CHALMERS
UNIVERSITY OF TECHNOLOGY

Joint state and parameter estimation for Anaerobic Digestion using PSO - tuned EKF based on the modified AMOCO model

Downloaded from: https://research.chalmers.se, 2026-09-11 10:36 UTC

Citation for the original published paper (version of record):
Abera, B., Mamo, M., Bekele, G. et al (2026). Joint state and parameter estimation for Anaerobic Digestion using PSO - tuned EKF based on the
modified AMOCO model. Biomass and Bioenergy, 213.
http://dx.doi.org/10.1016/j.biombioe.2026.109365

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Biomass and Bioenergy 213 (2026) 109365

Contents lists available at ScienceDirect

Biomass and Bioenergy

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

Research paper

Joint state and parameter estimation for Anaerobic Digestion using PSO - tuned EKF based on the modified AMOCO model

Bethlehem Abera [a,b,*], Mengesha Mamo [a], Getachew Bekele [a], Torsten Wik [b]

[a] School of Electrical and Computer Engineering, Addis Ababa University, Addis Ababa, Ethiopia
[b] Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Anaerobic digestion
Extended Kalman Filter
Particle swarm optimization
Sensitivity
State and parameter estimation

ABSTRACT

Biogas production through anaerobic digestion (AD) presents a sustainable energy alternative with significant potential to reduce global warming. However, AD is a complex, nonlinear, and dynamic process influenced by time-varying parameters and non-stationary disturbances. These challenges, together with the limited availability of reliable online measurements for key concentration variables, hinder effective real-time monitoring. To address these limitations, this study proposes a joint state and parameter estimation approach based on a modified Advanced Monitoring and Control (AMOCO) model with a Particle Swarm Optimization (PSO) - tuned Extended Kalman Filter (EKF), a combination not previously applied to anaerobic digestion processes. The modified AMOCO model, originally developed for control applications, is adapted to better align with both simulated and experimental data. Sensitivity analysis identifies three key parameters whose estimation significantly improves system reconstruction. To further enhance estimation performance, PSO is employed to tune the noise covariance matrices of a discrete EKF. Validation using the Anaerobic Digestion Model No. 1 (ADM1) as a benchmark plant confirms reliable state and parameter estimation and accurate output predictions. Robustness is assessed by applying the EKF tuned for nominal noise to different measurement-noise levels, demonstrating stable performance under moderate noise mismatch and limited degradation under severe mismatch. Results show that the proposed PSO-EKF approach achieves a 70%–80% reduction in augmented state-estimation RMSE compared with a conventionally tuned EKF. The methodology provides a foundation for monitoring and control in AD and has potential for adaptation to other complex, non-linear bioprocesses, thus supporting more sustainable and efficient waste-to-energy systems.

1. Introduction

Biogas, as of being one of the renewable and clean energy alternatives, is of high interest in the mitigation of global warming by reducing the use of fossil fuels and greenhouse gas emissions [1]. This is done by reducing methane emissions and utilizing it instead of letting it dissipate into the atmosphere. Methane is a most hazardous greenhouse gas and is the second-largest contributor to global warming [2]. The largest sources of methane are agriculture, fossil fuels, and decomposition of landfill waste. The need to reduce methane emissions through recovery and usage in waste management has therefore grown in interest, with a consequent increase of biogas plants based on anaerobic digestion [1].

Anaerobic digestion (AD) is now widely recognized as a viable and environmentally responsible method for managing organic waste while producing biogas, a gas mainly composed of methane and carbon dioxide [1]. The AD process involves a combination of multiple physical, chemical, and biological processes, producing biogas and digestate. One of the major challenges associated with monitoring and controlling anaerobic digestion processes is that they are highly complex and nonlinear dynamic processes that rely on a delicate balance of various factors for optimal performance [3,4].

Biological systems, such as those involved in anaerobic digestion, often exhibit high variability due to non-stationary disturbances, dynamic interactions between subsystems, and the inherent unpredictability of biological processes. This variability necessitates robust and accurate online estimation techniques for effective model-based monitoring and control. In the context of anaerobic digestion, optimizing biogas production requires reliable online measurement systems. However, since key concentration variables in the digester are difficult to measure online, it is necessary to develop reliable state and parameter estimators based on mathematical models. Mathematical models serve as essential instruments for the numerical simulation of the process, to predict how

* Corresponding author.
E-mail addresses: betlehem@chalmers.se, bethlehem.abera@aau.edu.et (B. Abera), mengesha.mamo@aau.edu.et (M. Mamo), getachew.bekele@aau.edu.et (G. Bekele), torsten.wik@chalmers.se (T. Wik).

https://doi.org/10.1016/j.biombioe.2026.109365
Received 14 January 2026; Received in revised form 31 March 2026; Accepted 31 March 2026
Available online 4 April 2026
0961-9534/© 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/).

a system will behave over time under different conditions, and also for the design of monitoring and control algorithms. The anaerobic digestion model no. 1 (ADM1) developed by the IWA task group [5], is recognized as the most comprehensive and sophisticated model. Though it can assist in the study of the process, it poses challenges in the design of monitoring and control systems. Due to this difficulty, researchers have been interested in the development of simpler models, including those by Hill [6], Dochain et al. [7] and Bernard et al. [8]. The Advanced MOnitoring and COntrol model for anaerobic processes (AMOCO) by Bernard et al. is a particularly promising alternative as it significantly reduces the number of states and parameters compared to ADM1 while still being physics-based. The original AMOCO has then been modified to account for the hydrolysis step, a decay term in the growth rates of the biomasses, and inorganic nitrogen [3,9]. This work highlights the first use of a modified AMOCO model from [3] for the joint estimation of states and parameters in the anaerobic digestion process, as explained in the next section.

Several studies have been conducted regarding the design of estimators for AD process. The Kalman Filter (KF) as in [10] and its nonlinear modifications, such as the Extended Kalman Filter (EKF) [11–14] and the Unscented Kalman Filter (UKF) [12,15] have commonly been used for estimation of states and parameters of the AD. Also other approaches, as the least square regression [8], asymptotic observer [16,17], nonlinear adaptive observer [4], unknown input observer (UIO) [11], extended Luenberger observer [18], particle filter and moving horizon estimator [12] have also been proposed. The estimators developed so far are based on models such as Hill's [11,13], the original AMOCO [4,12], and other reduced-order formulations. Among these, the EKF is the most widely utilized method for state estimation of an AD.

Despite its widespread use, it is well recognized that accounting for model and measurement uncertainties is essential for achieving reliable estimation performance in dynamic systems [19]. In the context of the EKF, these uncertainties are represented through the covariance matrices of process and measurement noises, denoted Q and R, respectively. However, a significant challenge in designing an EKF lies in the considerable impact of these matrices, as well as system parameters, on the estimation accuracy. The matrices Q and R should ideally be determined by considering the stochastic properties of the corresponding noises. However, since these properties are often unknown, the covariance matrices are typically used as tuning knobs. Notably, in previous studies, these matrices were manually tuned using trial-and-error methods, which are proven to be cumbersome and very time-consuming procedures because of the high model order. Automatic tuning is therefore an attractive alternative.

To address the limitations of the original AMOCO, the presence of time-varying parameters, and the challenges in observer tuning, this study explores a modified AMOCO model for joint state and parameter estimation in anaerobic digestion (AD) processes, using a discrete EKF. A sensitivity analysis is first conducted to identify the most influential parameters, from which three critical parameters are selected for online estimation to enhance accuracy and reduce computational complexity. Furthermore, Particle Swarm Optimization (PSO) is introduced to optimize noise covariance matrices (Q and R), thereby eliminating reliance on heuristic tuning.

In addition, robustness of the proposed approach is evaluated through structured measurement noise mismatch scenarios, enabling assessment of estimator performance under realistic uncertainty conditions. The key contribution lies in the integration of EKF joint estimation, PSO-based covariance tuning, and parameter selection based on sensitivity analysis within the modified AMOCO model framework. To the best of the author's knowledge, this combined approach has not been previously reported for this model, which presents increased nonlinear interactions for estimation. The proposed estimation scheme is validated using a high-fidelity benchmark model (ADM1), allowing assessment under realistic nonlinear dynamics and uncertainty.

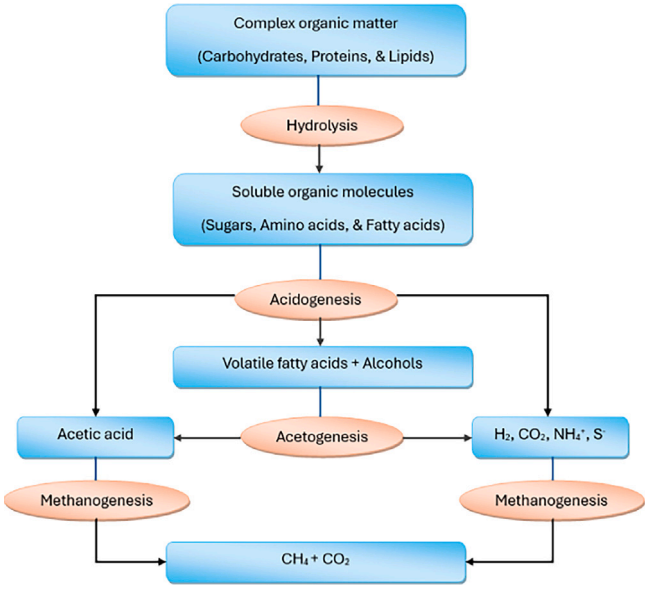

Fig. 1. The four basic stages of AD (Rectangle: intermediate products, oval: stages).

The remainder of this paper is organized as follows: Section 2 presents the anaerobic digestion process models that are used in this paper. Section 3 describes the theoretical background of combined state and parameter estimation. Section 4 details the methodology used in this work. Section 5 presents the simulation results and provides a discussion. Finally, Section 6 concludes the paper.

2. Anaerobic digestion process models

Biogas production from anaerobic digesters is widely used. Anaerobic digestion is a biochemical process where organic matter is decomposed in the absence of oxygen to produce biogas and digestate/bioslurry. Biogas produced in anaerobic digesters consists of 50%–75% methane, 25%–45% carbon dioxide, and small percentages of other gases, such as ammonia and Hydrogen Sulphide [1]. The biogas production is a complex process involving four stages (see Fig. 1), each being crucial for the complete breakdown of the initial organic material. The process begins with large, complex molecules being broken down into smaller ones in a stepwise manner. Specific groups of microorganisms play essential roles in each stage, working in succession to decompose the products generated from the preceding step [20].

