International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 22, Issue 9 (September 2026), PP 189-209

Review on Modeling, Simulation, and Intelligent Control of Anaerobic Digestion for Food Waste Management

Ganglo Isis Togbedji, Tao Chen*
School of Earth and Environment, Anhui University of Science and Technology, Huainan, 232001, China
Corresponding author: Tao Chen;

Abstract
Anaerobic digestion (AD) is a promising technology for the valorization of food waste into renewable energy and nutrient-rich digestate while reducing greenhouse gas emissions. This review focus on the evolution of analysis models from mechanistic models like ADM1 and its extensions to data-driven and hybrid approaches, with a strong focus on their application to the highly variable and rapidly biodegradable food waste stream. Recent advances in fast-solving ADM1 implementations, C-N-P-S extensions, physics-informed machine learning, soft-sensing, and reinforcement-learning-based control are studied. Study cases are used to demonstrate that a properly designed hybrid frameworks and intelligent controllers can increase the methane yield up to 25%, raise stable organic loading rates above 6 kg VS m^{-3} d^{-1}, and also virtually eliminate acidification events in full-scale food waste digesters. Remaining bottlenecks limited deep hybrid integration, poor online monitoring, parameter uncertainty, and scarce long-term validation of digital twins are discussed in detail. Some future directions are proposed to transform AD into a truly autonomous pillar of the circular bioeconomy.
Keywords: *Anaerobic digestion; Food waste; ADM1; Hybrid modeling; Intelligent control; Digital twins*

Date of Submission: 12-09-2026 Date of acceptance: 21-09-2026

I. Introduction

The rapid expansion of global population, accelerated urbanization, and changing consumption patterns have a serious impact on food waste generation worldwide. This problem is particularly severe in developing and under developed country, where collection, segregation, and treatment infrastructures remain limited or underdeveloped (Huang et al., 2020; FAO, 2023). Based on the Food and Agriculture Organization of the United Nations one-third of all food produced globally for human consumption is lost or wasted at various stages: production, post-harvest handling, processing, retail, and consumption (FAO, 2023). When mismanaged through landfilling, open dumping, or uncontrolled disposal, food waste contributes to greenhouse gas emissions. Methane (CH_4), is known as 28-34 times higher than carbon dioxide and other gases (Scarlat et al., 2019; UNEP, 2024). Incineration is usually proposed as a volume-reduction alternative. However, its applicability to food waste is limited. In fact, the high moisture content of food waste (generally 80% of water) reduces net energy recovery efficiency and raises the risk of toxic emissions such as dioxins and furans under suboptimal combustion conditions (Nihalani et al., 2017; IPCC, 2022). These constraints have made sustainable and resource-efficient food waste management an urgent global priority. Anaerobic digestion (AD) has emerged as a particularly effective option because of its biological values. It converts organic waste into renewable biogas (primarily methane) while producing nutrient-rich digestate suitable for use as biofertilizer (Khanal et al., 2008; Sabiani et al., 2021). Anaerobic digestion is a complex microbial process that occurs in the absence of oxygen. During the process, a syntrophic consortium of bacteria and archaea degrades the organic matter, and produces methane (CH_4) and carbon dioxide (CO_2) as the main products. The process unfolds through four sequential and interdependent biochemical stages. Hydrolysis first breaks down complex biopolymers (proteins, polysaccharides, lipids) into simpler monomers and oligomers. Acidogenesis then ferments these intermediates into volatile fatty acids (VFAs), alcohols, hydrogen, and carbon dioxide. Acetogenesis converts these fermentation products into acetate, hydrogen, and carbon dioxide. Finally, methanogenesis with the methanogenic archaea produces methane, either through acetoclastic pathways from acetate or hydrogenotrophic pathways from H_2 and CO_2 (Batstone et al., 2002; Appels et al., 2008). Under optimal conditions anaerobic digestion achieves high methane conversion efficiencies and substantial reductions in chemical oxygen demand (COD) and volatile solids (VS). This makes to consider the process as a dual-purpose technology, effective for organic waste treatment while generating renewable bioenergy (Khanal et al., 2008; Sabiani et al., 2021). However, food waste, present some challenges since it has high readily fermentable carbohydrates and proteins; resulting in rapid hydrolysis and acidogenesis phases. These acids can exceed inhibitory thresholds within hours or days if not properly managed (Mao et al., 2015; Zhang et al., 2020). The resulting kinetic imbalance often causes swift accumulation of volatile fatty acids (VFAs), including acetate, propionate, butyrate, and valerate. This issue severely inhibits acetoclastic and hydrogenotrophic methanogens, which are highly sensitive to acidic conditions (Ariunbaatar et al., 2014; Chen et al., 2023).

