Robin Eccleston[1]
Christian Wolf[1]
Matthias Balsam[2]
Franziska Schulte[3]
Michael Bongards[1]
Astrid Rehorek[2]

[1]Cologne University of Applied Sciences, Research group GECO-C, Gummersbach, Germany.

[2]Cologne University of Applied Sciences, PRA & PAT Centre, CHEMPARK Leverkusen, Leverkusen, Germany.

[3]art photonics GmbH, Berlin, Germany.

Mid-Infrared Spectroscopy for Monitoring of Anaerobic Digestion Processes – Prospects and Challenges

To develop an online probe that is not only sufficiently robust, but also able to measure crucial process variables in biogas plants is a tough challenge. Therefore, a mid-infrared (MIR) spectroscopic attenuated total reflection (ATR) probe and robust probe fitting were established. A fully automated probe control, calibration after probe cleaning, and analysis of the absorption spectra using machine learning were implemented in order to reduce maintenance of the probe to a minimum. The relevant wavelengths in the MIR spectrum for organic acids, total alkalinity, and ammonium nitrogen concentration were identified. Finally, intensive lab testing was carried out, followed by operation of the complete online measurement system at an industrial biogas plant. In order to improve signal strength and sensitivity, microelectronic mechanical system (MEMS)-based Fabry-Pérot interferometers were also investigated.

Keywords: Anaerobic digestion, Biogas, Machine learning, Microelectronic mechanical system interferometer, Mid-infrared spectroscopy

Received: June 10, 2015; *revised:* August 02, 2015; *accepted:* November 13, 2015

DOI: 10.1002/ceat.201500334

1 Introduction

The biogas industry has grown significantly for the past 20 years all over Europe, supported by very attractive feed-in tariffs [1–4]. In particular in Germany, until now nearly 8000 biogas plants [5] have been built and many of them are not operating at full capacity [6]. One of the main reasons for operational problems in biogas plants are caused by high volatile fatty acid (VFA) and low total alkalinity (TA) concentrations in combination with a lack of suitable online measurement systems for these process variables. As reported in [6], many biogas plants rely on a minimum of online process monitoring, measuring mainly temperature and energy production. Not even pH or the amounts of produced biogas are measured at small and medium-sized digesters. Relatively high prices for online analyzers for VFA (around 20 000 €) and the high and complex maintenance of these systems make them unfeasible for biogas plants. Furthermore, expert knowledge is required to successfully use and maintain such equipment, which an operator of a biogas plant does not have. Thus, accurate measurement of key variables of the anaerobic digestion (AD) process is presently mostly done by laboratory analysis of samples taken from the digester. This is a time-consuming and costly process, and requires a skilled operator to perform. Hence, is generally only done when there are concerns with digester stability.

If the conditions in a biogas digester become toxic to the bacteria, this can be catastrophic and may result in high losses as the digester may require draining and could suffer from reduced biogas production for several months, until the digester is operating at the previous level. Consequently, it is typical that biogas plants are operated conservatively with a safety margin. Therefore, the need for reliable and low-maintenance online instrumentation for biogas plants and anaerobic digestion processes in general is high and it can be considered one of the main challenges of the biogas sector. Not only for process monitoring but also for improved process control which will allow better utilization of the capacity of the digester and will in turn result in a higher level of biogas produced, leading to increased revenues and higher profits.

Tab. 1 gives an overview of the range of process variables that have an effect on the performance of the anaerobic digestion process and are of special interest for online monitoring. Furthermore, the corresponding concentration ranges are given.

Mid-infrared (MIR) spectroscopy is a very promising technique to characterize organic matter in AD processes. One major advantage over existing near-infrared (NIR) sensors is that process variables such as VFA, TA or total anorganic carbon (TAC), NH_4-N, and total solids show distinctive peaks in the MIR spectrum between 1800 and 800 cm^{-1}, which makes it easier to correlate peak intensity to actual concentrations. In previous work [7], Fourier transform infrared spectroscopy (FTIR) and fluorescence spectroscopy were used to character-

Correspondence: Dr. Christian Wolf (christian.wolf@fh-koeln.de), Cologne University of Applied Sciences, Research group GECO-C, Steinmüllerallee 1, 51643 Gummersbach, Germany.

Table 1. Overview of important AD process variables and their concentration ranges according to [10].

Parameter	Range				Fmc [mmol g^{-1}]
	from		to		
	[g L^{-1}]	[mmol L^{-1}]	[g L^{-1}]	[mmol L^{-1}]	
NH_4-N	1.80	77.78	3.66	158.3	55.56
TS [g kg^{-1}]	74	–	–	–	–
oTS [g kg^{-1}]	54	–	–	–	–
TAC ($CaCO_3$)	8	79.9	15	149.9	9.99
VFA (acetic acid)	2.05	34.1	6.5	108.2	16.65
Acids					
Acetic	0	0	2.99	49.79	16.65
Propionic	0	0	0.60	8.10	13.50
Butyric	0	0	0.05	0.57	11.35
Isobutyric	0	0	0	0	11.35
Valeric	0	0	0.11	1.08	9.79
Isovaleric	0	0	0	0	9.79
Carboxylic	0	0	0.02	0.17	8.61

ize the organic matter evolution during the AD and composting of pig slurry. Steyer et al. [8] also applied an MIR spectrometer for several years for the online measurement of chemical oxygen demand (COD), total organic carbon (TOC), VFA, total and partial alkalinity of an AD fixed bed treating industrial wine distillery wastewater. Spanjers et al. [9] employed the same technique at a full-scale plant for the online monitoring of VFA, COD, alkalinity, sulfate, total nitrogen, ammonia, and nitrate. However, main problems with MIR spectroscopy were always the fiber length which resulted in high signal losses, maintenance of the attenuated total reflectance probe and costs for the spectrometer. Furthermore, there are papers covering online monitoring in AD using NIR spectroscopy. VFAs were measured by partial least squares (PLS) regression with 0.9 g kg^{-1} root-mean-square error of prediction (RMSEP) [10]. Similarly, Total VFAs were measured with an RMSEP of 1.53 g L^{-1} [11] using a combination of PCA and PLS regression. PCA and PLS regression were also used for analysis of data from a simulation of an at-line process with a new embedded NIR sensor [12], with an R^2 value of 0.94 for total VFAs, with individual VFAs having lower accuracy.

