing is an important step of any qualitative or quantitative data analysis (Christen, 2012; Sun, 2009). Data pre-processing removes the noise and transforms the data for the better further information processing. There are numerous pre-processing methods in literature, depending on the task in data pre-processing. In this work the standard normal variate (SNV) was used to normalize the data. The calculation routine of SNV is shown in equations (2) and (3).

$$x_{i,norm} = (x_i - \bar{x})/\sigma \quad \text{Eq. (2)}$$

$$\sigma = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1)} \quad \text{Eq. (3)}$$

Here x_i is a simulated value including the addition of noise, while $\bar{x}$ is the mean value of x_i. The computed σ represents the standard deviation of all x_i.

2.5. Partial least square regression

The PLSR models with a leave-one-out cross-validation (cv) technique were applied to predict the biogas flow rate. The number of principal components (PC) was varied to get the best model dimension. In the present work only the best results are shown. The model evaluation was carried out using two model quality parameters, the cross-validated root mean square error (cv-RMSE) and the coefficient of determination (R^2). RMSE estimated the prediction accuracy of the calculated models, while R^2 represented their robustness. The calculation routines of the model quality parameters are shown in equations (4) and (5).

$$RMSE = \sqrt{\left(\sum_{k=1}^{n} (\hat{y_k} - y_{s,k})^2/n\right)} \quad \text{Eq. (4)}$$

$$R^2 = 1 - \left(\sum_{k=1}^{n} (\hat{y_k} - y_{s,k})^2\right) / \left(\sum_{k=1}^{n} (y_{s,k} - \bar{y})^2\right) \quad \text{Eq. (5)}$$

Here $y_{s,k}$ are the simulated values, $\bar{y}$ is the mean of $y_{s,k}$, while $\hat{y_k}$ are the predicted values.

2.6. LWR

LWR is a non-linear regression technique. Its main characteristic is the regression procedure on the local level (Cleveland et al., 1988), which removes of the non-linarites from the data set. Thus only the defined local samples remain in the model, which will be subsequently linearly approximated. From the mathematical point of view the local points are the closest points in the score place, calculated according to the “Mahalanobis distance” principle. The number of the local points and the principal components number were varied in order to define the best model dimension. The evaluation of the LWR models was done using two model quality parameters, cv-RMSE and R^2 (Eqs. (4) and (5)).

2.7. Artificial neural networks

The structure of the neural network used in this work is described in Beltramo et al., (2016). The implemented ANN model used a two-layer feedforward network with a sigmoid transfer function in the hidden layer and a linear transfer function at the output layer. Training used the Levenberg-Marquardt backpropagation algorithm imbedded in the Matlab ® Neural Networks Toolbox. The input neurons represented the independent input variables. The biogas flow rate was the dependent output neuron. The generated data were split into three data sets, training data (70%, days 1–35), validation data (15%, days 36–43) and test data (15%, days 44–50). The training data set adjusted the models, while the test data set was used for the prediction. The aim of the validation data set was to stop the training before overfitting. The number of the hidden layers and the number of neurons were varied. The evaluation of the calculated models was done using the model quality parameters RMSE and the R^2 (see Eq. (4) and Eq. (5)), which were computed for the training, validation and test models separately.

2.8. Metaheuristics

The main objective to apply metaheuristic methods in this work was to improve the model prediction performance. Due to a high number of process variables existing in biogas production process it is problematic to provide the relevant process information. The applied optimization tools performed a smart feature selection procedure searching for the best process variables and their combinations. In this work two optimization tools were applied, namely an ant colony optimization and a genetic algorithm. The implemented ACO tool is a discrete type, by which the variable selection procedure was executed using a random factor correlated with a pheromone density function (Mullen et al., 2009). The pheromone concentration, calculated for each process variable, was a key feature to define the significant process variables. The pheromone concentration was between 0 and 1. In this work the process variables with a pheromone concentration higher than 0.5 units were identified as process relevant. The ACO parameters set in this research are shown in Table 2. The sequence of the algorithm phases is presented in Supplementary, Table S1.

GA is another optimization tool used in this research. The G_0 generation included 256 individuals with a cross-over rate of 2 and a mutation rate of 0.005. The algorithm performed 100 runs. The variable selection procedure was based on the evaluation of the best cv-RMSE results of the calculated PLSR models. The GA parameters are represented in Table 3.

3. Results and diskussion

3.1. Data generation results

ADM1 was used to simulate the co-digestion of three agricultural substrates. For comparable results a typical input composition was chosen, namely a co-digestion of cow-manure, grass silage and corn-

Table 2
ACO parameters.

Number of ants	100
Number of iterations	50
Model type	PLS
Initial pheromone concentration	10^{-6}
Evaporation rate	0.5
Principal component number (PC)	15

Table 3
GA parameters.

Population size	256
Regression method	PLS
Response	Cross-over
Mutation rate	0.005
Cross-over	2
Number of runs	100
Number of individuals	10

silage. The simulation was performed for 50 fermentation days with a frequency of 20 simulation points per day. The dynamic evolution of the biogas flow rate is shown in Fig. 2.

The dynamic evolutions of soluble amino acids, soluble acetic acid, soluble long chain fatty acids (LCFA), soluble sugars, lipids, proteins and particulate acetic acid are shown in Figures S1 – S7 (Supplementary).

3.2. Model optimization

The metaheuristic tools used in this research to optimize the prediction models performed a feature selection, reducing the model dimension and providing the implemented models with the significant process information. The data set included 19 process variables, generated by ADM1. The process variables used for the prediction of the biogas flow rate are shown in Table 4.

Accordingly, ACO picked out seven significant process variables, which pheromone concentration was higher, than 0.5 units (Table S2). GA selected five significant process variables, which were included in the models with the lowest cv-RMSE. The process variables soluble acetic acid, particulate carbohydrates, proteins and lipids were selected by both optimization tools.

3.3. PLSR prediction results

A lot of PLSR models were calculated for the prediction of the biogas flow rate. The primary models included 19 process variables as the input. The optimized models however had seven, five and four significant process variables, selected by the optimization tools. For a better comparison, the PLSR models had the same principal components number (PC). The prediction results gained with the help of the PLSR models are shown in Table 6.

The implemented linear regression models showed good prediction results. All models were robust with an R^2 higher than 0.9. The

Table 4
Process variables. The significant process variables selected by the optimization tools are represented in Table 5.

Variable name	Abbreviation	Unit
Soluble sugar	S_{su}	kg COD/m^3
Soluble amino acids	S_{aa}	kg COD/m^3
Soluble LCFA	S_{fa}	kg COD/m^3
Soluble valeric acid	S_{va}	kg COD/m^3
Soluble butyric acid	S_{bu}	kg COD/m^3
Soluble propionic acid	S_{pro}	kg COD/m^3
Soluble acetic acid	S_{ac}	kg COD/m^3
Soluble inorganic carbon	S_{ic}	kmoleN/m^3
Soluble inorganic nitrogen	S_{in}	kmoleN/m^3
Soluble composites	S_i	kg COD/m^3
Particulate carbohydrates	X_{ch}	kg COD/m^3
Proteins	X_{pr}	kg COD/m^3
Lipids	X_{li}	kg COD/m^3
Particulate sugar	X_{su}	kg COD/m^3
Particulate amino acids	X_{aa}	kg COD/m^3
Particulate LCFA	X_{fa}	kg COD/m^3
Particulate propionic acid	X_{pro}	kg COD/m^3
Particulate acetic acid	X_{ac}	kg COD/m^3
Inert particulates	X_i	kg COD/m^3

Table 5
ACO and GA selected variables.

ACO-selected variables	GA-selected variables
Soluble sugar	LCFA
Soluble amino acids	Soluble acetic acid
Soluble acetic acid	Particulate carbohydrates
Particulate carbohydrates	Proteins
Proteins	Lipids
Lipids	
Particulate acetic acid	

Table 6
PLSR prediction results.