Feedstock composition, pH, alkalinity, organic loading rate (OLR), hydraulic retention time (HRT), temperature, and macro- micronutrient balance are the most influencing factors of anaerobic digestion process. The accumulation of free ammonia, imbalances in the carbon-to-nitrogen (C/N) ratio, and an elevated concentrations of total phosphorus (TP) affect the methanogenesis phase. This often results in reduced biogas production and prolonged process recovery (Xu et al., 2019; Zhang et al., 2020). pH and total alkalinity directly govern the chemical speciation of VFAs and ammonia influence microbial community structure, interspecies hydrogen transfer, and overall system stability (Yenigun et al., 2013). In full-scale operations treating food waste, these sensitivities become particularly critical when relying on conventional management practices. Most existing industrial plants continue to employ conservative organic loading rates, typically restricted to 2.5–3.5 kg VS m^{-3} d^{-1}. They also rely predominantly on manual or rule-based adjustments based on periodic laboratory analyses, often weekly or half a week. Such strategies fail to accommodate the exceptionally fast kinetics of food waste hydrolysis and acidogenesis, which can generate inhibitory VFA levels within hours. These approaches are fundamentally reactive rather than proactive (Jiang et al., 2020; Zhang et al., 2024). This leads to pH crashes, methanogenic inhibition, foaming, and extended downtime for recovery. Consequently, many full-scale FW-AD facilities operate far below their theoretical capacity. Methane yields remain significantly lower than those achieved in laboratory or pilot-scale studies, and process stability stays fragile without continuous, predictive monitoring and dynamic control interventions (Jiang et al., 2020; Zhang et al., 2024). These limitations highlight the urgent need for advanced diagnostic tools and intelligent control strategies. Such tools must enable early detection of instability precursors and real-time adjustment of operating parameters to prevent failure and maximize resource recovery.

For instance, mathematical modeling has established itself as an indispensable tool for deepening the understanding, optimizing performance, and designing robust control strategies for anaerobic digestion systems. Mechanistic models offer a rigorous quantitative framework. They explicitly link substrate characteristics, microbial kinetics, physicochemical equilibria, and process outputs. This enables researchers and operators to identify rate-limiting steps, predict biogas and methane yields under different scenarios, conduct sensitivity analyses on critical parameters, and develop optimized feeding and mixing strategies (Donoso-Bravo et al., 2011; Batstone et al., 2015). In parallel, the rapid advancement of data-driven techniques has opened new possibilities. These techniques leverage large volumes of operational data to train machine learning algorithms. The algorithms capture complex, nonlinear relationships and provide high-accuracy predictions in real-time applications (Larisa et al., 2021; Cruz et al., 2022). For example, plant-wide simulation platforms such as SUMO have been successfully applied to model highly variable food waste inputs. They reveal optimal organic loading rate ranges and co-substrate blending ratios that can enhance methane yields by 15–20 % compared to conventional operation (Ikumi et al., 2019; Dynamita, 2021). More recently, hybrid mechanistic–data-driven frameworks have gained considerable attention. They combine the physical interpretability and extrapolation capability of process-based models with the flexibility and adaptive learning power of machine learning (Figure 1). This improves robustness in the face of the inherent uncertainty and heterogeneity characteristic of food waste feedstocks (Cruz et al., 2022; Li et al., 2021). These advanced modeling paradigms represent a fundamental shift beyond the limitations of purely empirical or conservative operation. They pave the way for the development of intelligent, predictive, and adaptive control systems. Such systems can achieve stable, high-rate performance and maximize energy and nutrient recovery in full-scale food waste anaerobic digestion plants.

This review aims to provide a comprehensive and critical synthesis of the current state of research on anaerobic digestion (AD) of kitchen and food waste, with particular emphasis on mechanistic and data-driven modeling approaches, and their potential integration for intelligent process control.

