



Research article

Improving ADM1 predictions via Bayesian analysis for continuous anaerobic digestion

Yanxin Liu^{a,b}, Ying Jiang^a, Nasreen Nasar^a, Yadira Bajon-Fernandez^a, Philip Longhurst^a, Weisi Guo^a, Mei Lei^c, Immacolata Bortone^{a,*}

^a Faculty of Engineering and Applied Sciences, Cranfield University, College Road, Cranfield, MK43 0AL, UK

^b Faculty of Environment, Science and Economy, University of Exeter, Stocker Road, Exeter, EX4 4PY, UK

^c Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

ARTICLE INFO

Keywords:

Anaerobic digestion
ADM1
Steady-state performance
Global sensitivity analysis
Bayesian calibration

ABSTRACT

The Anaerobic Digestion Model No.1 (ADM1) application for continuous anaerobic digestion is often constrained by challenges in reliably calibrating model parameters, especially when long-term data are unavailable. This study presents a Bayesian inference-based framework that enables ADM1 calibration using only initial-stage digester performance data. A custom Python implementation was developed, integrating modules for global sensitivity analysis, Bayesian calibration and parameter identifiability evaluation. Key microbial and ionic parameters were refined through Random Balance Designs–Fourier Amplitude Sensitivity Test (RBD-FAST), identifying sugar/acetate degraders and cation/anion levels in the inoculum as critical drivers of steady-state performance.

With informative priors derived from similar ADM1 studies, the model was calibrated with less than two hydraulic retention times of data and validated against steady-state performance data. It predicted pH and total chemical oxygen demand (tCOD) with mean percentage errors of 1.10 % and 5.38 % respectively. Biogas production trends were captured within the 95 % credible interval for 63.14 % of observations. Compared to uniform priors, the Bayesian approach with informative priors improved predictive accuracy. Jensen-Shannon divergence revealed that hydrolysis rates is the most identifiable for thermally hydrolysed sludge.

Unlike conventional ADM1 calibration approaches that require long-term steady-state data, this Bayesian framework achieves reliable predictions using only early-stage observations. By enabling accurate simulation of organic contaminant degradation and system stability from limited data, the framework supports risk-informed design and operation of anaerobic digesters and offers a solution for data-scarce industrial settings to improve safety, sustainability, and optimisation.

1. Introduction

Anaerobic digestion (AD) is a process in which biogas is produced through the microbial degradation of organic materials under anaerobic conditions (Rocamora et al., 2020). AD resulting biogas typically contains over 60 % methane (Shen et al., 2013), positioning this process as a significant contributor to the goal of meeting 27 % of total energy demand with renewable sources by 2030 (European Council, 2018). AD feedstock can be a diverse range of organic waste materials, e.g., manure, crop residues, sewage sludge and food waste. As a technology

that provides solutions for both organic waste treatment and renewable energy production, AD has received increased investment in the past decades. This is highlighted by the growth of the number of AD plants in the EU from 6507 plants in 2009 to 18,843 in 2021 (European Biogas Association, 2022).

AD process modelling offers a cost-effective and time-efficient alternative to extended pilot trials, supporting operator decision-making and process optimisation. By simulating diverse operational scenarios, AD models enable evaluation of digester performance under varying conditions, facilitating improved process control and

* Corresponding author.

E-mail addresses: Y.Liu13@exeter.ac.uk (Y. Liu), Ying.Jiang@tii.ae (Y. Jiang), Nasreen.Nasar@cranfield.ac.uk (N. Nasar), y.bajonfernandez@cranfield.ac.uk (Y. Bajon-Fernandez), P.J.Longhurst@cranfield.ac.uk (P. Longhurst), Weisi.Guo@cranfield.ac.uk (W. Guo), leim@igsrr.ac.cn (M. Lei), imma.bortone@cranfield.ac.uk (I. Bortone).

<https://doi.org/10.1016/j.jenvman.2025.128537>

Received 25 July 2025; Received in revised form 26 November 2025; Accepted 31 December 2025

Available online 6 January 2026

0301-4797/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

troubleshooting strategies (Donoso-Bravo et al., 2011). The Anaerobic Digestion Model No.1 (ADM1), developed by the AD modelling task group of the International Water Association, is the industry standard for simulating continuous AD processes.

ADM1 comprehensively captures the sequential biological and physiochemical reactions involved in AD, ensuring mass balance of key stoichiometric parameters including chemical oxygen demand (COD), carbon (C), and nitrogen (N) (Batstone et al., 2002). Its versatility has been demonstrated in full-scale applications, including the optimisation of biogas production from diverse feedstocks, such as maize silage, pig manure, cattle manure, and digestate (Białobrzewski et al., 2024). More recently, ADM1 has proven valuable for investigating AD systems operating under atypical conditions, such as auto-generative or high-pressure environments, where its modelling framework enables deeper insights into kinetic behaviour (De Crescenzo et al., 2024).

At its core, ADM1 describes time-dependent concentration changes of reactants and products (referred to as state variables) through a system of ordinary differential equations (ODEs). Key operational parameters, such as organic loading rate (OLR) and hydraulic retention time (HRT), are explicitly embedded within its framework, allowing simulations to capture the dynamic interplay between process inputs and outputs.

While ADM1 exemplifies a mechanistic modelling approach, alternative data-driven frameworks have also been explored for AD performance prediction. These include long short-term memory networks enhanced with synthetic minority oversampling techniques (Han et al., 2023), three-layer feedforward backpropagation neural networks (Antwi et al., 2017), genetic algorithm-optimised neural networks (Qdais et al., 2010), and adaptive network-based fuzzy inference systems (Asadi et al., 2020). Such models are comparatively easier to implement but rely heavily on large, high-quality training datasets. Consequently, their predictive accuracy is often limited to the specific operational conditions and feedstock characteristics represented in the training data, reducing their generalisability across diverse AD systems. This is particularly valuable for wastewater utilities commissioning new digesters, where early-stage calibration reduces commissioning risks and operational uncertainty.

As a process-based biochemical model, ADM1 incorporates more than fifty kinetic and stoichiometric parameters (θ_{ks}) (Rosen and Jeppsson, 2006). Depending on the characteristics of the substrate used, different set of θ_{ks} may exert significant influence on simulation outcomes, reflecting the digester performance. Therefore, it is essential to first identify the most sensitive θ_{ks} , such as those with the greatest impact on model outputs and subsequently calibrate their values with experimental observations. Reliable prediction of digester performance under various operational conditions is only possible once such AD parameter calibration is achieved.

In addition to θ_{ks} parameter calibration, the initial values (θ_0) of ADM1 state variables must be defined to initiate simulation. θ_0 values represent the digester's initial environmental conditions and are primarily influenced by the inoculum characteristics. A common approach for determining θ_0 involves 'spinning up' the model under constant substrate inflow until the state variables reach stabilisation, with the resulting steady-state concentrations then used as initial conditions for subsequent simulations aimed at process optimisation (Blumensaat and Keller, 2005; Boubaker and Ridha, 2008; Catenacci et al., 2021; Ozgun, 2019). This method is based on the rationale that, under prolonged steady-state operation, the influence of the inoculum diminishes, and chemical concentrations stabilise. However, it requires observational data from digester steady-state conditions for θ_{ks} calibration, restricting ADM1's applicability to systems with long operational histories (Ruffino et al., 2015; Razaviarani and Buchanan, 2015). In contrast, this study aims to demonstrate that ADM1 can be reliably calibrated using minimal data from the initial stage of digester operation. By integrating Bayesian inference with targeted sensitivity analysis, the proposed framework enables prediction of steady-state performance and system behaviour

requiring long-term observational datasets, therefore expanding ADM1's applicability to data-constrained industrial contexts.

An alternative approach to determining θ_0 involves chemical analysis of inoculum characteristics, as demonstrated in several studies (Pastor-Poquet et al., 2019; Thamsirirot and Murphy, 2011). However, a practical limitation of this approach is the inability to accurately quantify the initial values of certain state variables, particularly the microbial group concentrations (e.g., sugar degraders, amino acid degraders), due to the lack of standardised measurement techniques. A potential solution is to calibrate these hard-to-measure initial values (θ_0) alongside the most sensitive model parameters (θ_{ks}), a practice well-established in small-scale ODE models such as the Lotka-Volterra system (McElreath, 2018). While theoretically valid, this approach increases the dimensionality of the calibration task and poses challenges when observational data are limited.

To address this, the present study implemented a sequential down-selection of the most influential kinetic parameters (θ_{ks}) and uncertain initial conditions (θ_0) for calibration. The process began with a pre-screening using the Morris method, followed by a variance-based global sensitivity analysis (GSA). Bayesian inference was then applied to calibrate the reduced parameter set. By incorporating informative prior distributions derived from an extensive literature review, the Bayesian framework enabled robust estimation of uncertain inputs under data-sparse conditions. Unlike conventional ADM1 calibration techniques, such as heuristic manual tuning or least squares optimisation which yield only local point estimates (Catenacci et al., 2021), Bayesian inference provides full posterior distributions that quantify parameter uncertainty (Gelman et al., 2013). This uncertainty is propagated through model predictions, supporting more risk-informed decision-making. Using this approach, the steady-state performance of a semi-continuous anaerobic digester treating sewage sludge was predicted using experimental data spanning less than two hydraulic retention times (HRTs).

2. Methodology

2.1. Experimental data

In this study, AD data were obtained from a laboratory-scale semi-continuous anaerobic digester treating sewage sludge pre-treated via thermal hydrolysis process (THP). The reactor had a working volume of 5.20 L and headspace 3.22 L, operating at 39.5 °C, a temperature characteristic of advanced anaerobic digestion systems designed to enhance microbial activity and process efficiency (Chew et al., 2021). The inoculum used in the digester was the digestate collected from a full-scale anaerobic digester processing THP sewage sludge at a domestic wastewater treatment plant (WWTP) in Milton Keynes, UK. The volatile solid (VS) content of the sludge feedstock over the course of the experiment was 0.76 kg m⁻³ ($\pm 11.82\%$). The daily feed flow was 2.06×10^{-4} m³/day ($\pm 11.5\%$) to maintain a stable OLR of 3 kg VS m³/day. The average HRT of this anaerobic digester was 25.24 day.

Experimental observations of daily biogas flow (m³/d) and pH, total ammoniacal nitrogen (TAN), total COD (tCOD) of digestate from a period of 323 days' operation were considered for the ADM1 calibration and validation. Biogas production was recorded automatically using tipping-bucket gas counters, as described in Walker et al. (2009). Outliers, defined as data points beyond three standard deviations, were removed. To smooth trends and address gas counter failures, a 10-day moving average of biogas flow was calculated. Digestate pH and TAN were monitored from day 1–50 and day 290–313, while the tCOD was only monitored from day 290–313. The operation data during digester acclimatisation stage (before day 50), were used for model calibration, while the remaining data were reserved for validation of the modelled results.

Table 1
Inoculum characteristics.