Optimization	None	ACO	GA	GA + ACO
Pre-processing	SNV	SNV	SNV	SNV
Input variables	19	7	5	4
PC number	3	3	3	3
cv-RMSE [%]	16.34	13.71	12.34	10.19
R^2 [-]	0.92	0.91	0.93	0.94

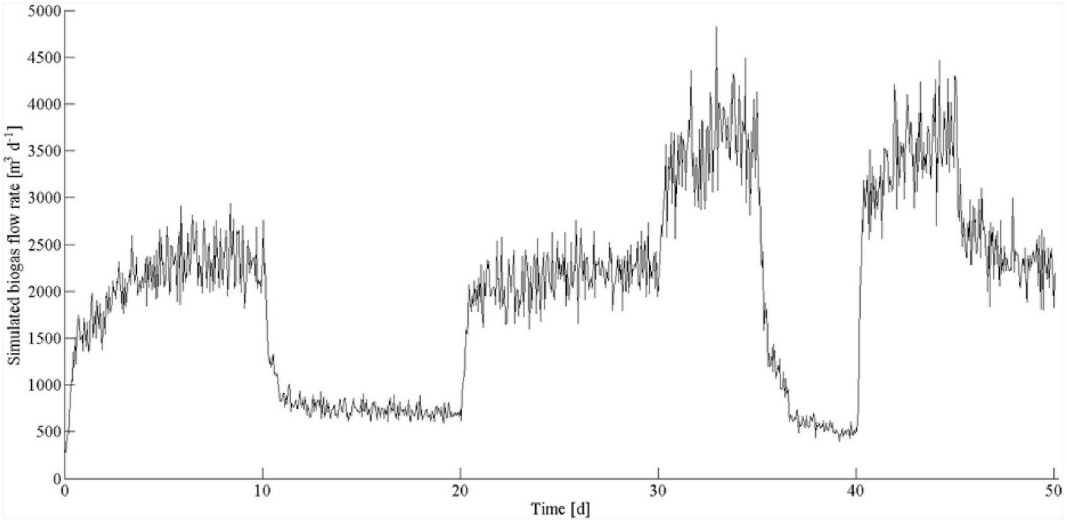

Fig. 2. Dynamic evolution of the biogas flow rate.

accuracy parameter cv-RMSE was higher for the non-optimized models (16.34%). The optimized PLSR models were more precise with a prediction error of 12.34% for the GA-optimized models and 13.71% for the ACO-optimized models. Regarding the optimization procedure, the used GA tool was more appropriate. The best prediction of the biogas flow rate was done with the four process variables selected by both optimization algorithms. Here cv-RMSE was 10.19%, while R^2 reached 0.94. Thus, for the PLSR models the best prediction of the output variable biogas flow rate was done using the small-dimensioned models.

3.4. LWR prediction results

The LWR model type is a non-linear prediction method used in this research to predict the output variable biogas flow rate. This type of modelling performs the prediction on a local level, searching for several local points in data, which will be linear approximated. In this work the number of local points was defined to 500 for all models, excepting the ACO-optimized LWR models. For the prediction of the biogas flow rate on the local level the PLSR models were used. For a reliable comparison, the number of PCs in the PLSR models was hold constant of 3 PCs, excepting the ACO-optimized models with 4 PCs. For the calculation of the models 19, seven, five and four input variables were used. The results of the biogas flow rate prediction are represented in Table 7.

As can be seen, all LWR models had a good robustness with an R^2 0.93–0.94, except the non-optimized large-dimensioned models with 19 input process variables. Here the R^2 was 0.89. In comparison to the PLSR models, the accuracy of the LWR models was less satisfactory. The non-optimized large-dimensioned models had cv-RMSE of 18.44%. The cv-RMSE of the optimized models was comparable with the PLSR models. The ACO-optimized models had a cv-RMSE of 15.01%, while GA-optimized models 13.06%. The most accurate prediction was gained using the ACO-GA optimized models with four input process variables. Here the cv-RMSE was 12.11%.

3.5. ANN prediction results

ANN represents another non-linear technique used in this research to predict the biogas flow rate. ANN included the training, validation and test models. The training models were used to adjust ANN, the validation models to eliminate the over-fitting and the test models represented the actual prediction of the biogas flow rate. For that reason, only the evaluation of the test models will be discussed. The input data contained 19, seven, five and four process variables. The ANN training, validation and test results are shown in Table 8.

As can be seen in Table 8, the implemented ANN models had convincing results. All ANN models were robust with an R^2 between 0.96 and 0.97. That proved the ANN models to be able to learn the relation between the input and output information. The test-RMSE was similar for the non-optimized and with ACO and GA optimized models. Here the non-optimized models had a test-RMSE 0f 11.07%, ACO-optimized 12.07% and GA-optimized 11.31%. The most accurate prediction was done using the ACO-GA selected process variables. Here the test-RMSE was 9.04%.

Table 7
LWR prediction results.

Optimization	None	ACO	GA	ACO + GA
Pre-processing	SNV	SNV	SNV	SNV
Variable number	19	7	5	4
Algorithm	PLS	PLS	PLS	PLS
Local points number	500	750	500	500
Principal components number	3	4	3	3
cv-RMSE [%]	18.44	15.01	13.06	12.11
R^2 [-]	0.89	0.93	0.93	0.94

Table 8
ANN prediction results.

Optimization	None	ACO	GA	ACO + GA
Pre-processing	SNV	SNV	SNV	SNV
Number of input neurons	19	7	5	4
Number of hidden neurons	5	5	5	3
Number of iterations	19	15	15	20
RMSE training [%]	9.46	10.46	9.51	9.02
RMSE validation [%]	10.53	11.53	12.8	13.15
RMSE test [%]	11.07	12.07	11.31	9.04
R^2 training [-]	0.97	0.97	0.97	0.98
R^2 validation [-]	0.96	0.96	0.97	0.97
R^2 test [-]	0.96	0.96	0.97	0.97

The prediction results done with the help of the optimized linear and non-linear models are represented in Fig. 3.

The gained results showed, that in this research used model types proved to be capable to predict the biogas flow rate (Perai et al., 2010; Panagou et al., 2011). In case of the PLSR models, a linear approximation was done. The LWR models performed a pseudo-linear approximation, selecting the local data points suitable for the linear approximation, while the non-linear features were excluded from the data set. The non-linear ANN models learned the correlations between the data. The used metaheuristic tools executed the feature selection, defining the most important process variables. Due to it, the model dimension could be reduced from 19 input variables to seven selected by ACO and five picked out by GA. Both optimization tools chose four process variables, namely carbohydrates, proteins, fats and acetic acid. The ACO-GA optimized models showed the best prediction results with the highest R^2 and the lowest RMSE. The selected carbohydrates, fats and proteins represent the essential components of the biomass, which will be converted into their intermediates during the disintegration step. In further fermentation steps the built intermediates will be converted into acetic acid, which will be directly digested by acetoclastic microorganisms into biogas and accordingly into methane. Thus acetic acid is the most essential intermediate, defining the amount of the biogas output.

4. Conclusion

The aim of this study was to evaluate the linear and non-linear modelling types used for the prediction of the biogas flow rate. The desired variable was successfully modeled using both linear and non-linear models. The PLSR models performed the linear approximation of the data considering the correlation between the output variable biogas flow rate and the input variables. This strategy was already used by other authors, for example Kandel et al. used the PLSR models to predict the specific biogas yield and its kinetics using the infrared reflectance spectroscopy measurements (Kandel et al., 2013). Here the R^2 was 0.68, while the ratio of performance to deviation (RPD) was 1.83. Comparing to the results of this study, the used PLSR models achieved more accurate prediction of the desired variable using a lower number of the independent variables. Thus the best prediction of the biogas flow rate was obtained using four input variables (ACO-GA optimized models) with an R^2 of 0.94 and a cv-RMSE of 10.14%.

The LWR modelling was another modelling tool, used in this work to predict the biogas flow rate. Its prediction procedure was based on the principle of the local level. This modelling type was used as well for the prediction of the process variables in the AD processes. Amongst others, Dahlbacka et al. used LWR models to predict ammonium and acetate in the AD processes. Herewith acetate could be predicted with an RMSEp of 21%, while ammonium with an RMSEp of 29% (Dahlbacka et al., 2013). In contrast to it, in this research implemented LWR models showed a better prediction capacity. Similarly to the PLSR models, the LWR models were robust with an R^2 of 0.93–0.94,