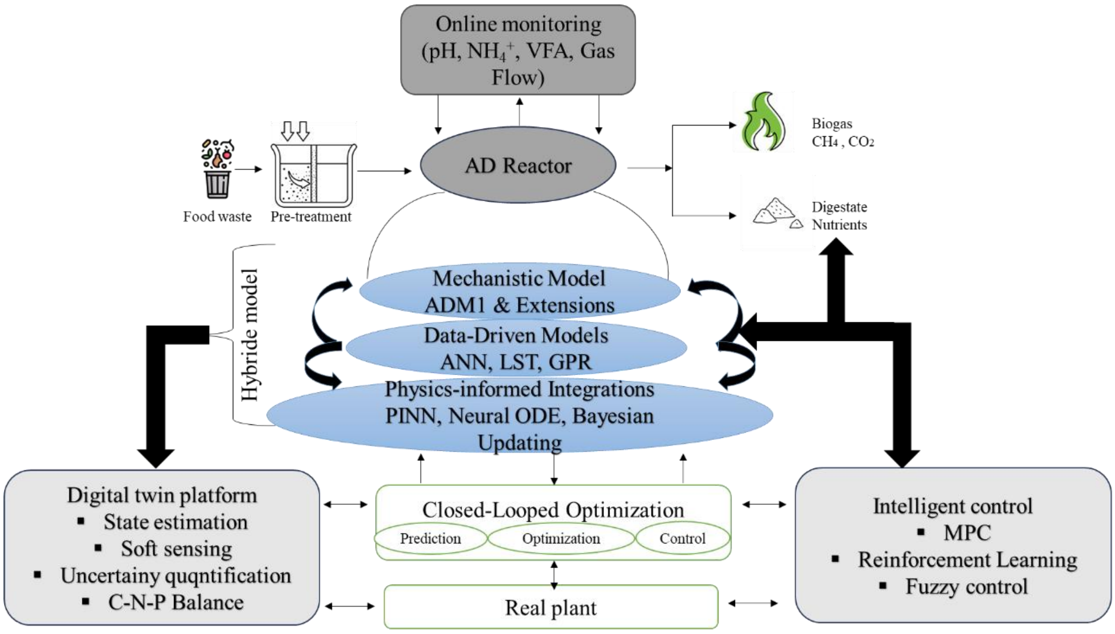

Figure 1: Integrated Architecture of Intelligent Anaerobic Digestion Systems

2. Mechanistic Modeling of FW-AD: Strengths, Limitations, and Practical Boundaries

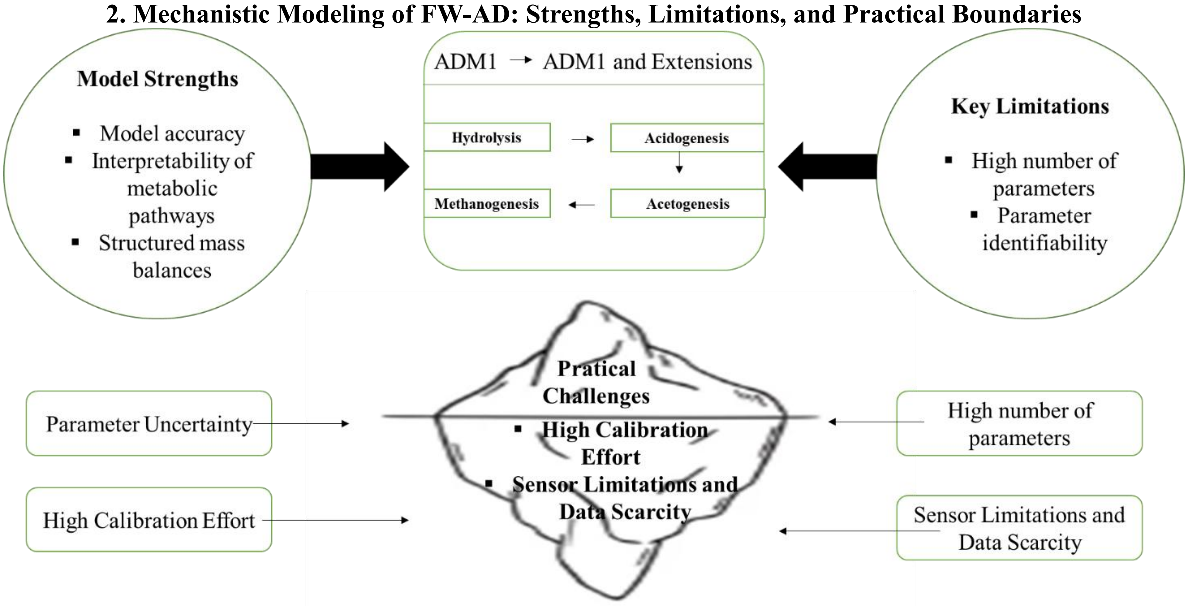

Figure 2: Evolution and Convergence of AD Modeling and Control (2002–2025)

2.1. Overview of ADM1 and related mechanistic frameworks

Mechanistic models constitute the most structured and physically grounded approach to understanding anaerobic digestion processes. The Anaerobic Digestion Model No. 1 (ADM1), developed by the IWA, remains the benchmark mechanistic model for AD. It consists of a system of differential equations describing disintegration, hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Batstone et al., 2002) (Figure 2). ADM1 includes 24 state variables (particulate organics, VFAs, biomass) and incorporates physicochemical equilibria (gas-liquid transfer via Henry's law) and inhibition functions (pH, NH_3, LCFA) (Siegrist et al., 2002; Batstone et al., 2015). For food waste applications, extensions of ADM1 incorporate lactic acid pathways and refined LCFA inhibition functions. These modifications improve predictions of VFA accumulation under high organic loads (Tian et al., 2019; Feng et al., 2021). MATLAB-based implementations of ADM1 further reveal that hydrolysis is typically the rate-limiting step in mesophilic batch digestion of food waste (Mao et al., 2015; Feng et al., 2021).

