



Vorhersage von Zeitserien der Biogasproduktion in anaeroben Faultürmen mit einem Temporal Fusion Transformer

Johannes Sappl · Matthias Harders · W. Rauch

Angenommen: 13. April 2021 / Online publiziert: 17. Mai 2021
 © Der/die Autor(en) 2021

Zusammenfassung Anaerobe Gärung in Faultürmen liefert als fixer Bestandteil vieler Abwasserreinigungsanlagen einen wertvollen Beitrag zur nachhaltigen Energieerzeugung. Der Einfluss verschiedener Parameter auf die gewonnene Menge an Biogas ist dabei noch nicht hinreichend erforscht. Mit Veränderungen am Faulprozess kann unter realen Bedingungen jedoch nicht experimentiert werden. Deshalb trainieren wir ein datengetriebenes Modell, einen sogenannten Temporal Fusion Transformer, welcher die biochemischen Vorgänge abbildet und Vorhersagen über die zu erwartende Produktionsrate trifft. Sowohl Zeitserien als auch nicht-numerische Merkmale werden dabei gemeinsam verarbeitet. Dies liefert ein umfassenderes Bild, weil externe Faktoren, wie z. B. Schwankungen bedingt durch den Tourismus, ebenfalls mit in Betracht gezogen werden. Das Modell stützt sich auf die vorliegende Information der letzten sieben Tage und erstellt daraus eine Vorhersage für die kommende Woche. Diese besteht zusätzlich zum Median aus den 2%-, 10%-, 25%-, 75%-, 90%- und 98%-Quantilen. Da-

mit unterstützt und verbessert unser Modell die betriebliche Entscheidungsfindung. Der zur Kalibrierung verwendete Datensatz stammt von zwei realen Fermentern in Tirol.

Schlüsselwörter Anaerobe Gärung · Faulgas · Maschinelles Lernen · Temporal Fusion Transformer · Quantilsregression · Zeitserie

Predicting biogas production rates in anaerobic digesters with a temporal fusion transformer

Abstract Anaerobic digestion in fermenters is a fixed component of many water resource recovery facilities. It makes a valuable contribution to sustainable energy production. The influence of various parameters on the amount of biogas produced has not yet been sufficiently researched. However, changes to the digestion process cannot be experimented with under real conditions. Therefore we train a data-driven model, a so-called Temporal Fusion Transformer, which learns to represent the biochemical processes and predicts the expected production rate. Both time series and non-numerical features are processed together. This provides a more comprehensive understanding because external factors such as fluctuations due to tourism are also taken into account. The model is based on the available information of the last seven days and produces a forecast for the coming week. In addition to the median it consists of the 2%, 10%, 25%, 75%, 90% and 98% quantiles. Our model thus supports and improves operational decision-making. The data set used for the calibration is taken from two real digesters in Tyrol.

Keywords Anaerobic digestion · Biogas · Machine learning · Temporal fusion transformer · Quantile regression · Time series

1 Einleitung

Anaerobe Gärung (AG) ist mittlerweile nicht nur auf größeren Abwasserreinigungsanlagen (ARA), sondern auch in eigenständigen Biogasanlagen etabliert und spielt als Quelle für erneuerbare Energie eine wichtige Rolle. Biologisch abbaubares organisches Material aus unterschiedlichen Bereichen wie z. B. Haushalten, landwirtschaftlichen Betrieben und der Industrie wird in anaeroben Reaktoren unter Ausschluss von gelöstem Sauerstoff zersetzt. Dieser komplexe nicht-lineare Abbauprozess umfasst unter anderem die Umwandlung des organischen Kohlenstoffs mittels Reduktions-Oxidationsreaktionen in seine am stärksten reduzierte Form Methan (CH₄) und jene Form mit dem höchsten Anteil an Sauerstoff, Kohlendioxid (CO₂). Das dabei entstehende Biogas, welches – nebst geringer Spuren von Verunreinigungen – eine Kombination aus beidem ist, kann anschließend in Blockheizkraftwerken zur ökologisch nachhaltigen Stromerzeugung genutzt werden. Der nach der vollständigen Vergärung verbleibende Gärrest wird oft in der Landwirtschaft als Dünger auf den Feldern ausgebracht und erfüllt somit eine weitere nützliche Funktion.

Die Bedingungen im Inneren eines Faulturms und die Komplexität der darin ablaufenden Prozesse selbst verhindern, dass Anlagenbetreiber qualitativ hochwertige Messungen vieler biochemischer Bestandteile erhalten. Meistens werden nur jene Parameter erfasst, welche entweder mit relativ geringem technischem und zeitlichem Aufwand oder besonders kostengünstig zu erhalten sind. Dazu zählen z. B. die Temperatur oder der pH-Wert des Faulschlammes. Infolgedessen ist es nicht möglich, den inneren Zustand eines Fermenters und die daraus resultierende Effizienz der Umwandlung organischer Bestandteile in Biogas quantitativ zu beurteilen. Dieser Mangel an Messdaten führt da-

DI J. Sappl (✉) ·
 Univ.-Prof. DI Dr. W. Rauch
 Arbeitsbereich Umwelttechnik,
 Universität Innsbruck,
 Technikerstraße 13, 6020 Innsbruck,
 Österreich
johannes.sappl@uibk.ac.at

Univ.-Prof. DI Dr. W. Rauch
wolfgang.rauch@uibk.ac.at

Univ.-Prof. Dipl.-Inf. Dr. M. Harders
 Forschungsgruppe Interaktive
 Grafik und Simulation,
 Universität Innsbruck,
 Technikerstraße 21, 6020 Innsbruck,
 Österreich
matthias.harders@uibk.ac.at

Tab. 1 Deskriptive Statistik mit Mittelwert μ und Standardabweichung σ der täglichen Betriebsparameter, welche über einen Zeitraum von zwei Jahren erhoben wurden. Die Bezeichnungen 1 und 2 beziehen sich auf zwei baulich identische eiförmige Fermenter auf der ARA