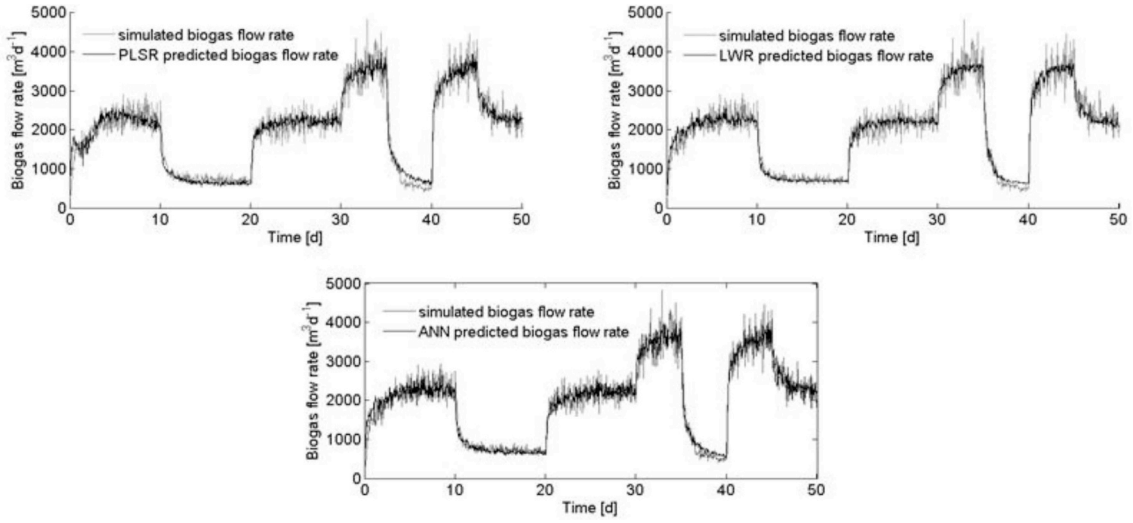

Fig. 3. Prediction of the biogas flow rate using the ACO-GA optimized PLS, LWR and ANN models.

excepting the non-optimized models (R^2 of 0.89). Regarding the accuracy of the prediction, a less good prediction was achieved using the non-optimized LWR models with 19 input variables. Here cv-RMSE was 18.44%. The optimized models showed good prediction capacity with a cv-RMSE of 13.06% for the GA-optimized models and 15.01% for the ACO-optimized models. The best prediction was done with the help of the ACO-GA optimized models; here the cv-RMSE was 12.11%.

The ANN with a back-propagation algorithm used in this research represented a non-linear modelling type as well. The implemented models could be successfully trained und utilized to provide an accurate prediction of the desired variable. In comparison to LWR and PLSR, the ANN models have an advantage to be capable to regard the non-linear features of the data. This approved modelling technique was widely used in the field of the AD processes. Hu Li et al. used ANN models for short- and long-term forecasting of CH_4, O_2 and CO_2 in landfill (Li et al., 2011). According to the results, the prediction of the CH_4 done with the help of ANN was successful with an R^2 of 0.76. Much better results were achieved due to the implemented GA algorithm; here the R^2 reached 0.9. In comparison to it, the prediction of the biogas flow rate done in this research was more robust and accurate. The R^2 was between 0.96 and 0.97 with an RMSE of 9.04% for the ACO-GA optimized models.

In this study three different modelling techniques were developed to predict the output process variable biogas flow rate. The results showed, that the implemented models could be successfully implemented providing an accurate prediction of the desired variable. In comparison to the PLSR models, ANN and LWR performed the prediction based on the non-linear approach. In LWR a data set was selected, which was appropriate for the linear prediction, while the non-linear features were excluded from the data set. This could result in loss of important information and accordingly less accurate prediction performance in comparison to the ANN models. The best and most accurate prediction could be achieved using the ANN models. Moreover, the use of the optimization tools proved to be helpful to identify the important process variables. The optimized models were more accurate in comparison to the non-optimized models. Nevertheless, the best prediction was achieved using the process variables, which were selected by both metaheuristic tools. Thus, a combination of the models and optimization techniques described a powerful approach for developing reliable models to plan and to predict the biogas process development.

The developed approach is unique in combining different model types and metaheuristics to predict the process development without any prior knowledge about the process. In the literature there are some examples of modelling in different scientific fields showing promising results. For example, Godin et al. compared linear and non-linear models used for prediction of biochemical methane potential of plant biomass based on near infrared spectrum data (Godin et al., 2015). Here, the multiple linear regression and the partial least squares regression were used. Jacobi et al. developed a monitoring tool based on near infrared spectroscopy data (Jacobi et al., 2012). Additionally, plant-feeding data, degradation kinetics were used to calculate the theoretically estimated biogas production of maize silage. The gained results proved this method to be successfully used for strong heterogeneous substrates. Moreover, the collection of the experimental data makes the tool to be complicated to use. Partial least squares regression was used, for example, in Kandel et al. to predict biogas yield and its kinetics (Kandel et al., 2013) and to reduce odours impact during anaerobic digestion (Orzi et al., 2018). In both articles the coefficient of determination was lower in comparison to the models used in this work. Neural networks were successfully used for prediction of the biogas production (Ghatak and Ghatak, 2018). The used models got a high coefficient of determination but a strong deviation factor of 10%. Other applications of mathematical modelling are shown in Chadel et al. Here the multiple linear regression and the neural networks were used to predict the operation parameters of a variable rate applicator of fluted roller metering mechanism (Chandel et al., 2019), while Taheri-Garavand used neural logic in combination with genetic algorithm for prediction of savory leaves drying process (Taheri-Garavand et al., 2018).

The represented results and the multidisciplinary application of the mathematical modelling showed the developed strategy to be a powerful tool to be used for the evaluation of the complex process. The used prediction models could successfully predict the process development. Metaheuristics improved model performance and simplified the model dimension. This approach can further used for the plant management and optimization of the anaerobic digestion processes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eaef.2019.06.001.

References

Abu Qdais, H., Bani Hani, K., Shatnawi, N., 2010. Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resour. Conserv. Recycl. 54 (6), 359–363. https://doi.org/10.1016/j.resconrec.2009.08.012.

Batstone, D.J., 2002. Anaerobic Digestion Model no. 1: [(ADM1)]. Scientific & technical report 13. IWA Publ, London.

Beltramo, T., Ranzan, C., Hinrichs, J., Hitzmann, B., 2016. Artificial neural network