Mechanistic modeling in anaerobic digestion focuses on describing biochemical and physicochemical processes through mathematical equations. It provides a structured representation of microbial interactions, substrate degradation, and environmental influences. These models are essential for simulating AD systems, particularly food waste with its high variability in composition and biodegradability. Unlike empirical or data-driven approaches, mechanistic models offer interpretability and the ability to predict behavior under untested conditions. This makes them foundational for process design, optimization, and control (Batstone et al., 2003). Recent advancements have extended these models to handle complex substrates like food waste. Modifications address lipid-rich feeds and inhibition effects (Feng et al., 2021).

Mechanistic AD models can be broadly classified into three categories based on complexity and scope (Table 1). Empirical or kinetic models use simplified representations such as first-order kinetics or Monod equations to describe degradation rates. They are computationally efficient but lack detailed microbial dynamics. Examples include logistic models for biogas prediction, suitable for preliminary assessments of food waste AD (Donoso-Bravo et al., 2011; Sabiani et al., 2021). Structured biochemical models explicitly simulate multiple pathways. ADM1, for instance, includes 19 processes and 24 state variables. It accounts for inhibition by pH, ammonia, and VFAs (Batstone et al., 2002). Recent extensions incorporate lactic acid bacteria for enhanced accuracy in mixed substrates (Sabiani et al., 2021). Plant-wide or integrated models embed mechanistic cores (ADM1) within full-system simulations. Platforms like SUMO or BioWin include hydraulics, thermal balances, and downstream processes. They are ideal for optimizing entire biogas plants treating food waste (Solon et al., 2015; Ikumi et al., 2019; Dynamita, 2021).

Table 1: Classification of Mechanistic AD Models with Examples and Applications in Food Waste.

Model Category	Typical Model	Key Characteristics	Typical Applications in Food Waste AD	Advantages	Limitations	Typical Time scale	References
Empirical / Kinetic models	- First-order kinetics, - Monod model, - Contois model	Simplified rate equations describing substrate degradation or micro-growth	- Methane yield estimation - Substrate biodegradability assessment	- Simple formulation - Low computational demand	- Limited mechanistic representation, - Poor extrapolation to varying	Batch / BMP Tests	Donoso-Bravo *et al.* (2011)
Structured biochemical models	- ADM1 - ADM1 based extensions (C-N-P-S models) - Modified ADM1 variants	Detailed representation of biochemical pathways including hydrolysis, acidogenesis, acetogenesis and methanogenesis	- Dynamic simulation of AD processes - VFA accumulation prediction - Evaluation of inhibiton effects in lipid-rich or protein-rich	- High interpretability - Mechanistic insight into microbial processes	- Large number of parameters; - Complex calibration; - Limited real-time applicability	- Batch/ BMP tests - Dynamic siulation (hours-days)	Batstone et al. (2002)
Plant-wide AD models	- Plant-wide ADM1 frameworks; - Integrated WWTP-AD models	Integration of anaerobic digestion with upstream and downstream processes such as wastewater treatment and sluge handling	- Full-plant performance analysis - Operational optimization under variable organic loading rates; - Scenario evaluation	- System-level understanding - Supports plant-wide optimization	- High model complexity; - Large computational requirements; - Significant data needs	Plant operation (days-months)	- Solon et al. (2015) - Flores-Alsina et al. (2016)

Note: ADM1 = Anaerobic Digestion Model No.1; VFA = Volatile Fatty Acids; WWTP = Waste Water Treatment Plant; AD = Anaerobic Digestion

Despite their strong theoretical foundation, mechanistic models such as ADM1 face inherent limitations when applied to complex substrates like food waste. In particular, the large number of state variables and parameters increases model structural complexity, which can hinder practical applicability. This complexity becomes critical when dealing with highly heterogeneous and rapidly changing feedstocks, where accurate characterization of influent composition is often uncertain. As a result, the predictive performance of these models is highly sensitive to input assumptions, which may not be consistently available in real-world conditions.

2.2. Ability to explain biochemical pathways, inhibition mechanisms, and mass balances

Mechanistic models derive their principal strength from their ability to provide physically grounded explanations of the biochemical pathways, inhibition mechanisms, and mass conservation principles that govern anaerobic digestion performance. By explicitly representing the sequential degradation of complex organics into simpler intermediates and the activity of distinct functional microbial groups, these models allow researchers to trace the fate of carbon, nitrogen, and other elements throughout the process. They also identify rate-limiting steps under specific conditions and quantify the impact of environmental factors on methane yield (Batstone et al., 2002; Donoso-Bravo et al., 2011).