The process starts with hydrolysis, where large organic polymers such as carbohydrates, proteins, and fats are broken down into simpler monomers like sugars, amino acids, and fatty acids. Following hydrolysis is acidogenesis, where these monomers are further converted into volatile fatty acids, alcohols, carbon dioxide, hydrogen, and ammonia by acidogenic bacteria [5,20,21].

Next, during acetogenesis, the volatile fatty acids and alcohols produced in the previous step are converted into acetic acid, carbon dioxide, and hydrogen. The final stage, methanogenesis, involves methanogenic bacteria converting acetic acid, hydrogen, and carbon dioxide into methane and water, producing biogas [5,20,21].

To better understand and design monitoring and control algorithms for anaerobic digesters, models based on mass-balance considerations of the above four stages are widely used. The anaerobic digestion model no. 1 (ADM1) developed by the IWA task group [5], is recognized as the most comprehensive and sophisticated such model. ADM1 has subsequently been refined by various authors to enhance accuracy, robustness, and applicability to specific scenarios [20–22]. Although ADM1 is widely used for numerically simulating process behavior, it

poses challenges for practical design and control, due to the need to estimate large number of parameters, approximately a hundred, depending on the specific substrate [9,23,24]. Such challenges has encouraged researchers to find simpler models.

The Advanced Monitoring and Control system for anaerobic processes (AMOCO) [8] has emerged as a promising alternative for monitoring and control system design. The AMOCO considers only two stages, specifically, the acidogenesis and methanogenesis stages. In the acidogenesis stage the acidogenic bacteria x_1 consume the organic substrate and produce CO_2 and VFAs (s_2). In the methanogenesis stage the methanogenic bacteria x_2 uses the VFAs as substrate and generate CO_2 and methane. To address shortcomings in the original AMOCO model though, it has been modified to account for the hydrolysis step, a decay term in the growth rates of the biomasses, and inorganic nitrogen [3,9]. The resulting modified AMOCO is defined by seven dynamic mass balance equations covering various aspects, such as hydrolysis dynamics, bacterial population balances (x_1 and x_2), organic substrate (s_1), VFAs (s_2), alkalinity (s_{alk}), and inorganic carbon (c) concentrations:

$$\frac{dx_1}{dt} = -Dx_1 + \mu_{1,\max}\frac{s_1}{s_1 + k_{s1}}x_1 - k_{d1}x_1 \tag{1}$$

$$\frac{dx_2}{dt} = -Dx_2 + \mu_{2,\max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 - k_{d2}x_2 \tag{2}$$

$$\frac{dx_0}{dt} = D\left(x_{0,in} - x_0\right) - \mu_0 x_0 \tag{3}$$

$$\frac{ds_1}{dt} = D\left(s_{1,in} - s_1\right) + k_0\mu_0 x_0 - k_1\mu_{1,\max}\frac{s_1}{s_1 + k_{s1}}x_1 \tag{4}$$

$$\frac{ds_2}{dt} = D\left(s_{2,in} - s_2\right) + k_2\mu_{1,\max}\frac{s_1}{s_1 + k_{s1}}x_1 - k_3\mu_{2,\max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 \tag{5}$$

$$\frac{ds_{alk}}{dt} = D\left(s_{alk,in} - s_{alk}\right) + \left(k_1 N_{s1} - N_{bac}\right)\mu_{1,\max} \times \frac{s_1}{s_1 + k_{s1}}x_1 - N_{bac}\mu_{2,\max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 + k_{d1}N_{bac}\mu_{1,\max}x_1 + k_{d2}N_{bac}\mu_{2,\max}x_2 \tag{6}$$

$$\frac{dc}{dt} = D\left(c_{in} - c\right) + k_4\mu_{1,\max}\frac{s_1}{s_1 + k_{s1}}x_1 + k_5\mu_{2,\max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 - r_c \tag{7}$$

$$\varphi = c + s_2 - s_{alk} + k_H p_T + \frac{k_6}{k_{La}}\mu_{2,max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 \tag{8}$$

$$r_c = k_{La}\left[\left(c + s_2 - s_{alk}\right) - \frac{\varphi - \sqrt{\varphi^2 - 4k_H p_T(c + s_2 - s_{alk})}}{2}\right] \tag{9}$$

$$r_{\mathrm{CH4}} = k_6\mu_{2,max}\frac{s_2}{s_2 + k_{s2} + \frac{s_2^2}{k_{I2}}}x_2 \tag{10}$$

where μ_0 (d^{-1}) is the specific hydrolysis rate, k_i ($i = 0, \ldots, 6$) represent stoichiometric coefficients, and $\mu_{i,\max}$ ($i = 1, 2$) (d^{-1}) denote the maximum specific growth rates of the bacteria. k_{s1} ($gCODL^{-1}$) and k_{s2} (mmol L^{-1}) are half-saturation constants, and k_{I2} (mmol L^{-1}) is an inhibition constant. k_H represents Henry's constant for CO_2 (mmol L^{-1} atm^{-1}), and, p_T, i.e. atmospheric pressure, is set to 1 atm. k_{d1} and k_{d2} are decay rate coefficients for the bacteria, N_{s1} is nitrogen content of s_1, and N_{bac} denotes the nitrogen content in the biomass. The parameter k_{La} (d^{-1}) denotes the liquid–gas transfer coefficient, while r_c (mmol $L^{-1}d^{-1}$) quantifies the CO_2 production rate and r_{CH4} (mmol $L^{-1}d^{-1}$) represents the methane production rate. The component concentrations in the influent stream are $x_{0,in}$, $s_{1,in}$, $s_{2,in}$, $s_{alk,in}$, and c_{in}.

3. Joint state and parameter estimation

Most of the states in both ADM1 and AMOCO based models are concentrations, which often cannot be directly measured by online sensors. The inherent uncertainties and inaccuracies in the parameters assumed in anaerobic digestion models are also challenges in the design of monitoring and control systems. This calls for a design of combined state and parameter estimators.

The state of a system typically consists of quantities that change rapidly, while the time-varying parameters in biological systems typically change slowly. This difference should open up for effective estimation methods. Traditionally, the state estimation and parameter identification problems have been seen as distinct problems [25], usually, handled in three ways. The most common approach is to first identify the parameters in an offline identification and then use the identified model for real-time state estimation. The second approach is to design dual filters [26], in which the estimation and identification are achieved through iterations between two filters; one for state estimation, and the other for parameter estimation. The third method is to design a joint estimator in which the state vector is augmented with parameters to form an augmented state space [4]. Hence, this approach treats the estimation and identification problems simultaneously in one single observer. The disadvantages of this are large matrix operations, due to higher dimensionality of the resulting augmented model, and potentially poor numeric conditioning due to vastly different time scales of the states (including parameters) in the augmented state vector. However, the joint estimation of states and parameters takes advantage of their interdependence, resulting in more accurate estimates than separate estimations. This is particularly advantageous in systems where parameters affect state dynamics and vice versa. It also improves robustness and flexibility in estimating parameters and states of nonlinear systems [27].

The EKF has shown promising results in state estimation in various applications related to anaerobic digestion [11–14]. For a nonlinear system defined by

$$\dot{x}(t) = f\left(x(t), u(t), \theta\right) + w_x(t) \tag{11}$$

$$y(t) = g\left(x(t), u(t), \theta\right) + v(t), \tag{12}$$

where x is the state, u is the input to the system, θ corresponds to the parameters of the system, f describes the dynamics of the system, g represents the output equation, and w and v are process and measurement noises with zero mean and covariances Q and R, respectively. It is assumed that the process and measurement noises are uncorrelated. Now, the parameter vector is split into parameters that vary with time, θ_z, and the remaining parameters $\bar{\theta}$ that are assumed to be unchanged. As the parameters of the system we want to estimate are assumed to have very slow variations in time, their dynamics is approximated by using

$$\dot{\theta}_z(t) = w_\theta(t). \tag{13}$$

Defining the augmented state $z(t) = \left[x(t)\ \ \theta_z(t)\right]^T$ results in new state and output equations given by

$$\dot{z}(t) = F(z, u, t) + w_z(t) \tag{14}$$

$$y(t) = G(z, u, t) + v(t), \tag{15}$$

where the constant parameters $\bar{\theta}$ are incorporated in F and G.

The system matrix $\phi(t)$ and the observation matrix $H(t)$ are the jacobians of $F(\cdot)$ and the measurement function $G(\cdot)$, i.e. $\phi(t) = \frac{\partial F(\cdot)}{\partial z}$ and $H(t) = \frac{\partial G(\cdot)}{\partial z}$. To implement the EKF in practical applications, where measurements and control inputs are typically available at discrete time intervals, the continuous-time model is discretized. The continuous-time state $z(t)$ is then sampled at fixed time intervals, resulting in a discrete-time state representation z_k, where k denotes the kth time-step. Similarly, the inputs $u(t)$ and outputs $y(t)$ are expressed as u_k and y_k, respectively.

The discretization of (14) and (15) then gives

$$z_k = F_k\left(z_{k-1}, u_{k-1}\right) + w_{k-1} \tag{16}$$

$$y_k = G_k(z_k, u_k) + v_k. \tag{17}$$

This discretization enables a recursive implementation of an EKF using discrete measurement data. The EKF mainly consists of two iterative steps. The first one is the time update (prediction) which propagates the state estimate forward in time. The apriori state estimate $\hat{z}_k^-$ and error covariance P_k^-, are computed through

$$\hat{z}_k^- = F_k\left(\hat{z}_{k-1}, u_k\right) \tag{18}$$

$$P_k^- = \phi_k P_{k-1} \phi_k^T + Q. \tag{19}$$

The second step is the measurement update (correction) updating the state estimate and error covariance through

$$\hat{z}_k = \hat{z}_k^- + K_k[y_k - G(\hat{z}_k^-, u_k)] \tag{20}$$

$$P_k = (I - K_k H) P_k^-, \tag{21}$$

where K_k is the Kalman gain, given by

$$K_k = P_k^- H^T [H P_k^- H^T + R]^{-1}. \tag{22}$$

The EKF initialization involves setting the initial state estimate $\hat{z}_0$ and error covariance matrix P_0, which are critical for ensuring filter convergence and stability [28].

The performance of the EKF depends critically on model accuracy, noise characterization, and the proper tuning of the covariance matrices Q and R to ensure stable and reliable estimates. The following section outlines the methodology employed to implement and assess this approach for an anaerobic digestion process.

4. Methodology

This section presents the methodological framework for developing and evaluating the proposed joint state and parameter estimation scheme using a discrete EKF. The procedure includes defining system inputs and outputs, selecting key parameters based on their influence, and designing the EKF using the modified AMOCO-based model, where the plant is represented by the same model but with introduced uncertainties. The process and measurement noise covariance matrices Q and R are then tuned to ensure robust performance. To assess its practical applicability, the designed EKF is further validated using the ADM1 as the "true plant". In this context, the "true" plant represents a simulated reference system, and the corresponding values are model-generated reference data.