prediction of the biogas flow rate optimised with an ant colony algorithm. Biosyst. Eng. 143, 68–78. https://doi.org/10.1016/j.biosystemseng.2016.01.006.
Beltramo, T., Theuerl, S., Klocke, M., Hitzmann, B., 2014. Estimation of the biogas production rate, a chemometrical approach. J. Cheminf. 6 (Suppl. 1), P26. https://doi.org/10.1186/1758-2946-6-S1-P26.
Biernacki, P., Steinigeweg, S., Borchert, A., Uhlenhut, F., 2013. Application of Anaerobic Digestion Model No. 1 for describing anaerobic digestion of grass, maize, green weed silage, and industrial glycerine. Bioresour. Technol. 127, 188–194. https://doi.org/10.1016/j.biortech.2012.09.128.
Björnsson, L., Murto, M., Mattiasson, B., 2000. Evaluation of parameters for monitoring an anaerobic co-digestion process. Appl. Microbiol. Biotechnol. 54 (6), 844–849. https://doi.org/10.1007/s002530000471.
Chandel, N.S., Tewari, V.K., Mehta, C.R., 2019. ANN model for predicting operating parameters of a variable rate applicator. Engineering in Agriculture, Environment and Food. . https://doi.org/10.1016/j.eaef.2019.04.001.
Christen, P., 2012. Data pre-processing. In: Christen, P. (Ed.), Data Matching. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 39–67.
Cleveland, W.S., Devlin, S.J., Grosse, E., 1988. Regression by local fitting. J. Econom. 37 (1), 87–114. https://doi.org/10.1016/0304-4076(88)90077-2.
Costa, J.C., Alves, M.M., Ferreira, E.C., 2009. Principal component analysis and quantitative image analysis to predict effects of toxics in anaerobic granular sludge. Bioresour. Technol. 100 (3), 1180–1185. https://doi.org/10.1016/j.biortech.2008.09.018.
Dahlbacka, J., Lillhonga, T., Döring, M., 2013. Quantitative measurements of anaerobic digestion process parameters using near infrared spectroscopy and local calibration models. J. Near Infrared Spectrosc. 21 (1), 23. https://doi.org/10.1255/jnirs.1033.
Demuth, H., Beale, M., 2000. Neural Network Toolbox User's Guide. The Mathworks, Natick.
Donoso-Bravo, A., Mailier, J., Martin, C., Rodríguez, J., Aceves-Lara, C.A., Vande Wouwer, A., 2011. Model selection, identification and validation in anaerobic digestion: a review. Water Res. 45 (17), 5347–5364. https://doi.org/10.1016/j.watres.2011.08.059.
Dorigo, M., Blum, C., 2005. Ant colony optimization theory: a survey. Theor. Comput. Sci. 344 (2–3), 243–278. https://doi.org/10.1016/j.tcs.2005.05.020.
Galí, A., Benabdallah, T., Astals, S., Mata-Alvarez, J., 2009. Modified version of ADM1 model for agro-waste application. Bioresour. Technol. 100 (11), 2783–2790. https://doi.org/10.1016/j.biortech.2008.12.052.
Ghatak, M.D., Ghatak, A., 2018. Artificial neural network model to predict behavior of biogas production curve from mixed lignocellulosic co-substrates. Fuel 232, 178–189. https://doi.org/10.1016/j.fuel.2018.05.051.
Godin, B., Mayer, F., Agneessens, R., Gerin, P., Dardenne, P., Delfosse, P., Delcarte, J., 2015. Biochemical methane potential prediction of plant biomasses: comparing chemical composition versus near infrared methods and linear versus non-linear models. Bioresour. Technol. 175, 382–390. https://doi.org/10.1016/j.biortech.2014.10.115.
Grote, B., Zense, T., Hitzmann, B., 2014. 2D-fluorescence and multivariate data analysis for monitoring of sourdough fermentation process. Food Control 38, 8–18. https://doi.org/10.1016/j.foodcont.2013.09.039.
Gueguim Kana, E.B., Oloke, J.K., Lateef, A., Adesiyan, M.O., 2012. Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm. Renew. Energy 46, 276–281. https://doi.org/10.1016/j.renene.2012.03.027.
Hitzmann, B., Ritzka, A., Ulber, R., Scheper, T., Schügerl, K., 1997. Computational neural networks for the evaluation of biosensor FIA measurements. Anal. Chim. Acta 348 (1–3), 135–141. https://doi.org/10.1016/S0003-2670(97)00153-0.
Holm-Nielsen, J.B., Esbensen, K.H., 2011. Monitoring of biogas test plants-a process analytical technology approach. J. Chemom. 25 (7), 357–365. https://doi.org/10.1002/cem.1344.
Holubar, P., 2002. Advanced controlling of anaerobic digestion by means of hierarchical neural networks. Water Res. 36 (10), 2582–2588. https://doi.org/10.1016/S0043-1354(01)00487-0.
Ihunegbo, F.N., Madsen, M., Esbensen, K.H., Holm-Nielsen, J.B., Halstensen, M., 2012. Acoustic chemometric prediction of total solids in bioslurry: a full-scale feasibility study for on-line biogas process monitoring. Chemometr. Intell. Lab. Syst. 110 (1), 135–143. https://doi.org/10.1016/j.chemolab.2011.10.009.
Jacobi, H.F., Ohl, S., Thiessen, E., Hartung, E., 2012. NIRS-aided monitoring and prediction of biogas yields from maize silage at a full-scale biogas plant applying lumped kinetics. Bioresour. Technol. 103 (1), 162–172. https://doi.org/10.1016/j.biortech.2011.10.012.
Kandel, T.P., Gislum, R., Jorgensen, U., Laerke, P.E., 2013. Prediction of biogas yield and its kinetics in reed canary grass using near infrared reflectance spectroscopy and chemometrics. Bioresour. Technol. 146, 282–287. https://doi.org/10.1016/j.biortech.2013.07.092.
Karpenko, O., Shi, J., Dai, Y., 2005. Prediction of MHC class II binders using the ant colony search strategy. Artif. Intell. Med. 35 (1–2), 147–156. https://doi.org/10.1016/j.artmed.2005.02.002.
Kessler, W., 2008. Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik: Ein Lehrbuch. 1. Aufl., 1. Nachdr. WILEY-VCH, Weinheim.
Lauwers, J., Appels, L., Thompson, I.P., Degrève, J., van Impe, J.F., Dewil, R., 2013. Mathematical modelling of anaerobic digestion of biomass and waste: power and limitations. Prog. Energy Combust. Sci. 39 (4), 383–402. https://doi.org/10.1016/j.pecs.2013.03.003.
Li, H., Sanchez, R., Joe Qin, S., Kavak, H.I., Webster, I.A., Tsotsis, T.T., Sahimi, M., 2011. Computer simulation of gas generation and transport in landfills. V: use of artificial neural network and the genetic algorithm for short- and long-term forecasting and planning. Chem. Eng. Sci. 66 (12), 2646–2659. https://doi.org/10.1016/j.ces.2011.03.013.
Lin, H., Wu, X., Miller, C., Zhu, J., Hadlocon, L.J., Manuzon, R., Zhao, L., 2014. Pilot-scale field study for ammonia removal from lagoon biogas using an acid wet scrubber. J. Environ. Sci. Health Part B Pestic. Food Contam. Agric. Wastes 49 (6), 439–448. https://doi.org/10.1080/03601234.2014.894781.
Lübken, M., Wichern, M., Schlattmann, M., Gronauer, A., Horn, H., 2007. Modelling the energy balance of an anaerobic digester fed with cattle manure and renewable energy crops. Water Res. 41 (18), 4085–4096. https://doi.org/10.1016/j.watres.2007.05.061.
Mu, Y., Yu, H.-Q., 2007. Simulation of biological hydrogen production in a UASB reactor using neural network and genetic algorithm. Int. J. Hydrogen Energy 32 (15), 3308–3314. https://doi.org/10.1016/j.ijhydene.2007.05.021.
Mullen, R.J., Monekosso, D., Barman, S., Remagnino, P., 2009. A review of ant algorithms. Expert Syst. Appl. 36 (6), 9608–9617. https://doi.org/10.1016/j.eswa.2009.01.020.
Nache, M., Scheier, R., Schmidt, H., Hitzmann, B., 2015. Non-invasive lactate- and pH-monitoring in porcine meat using Raman spectroscopy and chemometrics. Chemometr. Intell. Lab. Syst. 142, 197–205. https://doi.org/10.1016/j.chemolab.2015.02.002.
Orzi, V., Riva, C., Scaglia, B., D'Imporzano, G., Tambone, F., Adani, F., 2018. Anaerobic digestion coupled with digestate injection reduced odour emissions from soil during manure distribution. Sci. Total Environ. 621, 168–176. https://doi.org/10.1016/j.scitotenv.2017.11.249.
OZKAYA, B., DEMIR, A., BILGILI, M., 2007. Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environ. Model. Softw. 22 (6), 815–822. https://doi.org/10.1016/j.envsoft.2006.03.004.
Panagou, E.Z., Mohareb, F.R., Argyri, A.A., Bessant, C.M., Nychas, G.-J.E., 2011. A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. Food Microbiol. 28 (4), 782–790. https://doi.org/10.1016/j.fm.2010.05.014.
Perai, A.H., Nassiri Moghaddam, H., Asadpour, S., Bahrampour, J., Mansoori, G., 2010. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. Poultry Sci. 89 (7), 1562–1568. https://doi.org/10.3382/ps.2010-00639.
Ranzan, C., Strohm, A., Ranzan, L., Trierweiler, L.F., Hitzmann, B., Trierweiler, J.O., 2014. Wheat flour characterization using NIR and spectral filter based on Ant Colony Optimization. Chemometr. Intell. Lab. Syst. 132, 133–140. https://doi.org/10.1016/j.chemolab.2014.01.012.
Reed, J.P., Devlin, D., Esteves, S.R.R., Dinsdale, R., Guwy, A.J., 2013. Integration of NIRS and PCA techniques for the process monitoring of a sewage sludge anaerobic digester. Bioresour. Technol. 133, 398–404. https://doi.org/10.1016/j.biortech.2013.01.083.
Rosen, C., Vrecko, D., Gernaey, K.V., Pons, M.N., Jeppsson, U., 2006. Implementing ADM1 for plant-wide benchmark simulations in Matlab/Simulink. Water Sci. Technol. 54 (4), 11. https://doi.org/10.2166/wst.2006.521.
Rosen, C.Jeppsson, U. Aspects on ADM1 implementation within the BSM2 framework. http://www.iea.lth.se/WWTmodels_download/TR_ADM1.pdf.
Shmygelska, A., Hoos, H.H., 2005. An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. BMC Bioinf. 6, 30. https://doi.org/10.1186/1471-2105-6-30.
Strik, D.P.B.T.B., Domnanovich, A.M., Zani, L., Braun, R., Holubar, P., 2005. Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB Neural Network Toolbox. Environ. Model. Softw. 20 (6), 803–810. https://doi.org/10.1016/j.envsoft.2004.09.006.
Sun, D.-W., 2009. Infrared Spectroscopy for Food Quality Analysis and Control, vol. 1 Elsevier Academic Press, Amsterdam.
Taheri-Garavand, A., Meda, V., Naderloo, L., 2018. Artificial neural Network – Genetic algorithm modeling for moisture content prediction of savory leaves drying process in different drying conditions. Eng. Agric. Environ. Food 11 (4), 232–238. https://doi.org/10.1016/j.eaef.2018.08.001.
Wang, S., Yang, M., Li, J., Wu, X., Wang, H., Liu, B., Dong, Z., Zhang, Y., 2017. Texture analysis method based on fractional fourier entropy and fitness-scaling adaptive genetic algorithm for detecting left-sided and right-sided sensorineural hearing loss. FI 151 (1–4), 505–521. https://doi.org/10.3233/FI-2017-1507.
Wasserman, P.D., 1993. Advanced Methods in Neural Computing. VNR computer library. Van Nostrand Reinhold, New York.
Whitley, D., 1994. A genetic algorithm tutorial. Stat. Comput. 4 (2). https://doi.org/10.1007/BF00175354.
Wichern, M., Gehring, T., Fischer, K., Andrade, D., Lübken, M., Koch, K., Gronauer, A., Horn, H., 2009. Monofermentation of grass silage under mesophilic conditions: measurements and mathematical modeling with ADM 1. Bioresour. Technol. 100 (4), 1675–1681. https://doi.org/10.1016/j.biortech.2008.09.030.
Wold, S., Ruhe, A., Wold, H., Dunn III, W.J., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J. Sci. Stat. Comput. 5 (3), 735–743. https://doi.org/10.1137/0905052.
Zhang, Y., Wang, S., Ji, G., 2013. A rule-based model for bankruptcy prediction based on an improved genetic ant colony algorithm. Math. Probl. Eng. 2013 (3), 1–10. https://doi.org/10.1155/2013/753251.
Zhang, Y., Zhang, Z., Sugiura, N., Maekawa, T., 2002. Monitoring of methanogen density using near-infrared spectroscopy. Biomass Bioenergy 22 (6), 489–495. https://doi.org/10.1016/S0961-9534(02)00019-3.
Zhou, H., Löffler, D., Kranert, M., 2011. Model-based predictions of anaerobic digestion of agricultural substrates for biogas production. Bioresour. Technol. 102 (23), 10819–10828. https://doi.org/10.1016/j.biortech.2011.09.014.