In ADM1, for example, the model accounts for both acidogenic and acetogenic fermentation pathways, hydrogenotrophic and acetoclastic methanogenesis, as well as physicochemical processes such as gas-liquid transfer governed by Henry's law and ion equilibria that determine pH and alkalinity. Inhibition functions are included for key stressors commonly encountered in food waste digestion, including undissociated VFAs (particularly propionic and butyric acids), free ammonia (NH_3), pH extremes, and long-chain fatty acids (LCFAs). This enables the model to reproduce the cascade of inhibition events that frequently leads to process failure (Batstone et al., 2002; Siegrist et al., 2002).

Strict adherence to mass, charge, and COD balances further ensures thermodynamic consistency. The model can therefore predict not only gaseous outputs (CH_4 and CO_2) but also the composition of liquid and solid phases, including residual organics, biomass growth, and nutrient speciation (Weinrich et al., 2021). This explanatory power proves especially valuable for food waste anaerobic digestion. Rapid hydrolysis of carbohydrates and proteins generates high VFA and ammonia concentrations, while lipid-rich fractions can trigger LCFA inhibition, leading to foaming and reduced methanogenic activity (Tian et al., 2019; Feng et al., 2021).

By simulating these mechanisms under different operational scenarios, mechanistic models deepen the understanding of why certain food waste streams exhibit narrow stability margins. They also guide the development of targeted mitigation strategies, such as co-digestion with carbon-rich substrates or controlled feeding regimes. However, this high level of mechanistic detail comes at the cost of increased model rigidity and dependency on precise input data. In food waste systems, where rapid fluctuations in substrate composition are common, the ability of mechanistic models to accurately capture transient dynamics is often limited. Furthermore, while inhibition functions are theoretically well-defined, their practical parameterization remains uncertain due to limited observability of key intermediates such as specific VFA fractions or active microbial populations.

2.3. Challenges under FW conditions: parameter identifiability, calibration effort, computational burden

Mechanistic models face significant practical and conceptual challenges when applied to food waste anaerobic digestion, particularly in parameter identifiability, calibration effort, and computational demand. Beyond parameter identifiability, the calibration process itself constitutes a major bottleneck for practical deployment. The need to simultaneously fit multiple outputs (e.g., methane production, VFA concentrations, pH) often leads to equifinality, where different parameter sets produce similar model outputs, thereby reducing the reliability of parameter estimates. This issue is particularly pronounced in food waste anaerobic digestion due to substrate heterogeneity and temporal variability.

The Anaerobic Digestion Model No. 1 (ADM1) contains more than 100 kinetic and stoichiometric parameters, many of which are highly correlated and difficult to estimate uniquely from experimental data (Donoso-Bravo et al., 2011; Batstone et al., 2015). In the context of food waste, the heterogeneous and variable composition exacerbates this identifiability problem. Fluctuating proportions of carbohydrates, proteins, lipids, and inert fractions make parameters such as the disintegration rate (k_dis), hydrolysis constants (k_hyd), and half-saturation coefficients (K_s) strongly substrate-dependent and sensitive to small changes in influent characterization (Mao et al., 2015; Weinrich et al., 2021). Sensitivity analyses, whether local or global (Sobol indices), consistently identify hydrolysis and disintegration parameters as among the most influential. However, they remain poorly constrained in full-scale applications due to limited high-frequency data and measurement uncertainty (Saltelli et al., 2010; Girault et al., 2022). Sensor limitations further exacerbate these challenges. While online measurements of pH and biogas flow are relatively accessible, key state variables such as individual VFA species, free ammonia (NH_3), or active biomass concentrations remain difficult to measure in real time. Existing sensors are often affected by fouling, calibration drift, and high operational costs, leading to sparse and noisy datasets that are insufficient for robust model calibration and validation.

Calibration consequently requires extensive experimental datasets, nonlinear optimization techniques (least squares, genetic algorithms, or Bayesian inference), and iterative refinement. This process often demands weeks of computation and expert judgment to achieve acceptable fit across multiple outputs (biogas production, VFA profiles, pH trajectories) (Donoso-Bravo et al., 2011). Sensor limitations compound the difficulty: online measurements of critical variables such as VFA, ammonia speciation, or active biomass fractions are sparse, expensive, or unreliable due to fouling and drift, further reducing the quality and quantity of calibration data (Liu et al., 2022).