4.1. Inputs and outputs

The input signal used in an identification experiment can have a significant influence on the resulting parameter estimates [29]. In this work, the concentrations of each component in the influent stream, denoted as $x_{0,in}$, $s_{1,in}$, $s_{2,in}$, $s_{alk,in}$, and c_{in}, serve as the model inputs. Unfortunately, these model inputs can typically not be manipulated individually. Therefore, we apply a simultaneous pulsating square wave, corresponding to two different dilution levels of the feed to the digester. To decide on the frequency of the square wave, we investigated the speed of response of the states under the effect of two different input frequencies, corresponding to square wave pulses with periods of 0.5 days and 5 days, as shown in Fig. 2. It is seen that the concentrations of the acidogenic and methanogenic bacteria are the slowest, hardly showing any effect of high frequency input (short period). The alkalinity and inorganic carbon exhibit slightly faster response than the bacterial biomasses, but slower than the particulate organic matter, soluble organic matter, and VFAs. In most cases the input should emphasize the low-frequency properties of the system and hence should have a rich content of low frequencies [29].

From this we take the input that corresponds to addition of 20 percent to the initial concentration of substrates every five days, where the initial steady-state inputs are $u_0 = [x_{0,\ in} \quad s_{1,in} \quad s_{2,in} \quad s_{alk,in} \quad c_{in}]^T = [55 \quad 73 \quad 380 \quad 580 \quad 200]^T$, corresponding to influent composition of a cow manure [30].

The measurable outputs considered here are limited to the carbon dioxide production rate (r_c), the methane production rate (r_{CH4}), and the alkalinity (s_{alk}).

4.2. Selection of parameters for estimation

Even though the modified AMOCO is a simplified model, it still has seven states and twenty parameters. The designed joint estimator will not be able to accurately estimate all the parameters together with the seven states, i.e. if the augmented state have a total of twenty-seven variables, from the output measurements considered in this paper. Hence, to increase the robustness and accuracy of the estimator, we performed a sensitivity analysis to determine the most influential parameters of the model.

Sensitivity analysis involves a sequence of experiments where input values are varied around a central value, within specified limits, to observe how alterations in inputs influence changes in the model outputs [31]. There are several methods for performing sensitivity analysis. According to Zhou and Lin [31], sensitivity analysis methods are classified into three types: screening design method, local sensitivity analysis, and global sensitivity analysis. Various sensitivity analysis methods to determine the most influential parameters in anaerobic digestion models have been applied. Global sensitivity analysis techniques, such as the Sobol index, have been used to quantify the impact of individual parameters and their interactions on model outputs, like methane production [32]. Screening methods, such as Morris screening and standardized regression coefficient analysis, have been utilized to detect the most significant parameters in complex, plant-wide models, thereby supporting model calibration and optimization [33].

To explore the sensitivity of the system to individual parameters, a local one-at-a-time perturbation method with a 50% variation was adopted to limit computational cost, although it does not capture interactions between parameters. To partially account for such interactions, the off-diagonal elements of the process noise covariance matrix Q corresponding to the parameters were tuned, allowing the EKF to consider correlated variations between parameters. We solved the state and output equations before and after perturbation and calculated the normalized sensitivity values. In online estimation of parameters we focus on the ones that have impact on the variables r_C and r_{CH_4}. These two variables show high sensitivity to the yield coefficients k_1, k_2, k_3, k_5, and k_6. As we are trying to reduce the number of parameters, we used period averaging of sensitivity values over the last complete period of the input signal. By defining a threshold value of 0.5, i.e. 50% of the highest sensitivity values, we identified k_3, k_5, and k_6 to have a high influence on the production rates of CO_2 and CH_4. Details of the parametric sensitivity analysis of the modified AMOCO model are presented in [34].

To complement the sensitivity analysis, we assessed the practical identifiablility of the parameters k_3, k_5, and k_6 using profile likelihoods.

4.3. Implementation of the discrete joint EKF

An initial attempt was made using dual EKF, which decouples estimation of states and parameters by employing two EKFs. This typically involved four steps: parameter prediction, state prediction, state correction, and parameter correction. However, preliminary results showed limited performance, primarily because the parameters were not sufficiently updated by the available measurements, i.e. r_c and r_{CH4}, in the context of this specific process. Thus, a joint estimation was implemented, instead.

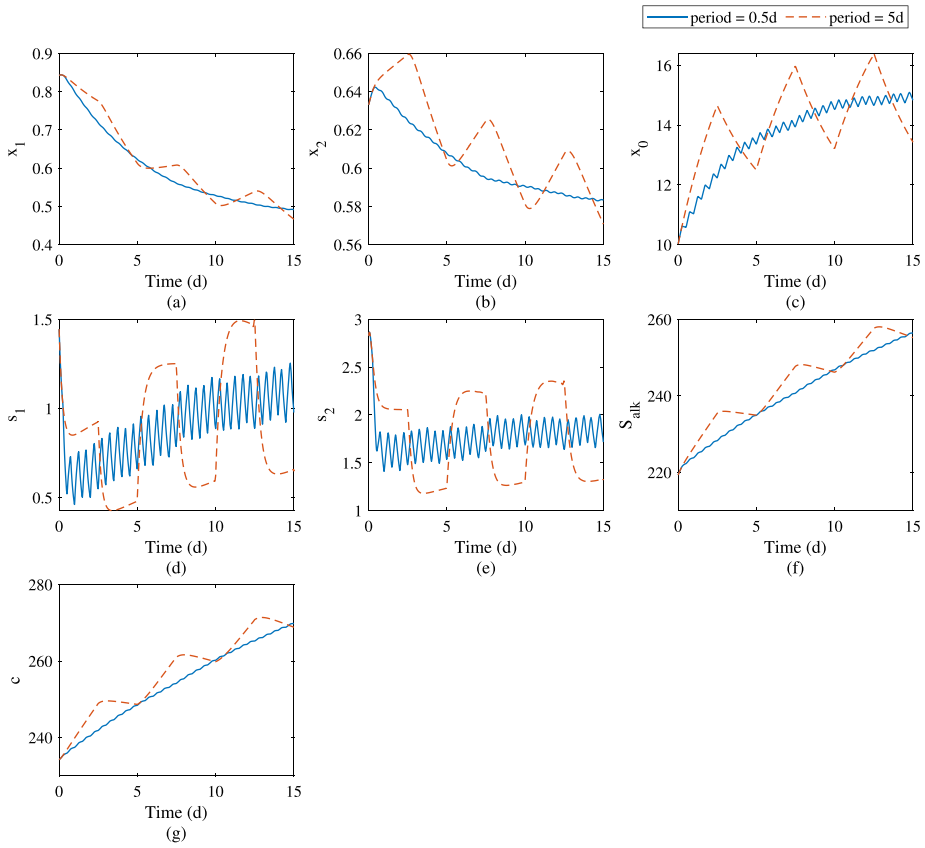

Fig. 2. Comparison of state responses under the influence of square-wave inputs with periods of 0.5 day and 5 days.

As described in Section 3, the implementation of the EKF consists of two steps: prediction of the next state estimate and covariance based on the dynamic system model, and update based on measurement information. Prior to these steps one needs to specify the initial augmented state vector ($\hat{z}_0$) and initial state estimation error (P_0), process noise (Q_{init}), and measurement noise (R_{init}) covariance matrices. The initial states are adapted from the work of Rodriguez et al. [4] and parameters used in the simulation are taken partially from Bernard et al. [8] and Della Bona et al. [3]. The choice of initial state estimation error covariance requires careful consideration. A small P_0 results in a low Kalman gain (K), causing the parameter estimates to evolve slowly from $\hat{z}_0$. Conversely, a large P_0 leads to a higher Kalman gain, allowing estimates to quickly jump away from $\hat{z}_0$, potentially promoting a faster initial convergence. The matrices P_0, Q_{init} and R_{init} are selected based on the formulas suggested by [15,35]:

$$P_{init} = \text{diag}\left(k_p \hat{z}_0^2\right) \tag{23}$$

$$Q_{init} = \text{diag}\left(k_q^2\, m_i \hat{z}_0^2\right) \tag{24}$$

$$R_{init} = \text{diag}\left(k_R^2 y_{init}^2\right) \tag{25}$$

where the values of k_p, k_q, k_R, and m_i are tuned by trial and error. y_{init} is the first instance of the measured output.

In the prediction and update step the nonlinear system is linearized at the estimate and then discretized. The details are presented as follows.

4.3.1. Problem formulation

The parametrized modified AMOCO model (1)–(7) can be expressed as a general nonlinear dynamic system (14), where the parameter vector is defined as

$$\theta_z = [k_3 \quad k_5 \quad k_6]^T,$$

and the augmented state to be estimated is consequently

$$z(t) = [x_1 \quad x_2 \quad x_0 \quad s_1 \quad s_2 \quad s_{alk} \quad c \quad k_3 \quad k_5 \quad k_6]^T.$$

The output of the system, $y(t)$, consists of three components: the carbon dioxide production rate (r_c), the methane production rate (r_{ch4}), and the alkalinity (z). These rates are calculated using (9) and (10), generally represented as (15), based on the current state and parameter estimates.

4.3.2. Linearization about an estimate

The EKF relies on linearization about the estimated trajectory. To achieve this, we apply Jacobian linearization about the estimate.

If the solution of a nonlinear state-space model remains near an operating point ($\bar{z}$, $\bar{u}$), then the dynamics can be approximated by a linear state space model of the form

$$\dot{z}(t) = A\ z(t) + B\ u(t) \tag{26}$$

$$y(t) = C\ z(t) + D\ u(t), \tag{27}$$

where

$$A = \frac{\partial F}{\partial z}\Big|_{u(t)=\bar{u}(t)}^{z(t)=\bar{z}(t)}, \qquad B = \frac{\partial F}{\partial u}\Big|_{u(t)=\bar{u}(t)}^{z(t)=\bar{z}(t)},$$

$$C = \frac{\partial G}{\partial z}\Big|_{u(t)=\bar{u}(t)}^{z(t)=\bar{z}(t)}, \quad \text{and} \qquad D = \frac{\partial G}{\partial u}\Big|_{u(t)=\bar{u}(t)}^{z(t)=\bar{z}(t)}$$

Here, the operating point is the current estimate $\bar{z} = \hat{z}$ and the input $\bar{u} = \hat{u}$. The Jacobian function in MATLAB's Symbolic Toolbox is used to linearize the process model and obtain the matrices A, B, C, and D at every sample.