4 Diskussion und Ausblick

Das Ziel dieser Dissertation ist es, ein schnelles und zuverlässiges Werkzeug zur Analyse von Biogasprozessen zu entwickeln, um die Ausbeute in Biogasanlagen anhand der Prozessgrößenidentifikation und einer effizienten Modellanpassung optimieren zu können. Die „VDI Norm 4631 zur Bestimmung der Gütekriterien für Biogasanlagen“ ist die Grundlage für die Planung, Umsetzung und Analyse der Biogasprozesse. Diese ist eine anerkannte Norm, die in der Regel für solche analytischen Zwecke verwendet wird. Die Umsetzung der VDI Norm 4631 setzt dafür allerdings ausgebildetes Personal, komplexe analytische Gerätschaften und langwierige Analysen voraus. Dies resultiert in hohen Kosten und zu späten Untersuchungsergebnissen. Dazu kommen gekürzte Zuschüsse für den Ausbau der regenerativen Energien, was die Umsetzung der Norm zumeist aufwendig und kostenintensiv macht. Diese Gründe führen zu einem Bedarf an der Entwicklung effizienter, praxisrelevanter und günstiger Alternativen für Anlagenplanung und Prozessoptimierung.

Aktuell stellen mathematische Modelle eine Alternative zur Prozessoptimierung in vielen wissenschaftlichen Bereichen dar, wie zum Beispiel in den Naturwissenschaften, Medizin und Wirtschaftswissenschaften. Im Vergleich dazu gibt es im Bereich der Biogasherstellung seit Ende der 60er Jahren eine Vielzahl verschiedener Modelle, welche die Biogasprozesse beschreiben. Die Modelle unterscheiden sich in der Anzahl der Prozessstufen und der Anzahl verwendeter Prozessgrößen. Die einfachen Modelle können in der Regel nur für bestimmte Prozessstufen oder einzelne Prozesse angewendet werden, während die komplexen Modelle näherungsweise den gesamten Prozess darstellen. Die meisten Modelle werden ausschließlich für einzelne Prozessabschnitte angewendet und können schwer auf andere Substrate oder Prozessbedingungen übertragen werden. Das bekannteste und anerkannteste Model im Bereich der Biogasmodellierung ist das ADM1 Modell. ADM1 beschreibt alle vier Stufen der Biogasherstellung und umfasst eine Vielzahl an Prozess- und Zustandsgrößen sowie Modellparameter, die für die Modellanpassung experimentell ermittelt werden müssen. Wegen seiner hohen Komplexität ist ADM1 für die praxisrelevanten Prozessführungen zumeist nicht einfach

einsetzbar. Zudem braucht ADM1 eine Menge an experimentellen Daten, was die Anwendung umständlich macht.

Eine andere Art der Prozessmodellierung stellen Regressionsmodelle und neuronale Netze dar. Diese Modelle sind datengetrieben und benötigen für die Vorhersage keine zusätzlich experimentell ermittelten Daten. In der Literatur gibt es bereits eine Vielzahl an Beispielen, welche eine aussagekräftige Implementierung von Vorhersagemodelle für Biogasprozesse beschreiben. So verwendete H. Abu Quidas neuronale Netze, um den Einfluss unterschiedlicher Prozessgrößen, wie zum Beispiel Temperatur, Trockensubstanz und pH-Wert, auf die Biogasausbeute zu untersuchen (Abu Qdais et al. 2010). Die verwendeten Modelle haben eine gute Vorhersagekraft mit einem Korrelationskoeffizient von 0,87. Kana et al. verwendete eine Kombination aus neuronalen Netzen und genetischem Algorithmus, um den Prozessverlauf der anaeroben Fermentation für verschiedene Substrate vorherzusagen (Gueguim Kana et al. 2012). Die mit GA optimierten ANN Modelle zeigten eine bessere Vorhersagekraft im Vergleich zu den anderen Verfahren. In dem Artikel von Pappu et al. wurde die Vorhersagekraft der GA optimierten neuralen Netze und der Regressionsmethoden verglichen (Pappu und Gummadi 2017). Diese und andere Beispiele zeigen, dass eine mathematische Modellierung eine gute Alternative für die Beschreibung der Biogasprozesse darstellt.

In dieser Dissertation wurden Kombinationen verschiedener Modellarten verwendet, mit zwei unterschiedlichen Algorithmen optimiert und getestet, um geeignete Modellarten zur Simulation und Optimierung von Biogasprozessen zu ermitteln. Mithilfe der metaheuristischen Verfahren sollen verwendete Prozessgrößen auf ihre Relevanz für die Schätzgüte hin untersucht werden. Dabei sollte die Komplexität der Modelle möglichst reduziert werden, um später die Anlagenführung der realen Biogasanlagen zu vereinfachen.

Für die Modellberechnung wurden zwei Datenarten verwendet: reale Messdaten, gemessen bei einer Biogasanlage, und simulierte Prozessdaten, berechnet mit Hilfe des ADM1 Modells. Da ADM1 den Prozessverlauf nur für einzelne Substrate beschreibt, was in Wirklichkeit nicht üblich ist, wurde in dieser Dissertation ADM1 an reale Bedingungen angepasst. In dem Modell wurde zusätzlich ein

Substratwechsel integriert. Diese Ergänzung in ADM1 hilft das Modell flexibel für verschiedene Substrate und Substratzusammensetzungen anzuwenden, um den Prozessverlauf darzustellen.

Für die Prozesssimulation wurden lineare und nicht lineare mathematische Verfahren untersucht. Das Ziel war die Modelle vergleichend zu bewerten, um die besten Modelle zu identifizieren. Als Zielgröße für die Schätzung wurde die Biogasausbeute ausgewählt. Die Biogasausbeute kann eine direkte Aussage über den Prozessverlauf abbilden, indem Prozessstörungen durch eine suboptimale Biogasmenge identifiziert werden.

Um die für die Modellierung prozessrelevanten Größen zu ermitteln, wurden die metaheuristischen Methoden, ACO und GA implementiert. Mit Hilfe beider Verfahren konnten einzelne Prozessgrößen, wie zum Beispiel Konzentration der Zwischenprodukte, in Bezug auf ihre Bedeutung für die Schätzung der Biogasausbeute untersucht werden. Das kann zur Ermittlung von Störgrößen im Prozess beitragen. Für die Vereinfachung der Modellgröße wurden die nicht relevanten Prozessgrößen aus dem Model entfernt, somit konnte die Anzahl unabhängiger Prozessgrößen verkleinert werden. Die erzielten Ergebnisse zeigen eine Verbesserung der Vorhersagekraft der Modelle trotz der reduzierten Anzahl an unabhängiger Variablen. Die Evaluierung beider metaheuristischen Methoden zeigt, dass ACO im Vergleich zu GA besser zur Selektion der Prozessgrößen geeignet ist.