In order to overcome the problems related to MIR spectroscopy and to investigate its potential for the use in biogas plants, the Cologne University of Applied Sciences and art photonics developed a diamond-tipped attenuated total reflection (ATR) probe which uses a specifically designed polycrystalline fiber. In conjunction with a fully automated process interface provided by the Knick GmbH, a complete online measurement system was developed and tested in a laboratory environment as well as at the industrial biogas plant "Leppe" operated by the AVEA GmbH & Co. in Lindlar. This plant is an industrial biogas plant for digestion of organic waste from urban households with a capacity of 1.1 Mkg m^2s^{-3} (MW) and operates according to the Valorga method [13] with a high dry matter content of 20–25 %.

This paper is organized as follows. Sect. 2 describes both experimental setups, in the lab and at the full-scale biogas plant and the used equipment in detail. Sect. 3 provides background information on the machine learning methods used for spectral data analysis. The results obtained by laboratory measurements and full-scale testing will be described subsequently in Sect. 4. Newly emerging micro-electronic mechanical system (MEMS)-based spectrometric sensor systems are introduced as a valid alternative for future sensor systems in Sect. 5 which is followed by a conclusion.

2 Experimental Setups

Two setups were used for testing, namely, a laboratory setup and a full-scale setup. Due to the various requirements of the two applications in terms of available space and robustness against vibrations, two different spectrometers were applied for the measurements, whose measurement parameters were adjusted according to the situation to ensure that an adequate spectrum in terms of resolution and signal-to-noise ratio (SNR) could be recorded. Following further development work, MEMS-based spectrometer systems could be a possible alternative, since depending on the sensor construction the sensor can be more compact and far more robust to vibration which would be more suitable for industrial environments.

2.1 Laboratory Setup

For the laboratory testing, an FTIR1750 Perkin-Elmer spectrometer was used for the samples. The spectrometer was connected to the diamond ATR probe via a 1-m fiber, giving a total path length of 2 m. The short path length is a major advantage because it reduces signal loss significantly improving the SNR. The FTIR1750 has two detectors, i.e., a triglycine sulfate (TGS) and a mercury cadmium telluride (MCT) detector, and can be switched internally. Due to the higher sensitivity the MCT detector was taken for all measurements. Fig. 1 illustrates the laboratory setup.

The parameters of the spectrometer for the conducted measurements were adapted manually to give the best performance in terms of signal quality, determined by SNR and resolution, and measurement duration. The best performance was achieved by averaging 64 scans for one measurement with a resolution of 4 cm^{-1} and an amplification of 8.

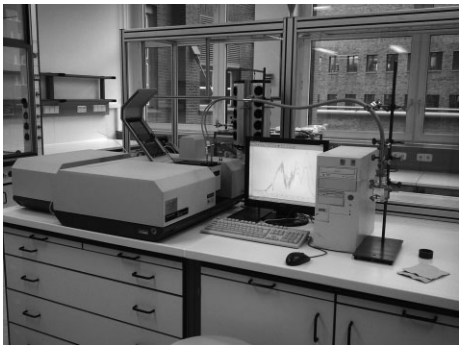

Figure 1. Laboratory setup with the Perkin Elmer FTIR 1750 and the diamond-tipped ATR probe connected by a PIR fiber.

In order to test the sensitivity of the spectrometer system, dilutions in deionized water were performed for acetic acid, sodium bicarbonate, and ammonium chloride. The diluted solutions were then measured. The data from these diluted samples were employed for the machine learning process and calibration. These samples are referred in this text to as dataset U_1.

A further sample of the substrate from the biogas plant was taken and spiked with acetic acid. Spectral data was then compared from the raw substrate, the spiked substrate, and the original pure acid. This data was employed to verify that peaks due the absorption of the acetic acid could be measured when using the spectrometer system. Additionally, comparisons were done between the absorption spectra for ammonium chloride, sodium bicarbonate, and raw substrate. These spiked results constituted dataset U_2.

Laboratory measurements were performed on samples from the biogas plant to analyze the composition, including acetic acid, bicarbonate content as well as NH_4-N concentration. For determination of the concentrations, the Anderson and Yang titration method [14] was carried out for VFA and TA. The level of dissolved ammonium was measured using the LCK303 ammonium cuvette test by Hach-Lange.

2.2 Full-Scale Application

For the full-scale biogas plant, a measurement system collaboratively developed between Cologne University of Applied Sciences and art photonics was applied. The measurement system involved a Thermo Scientific Nicolet iS5 FTIR spectrometer for the spectral measurements as well as a deuterated triglycine sulfate (DTGS) detector. To allow for direct connection of the fiber to the spectrometer, the iS5 was also fitted with a customized interface module developed by art photonics, which enables the optical coupling of the infrared (IR) signal to a polycrystalline infrared (PIR) fiber by SMA905 connectors.

For the full-scale application, it would not be practical to manually clean the probe after every measurement, and so an automated cleaning system was utilized. The Ceramat FOS® is an automatic sensor gate designed with a ceramic seal between the process and the cleaning chamber. The sensor gate forms an effective seal preventing any further material passing through the gate.

The Unical 9000-FOS enables clean-in-place functionality for ATR immersion probes. To clean the sensor, the sensor is first withdrawn from the recirculation pipe and moved into the ceramic housing of the sensor gate. Sealed off from the recirculating substrate, the probe can be cleaned with three cleaning media that can be freely chosen, and dried using compressed air. After cleaning, recalibration through baseline measurement can be performed before the probe is returned into the process. As part of this work, testing was carried out to evaluate the effectiveness of the different cleaning media by performing a measurement with the probe in milk, flour or sunflower oil, and then rinsing the probe with the different cleaning media and recording a further spectrum. The spectra were then compared to evaluate the cleaning efficiency of different chemicals.

Cleaning effectiveness tests were accomplished by collecting a baseline spectrum before performing a measurement, then measuring the substance and applying the Unical 9000 flushing functionality to clean the probe. The tests were done with milk, flour, sunflower oil, and a mix of sunflower oil and milk. The cleaning media tested were deionized water, soap solution, and acetone.

Due to the usage of the Unical system, the entire probe body would move when the probe was withdrawn for cleaning. In order to allow for this movement of the probe relative to the spectrometer, it was necessary to employ a 3-m PIR fiber for the connection of the diamond-tipped ATR probe to the iS5 spectrometer. Using this longer PIR fiber resulted in stronger attenuation of the IR signal, which reduced the SNR of the measured signal.