4.3.3. Discretization of the model

As the implemented estimator is a discrete extended Kalman filter, the modified AMOCO model also needs to be discretized. The linearized

model about an estimate is linearized using the Euler discretization method with a constant sampling time t_s and is given by

$$z_{k+1} = z_k + F\left(\hat{z}_k, u_k\right) t_s \tag{28}$$

$$y_k = G\left(z_k\right) + v_k \tag{29}$$

$$\phi_k = I_n + t_s A \tag{30}$$

A sampling time (t_s) of 0.01 days is selected to ensure accurate tracking of the system's dynamics with the convergence requirements of the EKF, without introducing unnecessary computational load.

4.4. Tuning of Q and R

In the design of the EKF for nonlinear systems, the choice of process noise covariance matrix Q and the measurement noise covariance matrix R should be done carefully to ensure good performance of the estimator, commonly referred to as tuning. The process noise covariance matrix Q represents the uncertainty in the system model or process dynamics. It accounts for unmodeled effects or disturbances that affect the state transition. The measurement noise covariance R represents the uncertainty in the sensor measurements. It determines how much the filter trusts the sensor measurements compared to the predicted state estimates.

Roughly, when Q is large relative to R, the filter assumes large uncertainty in the model and will trust the measurements more, even if they are noisy, which can lead to jumpy estimates. This is an undesirable situation as the purpose of the EKF is to optimally combine the model predictions and measurements to obtain smooth and accurate state estimates. Conversely, if R is large relative to Q, the filter will trust the process model more than the measurements. This leads to smooth estimates, but incorrect if the model is incorrect.

The ideal scenario is to have well-tuned Q and R matrices that accurately reflect the real noise characteristics of the sensors and process model. This allows the EKF to strike the right balance between following the measurements and relying on the model predictions, resulting in smooth state estimates that reject measurement noise while still being responsive to real changes in the state [36].

The process and noise covariance matrices, Q and R respectively, were first manually tuned using trial-and-error methods, which proved to be hard and very time-consuming procedures. To avoid this, employing modern optimization methods such as particle swarm optimization (PSO) for the tuning of Q and R can be a solution.

PSO is a powerful algorithm that employs a swarm of particles to search for the optimal solution in a multidimensional space [37]. The algorithm initializes N randomly positioned particles within a defined range of possible solutions, typically denoted as $[lb\ ub]$ for the lower and upper bounds. Each particle i in the swarm serves as an agent that explores the solution space, with its position p_i and velocity v_i updated at each iteration q of the optimization process. During each iteration, a particle evaluates the objective function V at its current position p. Basically, each particle maintains a memory of its own best p_{best}. This personal best is replaced whenever the particle finds a position that yields a better fitness value, that is, if $V(p) < V(p_{best})$. The algorithm also typically tracks the globally best position (G_{best}) achieved by any particle in the swarm. The movement of a particle is governed by a velocity update formula that considers the particle's current velocity, its personal best position, and the global best position of the swarm. The standard velocity update equation for particle i is

$$v_i\left(q+1\right) = w v_i\left(q\right) + c_1 r_1\left(p_{best,i}\left(q\right) - p_i\left(q\right)\right) + c_2 r_2\left(G_{best}\left(q\right) - p_i\left(q\right)\right),$$

where w is the inertia coefficient, c_1 and c_2 are acceleration coefficients, and r_1 and r_2 are random values [38,39].

After updating the velocity, the position of each particle is adjusted using the following equation:

$$p_i(q+1) = p_i(q) + v_i(q+1).$$

This process continues for a predetermined number of iterations or until a stopping criterion is met, at which point the algorithm returns the best solution found. The effectiveness of PSO stems from its ability to balance exploration of the search space with exploitation of known good solutions, making it adaptable to a wide range of optimization problems.

In this work, the novel approach is the integration of particle swarm optimization (PSO) with EKF for application on anaerobic digestion process. The implementation utilizes the PSO solver in MATLAB to tune the individual elements of the covariance matrices Q (process noise) and R (measurement noise) offline. Specifically, two PSO runs are employed: the first PSO optimizes the diagonal entries of Q and R, while the second focuses on the off-diagonal elements of Q associated with the parameters. To obtain suitable values for Q and R used in the estimator, synthetic data was generated using the nonlinear dynamic model under steady-state operation and time-varying influent conditions, with known parameter values and input profiles. The simulation was conducted over a time horizon of 100 days with a sampling interval of 0.01 day, resulting in 10,000 samples. To emulate realistic measurement conditions, Gaussian noise was added to the simulated outputs.

Each particle i in the swarm represents a candidate set of entries for Q and R, which are critical to balancing prediction and measurement updates in the EKF framework. The PSO configuration adopted in this work uses MATLAB's `particleswarm` function with a swarm size of 100 particles and a maximum of 400 iterations. The inertia is controlled using an inertia range of 0.1 – 1.1, and both self- and social- adjustment coefficients set to 1.49.

When setting up the PSO, the formulation of the cost function plays a critical role in the estimation performance. In particular, proper scaling of the variables is necessary to prevent any single variable from dominating the cost function due to differences in magnitude. To address this, normalization is applied to ensure balanced contributions across variables.

In addition, appropriate lower and upper bounds must be defined for the search space. The estimator was found to be sensitive to values of Q and R. Improper bounds may easily lead to numerical instability and unreliable estimates, whereas appropriately chosen bounds improve convergence and robustness of the optimization process. Additional details on scaling and bounds selection are provided in the Supplementary Material (Table S1).

The objective function is defined as:

$$\begin{aligned} &\min \quad V\left(p\right) \\ &\text{subject to} \quad lb \le p \le ub \end{aligned} \tag{31}$$

where V is defined as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{10}\sum_{j=1}^{10}\frac{\left|\hat{z}_j - z_{ref,j}\right|}{\max\left(z_{ref,j}\right) - \min\left(z_{ref,j}\right)},$$

where $\hat{z}_j$ is the estimated augmented state, based on the candidate Q and R, and $z_{ref,j}$ is the corresponding model-generated reference trajectory generated by the simulation model. Each error term is therefore scaled by the range of the reference trajectory. During optimization, the PSO algorithm evaluates these diagonal entries by propagating their corresponding covariance matrices through the prediction and correction steps of the EKF. The best performing particles, those producing the smallest state and parameter estimation errors, guide the swarm's search towards optimal matrices Q and R. The numerical values of the optimal covariance matrices are provided in Supplementary table S2.

4.5. Validation by considering ADM1 as a true plant

In the previous section, an EKF was designed using the modified AMOCO model, with the real plant being taken as the same model but with some uncertainties. However, it is crucial to validate the designed

EKF to determine its practicality for implementation in a real anaerobic digester facility. Therefore, this section discusses the validation of the EKF.

As a replacement for a real plant, the ADM1 is used, specifically the version in [20]. The Anaerobic Digestion Model no. 1 (ADM1) is the most well-known and complex model developed to describe the anaerobic digestion process, and hence can be used as a benchmark for a real plant. This ADM1 implementation includes a system of equation with 35 differential and 1 algebraic equation. In contrast, the modified AMOCO model consists of only 7 differential equations, corresponding to the seven states and 2 algebraic equations. This reduction is achieved through state aggregation and simplification of the underlying biochemical processes, where several state variables and interactions present in ADM1 are lumped together in the modified AMOCO formulation. Consequently, certain dynamic pathways and nonlinear interactions captured in ADM1 are not explicitly represented, resulting in structural and parametric differences between the model used for estimation and the reference system employed for simulation.

To interface the two models, we first simulated the ADM1 model using the complete set of parameters and influent conditions reported in [20]. The resulting ADM1 state trajectories served as the basis for constructing the corresponding AMOCO state variables. These AMOCO-equivalent states were obtained by applying the conversion formulas given in [9] with a slight modification to account for the hydrolysis stage [3] (see Appendix A).

Next, we obtain equivalent parameters of the AMOCO from the ADM1 simulation. Many AMOCO parameters, like the liquid–gas transfer coefficient (k_{La}) and biomass decay rate can be directly transferred from the ADM1 parameter set. Other parameters like N_{s1} and k_H are adapted from [3]. The remaining set of parameters, for which no direct correspondence exists, are identified using a hybrid optimization approach, to solve the following problem:

$$\min_{\theta} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{1}{9}\sum_{j=1}^{9}\frac{|m_{amoco}-m_{adm1}|}{\Delta m_j} \tag{32}$$

subject to $lb \le \theta \le ub$

where

$$\theta = [k_1 \quad k_2 \quad k_5 \quad k_6 \quad k_{I2} \quad k_{s1} \quad k_{s2} \quad \mu_{1,max} \quad \mu_{2,max}]^T,$$
$$m = [x_1 \quad x_2 \quad x_0 \quad s_1 \quad s_2 \quad s_{alk} \quad c \quad r_c \quad r_{CH4}]^T$$

with m_{amoco} and m_{adm1} being the simultaneous corresponding vectors of m by the modified AMOCO and by the ADM1. n is the number of samples, Δm_j is a normalizing vector used to handle that the magnitude of m vary across the variables. In this case the range of m_{adm1} is used, i.e. $\Delta m_j = \max(m_{adm1,j}) - \min(m_{adm1,j})$, to normalize each variable in m.

The lower and upper bounds, lb and ub, for the search space of the parameters, θ, are based on the results of the parameter identification by Della Bona et al. [3].

The modified AMOCO is simulated using the `ode15s` solver in MATLAB taking initial states equivalent to the ones used in the ADM1 simulation, using the conversion formulas in Appendix A. The inputs for simulation of ADM1 are the steady-state inputs presented by Rosen et al. [20] and with 20% increased concentrations for five days every ten days. The inputs used in simulations of AMOCO are calculated using the corresponding formulas in Appendix A. Hence, the first seven elements of m_{amoco} are the solutions of the `ode15s` solver and the last two variables are obtained using the algebraic equations stated in the modified AMOCO model.

Some mismatch between the modified AMOCO model and the ADM1 simulation may be explained by differences in the initial states and the nature of the input concentrations used. These differences can affect the performance of the estimator and should be taken into account when interpreting the results.

Table 1
Optimal values of modified AMOCO equivalent ADM1 parameters.