Components	Unit	Values
Total solids (TS)	kg/m ³	52.40
Volatile solids (VS)	kg/m ³	32.90
Density	kg/m ³	1
tCOD	kg/m ³	55.47
sCOD	kg/m ³	3.99
Total soluble sugars	g/100g dry matter (DM)	<0.1 (0.05 adopted)
Crude protein	g/100g DM	30.90 (Dumas Method)
Free long-chain fatty acids (LCFA)	g/100g DM	<0.5 (0.30 adopted)
Total carbohydrate	g/100g DM	<1 (0.50 adopted)
Total fat	g/100g DM	9.80
Valeric acid	kg/m ³	0.59
Butyric acid	kg/m ³	0.02
Propionic acid	kg/m ³	0.04
Acetic acid	kg/m ³	0.006
Total ammoniacal nitrogen (TAN)	kmole N/m ³	0.19
Total carbon (TC)	kmole C/m ³	1.52
Total organic carbon (TOC)	kmole C/m ³	1.49
Lignin	g/100g DM	7.16

2.2. Python-based ADM1 implementation

The ADM1 model comprises a set of differential equations (ODEs) that describe the temporal evolution of key chemical and biological species within the anaerobic digestion process. In this study, these equations were implemented in a custom Python framework to simulate 36 state variables (Table 2) (Rosen and Jeppsson, 2006), with the reaction and interaction terms structured via a Petersen matrix reported in Appendix A of Supplementary Material (SM).

Specifically, S_{H^+} evolution in this study was also described by an ODE (SM: Appendix A, Equation A-1), which is derived from the charge balance of ionic constituents reported in Equation (1), following the approach of Thamsiroj and Murphy (2011). This approach improves numerical stability and allows for the use of robust ODE solvers available in Python's scientific ecosystem (Sadrimajd et al., 2021).

$$S_{H^+} + S_{NH_4^+} - S_{HCO_3^-} - \frac{S_{ac^-}}{64} - \frac{S_{pro^-}}{112} - \frac{S_{bu^-}}{160} - \frac{S_{va^-}}{208} - S_{OH^-} - S_{an^-} = 0 \quad (\text{Equation 1})$$

The parameters used in Equation (1), including ionic species involved in the charge balance, correspond to those listed in Table 2. The simulations were performed using the backward differentiation formula (BDF) solver available in the SciPy package (Jones et al., 2001).

2.3. Inflow values (θ_f) and initial values (θ_0) of AD model state variables

To determine the inflow values (θ_f) of the process state variables, the transformer tool developed in Simulink (MathWorks, 2022) by Zaher et al. (2009) was employed. This tool estimates AD equations inputs values from basic feedstock characteristics, based on elemental mass balances, theoretical COD, charge neutrality and carbon covalent bonding principles.

θ_f define the composition of the substrate entering the digester and serve as essential boundary conditions for simulating substrate degradation, microbial activity, and biogas production within the ADM1 framework. The required feedstock characteristics and the resulting ADM1 input values generated by the Transformer tool are provided in SM: Appendix B.

Notably, the lipid fraction (X_{li}) in the substrate was estimated to be 0 kg COD/m³, consistent with the findings of Pilli et al. (2015), reporting that lipids are typically not a dominant component in thermally pre-treated sludge, whilst proteins and carbohydrates constitute its

Table 2
Initial values of state variables (θ_0) and their derivation from inoculum characteristics.

No.	Name	Description	Unit	Values	Estimated from
1	S_{su}^0	monosaccharides	kg	0.028	Total soluble sugars
2	S_{aa}^0	amino acid	COD/m ³	0.464	Crude Protein (Dumas)/6.25 × (1–8.54 %) × 17.75 %
3	S_{fa}^0	long-chain fatty acid		0.016	Free LCFA
4	S_{va}^0	total valerate	kg	0.006	VFA analysis
5	S_{bu}^0	total butyrate	COD/m ³	0.064	
6	S_{pro}^0	total propionate		0.037	
7	S_{ac}^0	total acetate		0.121	
8	$S_{H_2}^0$	hydrogen in liquid phase	kg COD/m ³	SA	/
9	$S_{CH_4}^0$	methane in liquid phase		SA	/
10	S_{IC}^0	inorganic carbon	kmole C/m ³	0.036	Total Inorganic Carbon (TIC)
11	S_{IN}^0	inorganic nitrogen	kmole N/m ³	0.190	TAN
12	S_I^0	soluble inert	kg COD/m ³	3.213	$sCOD - S_{su}^0 - S_{aa}^0 - S_{fa}^0 - S_{VFA}^0$ (No. 4–7) - $S_{H_2}^0 - S_{CH_4}^0$
13	X_c^0	particulate composites		SA	/
14	X_{ch}^0	carbohydrates		7.898	Total carbohydrate – Total soluble sugar
15	X_{pr}^0	proteins		2.044	Crude Protein (Dumas)/6.25 × (1–8.54 %) × 82.25 %
16	X_{li}^0	lipids		5.419	Total fat – Free LCFA
17	X_{su}^0	sugar degraders	kg COD/m ³	SA	$X_{bac}^0 = \mu_{max} \text{ pCOD} - X_c^0 - X_{ch}^0 - X_{pr}^0 - X_{li}^0 - X_I^0 = 0.1$
18	X_{aa}^0	amino acid degraders			$\mu_{max} (X_{aa}) = 0.08$
19	X_{fa}^0	LCFA degraders			$\mu_{max} (X_{fa}) = 0.06$
20	X_{c4}^0	valerate and butyrate degraders			$\mu_{max} (X_{c4}) = 0.06$
21	X_{pro}^0	propionate degraders			$\mu_{max} (X_{pro}) = 0.04$
22	X_{ac}^0	acetate degraders			$\mu_{max} (X_{ac}) = 0.05$
23	$X_{H_2}^0$	hydrogen degrader			$\mu_{max} (X_{H_2}) = 0.06$
24	X_I^0	particulate inert	kg COD/m ³	3.971 (SA)	Lignin
25	S_{cat}^0	cations	kmole/m ³	SA	/
26	S_{an}^0	anions		SA	/
27	S_{va}^0	base form valerate	kg COD/m ³	0.00599	$pK_a = 4.86$ (Sillen and Martell, 1964)
28	S_{bu}^0	base form butyrate		0.06393	$pK_a = 4.86$ (Sillen and Martell, 1964)
29	S_{pro}^0	base form propionate		0.03696	$pK_a = 4.88$ (Sillen and Martell, 1964)
30	S_{ac}^0	base form acetate		0.12089	$pK_a = 4.76$ (Sillen and Martell, 1964)

(continued on next page)

Table 2 (continued)

No.	Name	Description	Unit	Values	Estimated from
31	$S_{HCO_3^-}^0$	bicarbonate	kmole C/m ³	0.0345	pK _a = 6.35 (Lide, 2001)
32	$S_{NH_3}^0$	ammonia in liquid phase	kmole N/m ³	0.00647	pK _a = 9.25 (for NH ₄ ⁺) (Lide, 2001)
33	S_{gas,H_2}^0	hydrogen in gas phase	kg COD/m ³	1.588e-5	Manual fitting with reference to benchmark value
34	S_{gas,CH_4}^0	methane in gas phase	m ³	1.30022	
35	S_{gas,CO_2}^0	carbon dioxide in gas phase	kmole C/m ³	0.01591	
36	$S_{H^+}^0$	hydrogen ion	kmole H ⁺ /m ³	1.585e-8	pH

Note: The superscript '0' indicates initial values for state variables. The values for state variables No. 27–36 were reported to five decimal places to ensure work reproducibility.

Table 3

Variability ranges of uncertain values of θ_{ks} and θ_0 used in the sensitivity analysis. 'U' represents uniform distribution.

Name	θ_{ks}			θ_0		
	Description	Unit	Range	Name	Unit	Range
k_{dis}	Disintegration rate	d ⁻¹	U(0.101, 5.6)	$S_{H_2}^0$		U(0, 4.2 × 10 ⁻⁷)
$k_{hyd,ch}$	Carbohydrate hydrolysis rate		U(0.092, 10)	$S_{CH_4}^0$		U(10 ⁻⁸ , 0.055)
$k_{hyd,pr}$	Protein hydrolysis rate		U(0.03, 10)	X_c^0		U(6.648, 26.648)
$k_{hyd,li}$	Lipid hydrolysis rate		U(0.05, 10)	X_{su}^0		U(1.348 ÷ 5.792)
$k_{m,pro}$	Maximum uptake rate of propionate degraders		U(6, 20)	X_{aa}^0		U(1.079, 4.635)
$k_{m,ac}$	Maximum uptake rate of acetate degraders		U(4.57, 12)	X_{fa}^0	kg COD/m ³	U(0.809, 3.475)
k_{m,H_2}	Maximum uptake rate of H ₂ degraders		U(20, 35)	X_{c4}^0		U(0.809, 3.457)
				X_{pro}^0		U(1.348, 5.792)
$K_{S,pro}$	Half saturation coefficient of propionate uptake	kg COD/m ³	U(0.1, 0.2)	X_{ac}^0		U(0.809, 3.475)
$K_{S,ac}$	Half saturation coefficient of acetate uptake		U(0.1, 0.25)	$X_{H_2}^0$		U(0.809, 3.457)
K_{S,H_2}	Half saturation coefficient of H ₂ uptake		U(7 × 10 ⁻⁶ , 9 × 10 ⁻⁵)	X_l^0		U(3, 20)
K_{I,NH_3}	50 % inhibitory concentration of free NH ₃ to acetate uptake		U(0.0011, 0.00234)			
				S_{cat}^0	kmole/m ³	U(0.01, 0.166)
				S_{an}^0		U(0.02, 0.6)

primary biodegradable fractions.

As discussed earlier, most initial values θ_0 of ADM1 state variables were derived from a detailed physiochemical characterisation of the inoculum (Table 1). Table 2 presents all θ_0 values considered in the

model, along with the corresponding derivation methods, including analytical estimation, literature-informed assumptions, or calibration supported by sensitivity analysis, accordingly, indicated with SA, where direct measurement was not feasible.

Soluble sugars and carbohydrates were converted to COD equivalents using the default ADM1 molecular formulas C₆H₁₂O₆ and C₆H₁₀O₅ (Batstone et al., 2002), respectively. The amino acid (S_{aa}^0) and proteins (X_{pr}^0) fractions were estimated from the total N content (both organic and inorganic forms) determined via the Dumas' method (Nielsen, 2014). Specifically, 8.54 % of N was assumed to be inorganic based on the empirical value for wastewater sludge (Song et al., 2019), whilst, the remaining N was allocated as S_{aa}^0 at 17.75 %, based on the average values from three references (Pehlivanoglu and Sedlak, 2004; Song et al., 2019; Westgate and Park, 2010), and as X_{pr}^0 at 82.25 %. The initial values of S_{gas,H_2} , S_{gas,CH_4} , S_{gas,CO_2} were obtained by fitting reference values from the benchmark model (Rosen and Jeppsson, 2006) to the initial daily biogas production data. Total concentrations of valerate (S_{va}^0), butyrate (S_{bu}^0), propionate (S_{pro}^0), and acetate (S_{ac}^0) were quantified through volatile fatty acid (VFA) analysis (Su et al., 2021).