The complete measurement system is presented in Fig. 2. Section A shows the Ceramat FOS® probe fitting and sensor gate, fixed in position to the recirculation pipe of the two digesters. Section B displays the MIR spectrometer mounted in a secure enclosure and C shows the Unical 9000 control unit which is used for cleaning the probe head. With this system, it was possible to record spectra from the digestate automatically. These spectra samples are referred to as dataset U_3.

Figure 2. Measurement system installation at the full-scale industrial biogas plant. (A) MIR Probe, (B) MIR Spectrometer, (C) UNICAL cleaning system.

Compared to the laboratory setup, spectrometer parameters had to be chosen differently in order to compensate for the

lower sensitivity of the DTGS detector and the stronger attenuation caused by the longer PIR fiber and the interface module used for fiber coupling. Therefore, the average of 256 scans was calculated which increased the measurement duration substantially. In order to shorten this duration, the resolution was reduced to 16 cm^{-1}. An amplification of 8 was used, which is the same as for the laboratory spectrometer tests.

2.3 ATR Probe and PIR Fibers

The FlexiSpec diamond-tipped ATR probe from art photonics that was taken for the laboratory and full-scale measurements utilizes the ATR principle, where the absorption of the light occurs due to evanescent waves travelling into the sample. The probe features a single measurement face, which helps address the issue of soiling from the process material. Soiling occurs when the material that was previously measured, remains on the probe, and thus future measurements will also be influenced by the material. For sensors which pass light through a gap to measure the absorption, this gap is often easily soiled. Additionally, the high absorption of substrate in AD means that the gap has to be small (1 mm), which is even more difficult to clean. The ATR probes have a significantly shorter path length, which is suitable for measuring digestate, and the single measurement face is easier to clean than transmission sensors.

Different materials can be used to fabricate the ATR tip. Typical materials are diamond, ZnSe, and cubic zirconium. Depending on tip and fiber material, the resulting probes operate in different spectral ranges. Diamond was chosen due to its hardness, resulting in a spectral range from 600 to 1900 cm^{-1} which is congruent with the fingerprint region. Signals in the region from 2300 to 3100 cm^{-1} can still be detected but show a lower intensity. Even though diamonds are well-known for their durability, a protective top as can be seen in Fig. 3 was added. By this crown, the diamond is protected from stones and hard particles which can be found in the dry mass of an anaerobic digester.

Figure 3. Different sizes and designs of the FibreSpec, with and without protective crown, diamond-tipped ATR probe made of stainless steel, manufactured by art photonics.

In order to measure in the interesting fingerprint region in the MIR spectral range and to allow for a distance of 3 m between the spectrometer and diamond ATR probe, PIR fibers are used for the connection. They enable transmission from 3300 to 550 cm^{-1} and are obtained by extrusion from a solution of $AgCl_{1-x}Br_x$ where $0 < x < 1$ [15]. In contrast to most IR materials, AgCl:AgBr crystals are nontoxic and nonhygroscopic. Core and cladding are made from the same material, but the refractive index is different. After extrusion the PIR fibers are inserted into a loose polyether ether ketone (PEEK) tubing for mechanical protection and bending within the elasticity limits. Due to the size and weight of the FTIR spectrometers and to allow for movement of the retractable probe fitting, it was not possible to use a smaller PIR fiber length than 3 m. MEMS-based spectrometers are a feasible solution, where a smaller sensor could be integrated into the probe, which would substantially reduce the fiber length and result in a higher SNR and thus measurement accuracy.

3 Data Analysis and Machine Learning Methods

For interpretation of the spectral data, the data was analyzed by machine learning methods that have proven to be valuable tools for spectral data analysis [16, 17]. In this case, the software package Un-scrambler R V9.2 was employed for the analysis. Data preprocessing was performed to smooth the spectra by Savitzky-Golay filtering [18] and by a linear correction of the baseline. All further data analysis was then carried out on this preprocessed data. Principal component analysis (PCA) was performed using the exploratory data analysis. A further regression was carried out using PLS and v-support vector regression (SVR).

The version of the Un-scrambler software did not include the ν-SVR calculation, and so this was performed with the R software environment. Version 1.6 of the package E1071 was applied for this purpose.

The entire measured spectra consisted of 152 wavenumbers, but in order to facilitate the analysis, 59 selected wavenumbers were used, covering the spectral regions where the process variables show the strongest absorption. In total, three different datasets were taken for the analysis (Tab. 2):

- U_1: pure samples of different concentrations of ammonium, TA, and VFA in ultrapure water
- U_2: samples from the digestate spiked with a range of concentrations for ammonium, TA, and VFA
- U_3: samples collected from the full-scale plant in parallel to the online measurements.

3.1 PLS

In PLS, the original input matrix $\mathbf{X}$[1)] containing the spectral data and the target concentration variable $\mathbf{Y}$ are both projected into a new space in which the covariance between the projected $\mathbf{X}^*$ and $\mathbf{Y}^*$ is maximal according to the following underlying general model:

1) List of symbols at the end of the paper.

Table 2. Overview of the three datasets used for calibration and validation of the laboratory and full-scale MIR measurement systems.

	Training	Validation	Complete
Dataset, $U_1(\mathbf{X}_1, \mathbf{Y}_1)$	$\mathbf{X}_{1,\text{train}} = [m-n]_{54\times 59}$	$\mathbf{X}_{1,\text{val}} = [m-n]_{44\times 59}$	$\mathbf{X}_1 = [m-n]_{98\times 59}$
	$\mathbf{Y}_{1,\text{train}} = [m-p]_{54\times 3}$	$\mathbf{Y}_{1,\text{val}} = [m-p]_{44\times 3}$	$\mathbf{Y}_1 = [m-p]_{98\times 3}$
Dataset 2, $U_2(\mathbf{X}_2, \mathbf{Y}_2)$	$\mathbf{X}_{2,\text{train}} = [m-n]_{50\times 59}$	$\mathbf{X}_{2,\text{val}} = [m-n]_{25\times 59}$	$\mathbf{X}_2 = [m-n]_{75\times 59}$
	$\mathbf{Y}_{2,\text{train}} = [m-p]_{50\times 3}$	$\mathbf{Y}_{2,\text{val}} = [m-p]_{25\times 3}$	$\mathbf{Y}_2 = [m-p]_{75\times 3}$
Dataset 3, $U_3(\mathbf{X}_3, \mathbf{Y}_3)$	$\mathbf{X}_{3,\text{train}} = [m-n]_{117\times 59}$	$\mathbf{X}_{3,\text{val}} = [m-n]_{60\times 59}$	$\mathbf{X}_3 = [m-n]_{177\times 59}$
	$\mathbf{Y}_{3,\text{train}} = [m-p]_{117\times 3}$	$\mathbf{Y}_{3,\text{val}} = [m-p]_{60\times 3}$	$\mathbf{Y}_3 = [m-p]_{177\times 3}$

m, number of samples; *n*, number of wavenumbers; *p*, number of target variables.