Parameters	k_1	k_2	k_5	k_6	k_{I2}
Optimal values	11.82	167.7	105.96	482.87	20.09
Parameters	k_{s1}	k_{s2}	$\mu_{1,max}$	$\mu_{2,max}$	
Optimal values	1.6	8.82	1.01	0.27	

The parameters in Table 1 were obtained using a hybrid optimization strategy. A population-based Genetic Algorithm (GA) was first employed to perform a global search of the parameter space and identify a promising region. The solution obtained from the GA was then used as the initial guess for the constrained optimization solver `fmincon`, which refined the parameters to a local optimum of the problem in (32). This approach mitigates sensitivity to initialization and improves convergence reliability for the nonlinear optimization problem. The consistency of the obtained parameter set was confirmed by the fact that the GA and the subsequent `fmincon` refinement converged to very similar parameter values.

Once the modified AMOCO equivalent ADM1 state variables and solutions of the algebraic equations, which are taken as measured outputs, are obtained, the next step is to validate the EKF estimator. This validation involves three main steps: predicting the next state using the modified AMOCO model, updating the estimate with new measurements, and correcting it based on the difference between the prediction and the 'measured' outputs from the ADM1 simulation.

4.6. PSO-based covariance tuning and robustness analysis

In practical systems, measurement noise may vary due to sensor quality, environmental conditions, or instrumentation errors. Therefore, it is important to assess the estimator's performance under different noise levels.

The measurement noise is modeled as a zero-mean Gaussian noise:

$$v = \alpha w, \qquad w \sim \mathcal{N}(0, I)$$

where the scaling factor α determines the covariance $R = \alpha^2 I$.

4.6.1. PSO-based covariance tuning

To evaluate the capability of the proposed PSO-based framework to systematically tune the EKF covariance matrices for a given noise scenario, three measurement noise levels were considered:

Case 1: Nominal Noise

$$\alpha = t_s k_r / \sqrt{2}$$

This case corresponds to the nominal measurement noise level used in the baseline simulations.

Case 2: Moderate Noise

$$\alpha = t_s / \sqrt{2}$$

This scenario represents a moderate increase in measurement noise relative to the nominal assumption.

Case 3: Severe Noise

$$\alpha = 1 / \sqrt{2}$$

This case introduces a significantly larger noise level to assess estimator performance under challenging conditions.

For each scenario, the noisy measurements were used as input to the EKF, and the covariance matrices Q and R were re-optimized using PSO. Unlike conventional EKF approaches, where covariance matrices are empirically chosen and fixed, the proposed PSO-based framework systematically adjusts the covariance matrices for each noise scenario. The obtained simulation results, including state estimates, output predictions, and residual analysis, demonstrate that the estimator remains stable and accurate across all tested noise intensities.

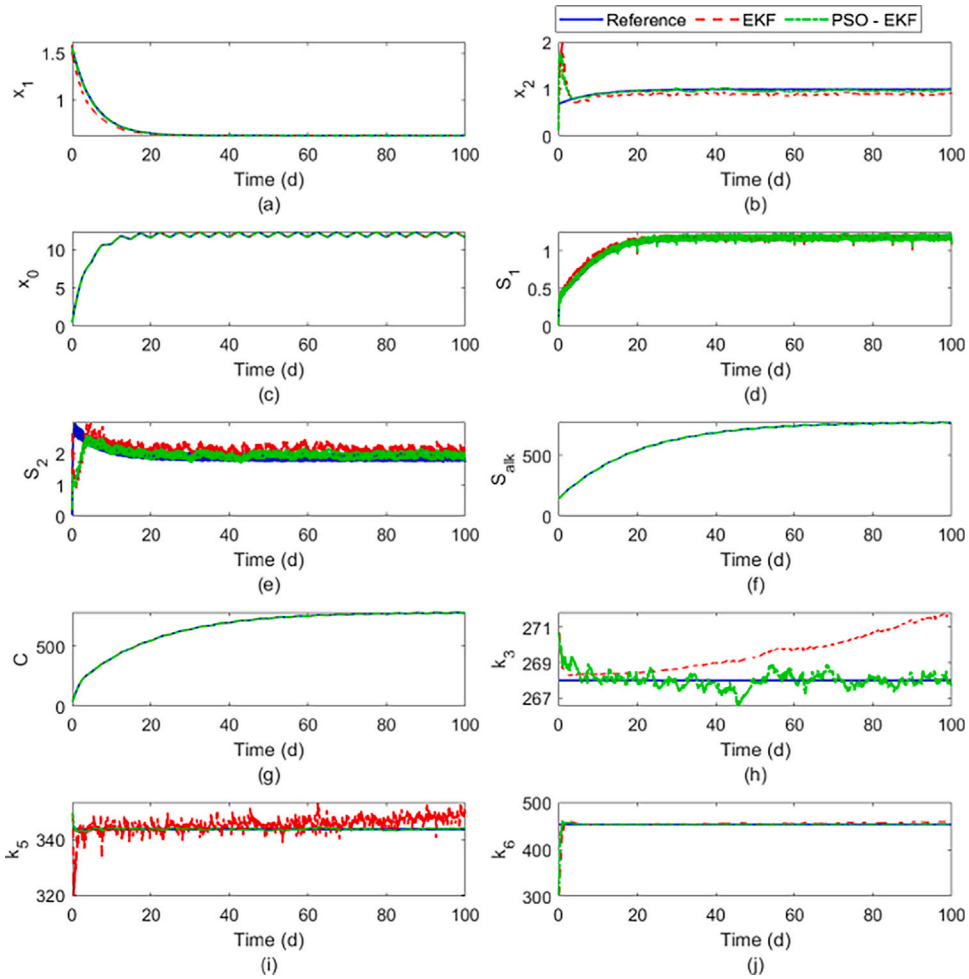

Fig. 3. Estimated states and parameters obtained by simulating the modified AMOCO model as a real plant. Subfigures (a) – (g) show state estimates, and (h) – (j) show parameter estimates. model-generated reference values represented by solid blue lines, manually tuned EKF estimates by red dashed lines, and PSO - tuned EKF estimates by green dash-dotted lines.

4.6.2. Robustness to noise mismatch

To evaluate estimator robustness, the EKF tuned for the nominal noise level was tested under different noise intensities without re-optimizing Q and R. This experiment assesses the robustness of the estimator to a mismatch between assumed and actual noise statistics. The same nominal noise definition to Section 4.6.1 is used for this robustness analysis, and different noise scenarios were generated by scaling the nominal noise by a factor of 0.5, 1, 3, and 10, corresponding to lower than nominal noise, nominal noise, moderate noise increase, and high noise intensity, respectively.

5. Results and discussion

In this section, the results of the state and parameter estimation using the modified AMOCO model as a true plant are presented first and then the results for the validation of the estimator applied to the ADM1 model as a true plant are discussed. The profile likelihood analysis showed that k_3 and k_6 are practically identifiable, exhibiting clear and well-defined minima. However, k_5 is weekly identifiable, with a flat profile region indicating that the current measurements do not fully constrain its value. All parameters were restricted to positive and physically realistic bounds, which prevents nonphysical parameter drift and ensures that parameter estimates remain within biologically meaningful ranges. The state and parameter estimation results are obtained under these identifiability and constraint considerations.

As the EKF is derived from the modified AMOCO model, the first implementation is based on simulations of the modified AMOCO model as a real plant. A small amount of measurement noise is added to test the robustness of the estimator.

In Fig. 3, estimation of the augmented state variables is presented with the comparison of the estimation based on manual tuning and PSO tuning. These plots confirm that the EKF is capable of estimating both the states and parameters of the anaerobic digestion process using the modified AMOCO model. As shown, manual tuning of the noise covariance matrices requires several trial-and-error iterations to achieve convergence across all estimations. Even after multiple attempts, the estimation of k_3 fails to converge to the model-generated reference value, and instead diverges. Although further manual tuning could improve this, it is difficult and highly time consuming. In contrast, using PSO not only reduces the estimation time, but also enhances estimation accuracy.

Although these results are promising, several important considerations must be kept in mind. It is generally advisable to keep the number of variables to be estimated as small as possible because estimating too many variables at once usually reduces robustness and may risk instability. In addition, the effect of sampling time should also be taken into account. Larger sampling intervals risk missing key system dynamics, while very small intervals can unnecessarily increase computational load without significant gains in accuracy.

To evaluate the accuracy of the estimation, the plots of the residuals, i.e. the difference between the true and estimated augmented state,

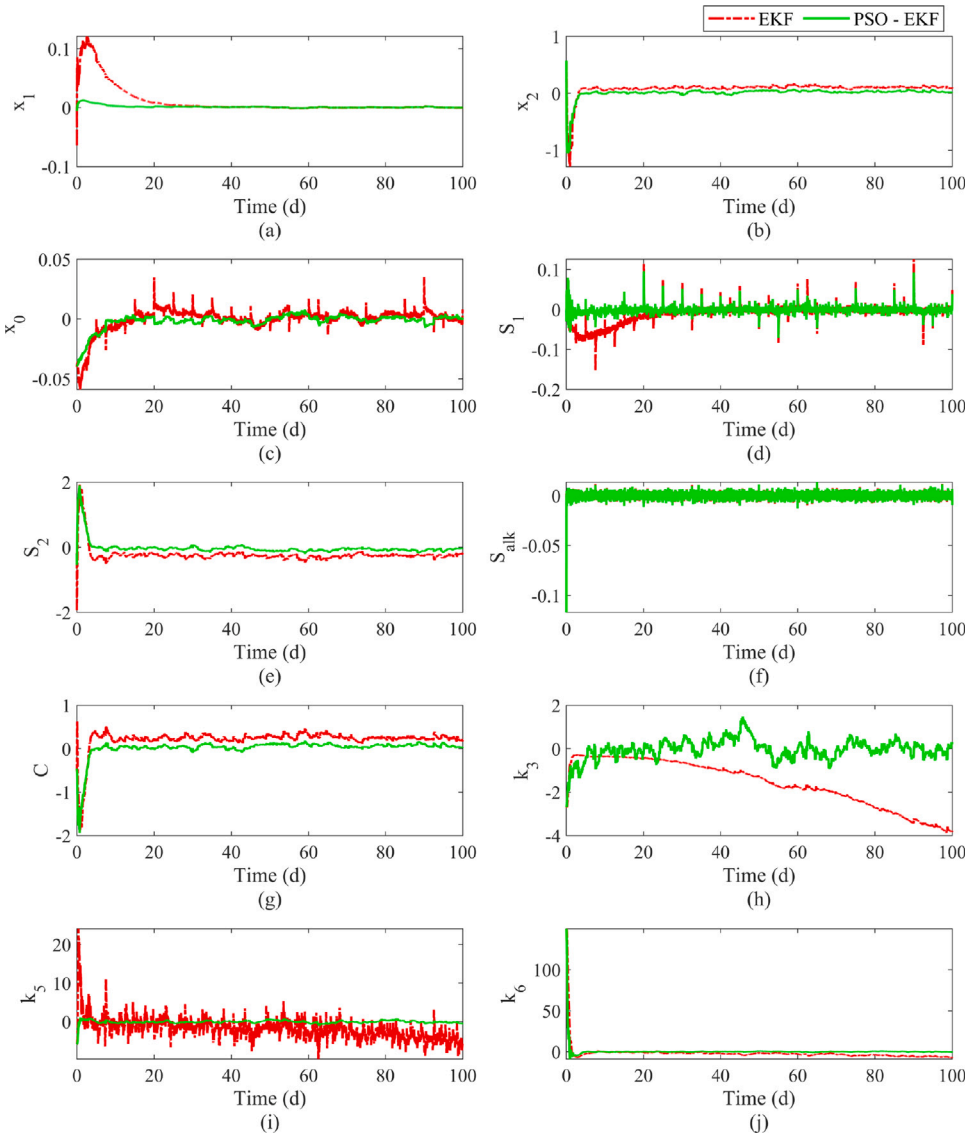

Fig. 4. Estimation error with respect to the model-generated reference values generated by the modified AMOCO model. The red dash-dotted lines represent error using the manually tuned EKF, whereas solid green lines show errors using the PSO-tuned EKF.