Parameter	Einheit	Fehlend	Min	Max	μ	σ
<i>Exogene Variablen</i>						
Rohschlamm 1	m ³	0	0	146,88	85,12	37,77
Rohschlamm 2	m ³	0	0	213,41	104,48	33,56
Rohschlamm gesamt	m ³	0	92,45	287,71	189,61	40,81
TS Rohschlamm	gl ⁻¹	253	27,90	119,50	57,71	8,09
Rohschlamm TS-Fracht	kg d ⁻¹	253	4441,65	24.469,80	11.014,53	2748,51
Rohschlamm oTS-Fracht	kg d ⁻¹	532	3962,14	13.577,40	8331,04	2018,07
Faulschlamm 1	m ³	0	2,59	191,81	83,34	32,70
Faulschlamm 2	m ³	0	1,73	200,45	99,02	33,46
Faulschlamm gesamt	m ³	0	55,30	337,82	182,35	43,31
Temperatur 1	°C	0	33,55	39,01	38,39	0,62
Temperatur 2	°C	0	36,09	38,98	38,60	0,42
pH-Wert 1	–	0	6,75	9,17	7,54	0,45
pH-Wert 2	–	0	6,91	8,30	7,53	0,36
Faulzeit	d	0	17,38	54,08	27,65	6,15
TS Faulschlamm	gl ⁻¹	237	24	50,50	31,02	4,29
Faulschlamm TS-Fracht	kg d ⁻¹	237	1625,70	14.121,00	5672,05	1500,90
Feststoffbelastung	kg m ⁻³ d ⁻¹	253	0,89	4,89	2,20	0,55
GV Faulschlamm	%	521	51,55	64,40	59,21	2,67
Faulschlamm oTS-Fracht	kg d ⁻¹	526	953,47	6034,82	3338,33	899,44
Kofermentation Bioabfälle	m ³	0	0	52,20	12,10	10,05
Kofermentation CSB-Fracht	kg	730	–	–	–	–
<i>Endogene Variablen</i>						
Faulgas 1	m ³	0	0,00	5479,44	2276,79	1139,95
Faulgas 2	m ³	0	185,52	4995,12	2822,47	1139,95

zu, dass die zu regulierenden Variablen während des aktiven Betriebs in der Regel gemäß dem Wissens- und Erfahrungsschatz von ExpertInnen adjustiert werden. Das möglicherweise dadurch vorhandene ungenutzte Potenzial kann mit Unterstützung von datengetriebenen Modellen ausgeschöpft werden und führt dadurch zu einer weiteren Steigerung der Biogasproduktionsrate.

Aus offensichtlichen Gründen kann man an einem aktiven Fermenter keine Experimente vornehmen, indem unterschiedliche Einstellungen der Parameter des Faulprozesses getestet und ihre Auswirkungen auf die Faulgasproduktion beobachtet werden. Ein weithin verbreiteter Ansatz, um dieses Problem zu umgehen, ist die Verwendung sogenannter Ersatzmodelle bzw. „surrogate models“. Dabei werden zuerst möglichst realitätsnahe Modelle mittels Labordaten oder echter Messreihen kalibriert. Anschließend kann die zu erwartende Biogasproduktion für die kommenden Tage nur basierend auf dem stellvertretenden Modell und nicht dem eigentlichen Faulturm vorhergesagt werden. Die bekannten Messdaten

der vergangenen Tage dienen dabei als Input. Bevor solche Modelle allerdings als Entscheidungshilfen für den Realfall herangezogen werden können, ist eine eingehende Untersuchung deren Genauigkeit erforderlich. Denn aufgrund der Äquifinalität kann derselbe Endzustand von verschiedenen, möglicherweise biochemisch nicht plausiblen Anfangsbedingungen aus erreicht werden.

1.1 Auszug verwandter Arbeiten

Die folgende Literaturrecherche erhebt keinen Anspruch auf Vollständigkeit und zielt nicht darauf ab, eine detaillierte Analyse aller datengetriebenen Ansätze zu präsentieren. Eine gelungene historische Übersicht der AG-Modellierung und ein Vergleich verschiedener mathematischer Modelle finden sich stattdessen in der ausführlichen Arbeit von Wade (2020).

Eines der bekanntesten verallgemeinerten (deterministischen) Modelle der AG ist das in Batstone et al. (2002) entwickelte sogenannte anaerobic digestion model No. 1 (ADM1). Mehrere Diffe-

rential- sowie algebraische Gleichungen beschreiben den Zusammenhang zwischen nicht weniger als 26 Konzentrationsvariablen. In einem komplett durchmischten System bestehen diese aus biochemischen Prozessraten, Massenbilanzen und Differentialgleichungen für die Flüssig- bzw. Gasphase. In Gaida et al. (2012) wird mithilfe von ADM1 ein künstlicher Trainings- und Testdatensatz für drei verschiedene Modelle aus dem Bereich des Maschinellen Lernens (ML) erstellt (vgl. Feigl et al. 2021). Darunter befindet sich auch der äußerst populäre Random Forest, welcher eine Einschätzung der Betriebszustände von Biogasanlagen in Echtzeit ermöglicht. Zwei große Nachteile des analytischen ADM1 sind jedoch, dass der Einfluss der Hydrodynamik auf AG vollständig vernachlässigt wird und dass ARA die meisten ADM1-Variablen nicht standardmäßig erfassen. Dies schmälert die praktische Anwendbarkeit von ADM1 und seine Bedeutung außerhalb des akademischen Bereichs erheblich.

In einer chinesischen Biogasanlage im industriellen Maßstab werden aufgezeichnete Daten analysiert, um die

betriebliche Entscheidungsfindung zu verbessern (De Clercq et al. 2019). Die verwendeten ML-Modelle umfassen unter anderem logistische Regression, Support Vector Machine und die k -nearest neighbors (k -NN)-Methode. Anstelle von Fermenterparametern wie z. B. der Temperatur wird die Auswirkung verschiedener Zusammensetzungen des Faulschlammes auf den AG-Prozess untersucht. Eine grafische Benutzeroberfläche stellt den ARA-Ingenieuren Handlungsempfehlungen für den täglichen Betrieb zur Verfügung.

Obwohl bereits weiterentwickelte ML-Modelle die Fuzzylogik-Systeme allmählich verdrängt haben, liefern diese bei der Vorhersage der Methanproduktionsraten noch immer ausreichend genaue Ergebnisse. Auf Zeitreihen von ARAs in der Türkei bzw. in den Vereinigten Staaten werden adaptive Neuro-Fuzzy-Inferenzsysteme in großem Maßstab kalibriert und ausgewertet (Cakmakci 2007; Kusiak und Wei 2014). Auch für einen anaeroben Schlammdeckenreaktor mit Aufwärtsströmung im Pilotmaßstab von 90 l unter verschiedenen Beladungen werden sie angewendet (Turkdogan-Aydnol and Yetilmezsoy 2010). Der Einfluss von z. B. Alkalinität und pH-Wert wird mit einem Multi-Input-Multi-Output-Modell auf Fuzzylogik-Basis bewertet.