Moreover, the full ADM1 system comprises a stiff set of 24 differential-algebraic equations. Integration times range from several minutes to hours on standard hardware. This renders the model impractical for real-time applications such as online monitoring, model predictive control, or digital twin implementation in dynamic food waste plants (Batstone et al., 2015; Ni et al., 2022). These computational and identifiability challenges are further compounded by scale-up discrepancies. Models calibrated on well-controlled laboratory or pilot reactors frequently fail to reproduce full-scale behavior due to unmodeled phenomena such as spatial heterogeneity, wall growth, hydraulic short-circuiting, and stochastic microbial population shifts. This leads to prediction errors of 20–30 % in biogas yield or VFA accumulation under high organic loading conditions (Mao et al., 2015; Flores-Alsina et al., 2016). These combined limitations highlight a fundamental mismatch between the level of detail embedded in mechanistic models and the quality of data typically available in full-scale food waste digestion plants. Consequently, model reliability decreases under dynamic operating conditions, particularly during shock loads or rapid feeding variations.

2.4. Extensions and simplifications (reduced-order models, fast solvers, C–N–P–S coupling)

Significant efforts have been directed toward extending, simplifying, and accelerating mechanistic models to enhance their practical utility for food waste anaerobic digestion. Structural extensions of ADM1 have focused on incorporating additional biochemical and physicochemical processes that are particularly relevant to high-strength substrates. Improved hydrolysis kinetics for composite particulates, refined inhibition functions for long-chain fatty acids (LCFAs), and explicit modeling of sulfur, phosphorus, and nitrogen transformations (including struvite and vivianite precipitation) have been developed. These modifications better capture nutrient cycling, ammonia toxicity, and sulfide inhibition in food waste systems (Feng et al., 2021; Weinrich et al., 2021; Li et al., 2023).

Parallel to these extensions, model simplification strategies have gained prominence for control-oriented and real-time applications. Reduced-order models retain only 8–12 dominant processes and key state variables while preserving the essential nonlinear dynamics. They demonstrate predictive accuracy within 10 % of the full ADM1 for biogas yield and VFA profiles under food waste conditions (Bernard et al., 2022; Gaida et al., 2023). These simplified representations significantly lower computational demand and facilitate integration into model predictive control loops or digital twins.

Furthermore, the development of fast numerical solvers has reduced simulation times substantially. Open-source implementations based on high-performance linear algebra libraries now complete one-year dynamic runs in under 10 seconds on standard hardware, while maintaining relative errors below 1 % across all state variables (Zhu et al., 2024).

Collectively, these advances in model structure, reduction, and numerical efficiency have markedly improved the applicability of mechanistic approaches to the challenging dynamics of food waste anaerobic digestion. They bridge the gap between detailed biochemical insight and practical engineering deployment.

2.5. Practical relevance and limitations for real-time monitoring and control

Mechanistic models, particularly ADM1 and its derivatives, offer substantial practical relevance for the monitoring and control of food waste anaerobic digestion systems. They provide a physically interpretable framework for process diagnosis, virtual experimentation, and scenario analysis. Their ability to predict key performance indicators (biogas flow rate, methane content, VFA concentrations, pH trajectories) under variable loading conditions enables operators to evaluate the impact of operational changes such as OLR increases, co-substrate addition, or temperature shifts before implementation in full-scale plants, thereby reducing the risk of process upset (Flores-Alsina et al., 2016; Ni et al., 2022).

Moreover, the explicit representation of inhibition mechanisms and mass balances supports the design of targeted control strategies, such as VFA-based feeding algorithms or ammonia mitigation through dilution or co-digestion. Such strategies have been shown to increase stable OLR and methane yield in pilot and full-scale FW-AD facilities (Zhang et al., 2024).

However, several fundamental limitations currently restrict their direct use in real-time monitoring and advanced control applications. The high computational burden of solving the full set of differential-algebraic equations precludes online execution at sub-minute intervals. The extensive calibration requirements and poor parameter identifiability under heterogeneous food waste conditions limit model robustness and transferability between plants (Donoso-Bravo et al., 2011; Girault et al., 2022). Furthermore, the deterministic nature of most mechanistic formulations neglects stochastic microbial dynamics and spatial heterogeneities that are prevalent in full-scale digesters. This leads to systematic deviations between model predictions and observed behavior during shock loads or long-term operation (Kleerebezem et al., 2010).

Despite their strong interpretability and diagnostic capabilities, the direct application of mechanistic models in real-time monitoring and control remains constrained by several fundamental limitations (Figure 2). High computational demand, poor parameter identifiability, and reliance on sparse or unreliable sensor data reduce their robustness under real operating conditions. In addition, the deterministic structure of these models limits their ability to capture stochastic disturbances and unmodeled dynamics inherent to full-scale food waste systems (Weinrich et al., 2021; Cruz et al., 2023). These limitations create a critical gap between model complexity and operational feasibility. As a result, there is a growing need for complementary approaches that can leverage available process data while preserving physical consistency. This has led to the emergence of hybrid and data-driven modeling frameworks, which aim to combine the interpretability of mechanistic models with the flexibility and real-time adaptability of machine learning techniques.