$$\begin{aligned}\mathbf{X} &= \mathbf{X}^{\star}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}} \\ \mathbf{Y} &= \mathbf{Y}^{\star}\mathbf{R}^{\mathrm{T}} + \mathbf{E}_{\mathbf{Y}}\end{aligned} \tag{1}$$

where $\mathbf{P}$ and $\mathbf{R}$ stand for orthogonal loading matrices and $\mathbf{E}_{\mathbf{X}}$ and $\mathbf{E}_{\mathbf{Y}}$ represent the errors. Thus, $\mathbf{X}^{\star}$ may contain less predictors than $\mathbf{X}$; furthermore, all predictors in $\mathbf{X}$ are orthogonal to the preceding predictors in $\mathbf{X}^{\star}$.

3.2 SVR

The most commonly used form of SVR is called ε-SVR and was introduced by Vapnik [19]. In this case, ν-SVR was used as the parameter $\nu \in \{0,1\}$ is independent from the dimensions of the absolute values of $\mathbf{X}$. The parameter ν is called sparsity parameter and limits the allowed error ε and the overall number of support vectors which improves computation time of the fully trained model. Furthermore, the parameter C, which is a tradeoff parameter between error and margin, needs to be adequately chosen for each application.

In its simplest form an SVR is a linear regression model but usually it is nonlinear by introducing so-called kernel functions. The kernel function maps the predictors into a high-dimensional feature space so that also highly nonlinear regression models can be learned. For the SVR used for COD and NH_4-N estimation, a radial basis function (RBF) kernel with the free parameter λ is used as it is perfectly suited for a nonlinear relation between $\mathbf{X}$ and $\mathbf{Y}$.

$$K(\mathbf{x_i}, \mathbf{x_j}) = \exp\left(-\lambda \left\|\mathbf{x_i} - \mathbf{x_j}\right\|^2\right) \text{ with } \lambda := \frac{1}{2\sigma^2} \tag{2}$$

and $\mathbf{x_i}, \mathbf{x_j} \in \mathbf{X}$

Another advantage of the RBF kernel is its ability to capture linear relations, as the linear kernel is a special case of the RBF kernel as proven by Keerthi and Lin [20].

The optimal parameters C, ν, and λ were determined by a grid search (Tab. 3), and 10-fold cross-validation was used in order to prevent overfitting.

4 Results

4.1 Cleaning

In order to assess efficiency of cleaning of the probe using the fully automated process interface, laboratory tests were conducted with milk, flour, and oil. In particular, oil has proven to be difficult to remove from optics and is considered to be the hardship case.

To quantitatively evaluate the performance, the mean signed deviation (*MSD*) was calculated for the spectra after cleaning when compared to the baseline spectra. This measure was used since for some of the test cases the cleaned result was an improvement over the baseline measurement, which is not shown when considering measures such as RMSE. As such, a positive *MSD* means that the spectrum is experiencing higher absorption than the baseline measurement, and conversely, a negative *MSD* indicates that the sensor is cleaner than the baseline. The calculation of *MSD* is given in Eq. (3), where $y_{\text{base},w}$ is the baseline absorption at wavenumber w, $y_{\text{clean},w}$ the absorption after cleaning at wavenumber w, and n denotes the total number of wavenumbers.

$$MSD = \sum_{w=1}^{n} \frac{y_{\text{base},w} - y_{\text{clean},w}}{n} \tag{3}$$

For cleaning milk from the probe, the first cleaning step with deionized water resulted in an *MSD* of 0.59 percentage points compared to the original baseline measurement. After cleaning with soap solution, the *MSD* was reduced to 0.23 percentage points. Finally, after cleaning with acetone, the variation was –0.008 percentage points, an improvement over the baseline measurement, and visually was no longer discernible from the measurement noise.

Table 3. Optimal ν-SVR parameter sets for the datasets U_1, U_2, and U_3.

	Ammonium	VFA	TA
Dataset 1, U_1	$C = 6.6\times10^{4}$ $\lambda = 6\times10^{-6}$ $\nu = 0.2$	$C = 8\times10^{6}$ $\lambda = 4\times10^{-8}$ $\nu = 0.3$	$C = 2\times10^{3}$ $\lambda = 1.5\times10^{-4}$ $\nu = 0.6$
Dataset 2, U_2	$C = 1.77\times10^{5}$ $\lambda = 1.1\times10^{-6}$ $\nu = 0.43$	$C = 3.83\times10^{4}$ $\lambda = 1.7\times10^{-6}$ $\nu = 0.37$	$C = 5.67\times10^{3}$ $\lambda = 1.7\times10^{-5}$ $\nu = 0.33$
Dataset 3, U_3	$C = 10$ $\lambda = 6.1\times10^{-3}$ $\nu = 0.23$	$C = 60$ $\lambda = 6.7\times10^{-3}$ $\nu = 0.47$	$C = 69$ $\lambda = 6.8\times10^{-4}$ $\nu = 0.40$

When testing cleaning of flour from the probe, this was more easily removed, with the deionized water being sufficient to give an *MSD* of 0.03 percentage points. Cleaning with soap solution gave a transmission that was slightly over the original baseline in some areas, with an *MSD* of –0.05 percentage points. This result is higher than the *MSD* for cleaning with deionized water, because the sensor was slightly cleaner than when the experiment started. Cleaning with acetone gave an *MSD* of –0.26 percentage points. Again, this was due to how effective acetone was in cleaning the sensor, and the *MSD* is higher because the sensor after the acetone measurement was considerably cleaner than at the start of the experiment.