are presented in Fig. 4. In addition to the overall reduction in residual magnitude, Fig. 4 reveals clear differences in the transient and steady-state behavior of the two estimators. The manually tuned EKF exhibits more pronounced oscillations in several states, particularly during the initial transient, and in some cases retains a small bias throughout the simulation. By contrast, the PSO-tuned EKF produces faster error attenuation and noticeably smoother trajectories, indicating improved numerical stability. These observations demonstrate that PSO tuning enhances both the convergence characteristics and robustness of the filter across all components of the augmented state. Fig. 5 shows how closely the outputs, based on the estimated augmented states, align with the true outputs. This close match confirms the effectiveness of the EKF, with further improvements seen when PSO tuning is applied.

Although three noise scenarios were tested, detailed time/series plots are presented only for the nominal noise scenario due to space limitations, while quantitative performance metrics for all noise levels are summarized in Table 2. When applying the EKF to ADM1 data, as shown in Fig. 6, the PSO-tuning significantly improves the convergence of the states towards the ADM1 simulation results. The estimated parameters using PSO-tuned EKF towards the 'true' values also demonstrates relatively faster convergence than the manually tuned EKF. The time series convergence is further clarified by plotting the residuals between the ADM1 simulations and the estimated states and parameters, as presented in Fig. 7. These error plots demonstrate the consistency of the estimation.

One possible reason for the lack of convergence in parameter k_5 (see Fig. 6i) estimation is the limited availability of data for parameter estimation. Specifically, the parameter identification in Section 4.5 is performed by using the seven states variables along with the production rates r_c and r_{CH4}, whereas, estimation relies only on alkalinity, s_{alk}, and the production rates, which may not provide sufficient information to accurately infer all parameters. Additionally, the parameters k_3, k_5, and k_6, are often assumed to be fixed stoichiometric coefficients, while in reality they depend on the substrate composition. This dependence introduces variability that adversely affects the accuracy of the parameter estimation.

Fig. 8 extends the comparison presented in Fig. 5 by showing how closely the outputs based on the estimated augmented states match those of the ADM1 model. While both filters reproduce the oscillatory behavior well, the PSO-tuned EKF provides a slightly more accurate fit. These results confirm that the improved estimation accuracy obtained via PSO tuning directly translates into more accurate output reconstruction under noisy measurement conditions.

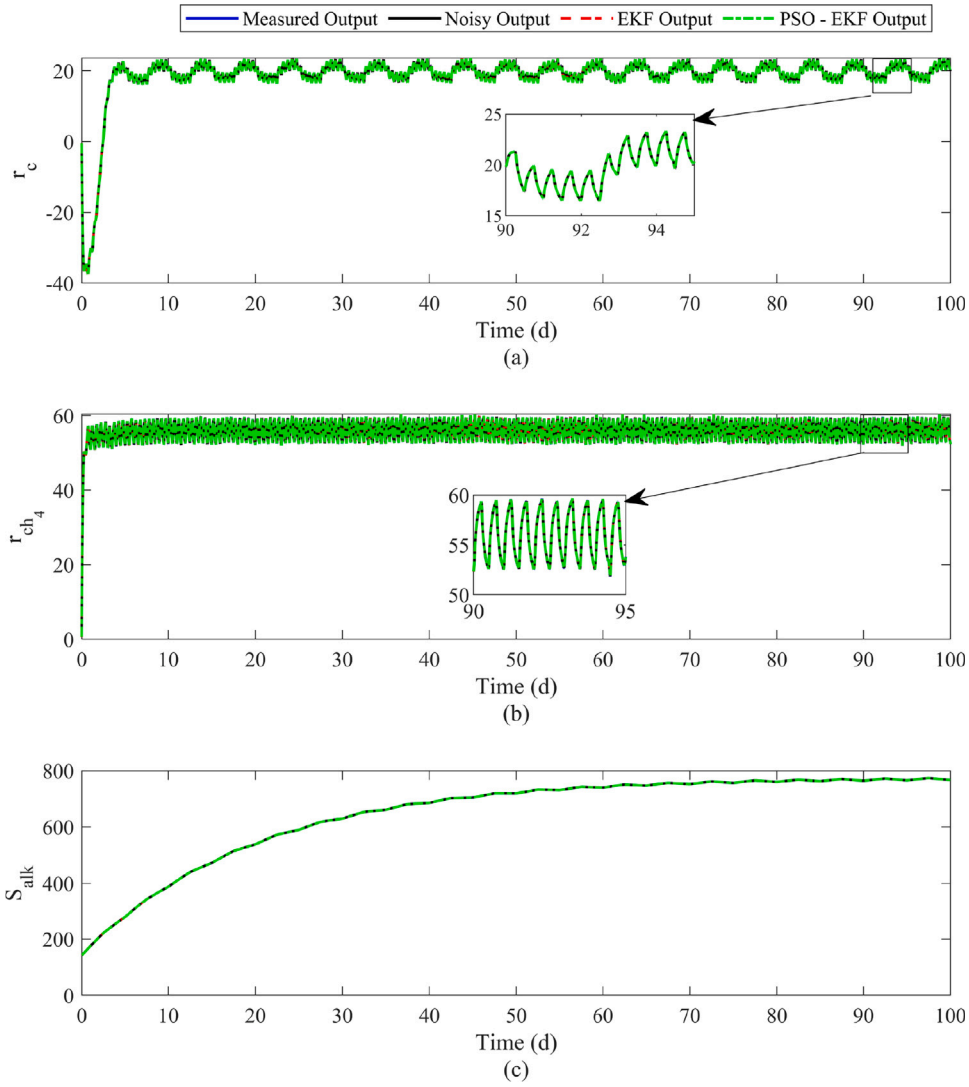

Fig. 5. Outputs derived from four sources: reference states (solid blue lines), noisy measurements (i.e. reference states with added noise (solid black lines)), and state estimates before (red dashed lines) and after PSO tuning (green dash-dotted lines).

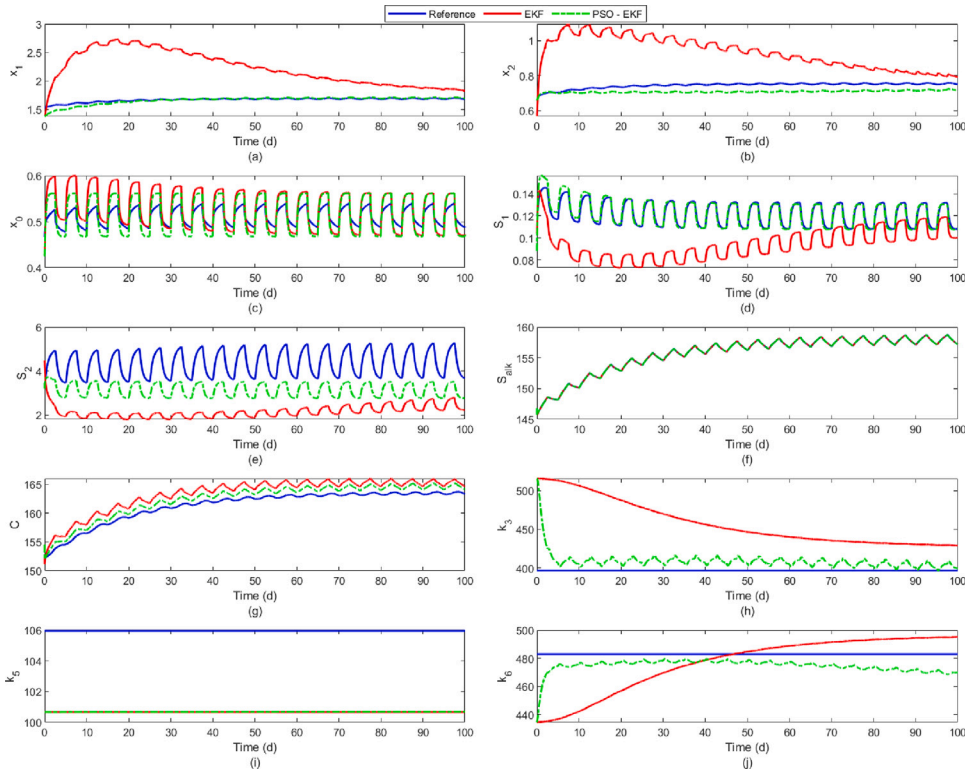

Fig. 6. Estimated states and parameters obtained when simulating the ADM1 model as a real plant. Sub-figures (a) – (g) show state estimates, and (h) – (j) show parameter estimates. model-generated reference values represented by solid blue lines, manually tuned EKF estimates by red dashed lines, and PSO - tuned EKF estimates by green dash-dotted lines).

Table 2
Quantitative comparison of EKF performance under varying measurement noise levels using manual and PSO-based covariance tuning.

Category	Case	Tuning	RMSE	MAE
Augmented States	Nominal	Manual	32.56	15.31
		PSO-based	**6.72**	**2.38**
	Moderate	Manual	32.48	15.23
		PSO-based	**6.89**	**2.32**
	High	Manual	26.35	11.81
		PSO-based	**9.66**	**2.96**
Outputs	Nominal	Manual	4.19	0.03
		PSO-based	**0.51**	**0.02**
	Moderate	Manual	4.19	0.04
		PSO-based	**0.50**	**0.05**
	High	Manual	0.67	0.40
		PSO-based	**0.51**	**0.30**

Table 3
Estimation performance of the EKF tuned for nominal noise under measurement noise mismatch.