Several state variables could not be experimentally quantified or readily converted to the units required by the model. These include the hydrogen and methane in liquid phase, ($S_{H_2}^0$, $S_{CH_4}^0$), the particulate composite (X_c^0), the concentrations of degraders X_{bac}^0 referring to X_{su}^0 , X_{aa}^0 , X_{fa}^0 , X_{c4}^0 , X_{pro}^0 , X_{ac}^0 , $X_{H_2}^0$, particulate inert (X_l^0), and cations and anions (S_{cat}^0 , S_{an}^0). The influence of these uncertain inputs on simulation outcomes was systematically assessed through sensitivity analysis, each flagged as 'SA' in Table 2.

2.4. ADM1 model calibration and validation

2.4.1. Sensitivity analysis (SA)

Sensitivity analyses (SA) were conducted to evaluate the influence of 11 kinetic parameters within θ_{ks} and 13 uncertain initial values of θ_0 (listed in Table 2) on selected three key model outputs, referred to as quantities of interest (QoIs), including: QoI₁: steady-state biogas production (m³/d), QoI₂: steady-state total ammoniacal nitrogen (TAN) concentration in digestate (kmole N/m³), and QoI₃: steady-state H⁺ concentration (pH) in digestate (kmole H⁺/m³).

The variation ranges of all θ_{ks} and a subset of θ_0 (including $S_{H_2}^0$, $S_{CH_4}^0$, S_{cat}^0 and S_{an}^0), were obtained from the literature and summarised in Table 3. The sources of these data are provided in the SM: Appendix C. For instance, the protein hydrolysis rate range ($k_{hyd,pr}$) was informed by Razaviarani and Buchanan (2015), while the half saturation coefficient for acetate uptake ($K_{S,ac}$) was derived from values reported in Rosen and Jeppsson (2006).

The variation ranges of the particulate composite (X_c^0) and anaerobic degraders (X_{bac}^0) were derived as follow. As components of particulate COD (pCOD), X_c^0 and X_{bac}^0 were set at a combined value of 32.13 kg COD/m³, calculated by subtracting sCOD (Table 1) and other pCOD fractions (carbohydrates (X_{ch}^0), proteins (X_{pr}^0), lipids (X_{li}^0) and lignin (X_l^0)) (Table 2) from tCOD (Table 1). Assuming 30 % of the tCOD enters the digester as particulate composite (Batstone et al., 2002; Rosen and Jeppsson, 2006; Catenacci et al., 2021), the mean of X_c^0 was set at 16.64 kg COD/m³, with variation range of 16.64 ± 10 kg COD/m³. The remaining 15.49 kg COD/m³ was allocated to X_{bac}^0 , with a variation range of 15.49 ± 10 kg COD/m³. X_{bac}^0 was further distributed among microbial groups (e.g., sugar degraders, acetate degraders) according to their maximum growth rates (μ_{max}), following the approach suggested by Souza et al. (2013a). This allocation was explored within plausible ranges to reflect inoculum variability and to capture potential uncertainty in biomass distribution.

Table 3 summarises the ranges of variability for sensitivity analysis of the 24 uncertain inputs within θ_{ks} and θ_0 .

Stage-wise global sensitivity analysis (GSA) was carried out to identify the most influential θ_{ks} and θ_0 for calibration. The Morris

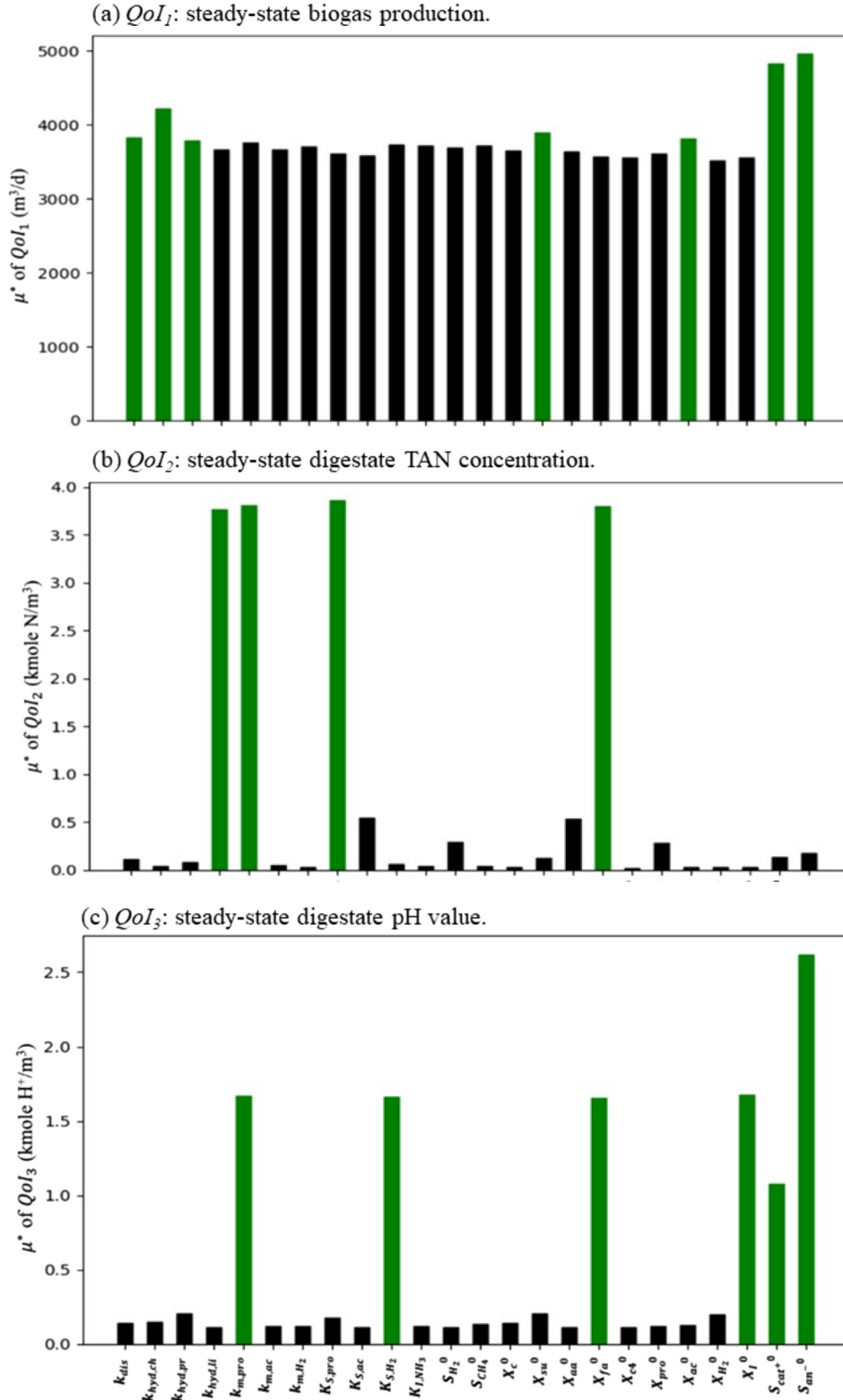


Fig. 1. Mean (μ_i^*) of the distributions of the elementary effects for the 24 uncertain inputs θ_{ks} and θ_o evaluated for (a) QoI_1 , (b) QoI_2 and (3) QoI_3 .

Elementary Effect (EE) method (Morris, 1991) was adopted to pre-screen the uncertain inputs (θ), offering a more stochastic alternative to traditional one-at-a-time local sensitivity analysis (LSA). For each input θ_i , 1000 combinations of the remaining parameters were sampled. The elementary effect (EE_i) for each combination was calculated using Equation (2), and the mean of the absolute values of EE_i , denoted as μ_i^* ,

was adopted as a sensitivity metric (Campolongo et al., 2007; Morris, 1991).

$$EE_i = \frac{y(\theta_1, \dots, \theta_i + \delta, \dots, \theta_d) - y(\theta_1, \dots, \theta_i, \dots, \theta_d)}{\delta} \quad (\text{Equation 2})$$

where $\theta = (\theta_1, \dots, \theta_i, \dots, \theta_d)$ is a randomly selected point in the d-

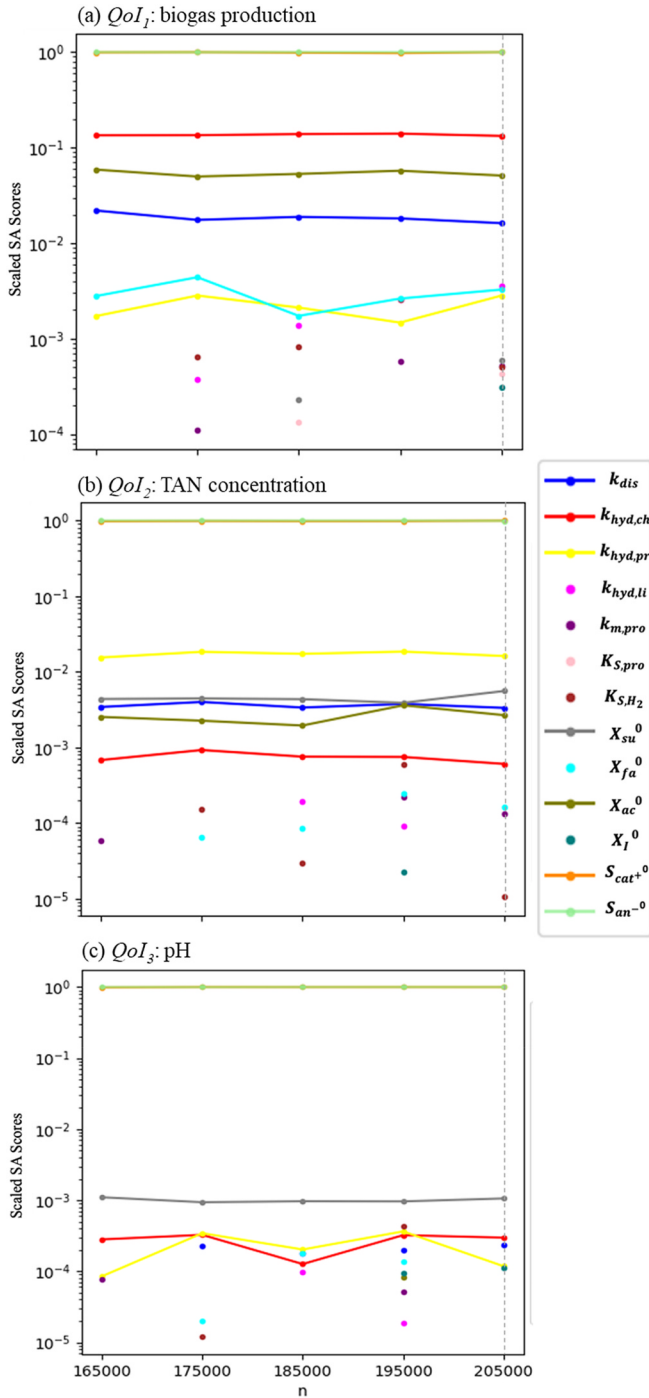


Fig. 2. RBD-FAST GSA outcomes: Main effect values of the 13 key uncertain inputs scaled relative to the largest main effect among the inputs for (a) QoI_1 , (b) QoI_2 , (c) QoI_3 (y-axis is in logarithmic scale).

dimensional hypercube parameter space Z_d . δ defines the perturbation step size applied to each input parameter and $\in \left\{ \frac{1}{k-1}, 1 - \frac{1}{k-1} \right\}$, where k is the discretisation level of θ_i .