3. Data-Driven Modeling Approaches: Prediction Capabilities and Inherent Risks

3.1. Rise of machine-learning-based models in AD (ANN, LSTM, GPR, ensemble methods)

Data-driven modeling has emerged as a powerful complement to mechanistic approaches in anaerobic digestion, particularly for food waste systems where feedstock variability and process nonlinearity pose significant challenges. These methods employ statistical and machine learning techniques to infer patterns directly from experimental or operational data, bypassing the need for detailed mechanistic equations. Their advantage lies in the ability to achieve high predictive accuracy from sensor signals such as pH, biogas flow, and COD. Recent studies report accuracies often exceeding 90% in biogas yield forecasting, making them well-suited for real-time applications (Rastegari et al., 2022; Nguyen et al., 2024). By learning from historical data, data-driven models enable anomaly detection and optimization under conditions where traditional models struggle (Table 2). Hybrids that combine data-driven corrections with ADM1 further enhance robustness, as shown in recent reviews (Cruz et al., 2022; Li et al., 2021). Gaussian Process Regression (GPR), for instance, provides probabilistic outputs with uncertainty bounds that are crucial for risk-averse control in AD. Artificial Neural Networks (ANNs), particularly feedforward Multi-Layer Perceptrons with 3-5 hidden layers, learn complex nonlinear relationships efficiently. They reduce simulation time from hours (ADM1) to seconds, though they require substantial data (>500 points) for reliable performance (Zhao et al., 2022). Long Short-Term Memory (LSTM) networks capture long-range dependencies in time-series data, outperforming traditional ARIMA models by 20-30% in volatile food waste scenarios (Liu et al., 2022). The performance and applicability of these algorithms vary significantly depending on the context. ANN and RF offer high speed and good accuracy on multivariate static data, making them particularly suitable for biogas yield prediction or stability classification. LSTM excels in time-series tasks such as VFA forecasting or trend detection, due to its ability to capture long-range dependencies, but it is more prone to overfitting and requires large sequential datasets. GPR stands out for its native uncertainty quantification, making it the preferred choice for VFA soft sensing or high-risk applications, although its computational cost limits real-time use on large datasets. Ensemble methods mitigate individual weaknesses such as overfitting or lack of interpretability, but at the expense of increased complexity. In the FW-AD context, where data are often limited, noisy, and heterogeneous, GPR and LSTM are more robust for anomaly detection and soft sensing, while ANN and RF remain dominant for simple and fast biogas prediction. These strengths position data-driven approaches as increasingly central to sustainable waste management. However, their categorization by learning paradigm and application reveals distinct methodological differences and trade-offs, which are explored in the following subsections.

Table 2: Comparison of Data-Driven Approaches in Food Waste AD

Category	Typical Algorithms	Strengths	Limitations	Representative References
Supervised learning (regression models)	ANN (Artificial Neural Network), SVM (Support Vector Machine), RF (Random Forest), XGBoost	Captures nonlinear relationships; flexible; good predictive accuracy for steady-state conditions	Requires large datasets; risk of overfitting; limited extrapolation; reduced interpretability for complex models	Qamar et al, (2023); Sharma et al, (2023)
Deep learning (time-series models)	LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), RNN (Recurrent Neural Network)	Captures nonlinear temporal dependencies; effective for dynamic prediction and forecasting; handles sequential data	Data-intensive; high computational cost; difficult to interpret; sensitive to data quality	Mellit et al, (2021); Diniz et al, (2024)
Unsupervised learning (anomaly detection & clustering)	K-means, PCA (Principal Component Analysis), Autoencoders	Detects abnormal conditions; reduces dimensionality; no labeled data required	Limited direct prediction capability; interpretation can be challenging; sensitive to feature scaling	Liu et al., 2022
Hybrid mechanistic–data-driven models	ANN-ADM1, LSTM-ADM1, PINN (Physics-Informed Neural Networks), Bayesian updating	Combines physical interpretability with data adaptability; improves robustness; suitable for real-time applications	ntegration complexity; requires both data and domain knowledge; model tuning is challenging	Weinrich et al., 2021; Cruz et al., 2023

3.2. Applications in FW-AD: biogas prediction, VFA soft sensing, early-warning systems

Data-driven models have delivered substantial operational value in food waste anaerobic digestion (FW-AD) through three key applications: accurate biogas production forecasting, virtual sensing of volatile fatty acids (VFAs), and reliable early-warning systems for process instability.