In einer aktuellen Studie wird die Biomethanproduktion in einer großtechnischen Co-Vergärungsanlage in China vorhergesagt. Der Zeithorizont erstreckt sich dabei von einem bis maximal 40 Tage (De Clercq et al. 2020). Insbesondere Random Forest und XGBoost liefern vielversprechende Ergebnisse, müssen aber für jeden einzelnen Zeithorizont individuell trainiert werden. Praktikabler ist im Gegensatz dazu ein einziger Outputvektor, welcher alle zukünftigen eintägigen Produktionsraten repräsentiert.

Die meisten Studien greifen auf die weit verbreiteten mehrlagigen künstlichen neuronalen Netzwerke (ANN) zurück, die in der Ingenieurgesellschaft sowohl sehr bekannt als auch eingehend erforscht sind. Mit einem ANN ist es möglich, gleichzeitig viele AG-Betriebsparameter zu verarbeiten und hochpräzise Ergebnisse mit Black-Box-Optimierung zu erhalten (Holubar et al. 2002; Kanat and Saral 2009; Tufaner and Demirci 2020). Nichtlineare autoregressive exogene Modelle sind in der Lage, mehrere Tage umfassende

zeitliche Dynamiken in den ARA-Daten zu reproduzieren (Dhussa et al. 2014).

Biochemische Prozesse im Allgemeinen und AG im Besonderen bestehen aus komplexen Wechselwirkungen zwischen verschiedenen Größen, die auch stark von externen Parametern beeinflusst werden. Die meisten Modelle in der Literatur berücksichtigen nur einen unvollständigen Satz dieser Variablen als Inputwerte. Die Erfassung von Schwankungen, die z. B. durch die Umgebungstemperatur, gesetzliche Feiertage oder den Tourismus verursacht werden, gestaltet sich daher schwieriger. Wir stützen uns auf eine sogenannte Temporal Fusion Transformer (TFT)-Architektur, welche in der Lage ist, sowohl kategorische als auch kontinuierliche Variable zu verarbeiten. Ein inhärenter Aufmerksamkeitsmechanismus erlaubt zudem, die Bedeutung einzelner Parameter, speziell im Gegensatz zu traditionellen ANNs, hervorzuheben und grafisch zu veranschaulichen. Zusätzlich zur mehrstufigen Vorhersage der Tageswerte verwendet unser TFT-Modell eine Quantilregression. Damit ist es möglich, zusätzlich zur Punktschätzung der Biogasproduktion, ein Konfidenzintervall mit beliebiger Bandbreite zu bestimmen. Unseres Wissens nach findet sich in der Literatur bisher keine Anwendung von TFTs in Verbindung mit einer Quantilregression im Kontext von AG.

2 Faulturm Datensatz

Wir verwenden in dieser Arbeit keine künstlich generierten oder labor-technischen Zeitserien, sondern die Messreihen aus einem Fermenter im industriellen Maßstab. Weitere Schritte umfassen einige Bereinigungen der Zeitserien von Messfehlern und Ausreißern. Die TFT-Modellarchitektur und ihre Vorteile werden beschrieben. Nach der Kalibrierung vergleichen wir die Ergebnisse mit drei etablierten Modellen als Benchmark, um den Zugewinn an Flexibilität und Vorhersagepotenzial unseres Modells aufzuzeigen.

2.1 Aufzeichnen der Daten

Die von uns betrachtete ARA wird vom Abwasserverband Achenal-Inntal-Zillertal (AIZ) betrieben und befindet sich in Strass im Zillertal, Tirol. Sie bewältigt ein Abwasservolumen von $28.270 \text{ m}^3 \text{ d}^{-1}$ bei einem Bemessungswert von 167.000 EW_{60} . Der

Klärschlamm setzt sich zusammen aus Abfällen, die größtenteils, nämlich zu 73%, aus Tourismus, Industrie und Landwirtschaft kommen. Die restlichen 27% stammen von den umliegenden Gemeinden und Haushalten. Insgesamt beläuft sich die BSB₅-Schmutzfracht auf 10.020 kg d^{-1} . Zwei baugleiche eiförmige Faultürme mit je 2500 m^3 Fassungsvermögen operieren im anaeroben Milieu bei 35°C . Nach Abschluss der Fermentation wird das Biogas in zwei Blockheizkraftwerken mit einer Gesamtleistung von 940 kW in ökologisch saubere Energie umgewandelt (AIZ 2020). Die Betriebsdaten des AG wurden über einen Zeitraum von zwei Jahren gesammelt und liegen als Tageswerte vor. Eine Zusammenfassung der statistischen Schlüsselindikatoren der noch unveränderten Zeitreihen ist in Tab. 1 aufgelistet.

2.2 Vorverarbeitung der Daten

Heutzutage sind die meisten ARAs mit moderner Elektronik und Sensoren ausgestattet, welche die Prozesse individuell überwachen und die entsprechenden Messdaten aufzeichnen. Viele dieser Messgeräte sind jedoch anfällig für technische Störungen und Ausfälle, da sie in unwirtlichen, z. B. sauren, feuchten oder heißen Umgebungen eingesetzt werden. Zusätzlich zu menschlichen Fehlern in der Behandlung von Messdaten führt dies zu qualitativ minderwertigen Datensätzen, die durch verrauschte oder verlorene Werte verfälscht sind. Aus diesem Grund ist das Aufbereiten der Datensätze ein wesentlicher Bestandteil jedes datengetriebenen Ansatzes, unabhängig davon, welche Modelle schlussendlich verwendet werden. Da es sich bei AG um einen dynamischen Prozess handelt, beeinflussen nicht nur historische Entwicklungen die tagesaktuellen Werte der beteiligten Variablen. Auch die Wechselwirkung der Parameter untereinander darf nicht vernachlässigt werden. Die Behandlung fehlender Werte und Ausreißer gestaltet sich dadurch jedoch noch schwieriger (Rustum and Adeloye 2007).