Inputs with higher μ_i^* values were considered more influential. Although the Morris method offers an efficient means of identifying key parameters, its reliability is limited by its exploration of only a fraction of the input space (Saltelli et al., 2019). To overcome this limitation and quantify the individual contribution of each uncertain input to output variance, referred to as its main effect, a variance-based GSA was performed. Rather than using the commonly applied Sobol' method, which

requires larger sample sizes for accurate main effect estimation in highly non-linear models (Jaxa-Rozen and Kwakkel, 2018; Tarantola et al., 2006), this study adopted the hybrid random balance design-Fourier amplitude sensitivity test (RBD-FAST) method. RBD-FAST estimates main effects using only n samples, in contrast to $n^*(d+2)$ evaluations required by Sobol', by using spectral decomposition to isolate each input's contribution without computing conditional variance, thereby offering a more computationally efficient alternative (Tarantola et al., 2006).

Convergence of main effects was assessed using the $\kappa_N < 0.05$ criterion (Equation (3)) which measures the stability of sensitivity estimates as sample size increases (Touzani and Busby, 2014). The sample size was initially set to $n = 165,000$ and incrementally increased by 10,000 until convergence was achieved.

$$\kappa_N = \frac{\frac{1}{t} \sum_{s=1}^t \|V_n - V_{n-s\Delta n}\|}{\|V_n\|} \quad (\text{Equation 3})$$

where V_n is the vector of estimated main effect score with n samples, $V_{n-s\Delta n}$ is the scores estimated with decreasing number of samples with Δn interval. $\|\cdot\|$ is the ℓ_2 norm of the vector (Touzani and Busby, 2014).

2.4.2. Bayesian calibration and model validation

Bayesian inference was applied to calibrate the most influential uncertain inputs identified through the GSA. Prior distributions were informed by parameter ranges reported in ADM1 studies involving THP sludge (SM: Appendix C), ensuring relevance to the substrate characteristics used in this work. Unlike the uniform distributions adopted in the SA section, Normal and Gamma distributions were applied to more accurately represent the probability density of the uncertain inputs.

The data likelihood distributions used to assess the discrepancy between model predictions and observations, were standard normal distributions. This approach assumes that the noise in the experimental observations follows a normal distribution. Different standard deviations ($\sigma_1 = 0.0015$; $\sigma_2 = 0.2$; $\sigma_3 = 0.01$) were assigned to account for the varying noise levels in the daily biogas flow, pH and TAN observations respectively (Equation (4)).

$$P(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \hat{y})^2}{2\sigma^2}\right) \quad (\text{Equation 4})$$

where y is experimental observations; $\hat{y}$ is the model evaluation for a specific set of uncertain inputs drawn from the prior distribution; σ is the standard deviation reflecting noise levels.

The calibration was performed by using the No-U-Turn Sampler (NUTS) in the PyMC3 package (Salvatier et al., 2016), and the posterior distribution was approximated by drawing 20 Markov Chain Monte Carlo (MCMC) chains of 5000 samples, with 20,000 burn-in samples. Convergence of the MCMC chains was assessed via the Gelman-Rubin statistic, with values close to one indicating successful convergence. Marginal posteriors were approximated with smoothed histogram of MCMC samples using kernel density estimation. The summary statistics of posterior distributions included the mean, standard deviation (SD), and high-density intervals (HDI) at 3 % and 97 %, representing the most probable parameter range after incorporating observations.

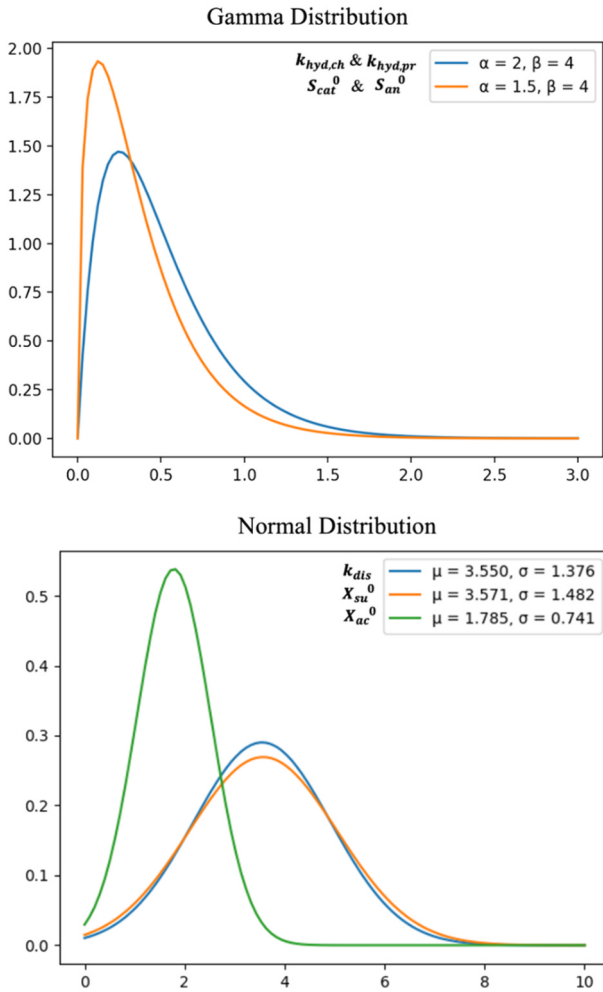
The identifiability of uncertain inputs was assessed by quantifying the disparity between the prior and posterior distributions. Two metrics were used including Location Difference and Jensen-Shannon (JS) divergence. The former measures the shift in the mean locations of the prior and posterior distributions, indicating how much the central belief about the parameter changed with new data. In comparison, the latter quantifies differences between two distributions in terms of both location and shape (e.g., variance and skewness), offering a robust and symmetric measure of disparity across the full domain (Kingetsu and Hamasuna, 2021).

To validate the model, a posterior predictive check was performed

Table 4

Statistical summary of the marginal prior and posterior distributions of the seven most influential uncertain inputs identified in GSA.

		Prior Distribution		Posterior Distribution				Location Difference	JS Divergence
		Type	Parameters	Mean	SD	HDI 3 %	HDI 97 %		
θ_{ks}	k_{dis}	Normal	$\mu = 3.55 \sigma = 1.37$	3.235	1.493	0.034	5.557	0.315	0.011
	$k_{hyd,ch}$	Gamma	$\alpha = 2 \beta = 4$	0.022	0.009	0.006	0.038	0.478	0.308
	$k_{hyd,pr}$	Gamma	$\alpha = 2 \beta = 4$	0.240	0.061	0.150	0.355	0.260	0.124
θ_0	X_{su}^0	Normal	$\mu = 3.57 \sigma = 1.48$	3.562	1.535	0.666	6.432	0.009	0.003
	X_{ac}^0	Normal	$\mu = 1.78 \sigma = 0.74$	1.977	0.580	0.950	3.000	0.197	0.043
	S_{cat}^0	Gamma	$\alpha = 1.5 \beta = 4$	0.198	0.147	0.003	0.488	0.177	0.067
	S_{an}^0	Gamma	$\alpha = 1.5 \beta = 4$	0.499	0.150	0.268	0.805	0.124	0.135

**Fig. 3.** Marginal prior distributions for k_{dis} , $k_{hyd,ch}$, $k_{hyd,pr}$, X_{su}^0 , X_{ac}^0 , S_{cat}^0 and S_{an}^0 .

using 3000 sets of θ_{ks} and θ_0 sampled from the posterior distributions. These were used to simulate daily biogas flow, pH, TAN and COD. The mean percentage error (MPE) between the posterior predictive means and observed values was calculated (Equation (5)). For daily biogas flow, which was affected by significant noise from the ultra-low flow gas counter, predictive performance was evaluated by calculating the percentage of the 10-day moving average trendline that fell within the 95 % credible interval of model predictions.

$$MPE = \frac{100\%}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right) \quad (\text{Equation 5})$$

where $\hat{y}_i$ is the model prediction. y_i is the experimental measurement. n is the number of steady-state observations for pH, TAN or COD.

To assess the impact of domain-informed priors, Bayesian calibration was repeated using uniform priors, consistent with those applied in the SA (Table 3), and followed by a second posterior predictive check.

3. Results and discussion

3.1. Sensitivity analysis results

3.1.1. Morris EE pre-screening results

As an initial screening step, the Morris method was applied to evaluate the influence of 24 uncertain kinetic parameters (θ_{ks}) and θ_0 on the identified key quantities of interest QoI_1 (steady-state biogas production), QoI_2 (steady-state TAN concentration) and QoI_3 (steady-state pH). Although the implicit backward Euler solver improved stability of the simulation results, certain parameter sets still produced unstable solutions due to the stiffness and non-linearity of the ADM1 ODE system (Nimmegeers et al., 2017). Extreme outlier values were observed for QoI_1 and QoI_2 , reaching 10^{-131} and 10^{-233} , respectively.

To focus the analysis on stable region of the parameter space, parameter sets producing outputs outside the range of 10^{-5} to 10^5 for QoI_1 and QoI_2 were excluded in the SA. The SA was repeated ten times. Fig. 1(a), (b) and (c) show the averaged μ_i^* values across ten runs for QoI_1 , QoI_2 and QoI_3 , respectively, where μ_i^* reflects the mean magnitude of a parameter's absolute EE_i values over 1000 parameter sets. The most influential uncertain parameters obtained from the analysis, highlighted with green bars, were subsequently examined through variance-based GSA.

These Morris pre-screening results revealed significant differences in parameter sensitivity. Key parameters, such as disintegration rate (k_{dis}) and hydrolysis rates of carbohydrate and protein ($k_{hyd,ch}$ and $k_{hyd,pr}$) were crucial for the steady-state biogas production. Initial concentrations of certain degraders, such as X_{su}^0 , X_{fa}^0 and X_{ac}^0 , may also be influential to the specific QoI s selected in this study. Interestingly, initial cation and anion concentrations (S_{an}^0 and S_{cat}^0) significantly influenced both steady-state biogas production and digestate pH, suggesting their importance role in maintaining ion balance and affecting digestate chemistry. In contrast, parameters like the 50 % inhibitory concentration of free NH_3 (K_{i,NH_3}), initial concentrations of hydrogen and methane in gas phase (S_{gas,H_2}^0 , S_{gas,CH_4}^0) and particulate composites concentrations (X_c^0) had minimal influence on all QoI s.