The test results for the cleaning tests with sunflower oil are illustrated in Fig. 4. The figure shows that cleaning with deionized water improved the spectrum, however, the characteristics of the sunflower oil spectrum are still clearly visible after cleaning with deionized water. The *MSD* was 4.14 percentage points. Cleaning with soap solution was not performed for this test.

Cleaning with acetone resulted in a spectrum that matched the original baseline very well, with an *MSD* of –0.036 percentage points. A detailed plot of the results for cleaning with acetone compared to the baseline measurement is presented in Fig. 5.

A final cleaning test was performed, in which a mixture was created from milk, flour, and sunflower oil. The results from this test are displayed in Fig. 6, with a detailed plot given in Fig. 7. After cleaning with water, there was still clearly probe fouling since an absorption profile with distinct peaks can be seen. The *MSD* was 3.74 percentage points. The cleaning with soap solution was more effective, with an *MSD* of 0.55 percentage points from the baseline measurement. Cleaning with acetone was clearly the most effective, with the spectra recorded after cleaning being very close to the original baseline measurement, and in some cases slightly higher than the original. The *MSD* for acetone was –0.16. Thus, cleaning of the probe using the process interface was successful in all three cases.

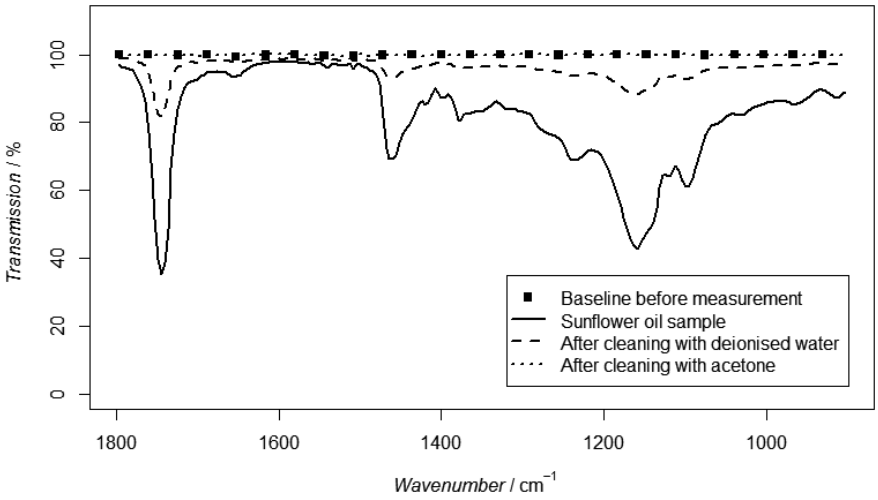

Figure 4. Comparison of cleaning sunflower oil with deionized water and acetone.

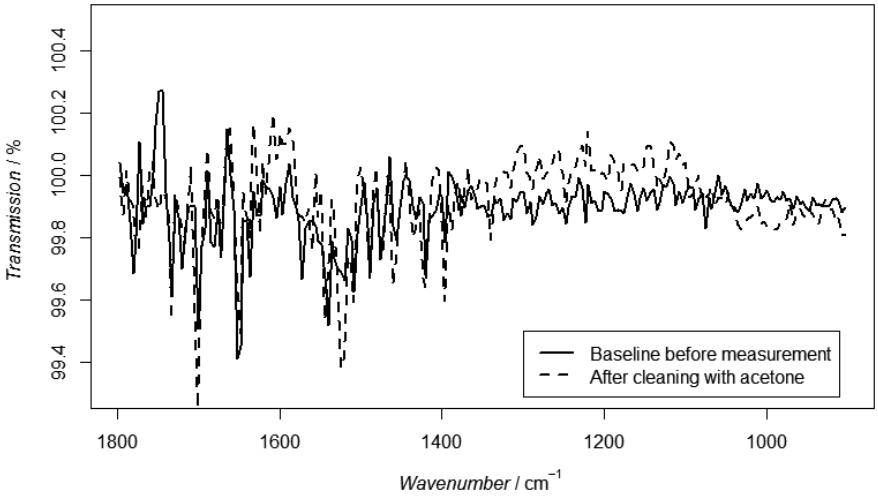

Figure 5. Effectiveness of cleaning with acetone after measuring sunflower oil.

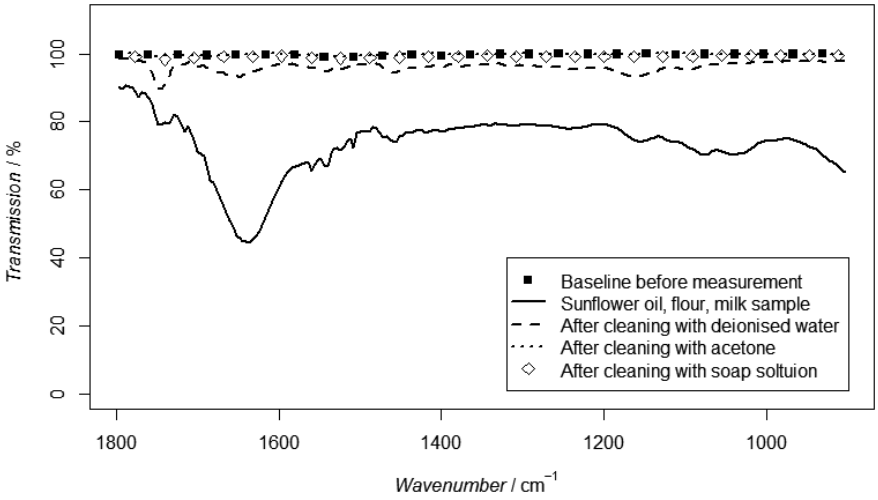

Figure 6. Effectiveness of cleaning a mixture of oil, flour, and milk using deionized water, acetone or soap solution.