Als ersten Schritt berechnen wir die Kreuzkorrelationen der Zeitserien im Datensatz, siehe Abb. 1. Beispielsweise korreliert die TS-Fracht sehr positiv mit der oTS-Fracht des Rohschlammes oder der Feststoffbelastung in beiden Fermentern. Merkmale mit hoher positiver Korrelation sind stark linear abhängig

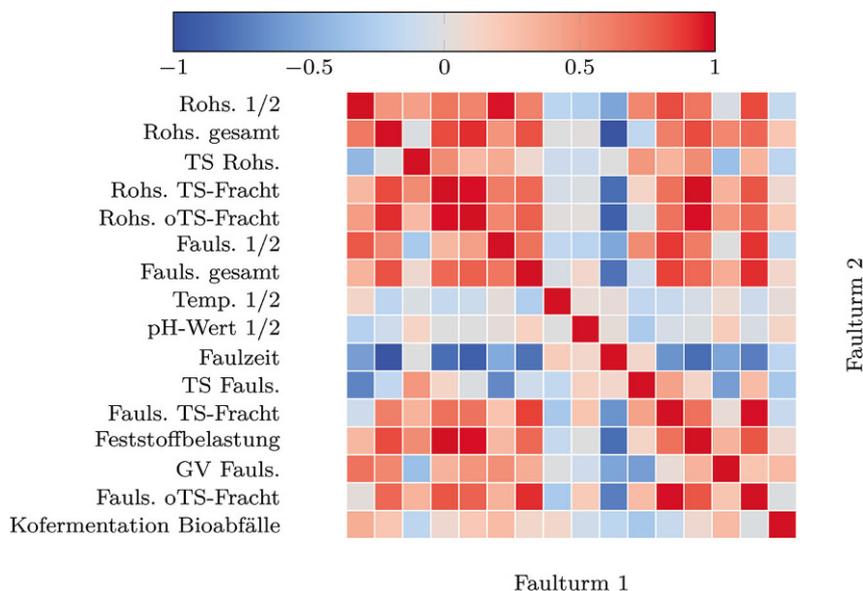


Abb. 1 Kreuzkorrelation der Zeitserien im Datensatz für Faulturm 1 (untere Dreiecksmatrix) und Faulturm 2 (obere Dreiecksmatrix). Merkmale mit hoher positiver Korrelation wie z. B. die TS- und oTS-Fracht des Rohschlammes liefern nahezu keine zusätzliche Information für das Modell. Sie werden ohne Einbußen bei der Modellgenauigkeit aus der Menge der exogenen Variablen entfernt

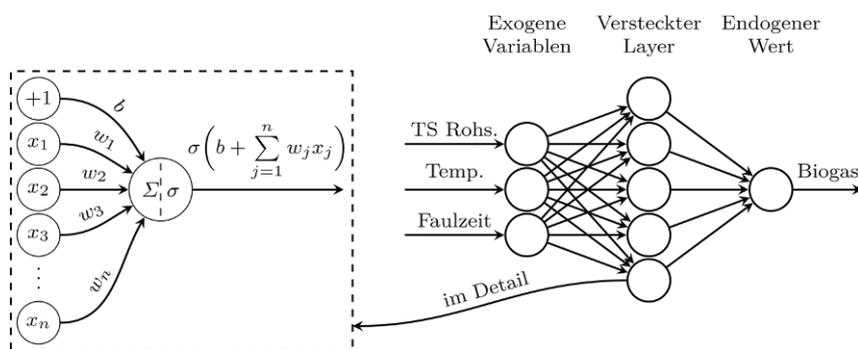


Abb. 2 Vereinfachte Darstellung eines ANN mit drei Variablen als Input. Die zu optimierenden Parameter sind w_1, w_2, \dots, w_n und ein Bias b in jedem Knotenpunkt. Wir trainieren das eigentliche ANN mit allen bereinigten Variablen, vgl. Abschn. 2.2. Der Output ist eine Vorhersage der zu erwartenden Produktionsrate von Biogas

und liefern daher wenig neue Information an das Modell. Sie können als Input vernachlässigt werden, wodurch der Rechenaufwand für ohnehin redundante Features verringert und die Optimierung verkürzt wird. Wir entfernen auf diese Weise die oTS-Fracht im Rohschlamm und Faulschlamm sowie die Feststoffbelastung, deren positive Korrelationen jeweils über einem Grenzwert von 0,90 liegen.

Die Ursachen für fehlende Messwerte in den Daten sind sehr unterschiedlich. In den meisten Fällen handelt es sich um Störungen der Messtechnik, Fehler während des Übertragungsprozesses oder eine korrumpierte Daten-

bank. Da die CSB-Fracht der Kofermentation weder 2018 noch 2019 von der AIZ erfasst wurde, lassen wir sie außen vor. Die restlichen Lücken im Datensatz werden linear zwischen dem letzten und nächsten aufgezeichneten Wert interpoliert.

Ausreißer in den Daten stellen für ein Modell ebenfalls ein Problem dar. Trotz ihrer geringen Anzahl fallen sie im Gegensatz zu normalen Werten überdurchschnittlich stark ins Gewicht und verzerrten statistische Kennzahlen wie Mittelwert und Standardabweichung. Dies wiederum beeinflusst die Güte des Modells negativ, da während des Trainingsprozesses zu viel Augenmerk

auf Werte außerhalb des eigentlichen Zeitreihentrends gerichtet wird. Wir berechnen zunächst die 25%- und 75%-Quartile (Q_{25} und Q_{75}) der exogenen Variablen sowie die Interquartilsabstände $Q_{75}-Q_{25}$. Alle Ausreißer, die sich um mehr als das 1,50-Fache dieses Abstands von Q_{25} bzw. Q_{75} unterscheiden, werden auf den Wert des entsprechenden Quartils gesetzt.

Die Inputvariablen in Tab. 1 werden durch zusätzliche Features ergänzt, welche die anfallende Schlammmenge oder den Faulprozess beeinflussen. Dazu zählen gesetzliche Feiertage in Österreich, die nicht zu vernachlässigende Anzahl der Nächtigungen in der Tourismusbranche im Einzugsgebiet der ARA sowie die ebenfalls saisonal schwankende Umgebungstemperatur.

3 Modelle und Methoden

Insgesamt zwei klassische Modelle aus der statistischen Datenanalyse sowie ein einfaches ANN liefern Referenzwerte für unseren Ansatz. Die Architektur des TFT und das Optimierungsverfahren werden beschrieben.