Eine weitere Reduktion der Anzahl unabhängiger Prozessgrößen wurde erzielt, indem nur die Prozessgrößen für eine Modellierung verwendet werden, die von beiden Optimierungsalgorithmen ausgewählt wurden. Dies zeigt, dass die Kombination von ACO und GA für die Selektion viel versprechend ist. Als Schnittmenge werden folgende Prozessgrößen für die Modellierung der Biogasausbeute identifiziert: hydraulische Verweilzeit, Trockensubstanz, Gehalt an Neutral-Detergentien-Faser, Säure-Detergentien-Lignin-Gehalt und n-Buttersäure.

In der vorliegenden Dissertation wurde ein schnelles und effizientes Verfahren zur Analyse von Biogasprozessen entwickelt. Die entwickelte Strategie kann flexibel für verschieden Arten der Prozessführung und unterschiedliche Substrate angewendet werden, um den Verlauf der produzierten Biogasausbeute darzustellen. Im Vergleich zu ADM1 Modell sind die Regressionsmodelle und neuronale Netze datengetrieben und brauchen keine Vorinformation über den Prozess, zum Beispiel

anhand der analytischen Voruntersuchungen. Dies spart in der Praxis eine Menge Zeit und Kosten. Die realitätsnahe Vorhersage der Prozesse stellt eine große Hilfe für die Anlagenbetreiber dar, da die Ergebnisse direkt auf verschiedene Prozessbedingungen übertragen werden können. Die entwickelten Tools können als Bausteine für modellbasierte Steuerungsverfahren eingesetzt und bei realen Biogasanlagen für die Prozessüberwachung verwendet werden. Außerdem können die Vorhersagemodelle für die Prozessplanung, zum Beispiel zur Bestimmung der optimalen Substratzusammensetzung, eingesetzt werden. Dieses Verfahren kann auch für die Vorhersage der Prozesse mit einer variablen Substratzusammensetzung und verschiedenen Prozessführungen gezielt eingesetzt werden.

Für die Verbesserung der Modellierung ist eine weitere Validierung mit Daten aus der Praxis notwendig. In dieser Dissertation wurden nur die Datensätze für die mesophile Prozessführung verwendet. Es wäre sinnvoll zusätzlich die Anwendung des Verfahrens für thermophile Prozessbedingungen zu untersuchen. Außerdem wäre es von Bedeutung verschiedene Fermenter-Typen zu testen, um den Einfluss der technologischen Prozessbedingungen zu quantifizieren. Dies hilft die Anpassung der Modelle zu verbessern und Fehlerquellen aufzudecken, um den Biogasbildungsprozess möglichst realistisch abzubilden. Somit können die Biogasherstellungsprozesse maximal effizient, ökologisch und ökonomisch gestaltet werden, um die Anwendung von Biogasprozessen voranzutreiben.

5. Literaturverzeichnis

Abu Qdais, H.; Bani Hani, K.; Shatnawi, N. (2010): Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. In: *Resources, Conservation and Recycling* 54 (6), S. 359–363. DOI: 10.1016/j.resconrec.2009.08.012.

Batstone, D. J. (2002): Anaerobic digestion model no. 1. (ADM1). London: IWA Publ (Scientific & technical report, 13).

Cheng, Jay (Hg.) (2018): Biomass to renewable energy processes. Second edition. Boca Raton, London, New York: Taylor & Francis CRC Press (Energy & clean technology).

Derbal, K.; Bencheikh-Lehocine, M.; Cecchi, F.; Meniai, A-H; Pavan, P. (2009): Application of the IWA ADM1 model to simulate anaerobic co-digestion of organic waste with waste activated sludge in mesophilic condition. In: *Bioresource technology* 100 (4), S. 1539–1543. DOI: 10.1016/j.biortech.2008.07.064.

Deublein, Dieter; Steinhauser, Angelika (2011): Biogas from Waste and Renewable Resources. An Introduction. 4th ed. Hoboken: John Wiley & Sons. Online verfügbar unter http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=510179.

Dorigo, Marco; Blum, Christian (2005): Ant colony optimization theory: A survey. In: *Theoretical Computer Science* 344 (2-3), S. 243–278. DOI: 10.1016/j.tcs.2005.05.020.

EEG, 2000: Erneuerbaren-Energien-Gesetz. Online verfügbar unter https://www.erneuerbare-energien.de/EE/Redaktion/DE/Dossier/eeg.html.

Francisci, Davide de; Kougias, Panagiotis G.; Treu, Laura; Campanaro, Stefano; Angelidaki, Irini (2015): Microbial diversity and dynamicity of biogas reactors due to radical changes of feedstock composition. In: *Bioresource technology* 176, S. 56–64. DOI: 10.1016/j.biortech.2014.10.126.

Goldberg, David E.; Holland, John H. (1988). In: *Machine Learning* 3 (2/3), S. 95–99. DOI: 10.1023/A:1022602019183.

Gueguim Kana, E. B.; Oloke, J. K.; Lateef, A.; Adesiyan, M. O. (2012): Modeling and optimization of biogas production on saw dust and other co-substrates using Artificial Neural network and Genetic Algorithm. In: *Renewable Energy* 46, S. 276–281. DOI: 10.1016/j.renene.2012.03.027.

Henze, Mogens (Hg.) (2007): Activated sludge models ASM1, ASM2, ASM2d and ASM3. Reprinted. London: IWA Publ (Scientific and technical report / IWA, 9).

Holm-Nielsen, J. B.; Al Seadi, T.; Oleskowicz-Popiel, P. (2009): The future of anaerobic digestion and biogas utilization. In: *Bioresource technology* 100 (22), S. 5478–5484. DOI: 10.1016/j.biortech.2008.12.046.

Karakashev, Dimitar; Batstone, Damien J.; Angelidaki, Irini (2005): Influence of environmental conditions on methanogenic compositions in anaerobic biogas reactors. In: *Applied and environmental microbiology* 71 (1), S. 331–338. DOI: 10.1128/AEM.71.1.331-338.2005.

Lauwers, Joost; Appels, Lise; Thompson, Ian P.; Degrève, Jan; van Impe, Jan F.; Dewil, Raf (2013): Mathematical modelling of anaerobic digestion of biomass and waste: Power and limitations. In: *Progress in Energy and Combustion Science* 39 (4), S. 383–402. DOI: 10.1016/j.pecs.2013.03.003.

Nallathambi Gunaseelan, V. (1997): Anaerobic digestion of biomass for methane production: A review. In: *Biomass and Bioenergy* 13 (1-2), S. 83–114. DOI: 10.1016/S0961-9534(97)00020-2.

Nielsen, Henrik Bangsø; Angelidaki, Irini (2008): Strategies for optimizing recovery of the biogas process following ammonia inhibition. In: *Bioresource technology* 99 (17), S. 7995–8001. DOI: 10.1016/j.biortech.2008.03.049.

Pappu, Sharon Mano J.; Gummadi, Sathyanarayana N. (2017): Artificial neural network and regression coupled genetic algorithm to optimize parameters for enhanced xylitol production by Debaryomyces nepalensis in bioreactor. In: *Biochemical Engineering Journal* 120, S. 136–145. DOI: 10.1016/j.bej.2017.01.010.

Ramirez, Ivan; Volcke, Eveline I. P.; Rajinikanth, Rajagopal; Steyer, Jean-Philippe (2009): Modeling microbial diversity in anaerobic digestion through an extended ADM1 model. In: *Water research* 43 (11), S. 2787–2800. DOI: 10.1016/j.watres.2009.03.034.

Rosen, C.; Vrecko, D.; Gernaey, K. V.; Pons, M. N.; Jeppsson, U. (2006): Implementing ADM1 for plant-wide benchmark simulations in Matlab/Simulink. In: *Water Science and Technology* 54 (4), S. 11–19. DOI: 10.2166/wst.2006.521.

Sekiguchi, Yuji; Kamagata, Yoichi; Harada, Hideki (2001): Recent advances in methane fermentation technology. In: *Current Opinion in Biotechnology* 12 (3), S. 277–282. DOI: 10.1016/S0958-1669(00)00210-X.

VDI-Fachbereich Energietechnik (2016): Fermentation of organic materials - Characterization of the substrate, sampling, collection of material data, fermentation tests.