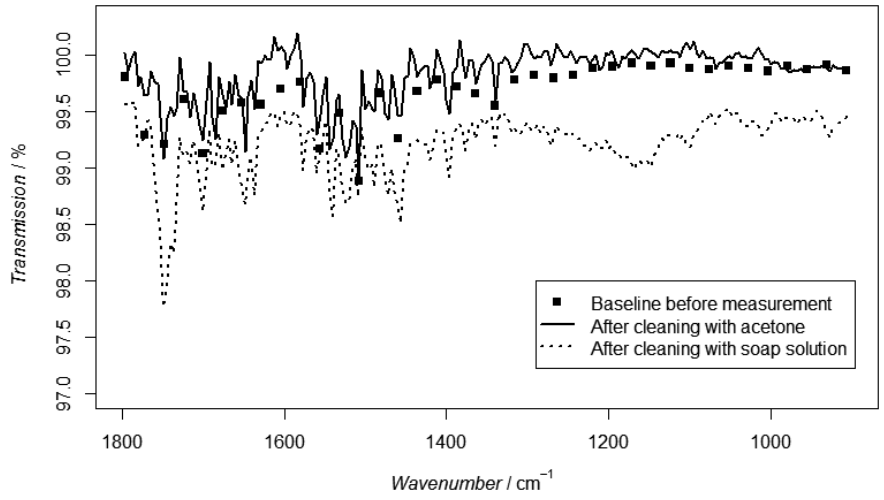

Figure 7. Different cleaning performances of acetone and soap solution when cleaning a mixture of oil, flour, and milk.

4.2 Lab Measurements and Calibration

The laboratory results for the calibration data from dataset U_1 are significant. The R^2 values are very high, showing that the majority of the variance is accounted for in the calibration models. For acetic acid, the limit of detection was 0.00389 mol L^{-1} and the limit of quantification was 0.00878 mol L^{-1}. A study concluded that acetic acid levels in excess of 800 mg L^{-1} indicate that the digester is unstable and could fail [21]. This corresponds to a concentration of 0.0133 mol L^{-1}. Thus, the laboratory results demonstrate that it is possible to quantify the levels of acetic acid as they are approaching levels that would threaten the digester operation.

When considering ammonium levels, experiments indicate that an ammonium level of ~ 0.15 mol L^{-1} resulted in biological failure [22]. From the laboratory testing, the limit of quantification was 0.063 mol L^{-1}. So, for the ammonium, it was also possible to quantify the present concentrations below the level that would result in failure in an anaerobic digester.

Levels of bicarbonate over 0.15 mol L^{-1} proved to inhibit methane production [24]. This is significantly higher than the limit of quantification, which was 0.00806 mol L^{-1}, and demonstrates that the sensitivity of the lab system would be high enough to perform measurements of the concentrations present under typical operating conditions, and not only when the digester was approaching failure.

Tab. 4 summarizes the results for the PLS regression model and Tab. 5 those for the ν-SVR model, with both models using dataset U_1. Generally, it can be concluded that the performance for ν -SVR is slightly better than that of the PLS model. For the total VFAs, the PLS model performs better, however, for all other parameters, the ν-SVR model has a smaller error. For propionic acid, the prediction accuracy is worse when compared to acetic acid or total VFAs, as seen by the lower R^2 value. This is due to the lower concentrations of propionic acid that are present. The prediction accuracy is closely related to the sensitivity of the sensor, and hence, an increase in the sensitivity of the sensor at lower concentrations would result in an improved accuracy of individual VFAs.

Figures previously published [23] show several of the spectra recorded as part of dataset U_2, which indicates the change in spectra due to spiking a substrate sample with acetic acid. From this plot, it can be seen that the spiked substrate spectrum follows the profile of the raw substrate spectrum, however, with additional peaks. The additional peaks correspond to the peaks from the spectrum of pure acetic acid. Furthermore, at the same wavenumbers, it is possible to see small peaks that are already present, which is expected as there was some acetic acid already present in the sample before it was spiked with additional acetic acid.

Table 4. Statistical analysis on PLS1 regression model using 59 wavenumbers from dataset U_1.

Parameter	PLS factors	R^2	RMSE [g L^{-1}]
Acetic acid	8	0.970	0.259
Propionic acid	1	0.544	0.260
Total VFAs	8	0.978	0.403
Sodium bicarbonate	3	0.991	0.594
Ammonium	2	0.990	0.108

Table 5. Statistical analysis on ν-SVR regression model using 59 wavenumbers from dataset U_1.

Parameter	R^2	RMSE [g L^{-1}]
Acetic acid	0.960	0.208
Propionic acid	0.637	0.196
Total VFAs	0.930	0.544
Sodium bicarbonate	0.995	0.400
Ammonium	0.993	0.068

When comparing the spectrum for the substrate sample to those of ammonium chloride and sodium bicarbonate from [23], clear similarities in the spectra become obvious. In the spectrum for the substrate, there are two strong peaks at ~ 1364 cm^{-1} and 1620 cm^{-1}, which corresponds to the two strong peaks in the spectrum for sodium bicarbonate. The peak at ~ 1456 cm^{-1} corresponds to the peak in ammonium chloride at the same wavenumber. Tab. 6 gives the wavelength maxima that were observed during testing.

Table 6. Absorption maxima from measured samples.

Substance	Wavenumber [cm^{-1}]	Wavelength [μm]
Hydrogen-carbon buffer	1620	6.17
Acetic acid	1544	6.48
Ammonium	1456	6.87
Acetic acid	1416	7.06
Hydrogen-carbon buffer	1364	7.33
TS	1050	9.52

4.3 Full-Scale Test

On the one hand, calibration and validation of the full-scale dataset U_3 show promising results (Tab. 7) for ammonium and TA with errors between 0.1 and 0.5 g L^{-1} and good R^2 values of 0.9 and 0.89 which is sufficient for online process monitoring of the AD process in biogas plants where measurement errors of up to 20 % are acceptable. However, results for VFA, one of the most interesting parameters, are very poor with R^2 values of 0.6 for PLS or 0.38 for ν-SVR, respectively.

Table 7. Validation results for dataset U_3 using PLS and ν-SVR.

	R^2 PLS	RMSE PLS [g L^{-1}]	R^2 SVR	RMSE SVR [g L^{-1}]
Ammonium	0.912	0.1	0.827	0.1
VFA	0.645	0.245	0.386	0.235
TA	0.890	0.480	0.768	0.508

On the other hand, long-term operation of the system indicated that despite the fully automated cleaning and calibration system, robustness is still one of the major problems. After five months of continuous operation, the probe fitting was severely damaged due to strong abrasion of the ceramics (Fig. 8 a) which eventually ended in a breaking of the probe fitting. Furthermore, the cleaning chamber contained deposits of the substrate and was blocked by digestate during operation on a regular basis which made cleaning impossible (Fig. 8 b).