3.1 Benchmark-Modelle

1. Der bekannte k -NN Algorithmus ist eine datengetriebene Methode, um für einen Messwert, basierend auf dessen Nachbarn, eine Aussage treffen zu können. Er kann sowohl zur Klassifizierung als auch zur Regression verwendet werden. Bei Letzterem werden die Werte der k Nachbarn gemittelt. Ist der Hyperparameter $k \in \mathbb{N} > 0$ groß, wird das Ergebnis stark geglättet, bei zu kleinen Werten ist k -NN anfällig für Überanpassung (overfitting).
2. Speziell zur Analyse und Vorhersage von Zeitserien wird oft autoregressive integrated moving average (ARIMA) eingesetzt. Das Verhalten des Modells ist bestimmt durch drei nicht-negative ganzzahlige Parameter. Die Anzahl der Schritte in die Vergangenheit, der Grad der Subtraktionen und die Anzahl der Werte für das gleitende Mittel wird mit p , d bzw. q bezeichnet. ARIMA-Modelle werden direkt auf den Zeitserien trainiert und bedürfen keinerlei exogener Variablen.
3. Als letzten Benchmark trainieren wir ein ANN mit drei fully-connected Layer wie in Abu Qdais et al. (2010) beschrieben. Eine vereinfachte Dar-

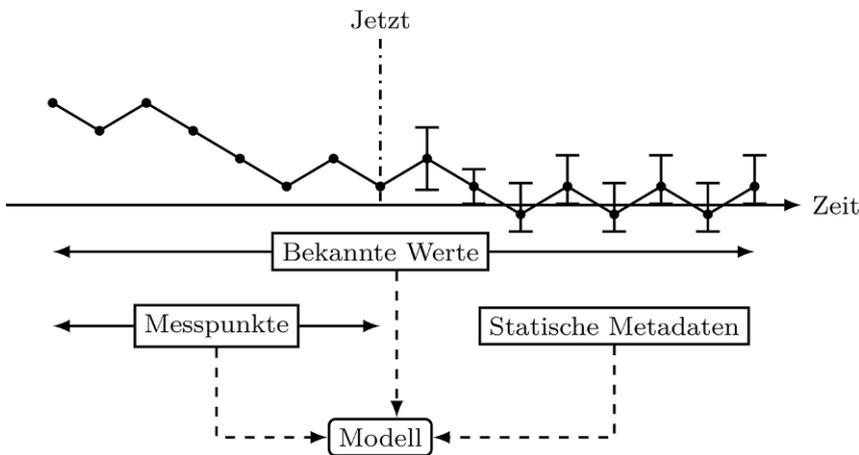


Abb. 3 Multi-Zeithorizont-Vorhersage einer Zeitserie mit einem komplexen Modell. Der Datensatz umfasst historische Werte (*Messpunkte*) sowie saisonale Features (*bekannte Werte*). Zusammen mit den unveränderlichen statischen Metadaten fließen beide als gekoppelte Information in das Modell mit ein (*gestrichelt*). Zusätzlich zu den Punktschätzungen wird die Dynamik des Systems mit Quantilen erfasst

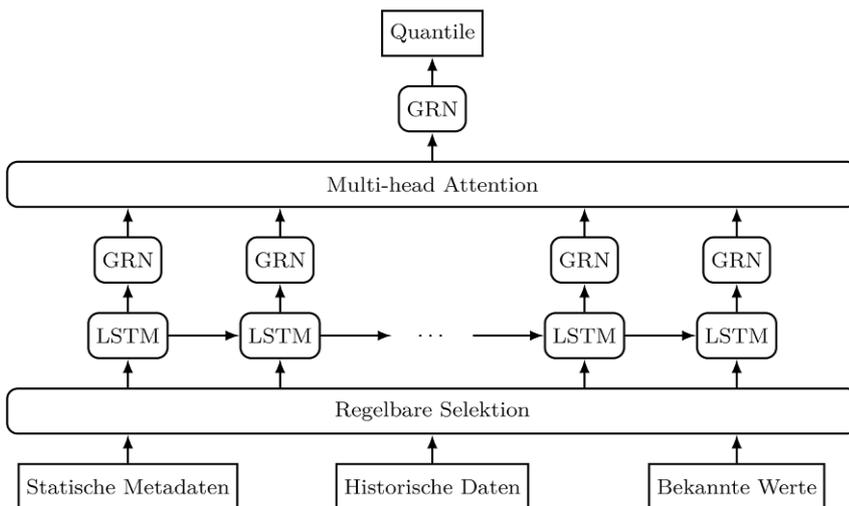


Abb. 4 Schematische Skizze der zugrundeliegenden Struktur eines TFT. Information fließt von unten nach oben. Modules sind mit abgerundeten, Input und Output mit eckigen Boxen gekennzeichnet

stellung der Architektur ist in Abb. 2 zu sehen. Das Potenzial von ANNs rührt von der großen Anzahl an zu optimierenden Parametern her. Zu den Nachteilen dieser Methode zählt die Undurchsichtigkeit des Entscheidungsprozesses in den einzelnen Knotenpunkten („black box“) und das oft langwierige Abstimmen der Hyperparameter wie Knotenanzahl, Nichtlinearitäten etc.

3.2 Der Temporal Fusion Transformer

Die Vorhersage der zukünftigen Entwicklung von Zeitserien basierend auf historischen Daten ist ein schon lange

bekanntes Problem. Anwendungsgebiete umfassen das Analysieren von Kundenströmen in Handel und Verkauf oder die (bisher erfolglosen) Versuche, Aktienkurse vorherzusagen. Meistens werden dabei sogenannte Punktschätzer eingesetzt. Diese Modelle verarbeiten entweder eine oder mehrere exogene Zeitserien und liefern anschließend einen einzigen Wert für den folgenden in der Zukunft liegenden Zeitschritt. Dadurch ergeben sich einige Nachteile, weil sie nicht entsprechend den folgenden drei Kriterien adaptiert werden können.

Information liegt in der Regel in verschiedenen Formen vor:

- **Zeitreihen:** eine Sammlung zeitabhängiger numerischer Werte, die gemessen und erfasst werden können (Firmengewinn, Flugpassagiere, Wetterdaten usw.).
- **Kategorische Features:** nicht-numerische Werte, die für das Modell relevant sein können (Augenfarbe, Güteklasse, Jahreszeit usw.).
- **Statische Metadaten:** sich in der Zeit nicht verändernde Merkmale in unterschiedlichen Formaten (Einwohnerzahl, geografische Lage, Verbindungen in einem Netzwerk usw.).

Klassische Modelle können ausschließlich numerische Merkmale verarbeiten, ohne sie mit nicht-numerischen zu verknüpfen. Durch eine Einbettung in einen entsprechenden Vektorraum ist es jedoch möglich, beide als Input zu verwenden und so vom zusätzlichen Informationsgewinn zu profitieren. Dabei wird der Vektor einer One-Hot-Kodierung $v \in \{0, 1\}$ mit einer kalibrierbaren Einbettungsmatrix $W \in \mathbb{R}^{K \times D}$ multipliziert. Der daraus resultierende Vektor vW dient als Modellinput. Weiters ermöglichen Multi-Zeithorizont-Vorhersagen, also das Schätzen einer relevanten Variablen für mehrere zukünftige Zeitschritte, höhere Planbarkeit über einen längeren Zeitraum. Modelle, die dazu in der Lage sind, finden speziell im industriellen Sektor Verwendung. Zuletzt sind Punktschätzungen zu restringiert, um Fluktuationen in zeitlich dynamischen Systemen ausreichend abbilden zu können. Stattdessen liefern Quantile, vergleichbar mit Konfidenzintervallen, Wertebereiche unterschiedlichen Ausmaßes, die es in der Praxis erlauben, formale Aussagen über eine spezifische Größe zu treffen (vgl. dazu Abb. 3).