Wang, Dianlong; Ai, Ping; Yu, Liang; Tan, Zhongxin; Zhang, Yanlin (2015): Comparing the hydrolysis and biogas production performance of alkali and acid pretreatments of rice straw using two-stage anaerobic fermentation. In: *Biosystems Engineering* 132, S. 47–55. DOI: 10.1016/j.biosystemseng.2015.02.007.

Weiland, Peter (2010): Biogas production: current state and perspectives. In: *Applied microbiology and biotechnology* 85 (4), S. 849–860. DOI: 10.1007/s00253-009-2246-7.

Wichern, Marc; Gehring, Tito; Fischer, Katrin; Andrade, Diana; Lübken, Manfred; Koch, Konrad et al. (2009): Monofermentation of grass silage under mesophilic conditions: measurements and mathematical modeling with ADM 1. In: *Bioresource technology* 100 (4), S. 1675–1681. DOI: 10.1016/j.biortech.2008.09.030.

Zhou, Haidong; Löffler, Daniel; Kranert, Martin (2011): Model-based predictions of anaerobic digestion of agricultural substrates for biogas production. In: *Bioresource technology* 102 (23), S. 10819–10828. DOI: 10.1016/j.biortech.2011.09.014.

Eidesstattliche Versicherung gemäß § 7 Absatz 7 der Promotionsordnung der Universität Hohenheim zum Dr. rer. nat.

(Anlage 2 zur Promotionsordnung der Universität Hohenheim zum Dr. rer. nat.)

1. Bei der eingereichten Dissertation zum Thema

„Entwicklung von datengetriebenen Auswerteverfahren zur Analyse und Schätzung der Reaktorleistung von Biogasanlagen“

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Ich habe nicht die Hilfe einer kommerziellen Promotionsvermittlung oder -beratung in Anspruch genommen.

4. Die Bedeutung der eidesstattlichen Versicherung und der strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Die Richtigkeit der vorstehenden Erklärung bestätige ich: Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____

Ort, Datum Unterschrift Doktorandin

Eidesstattliche Versicherung, Belehrung

(Anlage 3 zur Promotionsordnung der Universität Hohenheim zum Dr. rer. nat.)

Die Universität Hohenheim verlangt eine Eidesstattliche Versicherung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistungen, um sich glaubhaft zu versichern, dass die Promovendin bzw. der Promovend die wissenschaftlichen Leistungen eigenständig erbracht hat.

Weil der Gesetzgeber der Eidesstattlichen Versicherung eine besondere Bedeutung beimisst und sie erhebliche Folgen haben kann, hat der Gesetzgeber die Abgabe einer falschen eidesstattlichen Versicherung unter Strafe gestellt. Bei vorsätzlicher (also wissentlicher) Abgabe einer falschen Erklärung droht eine Freiheitsstrafe bis zu drei Jahren oder eine Geldstrafe.

Eine fahrlässige Abgabe (also Abgabe, obwohl Sie hätten erkennen müssen, dass die Erklärung nicht den Tatsachen entspricht) kann eine Freiheitsstrafe bis zu einem Jahr oder eine Geldstrafe nach sich ziehen.

Die entsprechenden Strafvorschriften sind in § 156 StGB (falsche Versicherung an Eidesstatt) und in § 161 StGB (Fahrlässiger Falscheid, fahrlässige falsche Versicherung an Eidesstatt) wiedergegeben.

§ 156 StGB: Falsche Versicherung an Eides statt

Wer vor einer zur Abnahme einer Versicherung an Eides statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

§ 161 StGB: Fahrlässiger Falscheid, fahrlässige falsche Versicherung an Eides statt

Abs. 1: Wenn eine der in den §§ 154 und 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

Abs. 2: Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Absätze 2 und 3 gelten entsprechend.

Ich habe die Belehrung zur Eidesstattlichen Versicherung zur Kenntnis genommen.

_____

Ort, Datum Unterschrift Doktorandin

Lebenslauf

Tanja Beltramo

Kontakt

Bittelbronner Steige 3/1
72160 Horb am Neckar
Email: tanja.beltramo@gmail.com

T. 07482 9136544
M. 0176 75884483

Persönliche Daten

geboren am	28. Dezember 1980 in Nishyn (Ukraine)
Nationalität	Deutsch
Familienstand	ledig

Beruflicher Werdegang

Ab 6/2017	*Manager Product Safety* Amcor Flexibles Singen GmbH
9/2016 – 12/2019	*Doktorandin* Universität Hohenheim
8/2012 – 8/2016	*Wissenschaftliche Mitarbeiterin, Doktorandin* Universität Hohenheim, Institut für Lebensmittelwissenschaft und Biotechnologie, Stuttgart
2/2011 – 6/2011	*Werkstudentin* Firma Robert Bosch GmbH, Abteilung Bosch Lab Systems
8/2010 -2/2011	*Praktikantin* Firma Bosch GmbH, Abteilung Bosch Lab Systems

Studien

10/2007 – 07/2012	*Lebensmitteltechnologie* (Dipl.-Lebensmittelingenieurin) Universität Hohenheim, Stuttgart Note: 2,1
10/2005 – 06/2007	*Lebensmittelchemie* (Vordiplom),

	Universität Stuttgart, Stuttgart-Vaihingen
09/1998 – 05/2003	*Lehramt an der Mittelschule für Deutsch und Englisch,*
	die Staatliche Pädagogische Universität, Ukraine
	Note: 1,0

Schulbildung

09/1996 – 06/1998	*Lyceum* an der Staatlichen Pädagogischen Universität, Ukraine
	Allgemeine Hochschulreife: 1,0
09/1987 – 06/1996	*Mittelschule*, Ukraine

Qualifikationen

Fremdsprachen	Ukrainisch (Muttersprache) Russisch (Muttersprache)
	Deutsch (fließend in Rede und Schrift) Englisch (fließend in Rede und Schrift) Chinesisch (B1)
European Business	Bilanz, Kostenrechnung, Wirtschaftsrecht
Competence Licence	Businessplan, Projektplanung, Marketing, Verkauf
(EBCL* Stufe A, B)	Finanzplanung, Finanzierung
EDV	MS Office
Programmiersprachen	Matlab, Simulink
	SPSS
	SAP
Spezielle technische	Gambit, TechPlot, Fluent
Programme	

Präsentationen, Vorträge, wissenschaftliche Publikationen

10. – 11. September 2014	International Conference Progress in Biogas III, Stuttgart
10. – 12. November 2013	9th German Conference on Chemoinformatics, Fulda

29. Oktober 2013 Infotag ,,Von dem Datenfluss zum Erkenntnisgewinn – der richtige Umgang mit (Bio-)-Prozessdaten‘‘, Frankfurt am Main

Prozessmikrobiologie in landwirtschaftlichen Biogasanlagen. Schlussbericht zum Forschungsprojekt Biogas-Biocoenosis. Teilvorhaben 4 – Biostatistische Prozessdatenauswertung. Veröffentlicht in *Bornimer Agrartechnische Berichte*, Heft 84 (2014), S. 110-190

Estimation of the biogas production rate, a chemometrical approach. Veröffentlicht in *Journal of Cheminformatics*, Band 6, S. 26

Supervision of food manufacturing process using optical process analyzers – an overview. Veröffentlicht in *ChemBioEng*, Band 3, Ausgabe 5, S. 219 – 228

Mathematisch aufgeklärt. Veröffentlicht in Labor Praxis, September 2016, S. 24 - 28

Artificial neural network prediction of the biogas flow rate optimized with an ant colony algorithm. Veröffentlicht in *Biosystems Engineering*, Band 143, S. 68-78

Prediction of Biogas production using GA and ACO input features selection methods for ANN model. Veröffentlicht in *Information Processing in Agriculture*, Band 6, Ausgabe 3, S. 349-356

Evaluation of linear and non-linear prediction models optimized with metaheuristics: application to anaerobic digestion process. Veröffentlicht in *Engineering in Agriculture, Environment and Food*, Band 12, S. 379-403.