Noise level	Augmented states		Outputs	
(relative to nominal)	RMSE	MAE	RMSE	MAE
0.5×	6.59	2.38	0.47	0.02
1×	6.59	2.38	0.47	0.02
3×	6.59	2.38	0.47	0.02
10×	6.90	2.57	0.47	0.02

The performance of the PSO-based tuning was found to be sensitive to the scaling of variables and the selection of bounds, which significantly influenced convergence and stability of the estimation process.

Table 2 shows that the PSO-tuned EKF consistently outperforms the manually tuned EKF across all noise scenarios. For the augmented states, the RMSE and MAE are reduced by approximately 70%–80%, indicating significantly improved state and parameter tracking. Similar improvements are observed for the outputs, where both the RMSE and MAE decrease significantly under all noise levels. Although estimation errors generally increase with measurement noise intensity, this trend is mild for the PSO-tuned EKF. In contrast, the manually tuned EKF exhibits an unusual behavior where RMSE and MAE change only slightly, and in some cases even decrease, as noise increases, making the filter comparatively insensitive to changes in the injected measurement noise. As a result, the manually tuned EKF does not adequately adapt to different noise conditions, whereas the PSO-based tuning maintains consistent performance across all scenarios.

A quantitative comparison in Table 3 shows that estimation errors remain nearly unchanged for noise levels up to three times the nominal value, with variations below 2% for all metrics. Under severe mismatch (10×nominal), the augmented state RMSE and MAE increase by 4.6% and 8.3%, respectively, while the output RMSE remains practically unchanged and the output MAE increases by 10.7%. Although performance degradation is observed under extreme noise conditions, the overall error levels remain within an acceptable range, confirming that the estimator retains practical robustness to significant covariance mismatch.

6. Conclusion

In this work, a joint estimator of anaerobic digestion process states and parameters using a discrete extended Kalman filter is designed based on the modified AMOCO model. Sensitivity analysis is performed to identify three critical parameters, improving estimator accuracy. To further improve filter performance, Particle Swarm Optimization (PSO) was used to optimize the process and measurement noise covariance matrices, resulting in a significant reduction in estimation errors, with a 70%–80% decrease in augmented state-estimation RMSE compared

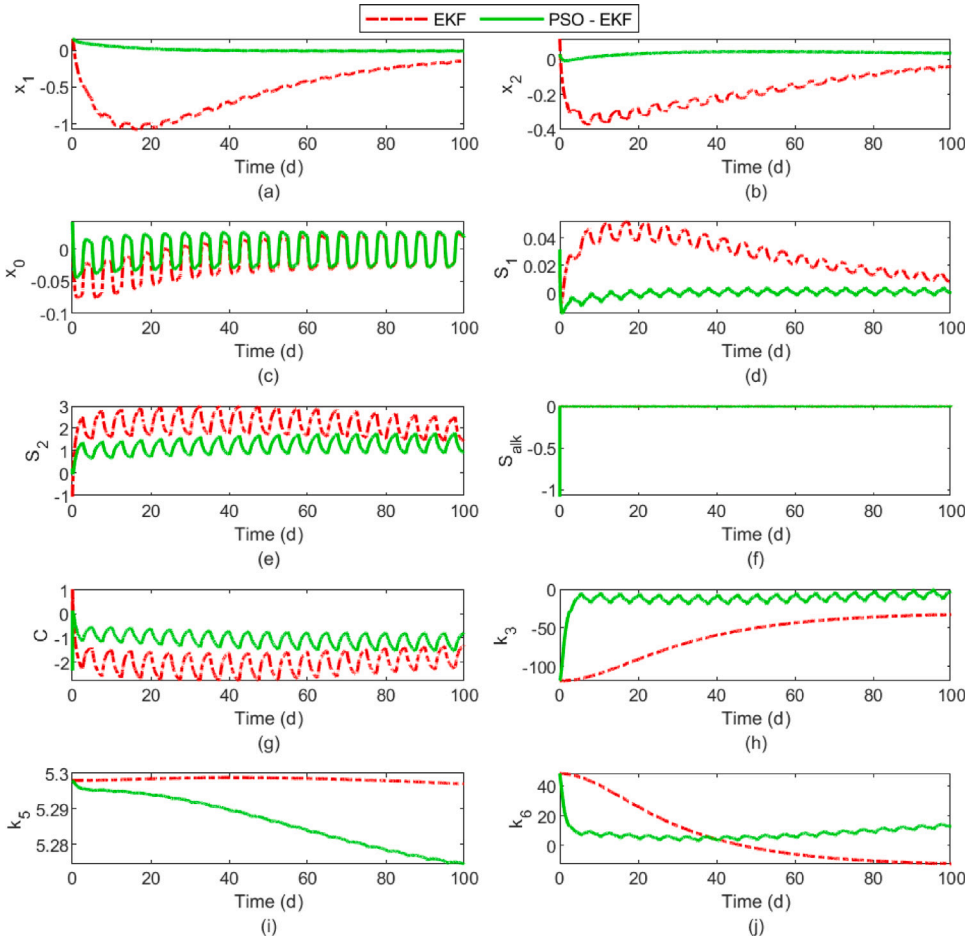

Fig. 7. Estimation error with respect to the model-generated reference values generated by the ADM1 model. The red dash-dotted lines represent error using the manually tuned EKF, whereas solid green lines show errors using the PSO - tuned EKF.

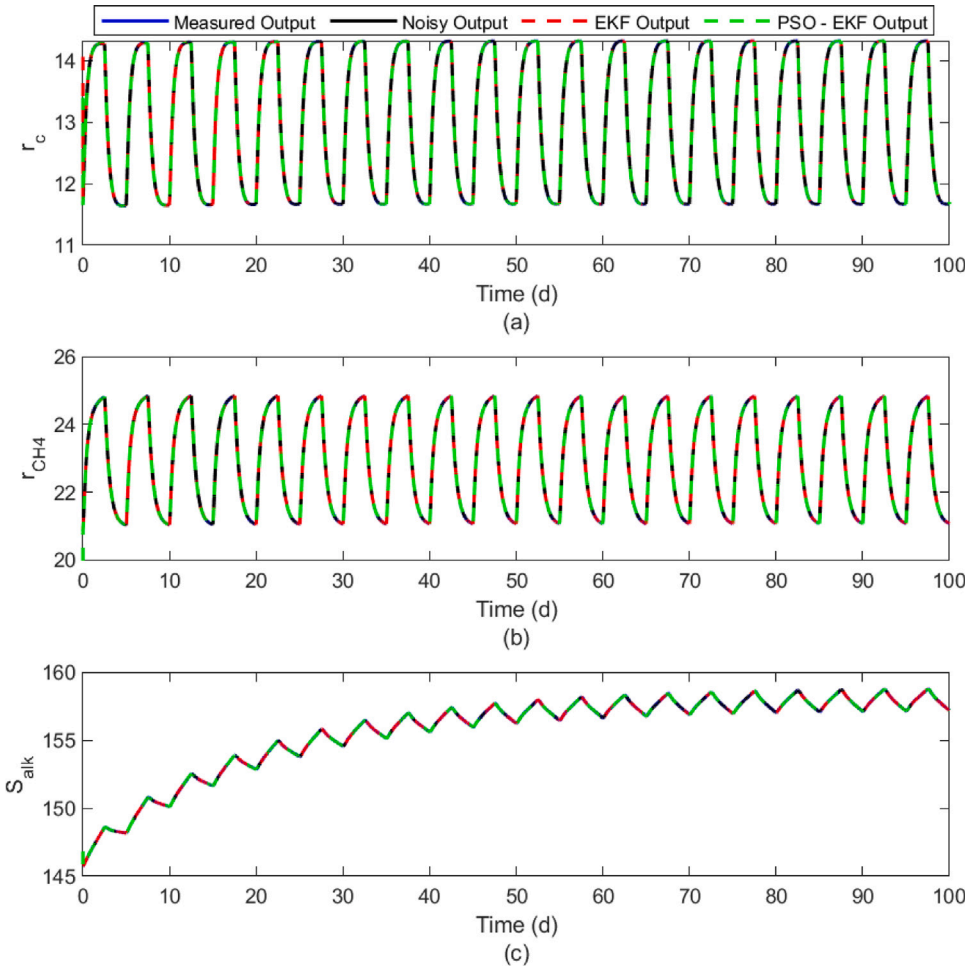

Fig. 8. Outputs derived from four sources: reference states generated by simulating the ADM1 model (solid blue lines), noisy measurements (i.e. reference states with added noise (solid black lines)), and manually tuned estimates (red dashed lines) and PSO - tuned EKF estimates (green dash-dotted lines).

with a conventionally tuned EKF. Validation of the designed estimator is presented by taking the ADM1 model as a real plant. It is shown that state estimation accuracy is relatively high. Predicted output also has a very high accuracy indicating that the model is capable of capturing overall system behavior.

Robustness analysis further demonstrated that the EKF tuned for nominal noise maintains reliable estimation performance under moderate covariance mismatch, with only limited degradation observed under severe noise conditions.

The integration of sensitivity analysis with a PSO-tuned-EKF provides a robust and efficient framework for real-time monitoring of AD processes, supporting more stable and efficient biogas production. In subsequent studies, this approach can be extended to real-time implementation and integrated with advanced control strategies. The methodology also holds potential for applications to other complex biological and physicochemical systems characterized by nonlinear dynamics and limited sensor availability.

CRediT authorship contribution statement

Bethlehem Abera: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Mengesha Mamo:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Getachew Bekele:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Torsten Wik:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Funding

This work was supported in part by the Swedish International Development Cooperation Agency (SIDA) under contribution number 51080124.

Appendix A. Equivalence between modified AMOCO and ADM1 variables

$$x_1 = (x_{su} + x_{aa} + x_{fa} + x_{C_4} + x_{pro})/1.55 \quad (A.1)$$

$$x_2 = (x_{ac} + x_{H_2})/1.55 \quad (A.2)$$

$$x_0 = x_c + x_{ch} + x_{pr} + x_{li} \quad (A.3)$$

$$s_1 = s_{su} + s_{aa} + s_{fa} \quad (A.4)$$

$$s_2 = 1000\left(\frac{s_{va}}{208} + \frac{s_{bu}}{160} + \frac{s_{pro}}{112} + \frac{s_{ac}}{64}\right) \quad (A.5)$$

$$s_{alk} = 1000\left(\frac{s_{va}}{208} + \frac{s_{bu}}{160} + \frac{s_{pro}}{112} + \frac{s_{ac}}{64} + s_{\mathrm{HCO_3}}\right) \quad (A.6)$$

$$c = 1000 s_{ic} \quad (A.7)$$

$$r_c = 1000 \rho_{T,10} \quad (A.8)$$

$$r_{\mathrm{CH_4}} = \frac{1000 \rho_{T,9}}{64} \quad (A.9)$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.biombioe.2026.109365.