Der TFT wurde speziell in Hinblick auf diese Voraussetzungen von Lim et al. (2019) entwickelt und ist daher auch für unseren konkreten Anwendungsfall, nämlich die Bestimmung der Produktionsrate von Biogas, geeignet. Eine schematische Darstellung der einzelnen Module sowie der Fluss von Information innerhalb der Modellarchitektur sind in Abb. 4 zu sehen.

3.2.1 Modellinput

Zu den kategorischen Features zählen der Wochentag, der Monat und die gesetzlichen Feiertage in Österreich. Wie beschrieben, werden diese Merkmale zuerst eingebettet und anschließend

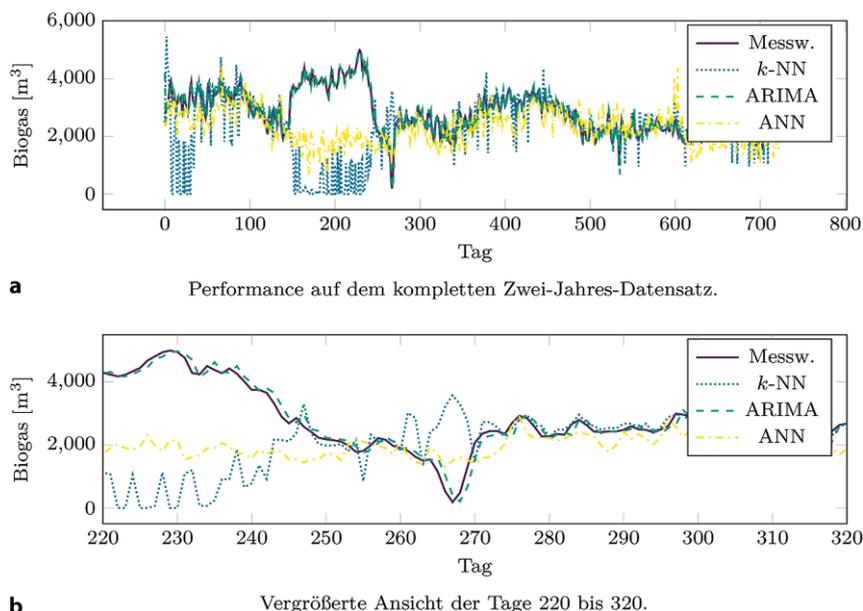


Abb. 5 Ergebnisse für die drei Benchmarks auf dem Testdatensatz, der aus den tatsächlichen Messwerten für das Biogas im zweiten Fermenter besteht

mit den exogenen Zeitserien reduziert über die Korrelationsanalysen (siehe Tab. 1) gemeinsam verarbeitet.

3.2.2 Regelbare Selektion

Besonders in der AG ist das Zusammenspiel der unterschiedlichen Komponenten nicht immer bekannt oder offensichtlich. Dieses Modul stuft die Inputparameter nach deren Wichtigkeit für die Vorhersage der Biogasproduktion ein. Es liefert somit wertvolle Erkenntnisse zum Entscheidungsprozess des TFT und dient dem besseren Verständnis des rechnerischen Vorgangs.

3.2.3 Long Short-Term Memory (LSTM)

Entwickelt von Hochreiter and Schmidhuber (1997) statten die LSTM-Module den TFT mit einem „Gedächtnis“ aus. Sogenannte Gates regulieren den Informationsfluss aus der Vergangenheit. Signifikante Events, wie z.B. ein Temperatursturz im Faulturm, der schon einige Tage zurückliegt, wirken sich damit nach wie vor auf die Vorhersagen aus. Umgekehrt vergisst der TFT Information wenn sie einen vom Modell bestimmten Schwellenwert nicht passieren kann und damit als nicht relevant eingestuft wird.

3.2.4 Gated Residual Network (GRN)

Im ML-Bereich sind Modelle mit zu vielen Veränderlichen oft überangepasst an die Daten. Die GRNs kontrollieren, ob aufgrund der Komplexität der gelieferten Information das Verarbeiten in zusätzlichen Layers notwendig ist. Falls nicht, können diese einfach übersprungen werden und die Daten passieren das GRN unverändert. Der TFT reguliert sich somit bei Bedarf von selbst.

3.2.5 Multi-head Attention

Ursprünglich wurde „attention“ für die Computerlinguistik entwickelt, um in einzelnen Sätzen Wortzusammenhänge zu erkennen. Der TFT analysiert damit verschiedene wiederkehrende zeitliche Muster. Weiters kann als Folge dessen bildlich veranschaulicht werden, auf welche Zeitschritte der TFT besonderes Augenmerk legt.

3.2.6 Output und Quantilsregression

Zusätzlich zu den Punktschätzungen für das Biogas liefert der TFT die 2%-, 10%-, 25%-, 75%-, 90%- und 98%-Quantile. Diese Form der Regression ist robuster gegenüber starken Schwankungen in den Daten und bietet umfangreichere Prognosen.

Da auf der ARA zwei baulich identische Fermenter in Betrieb sind, ergibt sich die Aufspaltung in einen Trainings-

und Testdatensatz auf natürliche Weise. Die historischen Daten der vergangenen sieben Tage werden verwendet, um eine Vorhersage für die folgende Woche zu erhalten. Insgesamt 546.924 Modellparameter werden mit einem verbesserten Gradientenabstiegsverfahren kalibriert. Ist ein Minimum erreicht, wird die Schrittweite automatisch verkleinert. Sofern sich die Kostenfunktion über mehrere Gradientenschritte nicht mehr verbessert, wird das Training abgebrochen.

4 Resultate und Diskussion

Der Code, mit dem die folgenden Ergebnisse generiert wurden, ist online frei verfügbar (<https://github.com/jsappl/OptiFaul>). Die Entwicklung der Kostenfunktion während des Optimierungsverfahrens sowie weitere grafische Auswertungen auf dem Testdatensatz können als TensorBoard¹ interaktiv eingesehen werden.