The removal of outliers and the stability provided by the backward Euler solver enhanced the reliability of these findings. The potential significance of certain initial values (θ_0) emphasised the importance of calibrating θ_0 together with θ_{ks} , when using ADM1 to predict the steady-state performance of a pilot-scale continuous digester with only start-up phase observations. The Morris pre-screening reduced the number of uncertain θ_{ks} and θ_0 for the subsequent variance-based GSA by half. Specifically, the retained 13 parameters are: k_{dis} , $k_{hyd,ch}$, $k_{hyd,pr}$, $k_{hyd,li}$, $k_{m,pro}$, $K_{S,pro}$ and K_{S,H_2} for θ_{ks} and X_{su}^0 , X_{fa}^0 , X_{ac}^0 , X_I^0 , S_{cat}^0 , and S_{an}^0 for θ_0 .

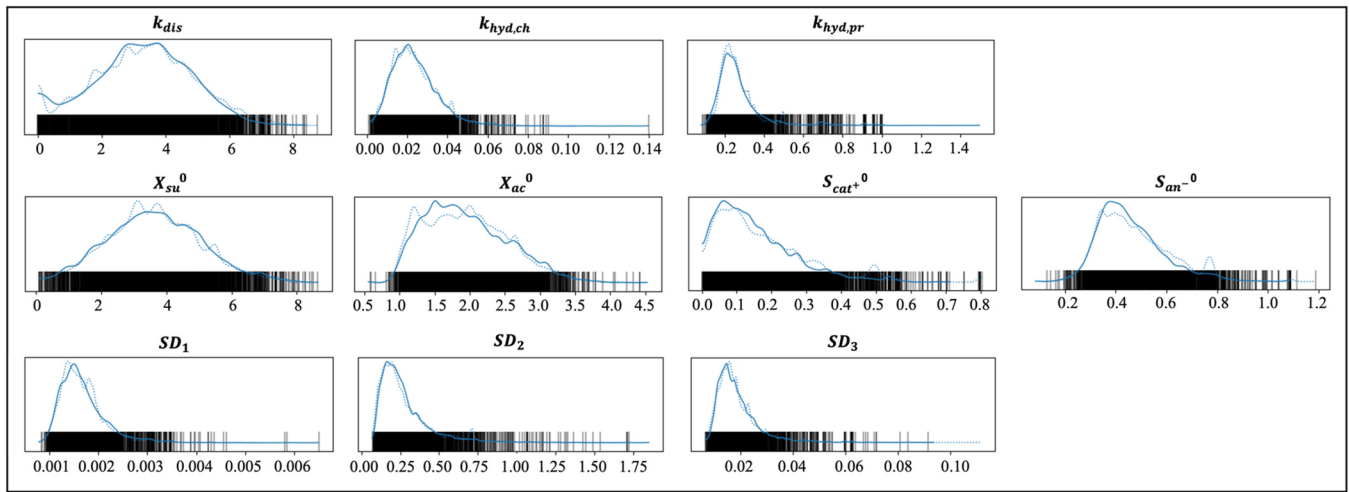


Fig. 4. Smoothed marginal posterior distributions for k_{dis} , $k_{hyd, ch}$, $k_{hyd, pr}$, X_{su}^0 , X_{ac}^0 , S_{cat}^0 and S_{an}^0 and estimated measurement errors for QoI_1 , QoI_2 and QoI_3 from two Markov Chain Monte Carlo chains.

θ_0 .

3.1.2. Hybrid RBD-FAST GSA results

Since the Morris method provides only a qualitative ranking of input importance, variance-based GSA was applied to assess main effects across the entire input domain. As described in section 2.4.1, the hybrid RBD-FAST method was chosen for its computational efficiency over the Sobol' method. Starting with $n = 165,000$, convergence ($\kappa_N < 0.05$) was reached at $n = 205,000$ for all $QoIs$, whereas the Sobol' method had yet to achieve convergence at 458,752 evaluations (due to space constraints, detailed results are not shown). This highlights the efficiency of RBD-FAST for sensitivity analysis of complex process-based model like ADM1.

Fig. 2(a), (b) and (c) present the scaled main effects of the 13 key uncertain inputs for QoI_1 , QoI_2 and QoI_3 , respectively, over five trials of increasing sample sizes (n). Results from trials with smaller sample sizes, where the convergence criterion were not met, were also included in the plots for cross-validation and comparison. The main effect of each θ_i was scaled relatively to the largest main effect among the 13 key θ_{ks} and θ_0 for each QoI , with y-axis presented in a logarithmic scale. Each dot in Fig. 2 represents the log-transformed scaled main effect values for θ_i . However, the main effects of the least important factors estimated by the RBD-FAST method tended to be negative (Tarantola et al., 2006) and could not be log-transformed. Thus, their main effects were not shown as dots in the plots. The θ_i with positive SA scores in all five trials were connected by a line, allowing for a reliable and discernible comparison of the relative influences of 13 values within θ_{ks} and θ_0 .

Notably, the initial concentrations of cation (S_{cat}^0) and anion (S_{an}^0) emerged as the most influential inputs across all three $QoIs$. According to the RBD-FAST method's results, for QoI_1 (steady-state biogas production), the carbohydrate hydrolysis rate ($k_{hyd, ch}$) was the next most significant input, with acetate degraders (X_{ac}^0) and the disintegration rate (k_{dis}) also demonstrated clear impacts. For QoI_2 (steady-state TAN concentration), the protein hydrolysis rate ($k_{hyd, pr}$) was dominant after cation and anion concentrations. For QoI_3 (steady-state pH), sugar degraders (X_{su}^0) exhibited moderate influence, along with impacts from cation and anion. In contrast, the lipid hydrolysis rate $k_{hyd, li}$ had no significant influence on any QoI , which aligns with the extremely low lipid content in the THP-treated sludge.

The spacing between lines in the plots also highlights the difference in the main effects among inputs. For QoI_2 and QoI_3 , the significant gap between the cation and anion lines and the next influential input underscores the dominant role of ionic concentrations in regulating TAN and pH levels. In contrast, the more evenly spaced lines for QoI_1 in Fig. 2

(a) suggest a distributed sensitivity across multiple parameters, reflecting the multifaceted nature of biogas production influenced by various microbial and substrate interactions.

The hybrid RBD-FAST method efficiently estimated main effects with lower computational demand. Notably, when modelling the approach to steady-state of the newly started digester treating THP sludge, not only the three kinetic parameters (θ_{ks}), k_{dis} , $k_{hyd, ch}$ and $k_{hyd, pr}$, commonly recognised as sensitive parameters in ADM1 calibration, but also the initial values (θ_0) of four state variables, X_{su}^0 , X_{ac}^0 , S_{cat}^0 and S_{an}^0 , significantly influenced model outputs. These findings emphasise the importance of initial values of microbial and ion concentrations in the inoculum during startup of a digester. Meanwhile, the negligible influence of $k_{hyd, li}$ identified by GSA underscores the value of identifying feedstock-specific sensitive parameters for further targeted model calibration.

3.2. Bayesian calibration and uncertain inputs identifiability

The Bayesian calibration targeted the seven most influential inputs identified through the SA (Table 3), including three θ_{ks} parameters (such as, the disintegration rate constant k_{dis} , and the carbohydrate and protein hydrolysis rate constants $k_{hyd, ch}$ and $k_{hyd, pr}$) and four θ_0 values (such as initial concentrations of soluble sugar degraders X_{su}^0 , acetate degraders, X_{ac}^0 , and cation and anion S_{cat}^0 and S_{an}^0). This reduced set captured the majority of output variance while maintaining computational efficiency.

3.2.1. Prior distribution

Normal and Gamma distributions were selected as prior distributions to capture the probability density information obtained from the literature specific to THP sludge (SM: Appendix C). Studies by Souza et al. (2013a, 2013b) indicated that while an increased disintegration rate of THP sludge due to the enhanced cell wall rupture and extracellular polymeric substance release, its hydrolysis rates were lower compared to raw sludge. Accordingly, right-skewed Gamma distribution priors were assigned to hydrolysis rates ($k_{hyd, ch}$ and $k_{hyd, pr}$) to reflect their tendency toward smaller values and similarly for the initial ion concentrations (S_{cat}^0 and S_{an}^0).

In contrast, Normal distributions were applied to the disintegration rate (k_{dis}) and the initial concentrations of degraders (X_{ac}^0 and X_{su}^0) to reflect the observed variability with a symmetric spread around the mean. For the Normal distributions, the mean (μ) was set at the midpoint of the observed lower and upper bounds for THP sludge in SM: Appendix

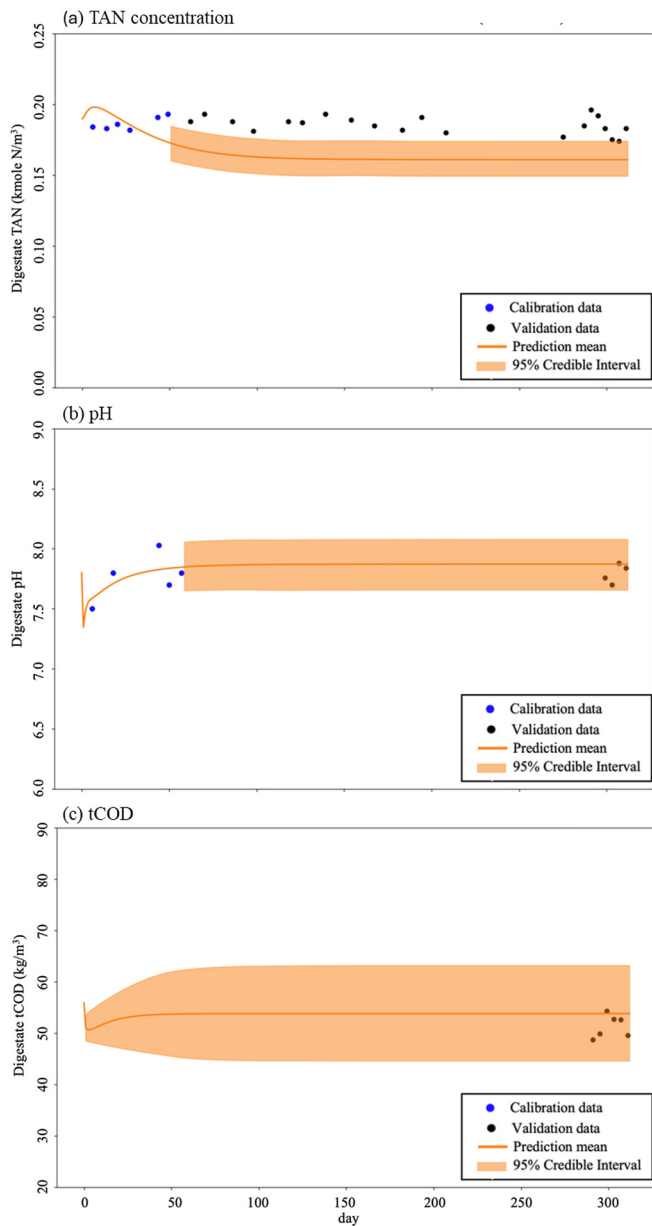


Fig. 5. Posterior predictive check on (a) digestate TAN, (b) digestate pH, and (c) digestate tCOD observations.