5 MEMS-Based Fabry-Pérot Spectrometers

There is a promising potential for future applications of MEMS-based spectrometers in the AD industry. Newly developed tunable MEMS-based Fabry-Pérot interferometers on a chip for the UV/vis, NIR, and MIR wavelength ranges are very small, i.e., 5×10 cm, and also relatively cheap if manufactured in great numbers. Currently, two different system designs exist. Neumann et al. [25] introduced a tunable MEMS interferometer for the middle- and long-infrared range using a pyrodetector. The different wavelengths can be generated by two bragg reflectors whose distance can be changed by a spring suspension. Although the presented performance results are good, the spring suspension is considered to be a drawback as it makes the spectrometer sensitive to vibration and wear. Therefore, VTT developed an interferometer design with piezo effect-based tuning of the gap between the reflectors [26, 27].

Problems using a PIR fiber with a length of 3 m to couple between the spectrometer and the sensor causing strong attenuation in the PIR fiber, attenuating the optical signal amplitude, and decreasing the SNR level significantly, can be overcome with these new spectrometers. In particular, the reduction of the length of the PIR fiber used for an installation by integrating the spectrometer into the probe housing could, therefore, substantially improve the SNR of the measured signal, and consequently the accuracy of the results and minimum quantifiable levels of the concentrations being measured. This combination of the probe and sensor into one complete unit would enable increased measurement accuracy whilst simultaneously reducing the cost of a spectroscopy-based monitoring system, and provides a potentially cheaper and less complex alternative to a system based on a traditional FTIR MIR spectrometer.

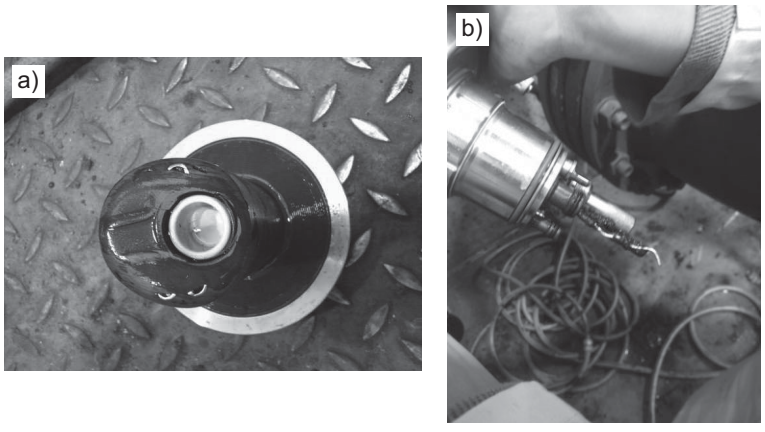

Figure 8. (a) Abrasion of the probe fitting after five months of operation; (b) soiling and blocking of the cleaning chamber with digestate.

6 Conclusion and Outlook

The results indicate that MIR spectroscopy is a valid method for online measurement of critical process variables in AD processes. The combination of an MIR spectroscopic probe with a fully automated system for cleaning proved to reduce maintenance to a minimum during full-scale testing, guaranteeing high measurement accuracy. Laboratory tests of the automated cleaning process showed that deionized water alone is not sufficient in most cases, particularly when the substance contains oil. At a minimum, cleaning with a soap solution will provide very good results, however, in order to minimize fouling, acetone or comparable substances are the most effective cleaning media.

Furthermore, the laboratory results demonstrate that it is possible to measure concentrations of ammonium and VFA at levels which would indicate that a digester was approaching failure. The performance for measurement of TA levels was better, and using the laboratory setup, it was possible to quantify concentrations at levels that would be present before the digester biocommunity was threatened. In full-scale application, ammonium and TA can be detected successfully, in contrast to VFAs which can only be identified reliably at higher levels. One of the main reasons for worse performance in full scale than in laboratory scale was the use of a PIR fiber that was three times longer which caused stronger attenuation and a decrease of the SNR, thus, a reduction of the fiber length would increase the SNR and hence allow measurement of lower concentration levels.

Furthermore, the number of wavenumbers used for analysis could be significantly reduced by 40 % guaranteeing the same prediction quality when applying machine learning methods. This reduction is significant for MEMS-based spectrometers using a Fabry-Pérot interferometer and can result in a reduction of measurement times.

In order to further enhance the measurement system performance, the use of MEMS-based spectrometers and their integration into the probe housing to reduce fiber length is the most promising solution. However, this would require further trials and testing to verify whether the performance is suitable for application in AD processes.

Acknowledgment

This research was funded by German Federal Ministry of Economic Affairs and Energy under the ZIM program. The project "INNO-MIR Biogas – Development and automation of an innovative online MIR measurement system for biogas plants" was funded under support code KF2137807AK1.

The authors have declared no conflict of interest.

Symbols used

C	[–]	SVR tradeoff parameter (margin vs. error)
K	[–]	kernel
$\mathbf{P}$	[–]	loading matrix for input data $\mathbf{X}$
$\mathbf{R}$	[–]	loading matrix for target data $\mathbf{Y}$
$\mathbf{E}$	[–]	error matrix
$\mathbf{X}$	[–]	input data
$\mathbf{Y}$	[–]	target data
$\mathbf{X}^*$	[–]	input data projected into a feature space
$\mathbf{Y}^*$	[–]	target data projected into a feature space

Greek letters

λ	[–]	free RBF kernel parameter
σ	[–]	free RBF kernel parameter
ν	[–]	SVR sparsity parameter

Abbreviations

AD	anaerobic digestion
ATR	attenuated total reflection
COD	chemical oxygen demand
DTGS	deuterated triglycine sulfate
FTIR	Fourier transform infrared spectroscopy
MCT	mercury cadmium telluride
MEMS	microelectronic mechanical system
MIR	mid-infrared
NH_4-N	ammonium nitrogen
NIR	near-infrared
oTS	organic total solids
PA	partial alkalinity
PEEK	polyether ether ketone
PIR	polycrystalline infrared
RBF	radial basis function
SMA	SubMiniature version A
SNR	signal-to-noise ratio
TA	total alkalinity
TAC	total anorganic carbon
TAN	total ammonia nitrogen
TGS	triglycine sulfate
TOC	total organic carbon
TS	total solids
UV/vis	ultraviolet/visible
VFA	volatile fatty acids

References

[1] *Erneuerbare-Energien-Gesetz 2012 (Renewable Energy Law 2012): EEG*, Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (Federal Ministry for the Environment, Nature Conservation and Nuclear Safety), Berlin **2011**.