In Abb. 5 wird die Performance der Benchmark-Modelle auf dem Testdatensatz geplottet. Der Input von k -NN sowie des ANN entspricht beinahe jenem des TFT. Wochentag und Monat werden in ganzen Zahlen ausgedrückt. Einzig die gesetzlichen Feiertage können die Modelle als rein kategorisches Merkmal nicht verarbeiten und wurden daher vernachlässigt. Das ARIMA-Modell benötigt keine exogenen Variablen. Es wird direkt auf den Biogasmessdaten trainiert. Der k -NN Algorithmus mit zwei Nachbarn erfasst die Dynamik ab Tag 250 relativ gut, tendiert jedoch dazu, die tagesaktuellen Werte weit zu unter- bzw. überschätzen. Im ersten Drittel scheitert er völlig an der Vorhersage, da nur die Daten aus dem ersten Fermenter zur Verfügung stehen. Zu eben diesem Zeitpunkt kam die Biogasproduktion im ersten Faulturm aber zum Erliegen. Das ARIMA-Modell scheint auf den ersten Blick nahezu perfekte Vorhersagen zu liefern. Bei genauerer Betrachtung stellt sich jedoch heraus, dass es nur gelernt hat, den Wert vom Vortag als Prediktion zu übernehmen. Die wahren Messwerte werden bloß kopiert und um einen Tag in die Zukunft verschoben. Daraus können keine qualitativen Aussagen über etwaige Veränderungen im Fermenter für die Folgetage getroffen werden. Das

¹ <https://tensorboard.dev/experiment/3KVF0zqmSpGbgPxtteeWJDQ>.

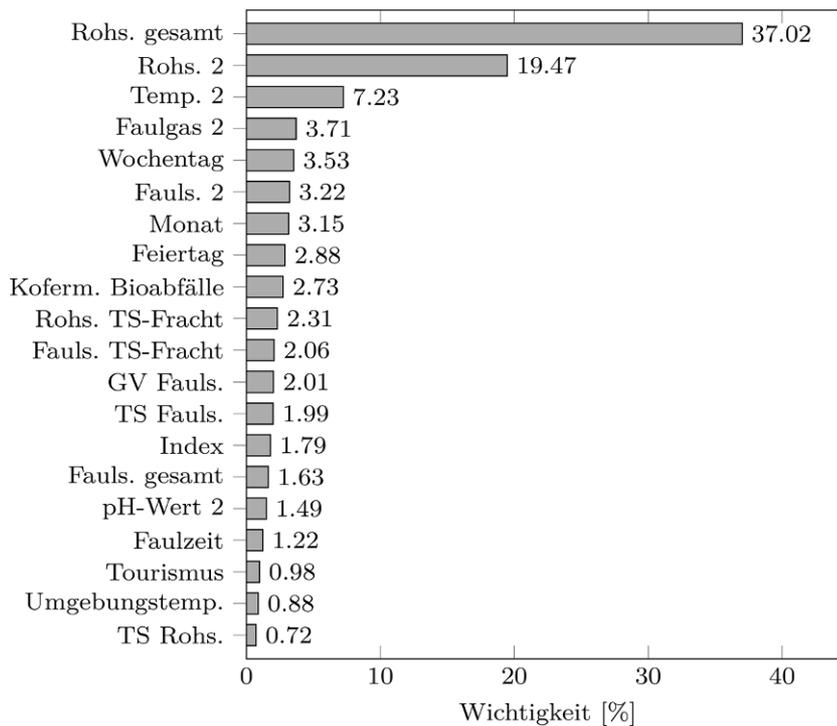


Abb. 6 Regelbare Selektion ermöglicht es dem TFT, einzelne Variable verschieden zu gewichten. Ihre Bedeutung in Prozent für die Vorhersage der Biogasproduktion lässt genauere Rückschlüsse auf die Entscheidungen des Modells zu

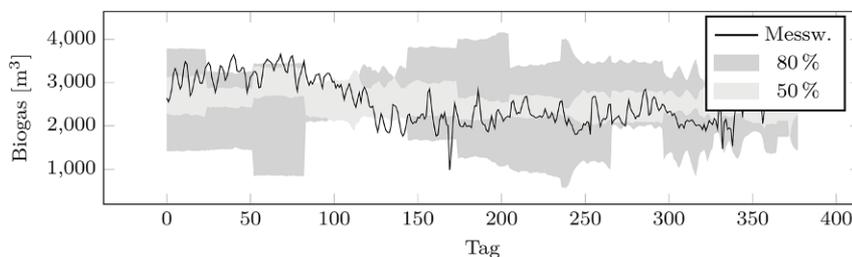


Abb. 7 Darstellung der Bereiche zwischen Q_{25} und Q_{75} , die 50 %, sowie zwischen Q_{10} und Q_{90} , welche 80 % der Daten enthalten. Die real gemessenen Werte liegen nur an wenigen Tagen außerhalb des schraffierten Bereichs

ANN performt im Gegensatz zu k -NN etwas besser bis Tag 250, unterschätzt aber nach wie vor die Biogasrate um mehr als 2000 m^3 . Auffällig sind die starken Tagesschwankungen, wobei der allgemeine Trend dennoch relativ gut vom ANN abgebildet wird.

Die oben beschriebene Regelbare Selektion des TFT erlaubt eine grafische Darstellung der Einstufung der Inputparametern (siehe Abb. 6). Je nachdem, wie viel Einfluss die Variablen auf die Biogasproduktion haben, werden sie vom TFT-Modell unterschiedlich gewichtet. Zusätzlich erfüllt die Regelbare Selektion eine Kontrollfunktion über den TFT. Der Einblick in diese Reihung erlaubt es den Betreibern der ARA zu

beurteilen, inwiefern sich die zugrunde liegenden biochemischen Prozesse darin widerspiegeln. Gegebenenfalls können so bei nicht sinnvollen Konstellationen die Hyperparameter, im Unterschied zu „black box“ Ansätzen, angepasst und das Modell neu antrainiert werden.

Da der TFT eine Prognose für die kommenden sieben Tage liefert, ist eine direkte Gegenüberstellung mit den Benchmarks, welche nur Punktschätzungen abgeben, schwierig. Stattdessen vergleichen wir die Vorhersagen direkt mit den tatsächlichen Messwerten von Biogas. Das Modell erstellt die täglichen Intervalle zwischen Q_{25} und Q_{75} , die 50 %, sowie zwischen Q_{10} und Q_{90} ,

welche 80 % der Daten enthalten (siehe Abb. 7). Schwankungen in der Biogasproduktion können somit bei der Vorhersage besser abgefangen werden. Nur vereinzelt liegen die realen Messwerte des Testdatensatzes außerhalb des 80%-Bereichs. Dabei handelt es sich teilweise um klare Ausreißer, wie z. B. an Tag 160 ersichtlich. Die Biogasmenge fiel in diesem Bereich innerhalb kürzester Zeit um ca. 1000 m^3 .