C (k_{dis} : 1.5–5.6, X_{su}^0 : 1.348–5.792, X_{ac}^0 : 0.809–3.457). Whilst the standard deviation (σ) was estimated by assuming that 86.64 % of the data (three σ) falls within these bounds.

Table 4 summarised the marginal prior distributions adopted, including the distribution type (i.e., Normal or Gamma) and its corresponding parameters (mean μ and standard deviation σ for Normal, and shape α and rate β for Gamma) used to encode the initial knowledge about the parameter distributions. Fig. 3 is a visualisation of these distributions.

3.2.2. Posterior distribution and uncertain inputs identifiability

The statistical summary of the marginal posterior distributions for the seven most influential parameters within θ_{ks} and θ_0 , calibrated against daily biogas production, digestate TAN, and pH data prior to day 50 (<2 HRTs), is reported in Table 4. The table presents posterior mean and standard deviation values alongside 95 % HDIs, location differences, and JS divergence, between marginal prior and posterior distributions, with units consistent with Table 3. Fig. 4 complements this by

showing the approximated marginal posterior distributions obtained from the MCMC draws.

To improve calibration efficiency, literature-informed priors were employed rather than non-informative uniform priors. As a result, identifiability of uncertain inputs could not be assessed solely by the sharpness of posterior distributions, as is typically done with uniform prior (Pastor-Poquet et al., 2019; Yi and Park, 2021). Instead, Location Difference and JS divergence were used to quantify the extent of prior-to-posterior updates. JS divergence showed to be particularly effective in capturing changes in both central tendency and distributional spread.

Among the kinetic parameters (θ_{ks}), $k_{hyd, ch}$ was the most identifiable, with both metrics indicating substantial updates. Its posterior mean (0.024 d^{-1} , Table 4) was lower than the values reported in other THP sludge studies (SM: Appendix C), yet comparable to the estimate calibrated by Pastor-Poquet et al. (2019), whose model distinguished between fast and slow carbohydrate hydrolysis. This supports the hypothesis that THP pretreatment partially hydrolyses complex carbohydrates into soluble sugars, leaving behind a more recalcitrant fraction, consistent with the stepwise degradation proposed by Shana et al. (2013). In contrast, k_{dis} exhibited limited identifiability, showing a substantial mean shift but minimal HDI change (Figs. 3 and 4), as confirmed by the small JS divergence, indicating persistent uncertainty in its estimation.

Among the four uncertain initial values (θ_0), both JS divergence and Location Difference metrics agreed that X_{su}^0 was the least identifiable input, whilst S_{cat}^0 and S_{an}^0 were the most identifiable. Notably, upon incorporating new observation, the HDI of the posterior for S_{cat}^0 narrowed while maintaining a right-skewed gamma-like distribution. In contrast, the peak of the posterior of S_{an}^0 shifted to the right. This led to a posterior mean difference of 0.301 kmol/m^3 between these two variables, aligning with the charge balance constraint as indicated by Equation (1), where $S_{an}^0 - S_{cat}^0 = 0.146 \text{ kmol/m}^3$, accounting for potential measurement errors in VFA, TC, TOC, and TAN.

Overall, although the calibrated values for S_{cat}^0 and S_{an}^0 are higher than the literature values (SM: Appendix C), they remain chemically plausible given the ionic strength of the inoculum. For example, the posterior mean cation concentration (S_{cat}^0) was 0.198 kmol/m^3 and the anion concentration ($S_{an}^0 = 0.499 \text{ kmol/m}^3$, Table 4), slightly above reported ranges of 0.01–0.16 and 0.02–0.60 kmol/m^3 , respectively (Song et al., 2019; Pehlivanoglu and Sedlak, 2004).

The analysis of degrader-groups influences was performed by varying the total biomass concentration (X_{bac}^0) while maintaining the relative distribution of degrader groups constant according to their maximum growth rates (μ_{max}). This analysis identified only the sugar degrader (X_{su}^0) and acetate degrader (X_{ac}^0) as influential. However, during the start-up phase, variability in the actual distribution of microbial functional groups may affect transient dynamics. These findings highlight the importance of more detailed inoculum characterisation to improve early-stage model calibration (Wu et al., 2016).

Beyond inoculum distribution, additional short-term mechanisms may also contribute to fluctuations in biogas production. Transient microbial inhibition, minor shifts in substrate availability, or gas storage dynamics in the digester headspace could all introduce variability not captured by the current ADM1 structure. Incorporating these processes into future ADM1 extensions, or collecting higher-resolution operational data, may improve predictive accuracy for dynamic biogas trends (Tisocco et al., 2024; Liu et al., 2022).

Based on data spanning less than two hydraulic retention times (HRT = 50 days), Bayesian inference effectively narrowed the uncertainty in parameter distributions. Under the specific calibration conditions employed in this study, the most identifiable parameters were the hydrolysis rates, $k_{hyd, ch}$ and $k_{hyd, pr}$, along with the initial ion concentrations, S_{cat}^0 and S_{an}^0 , as confirmed by JS divergence.

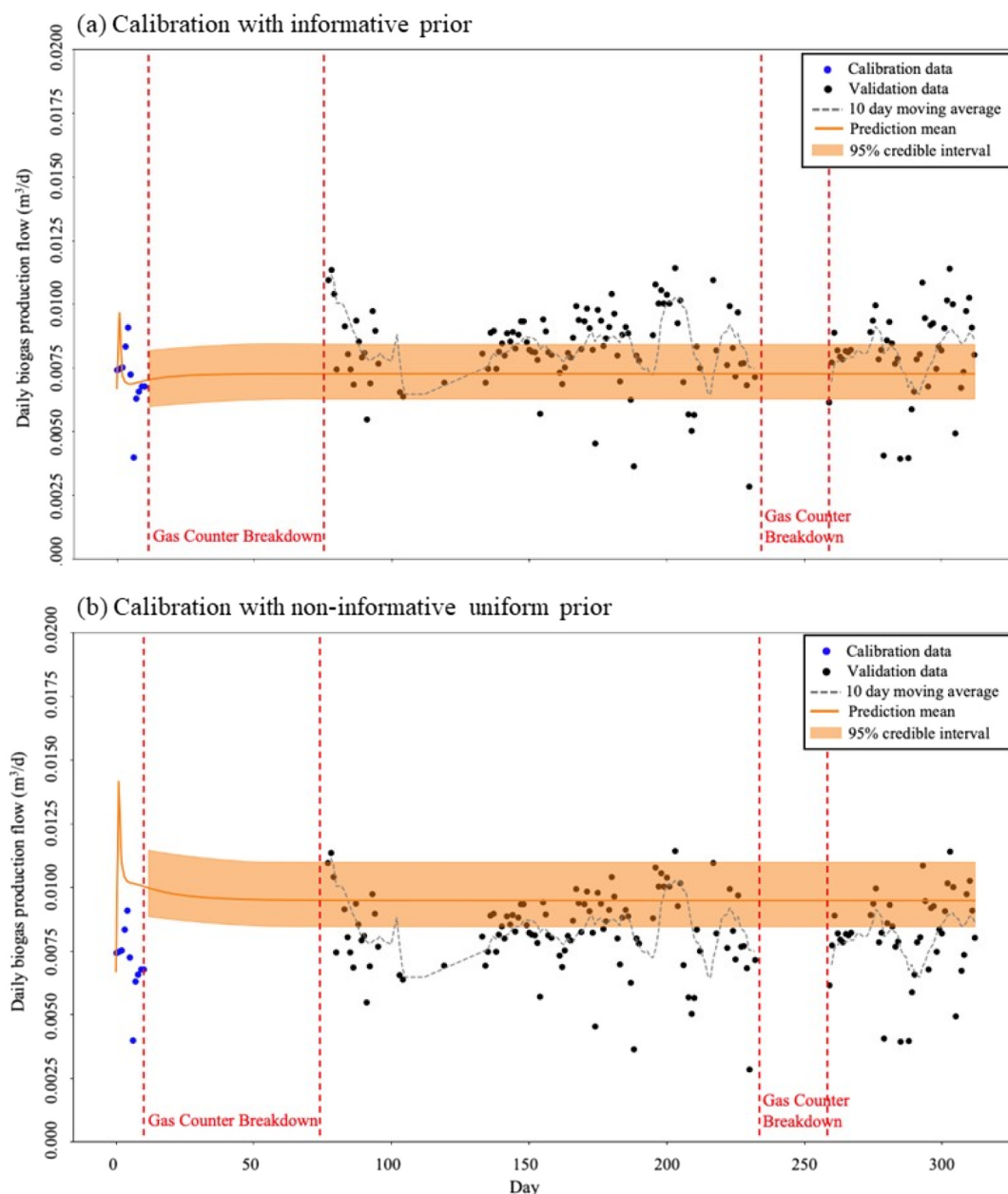


Fig. 6. Posterior predictive check of daily biogas flow for the model calibrated, respectively, with (a) informative and (b) non-informative uniform prior.

3.3. Validation of Python-based ADM1

The predictive performance of the calibrated model was subsequently evaluated using validation data covering key outputs such as biogas production, TAN, pH, and tCOD under steady-state conditions. Corresponding results are presented in Figs. 5 and 6 respectively. In these figures, the blue dots represent the experimental data used for calibration, while the black dots are the validation data. The orange line represents the mean of the model outputs from 3000 sets of the seven most influential θ_{ks} and θ_0 sampled from the posterior distribution, and the shaded area represents the 95 % credible interval, which is the range that has 95 % probability of containing the true values. In Fig. 6, the grey dotted line represents the 10-day moving average trendline of the experimental biogas production data.

As shown in Fig. 5(a), (b) and (c), the model calibrated using of acclimatisation stage data (<50 days) accurately predicted the steady-state digestate pH and tCOD concentrations, with 1.10 % and 5.38 % MPEs respectively, for the posterior mean predictions. Besides, the

calibrated model successfully captured the increasing pH trend during the early-stage of digestion. However, the steady-state digestate TAN level was underestimated by 10.60 %, which could be due to experimental measurement variability or slight differences between model assumptions and actual system behaviour. Simulation results indicated that TAN and pH concentrations stabilised around day 100 (~4 HRTs), while tCOD reached steady-state more rapidly, around day 50 (~2 HRTs).