Nebentätigkeiten wissenschaftliche Hilfskraft, Nachhilfe und Tutorat

12/2015 - 8/2016	*Betreuerin studentischer wissenschaftlicher Projekte am Institut für Lebensmittelwissenschaften und Biotechnologie*
04/2010 – 7/2010	*Tutorin Übungen in Biotechnologie für Studierenden*, Universität Hohenheim, Stuttgart
9/2009 – 12/2009	*Wissenschaftliche Hilfsmitarbeiterin*, Fraunhofer Gesellschaft, Institut für Arbeitswissenschaft und Technologiemanagement (IAT), Stuttgart

1/2009 – 6/2009	*Wissenschaftliche Hilfsmitarbeiterin*, Universität Hohenheim, Institut für Lebensmittelwissenschaft und Biotechnologie, Stuttgart
2006 – 2008	*Wissenschaftliche Hilfsmitarbeiterin*, Universität Stuttgart Institut für Automatisierung und Softwareentwicklung (IAS)
2006 - 2016	Nachhilfe in Mathematik, Physik, Chemie und Sprachen

_____

Ort, Datum Unterschrift Doktorandin

Anhang

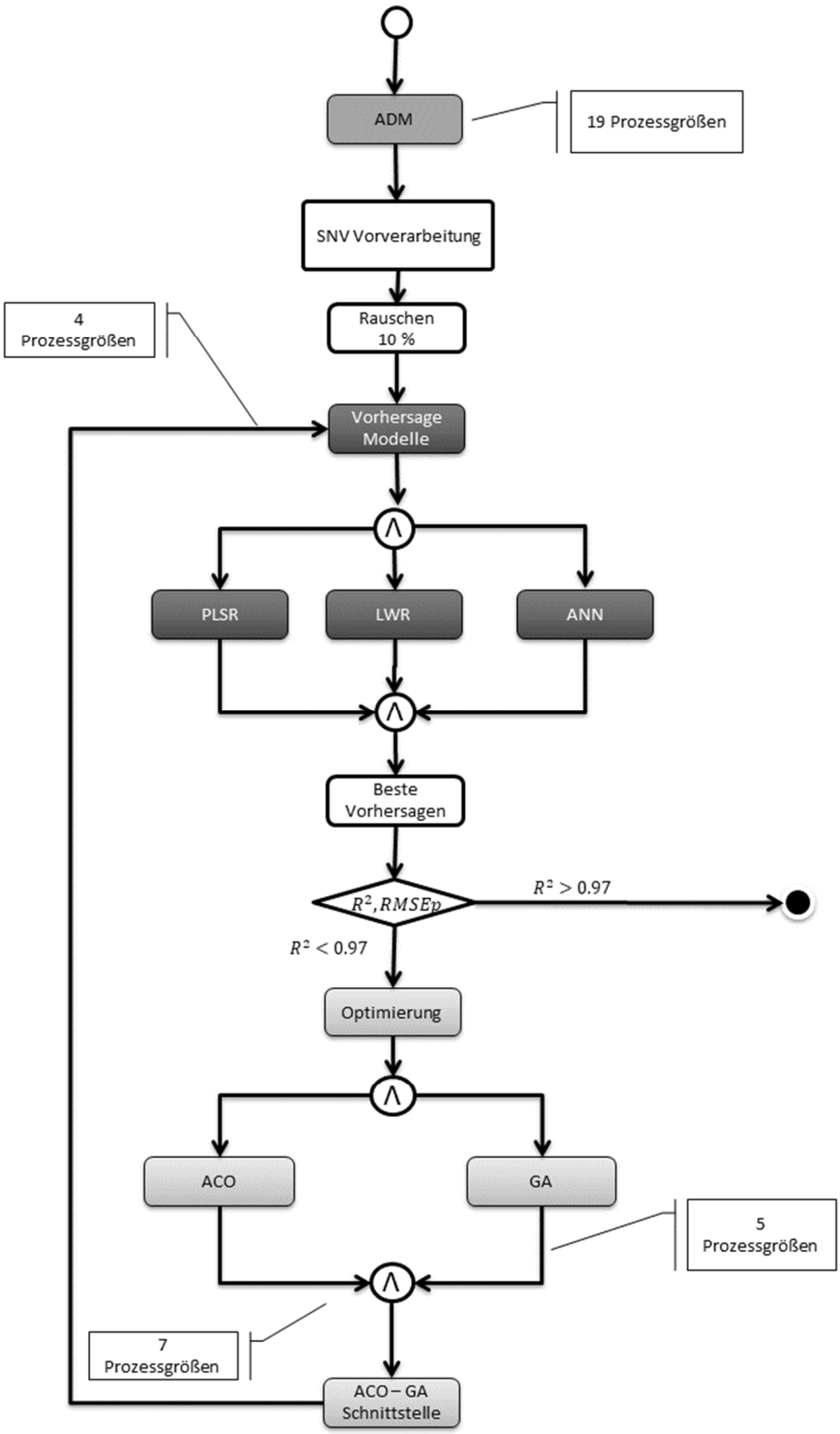

Abbildung 1: Vorhersage der Biogasausbeute anhand der simulierten Daten (ACO: Ant Colony OPtimization; ADM1: Anaerobic Digestion Model No.1; ANN: Artificial Neural Networks; GA: Genetic Algorithm; LWR: Locally Weighted Regression; PLSR: Partially Laest Square Regression; RMSE: Root Mean Square Error; SNV: Standard Normal Variate).

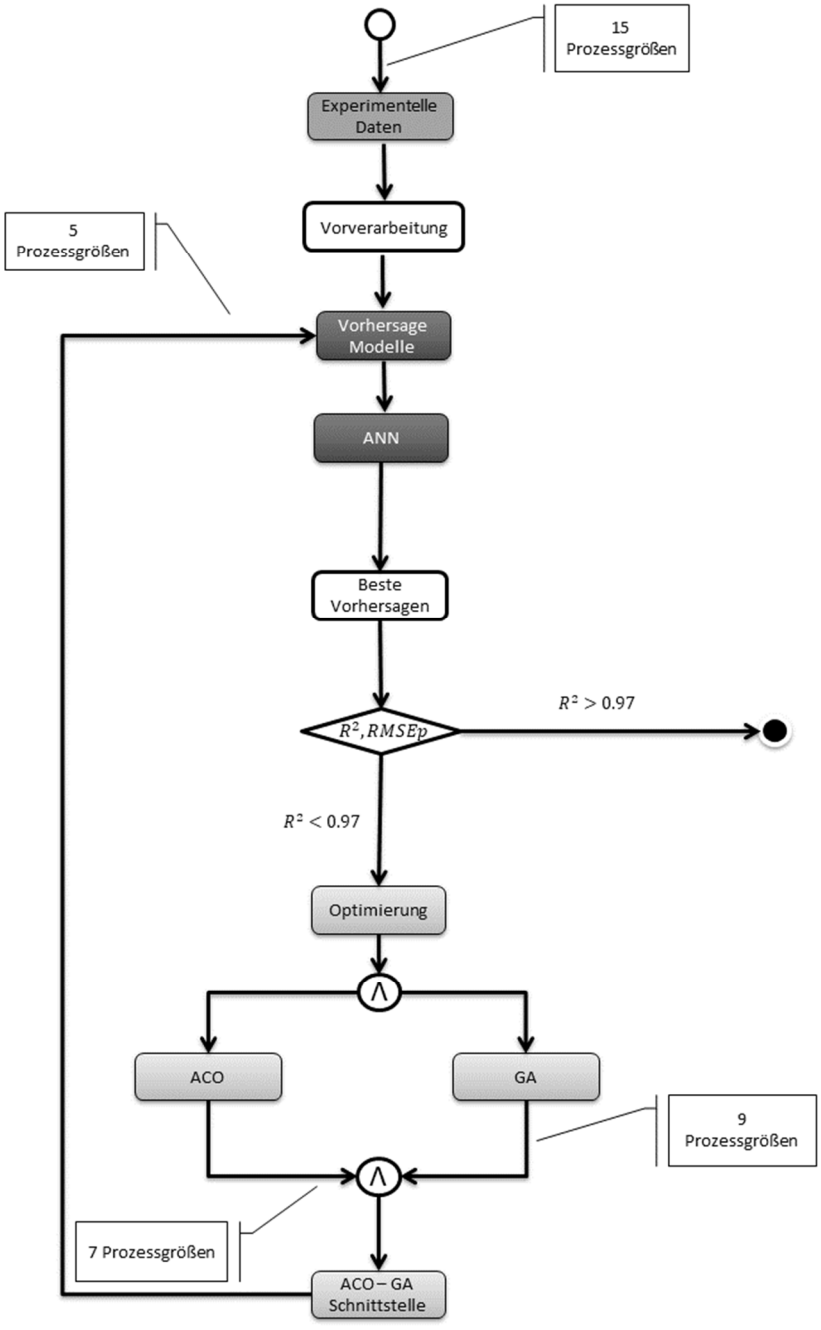

Abbildung 2: Vorhersage der Biogasausbeute anhand der experimentellen Daten (ACO: Ant Colony Optimization; ANN: Artificial Neural Networks; GA: Genetic Algorithm; RMSE: Root Mean Square Error).