5 Schlussfolgerung

Angesichts der aktuellen Klimakrise gewinnt die Erschließung und Optimierung von erneuerbaren Energiequellen immer mehr an Bedeutung. Die Stromerzeugung mittels Biogas, welches in Faultürmen auf ARAs gewonnen wird, kann dazu einen wichtigen Beitrag leisten. Nichtsdestotrotz sind viele Fragen zum biochemischen Vorgang selbst noch ungeklärt. Auch besteht noch kein allgemein gültiger Konsens darüber, welche technischen Umsetzungen den meisten Ertrag liefern. Das betrifft vor allem die Geometrie eines Faulturms sowie, falls überhaupt notwendig, die Methode der Wahl zur Umwälzung des Faulschlammes. Um diese Probleme adäquat adressieren zu können, werden Modelle benötigt, welche eine Zustandsbewertung der AG im Fermenter erlauben. Der TFT verarbeitet Zeitserien zusammen mit nicht-numerischen Merkmalen und ist daher in der Lage, genauere Prognosen als klassische Modelle zu erstellen. Neben den bekannten Vorteilen datengetriebener Methoden bietet der TFT auch verbesserte Möglichkeiten zur Interpretation der Vorhersagen. Die Machbarkeit und Genauigkeit unseres Ansatzes wurde anhand eines realen Beispiels gezeigt.

Danksagung Wir bedanken uns beim Abwasserverband AIZ für den Datensatz. Eine Titan X GPU wurde uns von der NVIDIA Corporation kostenlos zur Verfügung gestellt.

Förderung Diese Forschungsarbeit ist Teil des OptiFaul-Projekts B801259, das vom Österreichischen Bundesministerium für Landwirtschaft, Regionen und Tourismus in Zusammenarbeit mit der Kommunalkredit Public Consulting GmbH gefördert wird.

Funding Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und

angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betref-

fende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>. ■

Literatur

- 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. *Resources, Conservation and Recycling* 54(6):359–363, <https://doi.org/10.1016/j.resconrec.2009.08.012>
- AIZ (2020): Kläranlage. URL <http://www.aiz.at/CMS/index.php/ka-technodaten>. Zugegriffen: 10.05.2021
- Batstone, D., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H., Vavilin, V. (2002): *Anaerobic Digestion Model No.1 (ADM1)*, 1st edn. Scientific and Technical Report Series, IWA Publishing, London
- Cakmakci, M. (2007): Adaptive neuro-fuzzy modelling of anaerobic digestion of primary sedimentation sludge. *Bioprocess and Biosystems Engineering* 30(5):349–357, <https://doi.org/10.1007/s00449-007-0131-2>, URL <https://link.springer.com/article/10.1007/s00449-007-0131-2>
- De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., Wen, Z., Caicedo, L., Yuan, K. (2019): Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. *Journal of Cleaner Production* 218:390–399, <https://doi.org/10.1016/j.jclepro.2019.01.031>
- De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., Shang, R. (2020): Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. *Science of the Total Environment* 712:134574, <https://doi.org/10.1016/j.scitotenv.2019.134574>
- Dhussa, A.K., Sambhi, S.S., Kumar, S., Kumar, S., Kumar, S. (2014): Nonlinear Autoregressive Exogenous modeling of a large anaerobic digester producing biogas from cattle waste. *Bioresour. Technol.* 170:342–349, <https://doi.org/10.1016/j.biortech.2014.07.078>
- Feigl, M., Lebidzinski, K., Herrnegger, M., Schulz, K. (2021): Vorhersage der Fließgewässertemperaturen in österreichischen Einzugsgebieten mittels Machine Learning-Verfahren. *Österreichische Wasser- und Abfallwirtschaft* 73(7–8). <https://doi.org/10.1007/s00506-021-00770-4>
- Gaida, D., Wolf, C., Meyer, C., Stuhlsatz, A., Lippel, J., Bäck, T., Bongards, M., McLoone, S. (2012): State estimation for anaerobic digesters using the ADM1. *Water Science and Technology* 66(5):1088–1095, <https://doi.org/10.2166/wst.2012.286>, URL <http://iwaponline.com/wst/article-pdf/66/5/1088/443169/1088.pdf>
- Hochreiter, S., Schmidhuber, J. (1997): Long Short-Term Memory. *Neural Computation* 9(8):1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>, URL <https://www.mitpressjournals.org/doi/abs/10.1162/neco.1997.9.8.1735>
- 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(10):2582–2588, [https://doi.org/10.1016/S0043-1354\(01\)00487-0](https://doi.org/10.1016/S0043-1354(01)00487-0)
- Kanat, G., Saral, A. (2009): Estimation of Biogas Production Rate in a Thermophilic UASB Reactor Using Artificial Neural Networks. *Environmental Modeling and Assessment* 14(5):607–614, <https://doi.org/10.1007/s10666-008-9150-x>, URL <https://link.springer.com/article/10.1007/s10666-008-9150-x>
- Kusiak, A., Wei, X. (2014): Prediction of methane production in wastewater treatment facility: a data-mining approach. *Annals of Operations Research* 216(1):71–81, <https://doi.org/10.1007/s10479-011-1037-6>
- Lim, B., Arik, S.O., Loeff, N., Pfister, T. (2019): Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. arXiv e-prints URL <http://arxiv.org/abs/1912.09363>. Zugegriffen: 10.05.2021
- Rustum, R., Adeloje, A.J. (2007): Replacing Outliers and Missing Values from Activated Sludge Data Using Kohonen Self-Organizing Map. *Journal of Environmental Engineering* 133(9):909–916, [https://doi.org/10.1061/\(asce\)0733-9372\(2007\)133:9\(909\)](https://doi.org/10.1061/(asce)0733-9372(2007)133:9(909))
- Tufaner, E., Demirci, Y. (2020): Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. *Clean Technologies and Environmental Policy* 22(3):713–724, <https://doi.org/10.1007/s10098-020-01816-z>, URL <https://doi.org/10.1007/s10098-020-01816-z>
- Turkdogan-Aydinol, F.I., Yetilmezsoy, K. (2010): A fuzzy-logic-based model to predict biogas and methane production rates in a pilot-scale mesophilic UASB reactor treating molasses wastewater. *Journal of Hazardous Materials* 182(1–3):460–471, <https://doi.org/10.1016/j.jhazmat.2010.06.054>
- Wade, M.J. (2020): Not Just Numbers: Mathematical Modelling and Its Contribution to Anaerobic Digestion Processes. *Processes* 8(8):888, <https://doi.org/10.3390/pr8080888>

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.