For daily biogas production (Fig. 6(a)), the calibrated model with informative prior effectively captured the initial surge in biogas generation. The 10-day moving average of the validation data showed that 63.14 % of the points fell within the model's 95 % credible interval. As this study focuses on a newly initiated digester transitioning towards steady-state under a fixed OLR, short-term fluctuations arising from minor variations in feedstock inflow and measurement noise were not explicitly captured. Instead, the coverage metric was used as a benchmark of predictive accuracy, enabling comparison with the model calibrated using uniform priors (Section 3.4).

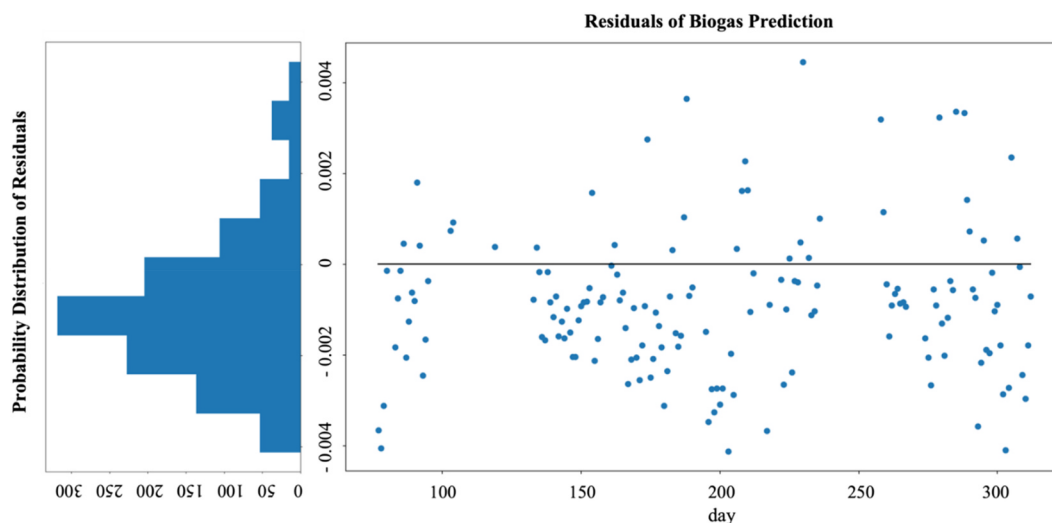


Fig. 7. Residual distribution between mean predictions of the model calibrated with informative prior and experimental validation data (from day 7).

Table 5

Comparison of steady-state performance predictions (mean of 3000 simulation outputs) between models calibrated with informative and uniform priors.

Categories	Informative Prior		Uniform Prior	
	Mean	MPE	Mean	MPE
Daily biogas flow (m ³ /d)	63.136 %	10-day moving average trendline within 95 % credible interval	38.983 %	10-day moving average trendline within 95 % credible interval
TAN (kmole N/m ³)	0.161	−10.602 %	0.164	−10.723 %
pH	7.891	+1.099 %	8.019	+2.8882 %
tCOD (kg/m ³)	53.793	+5.379 %	33.061	−35.483 %

Fig. 7 shows the residuals between the mean model predictions and the observed biogas production values in the validation dataset. The residuals' minimum and maximum values are approximately equidistant from zero, with a dense central cluster, confirming that the calibrated model achieved good predictive alignment overall. However, the distribution is slightly right-skewed, indicating that higher biogas flows were more frequently recorded by the gas counter, as observed in Fig. 6. This implies that the measurement device exhibited greater sensitivity at elevated flow rates.

Compared with traditional ADM1 calibration studies, which typically assess model performance through visual inspection (Blumensaat and Keller, 2005; Ersahin, 2018), this study adopts a more rigorous and statistically comparable evaluation using MPEs and the proportion of observations captured within the 95 % credible interval. Overall, these results highlight the effectiveness and robustness of the Bayesian calibration approach for ADM1, even when constrained by limited early-stage biogas and digestate data. The calibrated model successfully reproduced initial trends and delivered reliable predictions of steady-state digester performance.

3.4. Influence of prior distribution

To further assess the influence of prior assumptions on the effectiveness of Bayesian calibration, an additional calibration for the seven θ_{ks} and θ_0 was performed using non-informative uniform priors over the same ranges reported in Table 3. These uniform ranges were derived from a broader set of ADM1 parameters documented in the literature for sewage sludge, rather than being restricted to values specific to THP-treated sludge.

Table 5 compares the predictive accuracy of models calibrated with informative versus uniform priors, focusing on steady-state outputs for

biogas production, digestate TAN, pH, and tCOD. While both models underestimated steady-state TAN concentrations, the model calibrated with uniform priors consistently underperformed across the remaining metrics. Fig. 6 (a) and (b) further illustrates the impact of prior assumptions through posterior predictive checks of daily biogas flow. In Fig. 6 (b), the model with uniform priors overestimates biogas production, with only 38.98 % of the 10-day moving average trendline falling within the 95 % credible interval.

In contrast, Fig. 6(a) shows that the model calibrated with informative priors yields predictions more closely aligned with observed data and exhibits narrower uncertainty bounds, indicating better parameter identifiability and more reliable inference. These findings underscore the importance of incorporating prior knowledge to improve model robustness and predictive performance.

4. Conclusions

This study investigated the feasibility of calibrating the Anaerobic Digestion Model No. 1 (ADM1) using only early-stage data from a lab-scale semi-continuous anaerobic digester, with the aim of predicting its steady-state performance. Since reaching steady-state can require several months, this approach presents a practical and time-efficient alternative for evaluating feedstock suitability and operational strategies during start-up.

To overcome the challenge of calibrating ADM1 with limited data, given its extensive parameter set, a twofold strategy was adopted: global sensitivity analysis to identify influential parameters, and Bayesian calibration to incorporate domain knowledge via informative priors. These methods enabled effective model tuning despite data constraints.

ADM1 was implemented in Python to support high-throughput simulation, allowing seamless integration with sensitivity analysis and Bayesian workflows that require repeated model evaluation. This computational setup facilitated efficient exploration of parameter space and model calibration.

The findings highlighted the importance of jointly estimating initial state variables, such as ion concentrations and microbial biomass, alongside kinetic parameters. Under thermal hydrolysis pretreatment, hydrolysis rates were more influential than disintegration kinetics due to enhanced solubilisation. Jensen–Shannon divergence proved valuable for assessing parameter identifiability across different prior assumptions, and calibration performance significantly improved when informative priors were used.

Validation against steady-state data demonstrated that the calibrated model reliably reproduced key outputs including biogas production,

TAN, pH, and tCOD. By contrast, calibration with uniform priors resulted in reduced predictive accuracy, underscoring the importance of prior knowledge. The calibrated ADM1 successfully predicted daily biogas production, with 63.14 % of validation points captured within the 95 % credible interval, demonstrating reliable performance under data-limited conditions.

From an industry standpoint, the proposed framework offers a streamlined pathway to accelerate the design and commissioning of full-scale anaerobic digesters. By enabling reliable calibration and robust performance predictions from limited early-stage data, it supports timely decision-making, shortens evaluation periods, and enhances confidence in operational planning, especially for novel feedstocks or pretreatment technologies. Future extensions incorporating microbial inhibition kinetics and transient feedstock variability could further enhance predictive accuracy and broaden industrial applicability.

CRediT authorship contribution statement

Yanxin Liu: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Ying Jiang:** Writing – review & editing, Supervision, Conceptualization. **Nasreen Nasar:** Investigation. **Yadira Bajon-Fernandez:** Resources. **Philip Longhurst:** Writing – review & editing, Supervision. **Weisi Guo:** Supervision, Methodology. **Mei Lei:** Writing – review & editing. **Immacolata Bortone:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jenvman.2025.128537>.

Data availability

The dataset and Bayesian calibration implementation of ADM1 in Python are available upon request at: <https://github.com/Janice-Yanxin-Liu/ADM1-Calibration>.