Data availability

The data that is used is included in the article.

References

[1] T.A. Seadi, D. Rutz, H. Prassl, M. Köttner, T. Finsterwalder, S. Volk, R. Janssen, Biogas Handbook, University of Southern Denmark, Esbjerg, 2008.

[2] C. Pratt, K. Tate, Mitigating methane: Emerging technologies to combat climate change's second leading contributor, Environ. Sci. Technol. 52 (11) (2018) 6084–6097, http://dx.doi.org/10.1021/acs.est.7b04711.

[3] A. Della Bona, G. Ferretti, E. Ficara, F. Malpei, LFT modelling and identification of anaerobic digestion, Control Eng. Pract. 36 (2015) 1–11, http://dx.doi.org/10.1016/j.conengprac.2014.11.008.

[4] A. Rodríguez, G. Quiroz, J.D. León, R. Femat, State and parameter estimation of an anaerobic digester model, in: IEEE International Conference on Automation Science and Engineering, Trieste, Italy, 2011, pp. 690–695.

[5] D.J. Batstone, I. Angelidaki, V. Vavilin, The IWA anaerobic digestion model no 1 (ADM1), Water Sci. Technol. 45 (10) (2002) 65–73.

[6] D.T. Hill, Simplified monod kinetics of methane fermentation of animal wastes, Agric. Wastes 5 (1) (1983) 1–16, http://dx.doi.org/10.1016/0141-4607(83)90009-4.

[7] D. Dochain, M. Perrier, A. Pauss, Adaptive control of the hydrogen concentration in anaerobic digestion, Ind. Eng. Chem. Res. 30 (1) (1991) 129–136, http://dx.doi.org/10.1021/ie00049a020.

[8] O. Bernard, Z. Hadj-Sadok, D. Dochain, A. Genovesi, J.-P. Steyer, Dynamical model development and parameter identification for an anaerobic wastewater treatment process, Biotechnol. Bioeng. 75 (2001) 424–438, http://dx.doi.org/10.1002/bit.10036.

[9] E. Ficara, S. Hassam, A. Allegrini, A. Leva, F. Malpei, G. Ferretti, Anaerobic digestion models: a comparative study, IFAC Proc. Vol. 45 (2012) 1052–1057, http://dx.doi.org/10.3182/20120215-3-at-3016.00186.

[10] B. Kalchev, I. Simeonov, N. Christov, Kalman filter design for a second-order model of anaerobic digestion, Int. J. Bioautomation 15 (2011) 85–100.

[11] E. Rocha-Cózatl, M. Sbarciog, L. Dewasme, J.A. Moreno, A. VandeWouwer, State and input estimation of an anaerobic digestion reactor using a continuous-discrete unknown input observer, in: Proceedings of the 9th IFAC Symposium on Biological and Medical Systems, vol. 48, Berlin, Germany, 2015, pp. 129–134.

[12] S. Attar, F.A. Haugen, Comparison of different state estimator algorithms applied to a simulated anaerobic digestion reactor, in: Proceedings of the 59th Conference on Simulation and Modelling (SIMS 59), vol. 153, Linköping University Electronic Press, Oslo, Norway, 2018, pp. 118–125.

[13] L. Dewasme, M. Sbarciog, E. Rocha-Cózatl, F. Haugen, A. Vande Wouwer, State and unknown input estimation of an anaerobic digestion reactor with experimental validation, Control Eng. Pract. 85 (2019) 280–289, http://dx.doi.org/10.1016/j.conengprac.2019.02.003.

[14] I.F. Yupanqui Tello, D. Coutinho, A. Vande Wouwer, Extended Kalman filter design for semilinear distributed parameter systems with application to anaerobic digestion, Syst. Theory, Control. Comput. J. 1 (2021) 95–103, http://dx.doi.org/10.52846/stccj.2021.1.1.18.

[15] N. Raeyatdoost, M. Bongards, T. Bäck, C. Wolf, Robust state estimation of the anaerobic digestion process for municipal organic waste using an unscented Kalman filter, J. Process Control 121 (2023) 50–59, http://dx.doi.org/10.1016/j.jprocont.2022.11.013.

[16] C.A. Aceves-Lara, E. Aguilar-Garnica, V. Alcaraz-González, O. González-Reynoso, J.P. Steyer, J.L. Dominguez-Beltran, V. González-Álvarez, Kinetic parameters estimation in an anaerobic digestion process using successive quadratic programming, Water Sci. Technol. 52 (2005) 419–426, http://dx.doi.org/10.2166/wst.2005.0548.

[17] O. Bernard, B. Chachuat, J.-P. Steyer, State estimation for wastewater treatment processes, in: P. Quevauviller, O. Thomas, A. van der Beken (Eds.), Wastewater Quality Monitoring and Treatment, John Wiley and Sons, 2006, pp. 247–263, http://dx.doi.org/10.1002/9780470058725.ch14.

[18] S. Semcheddine, H. Bouchareb, Robust control and state estimation of a three-stage anaerobic digestion process, Ecol. Eng. Environ. Prot. (2019) 29–38, http://dx.doi.org/10.32006/eeep.2019.2.2938.

[19] X. Wu, X. Li, X. Liu, B. Huang, Uncertainty-based SOH prediction for lithium battery via improved kernel extreme learning machine, Proc. Inst. Mech. Eng. Part D: J. Automob. Eng. (2025) http://dx.doi.org/10.1177/09544070251366345.

[20] C. Rosén, U. Jeppsson, Aspects on ADM1 Implementation within the BSM2 Framework, Tech. Rep. 7224, Department of Industrial Electrical Engineering and Automation, Lund Institute of Technology, Lund, Sweden, 2005.

[21] F. Blumensaat, J. Keller, Modelling of two-stage anaerobic digestion using the IWA anaerobic digestion model no. 1 (ADM1), Water Res. 39 (2005) 171–183, http://dx.doi.org/10.1016/j.watres.2004.07.024.

[22] A. Galí, T. Benabdallah, S. Astals, J. Mata-Alvarez, Modified version of ADM1 model for agro-waste application, Bioresour. Technol. 100 (2009) 2783–2790, http://dx.doi.org/10.1016/j.biortech.2008.12.052.

[23] G. Antonopoulou, M. Alexandropoulou, C. Lytras, G. Lyberatos, Modeling of anaerobic digestion of food industry wastes in different bioreactor types, Waste Biomass Valorization 6 (2015) 335–341, http://dx.doi.org/10.1007/s12649-015-9362-7.

[24] N. Kythreotou, G. Florides, S.A. Tassou, A review of simple to scientific models for anaerobic digestion, Renew. Energy 71 (2014) 701–714, http://dx.doi.org/10.1016/j.renene.2014.05.055.

[25] X. Yang, Z. Li, B. Dahhou, Parameter and state estimation for uncertain nonlinear systems by adaptive observer based on differential evolution algorithm, Appl. Sci. (Switzerland) 10 (2020) http://dx.doi.org/10.3390/app10175857.

[26] T.A. Wenzel, K.J. Burnham, M.V. Blundell, R.A. Williams, Dual extended Kalman filter for vehicle state and parameter estimation, Veh. Syst. Dyn. 44 (2) (2006) 153–171, http://dx.doi.org/10.1080/00423110500385949.

[27] T.J. Meijer, V.S. Dolk, M.S. Chong, R. Postoyan, B.D. Jager, D. Nesic, W.P. Heemels, Joint parameter and state estimation of noisy discrete-time nonlinear systems: A supervisory multi-observer approach, in: 2021 60th IEEE Conference on Decision and Control, CDC, Austin, TX, USA, 2021, pp. 5163–5168, http://dx.doi.org/10.1109/CDC45484.2021.9683580.

[28] P.S. Madhukar, L.B. Prasad, State estimation using extended Kalman filter and unscented Kalman filter, in: Proceedings of the 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), IEEE, Moradabad, India, 2020, pp. 1–4.

[29] S. Torsten, S. Petre, System Identification, Prentice Hall International, Uppsala, 2001, pp. 96–125.

[30] H. Zhou, D. Löffler, M. Kranert, Model-based predictions of anaerobic digestion of agricultural substrates for biogas production, Bioresour. Technol. 102 (2011) 10819–10828, http://dx.doi.org/10.1016/j.biortech.2011.09.014.

[31] X. Zhou, H. Lin, Sensitivity analysis, in: Encyclopedia of GIS, Springer International Publishing, Cham, 2015, pp. 1850–1858, http://dx.doi.org/10.1007/978-3-319-23519-6_1191-2.

[32] A. Martinez, L. Vernieres-Hassimi, L. Abdelouahed, B. Taouk, C. Mohabeer, L. Estel, Modelling of an anaerobic digester: Identification of the main parameters influencing the production of methane using the sobol method, Fuels 3 (2022) 436–448, http://dx.doi.org/10.3390/fuels3030027.

[33] D. Ikumi, Sensitivity analysis on a three-phase plant-wide water and resource recovery facility model for identification of significant parameters, Water SA 46 (2020) http://dx.doi.org/10.17159/wsa/2020.v46.i3.8658.

[34] B. Abera, T. Wik, G. Bekele, M. Mamo, Parametric sensitivity analysis of the modified AMOCO model for an anaerobic digester, in: 2024 IEEE PES/IAS PowerAfrica, IEEE, Johannesburg, South Africa, 2024, pp. 1–5.

[35] F. Haugen, R. Bakke, B. Lie, State estimation and model-based control of a pilot anaerobic digestion reactor, J. Control Sci. Eng. 2014 (1) (2014) http://dx.doi.org/10.1155/2014/572621.

[36] B. Boulkroune, K. Geebelen, J. Wan, E.V. Nunen, Auto-tuning extended Kalman filters to improve state estimation, in: IEEE Intelligent Vehicles Symposium, Proceedings, IEEE, Anchorage, AK, USA, 2023, pp. 1–6, http://dx.doi.org/10.1109/IV55152.2023.10186760.

[37] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4, IEEE, 1995, pp. 1942–1948, http://dx.doi.org/10.1109/ICNN.1995.488968.

[38] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360), IEEE, 1998, pp. 69–73.

[39] T.M. Shami, A.A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M.A. Summakieh, S. Mirjalili, Particle swarm optimization: A comprehensive survey, IEEE Access 10 (2022) 10031–10061.