[2] *Electricity: The Feed-in Tariffs (Specified Maximum Capacity and Functions) Order 2010*, No. 678, UK Department of Energy and Climate Change, London **2010**. www.fitariffs.co.uk/library/regulation/uksi_20100678_en1.pdf

[3] *A Competition for Electricity Generation from Anaerobic Digestion, Biomass Powered High Efficiency CHP Ocean Energy and Offshore Wind Energy*, Department of Communications, Energy and Natural Resources Ireland, Dublin **2009**,

[4] Energy Policies of IEA Countries, *The Netherlands 2008 Review*, International Energy Agency, **2009**. www.iea.org/publications/freepublications/publication/Netherlands2008.pdf (accessed March 08, 2016)

[5] *Biogas Segment Statistics 2014*, Fachverband Biogas e.V., Freising **2015**. www.biogas.org/edcom/webfvb.nsf/id/DE_Branchenzahlen/$file/15-11-19_Biogas%20Branchenzahlen-2014_Prognose-2015_final.pdf (accessed May 28, 2015)

[6] *Biogas-Messprogramm II – 61 Biogasanlagen im Vergleich: (Biogas Measurement Program II – A Comparison of 61 German Biogas Plants)*, 1st ed., Fachagentur Nachwachsende Rohstoffe (Agency for Renewable Resources), Gülzow **2009**.

[7] M. R. Provenzano, A. D. Malerba, D. Pezzolla, G. Gigliotti, *Waste Manage.* **2014**, *34 (3)*, 653. DOI: 10.1016/j.wasman.2013.12.001

[8] P. Steyer, J. C. Bouvier, T. Conte, P. Gras, J. Harmand, J. P. Delgenes, *Water Sci. Technol.* **2002**, *45 (10)*, 133.
[9] H. Spanjers, J. Bouvier, P. Steenweg, I. Bisschops, W. van Gils, B. Versprille, *Water Sci. Technol.* **2006**, *53 (4–5)*, 55.
[10] L. C. Krapf, H. Heuwinkel, U. Schmidhalter, A. Gronauer, *Biomass Bioenergy* **2013**, *48*, 224–230. DOI: 10.1016/j.biombioe.2012.10.027
[11] A. J. Ward, E. Bruni, M. K. Lykkegaard, A. Feilberg, A. P. S. Adamsen, A. P. Jensen, A. K. Poulsen, *Bioresource Technol.* **2011**, *102*, 4098–4103. DOI: 10.1016/j.biortech.2010.12.052
[12] J. Holm-Nielsen, H. Andree, H. Lindorfer, K. Esbensen, *J. Near Infrared Spectrosc.* **2007**, *15*, 123. DOI: 10.1255/jnirs.719
[13] M. Bonhomme, A. Pavia, *Rev. Energ.* **1986**, *385*, 674.
[14] G. K. Anderson, G. Yang, *Water Environ. Res.* **1992**, *64*, 53–59.
[15] V. G. Artyushenko, A. Bocharnikov, G. Colquhoun, C. A. Leach, V. Lobachov, L. Pirogova, T. Sakharova, D. Savitskij, T. Ezhevskaya, A. Bublikov, in *Optics/Photonics in Security and Defence* (Eds: G. W. Kamerman, O. K. Steinvall, K. L. Lewis, K. A. Krapels, J. C. Carrano, A. Zukauskas), SPIE, Bellingham, WA **2007**.
[16] C. Wolf, D. Gaida, A. Stuhlsatz, T. Ludwig, S. McLoone, M. Bongards, *Trans. Inst. Meas. Control* **2013**, *35 (1)*, 5. DOI: 10.1177/0142331211403797
[17] J. Luts, F. Ojeda, R. van de Plas, B. de Moor, S. van Huffel, J. A. K. Suykens, *Anal. Chim. Acta* **2010**, *665 (2)*, 129. DOI: 10.1016/j.aca.2010.03.030
[18] A. Savitzky, M. J. E. Golay, *Anal. Chem.* **1964**, *36 (8)*, 1627. DOI: 10.1021/ac60214a047
[19] V. N. Vapnik, *Statistical Learning Theory, Adaptive and Learning Systems for Signal Processing, Communications, and Control*, Wiley, New York **1998**.
[20] S. S. Keerthi, C.-J. Lin, *Neural Comput.* **2003**, *15 (7)*, 1667. DOI: 10.1162/089976603321891855
[21] D. T. Hill, S. A. Cobb, J. P. Bolte, *Trans. ASAE* **1987**, *30 (2)*, 496–501. DOI: 10.13031/2013.31977
[22] H. M. Poggi-Varaldo, R. Rodriguez-Vazquez, G. Fernandez-Villagomez, F. Esparza-Garcia, *Appl. Microbiol. Biotechnol.* **1997**, *47 (3)*, 284–291. DOI: 10.1007/s002530050928
[23] M. Bongards, D. Gaida, O. Trauer, C. Wolf, *Energy Sustainability Soc.* **2014**, *4*, 19. DOI: 10.1186/s13705-014-0019-3
[24] Y. Lin, F. Lü, L. Shao, P. He, *Bioresource Technol.* **2013**, *137 (6)*, 245. DOI: 10.1016/j.biortech.2013.03.093
[25] N. Neumann, M. Ebermann, E. Gittler, M. Meinig, S. Kurth, K. Hiller, in *Proc of. IEEE Sensors 2010*, IEEE, Piscataway, NJ **2010**, 2383. DOI: 10.1109/ICSENS.2010.5690856
[26] J. Antila, M. Tuohiniemi, A. Rissanen, U. Kantojärvi, M. Lahti, K. Viherkanto, M. Kaarre, J. Malinen, in *Encyclopedia of Analytical Chemistry*, John Wiley and Sons, Hoboken, NJ **2014**. DOI: 10.1002/9780470027318.a9376
[27] J. H. Mäkynen, M. Tuohiniemi, A. Näsilä, R. Mannila, J. E. Antila, *SPIE Proc.* **2014**, *8977*, 1. DOI: 10.1117/12.2036272