References

- Antwi, P., Li, J., Boadi, P.O., Meng, J., Shi, E., Deng, K., Bondinuba, F.K., 2017. Estimation of biogas and methane yields in an UASB treating potato starch processing wastewater with backpropagation artificial neural network. *Bioresour. Technol.* 228, 106–115.
- Asadi, M., Guo, H., McPhedran, K., 2020. Biogas production estimation using data-driven approaches for cold region municipal wastewater anaerobic digestion. *J. Environ. Manag.* 253, 109708.
- Batstone, D.J., Keller, J., Angelidaki, I., 2002. The IWA anaerobic digestion model No.1 (ADM1). *Water Sci. Technol.* 45 (10), 65–73.
- Bialobrzewski, I., Waszkiel, K., Bulkowska, K., 2024. The application of anaerobic digestion model no. 1 for the optimization of biogas production from maize silage, pig manure, cattle manure, and digestate in a full-scale biogas plant. *Fuel* 357 (B), 129789. <https://doi.org/10.1016/j.fuel.2023.129789>.
- Blumensaat, F., Keller, J., 2005. Modelling of two-stage anaerobic digestion using the IWA anaerobic digestion model no. 1 (ADM1). *Water Res.* 39 (1), 171–183. <https://doi.org/10.1016/j.watres.2004.07.024>.
- Boubaker, F., Ridha, B., 2008. Modelling of the mesophilic anaerobic co-digestion of olive mill wastewater with olive mill solid waste using anaerobic digestion model no. 1 (ADM1). *Bioresour. Technol.* 99 (14), 6565–6577. <https://doi.org/10.1016/j.biortech.2007.11.035>.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environ. Model. Software* 22 (10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>.
- Catenacci, A., Azzellino, A., Malpei, F., 2021. Development of statistical predictive models for estimating the methane yield of Italian municipal sludges from chemical composition: a preliminary study. *Water Sci. Technol.* 79 (3), 435–447. <https://doi.org/10.2166/wst.2019.063>.
- Chew, K.R., Leong, H.Y., Khoo, K.S., Vo, D.N., Anjum, H., Chang, C.K., Show, P.L., 2021. Effects of anaerobic digestion of food waste on biogas production and environmental impacts: a review. *Environ. Chem. Lett.* 19, 2921–2939. <https://link.springer.com/article/10.1007/s10311-021-01220-z>.
- De Crescenzo, C., Marzocchella, A., Karatza, D., Chianese, S., Musmarra, D., 2024. Autogenerative high-pressure anaerobic digestion modelling of volatile fatty acids: effect of pressure variation and substrate composition on volumetric mass transfer coefficients, kinetic parameters, and process performance. *Fuel* 358 (A), 130144. <https://doi.org/10.1016/j.fuel.2023.130144>.
- Donoso-Bravo, A., García, G., Pérez-Elvira, S., Fdz-Polanco, F., 2011. Initial rates technique as a procedure to predict the anaerobic digester operation. *Biochemical Engineering Journal* 53 (3), 275–280. <https://doi.org/10.1016/j.bej.2010.11.007>.
- Ersahin, M.E., 2018. Modeling the dynamic performance of full-scale anaerobic primary sludge digester using Anaerobic Digestion Model No. 1 (ADM1). *Bioprocess and Biosystems Engineering* 41 (10), 1539–1545. <https://doi.org/10.1007/s00449-018-1981-5>.
- European Biogas Association, 2022. Statistical report 2022. Tracking biogas and biomethane deployment across Europe. <https://www.europeanbiogas.eu/SR-2022/EBA/#our>.
- European Council, 2018. Renewable energy: council confirms deal reached with the European Parliament 2018. <https://www.consilium.europa.eu/en/press/press-releases/2018/06/27/renewable-energy-council-confirms-deal-reached-with-the-european-parliament/>.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. *Bayesian Data Analysis*, third ed. CRC Press.
- Han, Y., Du, Z., Hu, X., Li, Y., Cai, D., Fan, J., Geng, Z., 2023. Production prediction modeling of food waste anaerobic digestion for resources saving based on SMOTE-LSTM. *Appl. Energy* 352, 122024. <https://doi.org/10.1016/j.apenergy.2023.122024>.
- Jaxa-Rozen, M., Kwakkel, J., 2018. Tree-based ensemble methods for sensitivity analysis of environmental models: a performance comparison with Sobol and Morris techniques. *Environ. Model. Software* 107, 245–266. <https://doi.org/10.1016/j.envsoft.2018.06.011>.
- Jones, E., Oliphant, T., Peterson, P., 2001. SciPy: open source scientific tools for python. <http://www.scipy.org/>.
- Kingetsu, Y., Hamasuna, Y., 2021. Jensen–Shannon divergence-based k-Medoids clustering. *J. Adv. Comput. Intell. Inf.* 25 (2), 226–233. <https://doi.org/10.20965/jacii.2021.p0226>.
- Lide, D.R., 2001. *CRC Handbook of Chemistry and Physics*, 82nd ed. CRC Press.
- Liu, Y., Jiang, Y., Bortone, I., 2022. A Scheme for Anaerobic Digestion Modelling and ADM1 Model Calibration. *Chemical Engineering Transactions* 96.
- MathWorks, 2022. MATLAB R2022a Documentation. The MathWorks, Inc., Natick, Massachusetts, United States.
- McElreath, R., 2018. Statistical Rethinking: a Bayesian Course with Examples in R and Stan. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>.
- Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33 (2), 161. <https://doi.org/10.2307/1269043>.
- Nielsen, S.S., 2014. *Food Analysis*, fifth ed. Springer. ISBN: 9781441914774.
- Nimmegeers, P., Lauwers, J., Telen, D., Logist, F., Impe, J. Van, 2017. Identifiability of large-scale non-linear dynamic network models applied to the ADM1-case study. *Math. Biosci.* 288, 21–34. <https://doi.org/10.1016/j.mbs.2017.02.008>.
- Ozgun, H., 2019. Anaerobic Digestion Model No. 1 (ADM1) for mathematical modeling of full-scale sludge digester performance in a municipal wastewater treatment plant. *Biodegradation* 30 (1), 27–36. <https://doi.org/10.1007/s10532-018-9859-4>.
- Pastor-Poquet, V., Papirio, S., Harmand, J., Steyer, J.-P., Trably, E., Escudé, R., Esposito, G., 2019. Assessing practical identifiability during calibration and cross-validation of a structured model for high-solids anaerobic digestion. *Water Res.* 164, 114932. <https://doi.org/10.1016/j.watres.2019.114932>.
- Pehlivanoglu, E., Sedlak, D.L., 2004. Bioavailability of wastewater-derived organic nitrogen to the alga *Selenastrum capricornutum*. *Water Res.* 38 (14–15), 3189–3196. <https://doi.org/10.1016/j.watres.2004.04.028>.
- Pilli, S., Yan, S., Tyagi, R.D., Surampalli, R.Y., 2015. Thermal pretreatment of sewage sludge to enhance anaerobic digestion: a review. *Crit. Rev. Environ. Sci. Technol.* 45 (6), 669–702. <https://doi.org/10.1080/10643389.2013.876527>.
- Qdais, H.A., 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>.
- Razaviarani, V., Buchanan, I.D., 2015. Calibration of the Anaerobic Digestion Model No. 1 (ADM1) for steady-state anaerobic co-digestion of municipal wastewater sludge with restaurant grease trap waste. *Chem. Eng. J.* 266, 91–99. <https://doi.org/10.1016/j.cej.2014.12.080>.
- Rocamora, I., Wagland, S.T., Villa, R., Simpson, E.W., Fernández, O., Bajón-Fernández, Y., 2020. Dry anaerobic digestion of organic waste: a review of operational parameters and their impact on process performance. *Bioresour. Technol.* 299, 122681. <https://doi.org/10.1016/j.biortech.2019.122681>.
- Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 implementation within the BSM2 framework. <https://www.researchgate.net/publication/228684581>.
- Ruffino, B., Fiore, S., Roati, C., Campo, G., Novarino, D., Zanetti, M., 2015. Scale effect of anaerobic digestion tests in fed-batch and semi-continuous mode for the technical and economic feasibility of a full-scale digester. *Bioresour. Technol.* 182, 302–313. <https://doi.org/10.1016/j.biortech.2015.02.021>.

- Sadrimajd, P., Mannion, P., Howley, E., Lens, P.N.L., 2021. PyADM1: a python implementation of anaerobic digestion model no. 1. bioRxiv. <https://doi.org/10.1101/2021.03.03.433746>.
- Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. *Environ. Model. Software* 114, 29–39. <https://doi.org/10.1016/j.envsoft.2019.01.012>.
- Salvatier, J., Wiecki, T.V., Fonnesbeck, C., 2016. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* 2, e55. <https://doi.org/10.7717/peerj-cs.55>.
- Shana, A., Ouki, S., Asaadi, M., Pearce, P., Mancini, G., 2013. The impact of intermediate thermal hydrolysis on the degradation kinetics of carbohydrates in sewage sludge. *Bioresour. Technol.* 137, 239–244. <https://doi.org/10.1016/j.biortech.2013.03.121>.
- Shen, F., Yuan, H., Pang, Y., Chen, S., Zhu, B., Zou, D., Liu, Y., Ma, J., Yu, L., Li, X., 2013. Performances of anaerobic codigestion of fruit & vegetable waste (FVW) and food waste (FW): single-phase vs. two-phase. *Bioresour. Technol.* 144, 80–85. <https://doi.org/10.1016/j.biortech.2013.06.099>.
- Sillen, S.G., Martell, A.E., 1964. *Stability constants of metal ion complexes*, spl. Chemistry Society.
- Song, X., Shi, Z., Li, X., Wang, X., Ren, Y., 2019. Fate of proteins of waste activated sludge during thermal alkali pretreatment in terms of sludge protein recovery. *Front. Environ. Sci. Eng.* 13 (2), 25. <https://doi.org/10.1007/s11783-019-1114-7>.
- Souza, T.S.O., Carvajal, A., Donoso-Bravo, A., Peña, M., Fdz-Polanco, F., 2013a. ADM1 calibration using BMP tests for modeling the effect of autohydrolysis pretreatment on the performance of continuous sludge digesters. *Water Res.* 47 (9), 3244–3254. <https://doi.org/10.1016/j.watres.2013.03.041>.
- Souza, T.S.O., Ferreira, L.C., Sapkaite, I., Pérez-Elvira, S.I., Fdz-Polanco, F., 2013b. Thermal pretreatment and hydraulic retention time in continuous digesters fed with sewage sludge: assessment using the ADM1. *Bioresour. Technol.* 148, 317–324. <https://doi.org/10.1016/j.biortech.2013.08.161>.
- Su, J., Zhang, L., Loh, K.-C., 2021. Review and perspectives of enhanced volatile fatty acids production from acidogenic fermentation of lignocellulosic biomass wastes. *Biorenewables and Bioprocessing* 8, 68. <https://doi.org/10.1186/s40643-021-00420-3>.
- Tarantola, S., Gatelli, D., Mara, T.A., 2006. Random balance designs for the estimation of first order global sensitivity indices. *Reliab. Eng. Syst. Saf.* 91 (6), 717–727. <https://doi.org/10.1016/j.ress.2005.06.003>.
- Thamsiriroj, T., Murphy, J.D., 2011. Modelling mono-digestion of grass silage in a 2-stage CSTR anaerobic digester using ADM1. *Bioresour. Technol.* 102 (2), 948–959. <https://doi.org/10.1016/j.biortech.2010.09.051>.
- Tisocco, S., Weinrich, S., Lyons, G., Wills, M., Zhan, X., Crosson, P., 2024. Application of a simplified ADM1 for full-scale anaerobic co-digestion of cattle slurry and grass silage: assessment of input variability. *Front. Environ. Sci. Eng.* 18 (4), 50.
- Touzani, S., Busby, D., 2014. Screening method using the derivative-based global sensitivity indices with application to Reservoir simulator. *Oil & Gas Science and Technology – Revue d'IFP Energies Nouvelles* 69 (4), 619–632. <https://doi.org/10.2516/ogst/2013195>.
- Walker, M., Zhang, Y., Heaven, S., Banks, C.J., 2009. Potential errors in the quantitative evaluation of biogas production in anaerobic digestion processes. *Bioresour. Technol.* 100 (24), 6339–6346. <https://doi.org/10.1016/j.biortech.2009.07.018>.
- Westgate, P.J., Park, C., 2010. Evaluation of proteins and organic nitrogen in wastewater treatment effluents. *Water Res.* 44 (16), 4883–4895. <https://doi.org/10.1016/j.watres.2010.07.029>.
- Wu, B., Wang, X., Deng, Y., He, X., Li, Z., Li, Q., Qin, H., Chen, J., He, M., Zhang, M., Hu, G., Yin, X., Yin, X.-B., 2016. Adaption of microbial community during the start-up stage of a thermophilic anaerobic digester treating food waste. *Biosci. Biotechnol. Biochem.* 80 (10), 2025–2032. <https://doi.org/10.1080/09168451.2016.1191326>.
- Yi, D.H., Park, C.S., 2021. Model selection for parameter identifiability problem in Bayesian inference of building energy model. *Energy Build.* 245, 111059. <https://doi.org/10.1016/j.enbuild.2021.111059>.
- Zaher, U., Buffiere, P., Steyer, J.-P., Chen, S., 2009. A procedure to estimate proximate analysis of mixed organic wastes. *Water Environ. Res.* 81 (4), 407–415. <https://doi.org/10.2175/106143008x370548>.



DSpace

<https://dspace.org/>

Improving ADM1 predictions via Bayesian analysis for continuous anaerobic digestion

Liu, Yanxin; Jiang, Ying; Nasar, Nasreen; Bajón Fernández, Yadira; Longhurst, Philip J. ...

2026-01-15

Elsevier

<https://doi.org/10.1016/j.jenvman.2025.128537>

Liu Y, Jiang Y, Nasar N, et al., (2026) Improving ADM1 predictions via Bayesian analysis for continuous anaerobic digestion. *Journal of Environmental Management*, Volume 398, January 2026, Article number 128537