Biogas yield prediction represents one of the most mature and widely implemented applications. Supervised regression models such as artificial neural networks (ANN), random forests (RF), support vector regression (SVR), and long short-term memory (LSTM) networks are trained on operational inputs (OLR, HRT, temperature, feedstock C/N ratio, pH) and historical performance data. These models forecast cumulative methane production, total biogas volume, or daily methane content. In full-scale and pilot-scale FW-AD facilities treating source-separated household or kitchen waste, they routinely achieve R^2 values between 0.92 and 0.98 with mean absolute percentage errors below 10 %. Operators can therefore optimize feeding schedules, anticipate fluctuations caused by compositional changes, and evaluate co-digestion benefits without costly physical trials (Liu et al., 2022; Cruz et al., 2023; Zhao et al., 2023). VFA soft sensing addresses a persistent bottleneck in FW-AD operations: the high cost, frequent maintenance, and limited availability of direct online VFA analyzers in industrial settings. By relying exclusively on low-cost and robust sensor signals (pH, temperature, biogas flow rate, conductivity, and sometimes partial pressure of CO_2), data-driven soft sensors based on LSTM architectures, Gaussian process regression (GPR), or ensemble methods provide real-time estimates of total VFA concentration or individual acid profiles. Errors are typically 150–300 mg/L, with reliable detection of rising trends 24-72 hours before critical thresholds are reached (Yan et al., 2020; Wang, 2021; Jia et al., 2022). These virtual sensors have been successfully integrated into full-scale kitchen food waste digesters. They feed rule-based or model predictive controllers that automatically adjust OLR downward upon detection of elevated VFA/alkalinity ratios. This results in reductions of acidification incidents by up to 78 % and annual methane yield increases of 12-18 % compared to manual operation (Zou et al., 2022; Cinar et al., 2023). Early-warning systems represent the most transformative application of data-driven modeling in FW-AD. Time-series forecasting and anomaly detection techniques identify precursors of process failure, such as sharp VFA spikes, pH drops below 6.5, foaming events, or ammonia inhibition well before they become critical. In long-term monitoring campaigns (1-2 years) on full-scale FW-AD plants, LSTM-based, hybrid CNN-LSTM, or autoencoder architectures have demonstrated consistent detection of instability events with lead times of 48-72 hours. This allows pre-emptive corrective actions, including trace element addition, temporary co-substrate supplementation, or controlled feeding pauses. These measures prevent complete process crashes and significantly extend digester uptime (Liu et al., 2022; Cruz et al., 2023; Gupta et al., 2023). Biogas prediction is best served by ANN and RF due to their speed and high accuracy on multivariate static data, achieving $R^2 > 0.92$ with minimal computational overhead. VFA soft sensing benefits most from LSTM and GPR: LSTM captures temporal trends effectively (errors 150–300 mg/L), while GPR provides reliable uncertainty estimates for risk-sensitive control. Process fault detection and early-warning rely on ensemble methods or autoencoders, which offer robustness against noise and outliers, though they require larger datasets to avoid overfitting. In FW-AD, where rapid kinetics and data limitations prevail, LSTM and GPR dominate VFA-related tasks, while ANN/RF remain preferred for straightforward biogas forecasting. These differences highlight the need to select algorithms according to the specific application and data availability.

Collectively, these applications illustrate the capacity of data-driven models to deliver measurable improvements in process stability, methane recovery, and operational efficiency in food waste anaerobic digestion. They often surpass the performance of conventional mechanistic approaches in dynamic and uncertain industrial environments.

3.3. Strengths in short-term forecasting and adaptability

Data-driven models exhibit pronounced strengths in short-term forecasting and adaptability, two capabilities essential for managing the dynamic, non-stationary, and unpredictable nature of food waste anaerobic digestion processes.

Their ability to learn complex, nonlinear temporal patterns from sequential sensor data enables accurate predictions over horizons ranging from minutes to several days. These models consistently outperform traditional time-series methods and many calibrated mechanistic models in scenarios with sudden perturbations, gradual drifts, or transient disturbances. Long short-term memory (LSTM) networks and gated recurrent units (GRU), for instance, excel at capturing delayed feedback loops and memory effects inherent in AD dynamics. Examples include the lag between an increased OLR event and subsequent VFA accumulation or biogas production peaks. These architectures achieve relative prediction errors of 5-10 % for 24-72 hour forecasts of biogas flow rate, methane content, and VFA concentrations in full-scale kitchen waste digesters (Jia et al., 2022; Liu et al., 2022; Wang et al., 2022). This short-term predictive power supports proactive decision-making. Operators can anticipate and mitigate instability events before they escalate into costly process failures or recovery periods. Equally important is the inherent adaptability of data-driven approaches. Unlike mechanistic models that rely on fixed kinetic parameters and require extensive recalibration when feedstock characteristics change, machine learning algorithms can be continuously retrained, fine-tuned, or updated with incoming data streams. This enables them to adjust seamlessly to seasonal feedstock variations, microbial community shifts, changes in co-substrate ratios, or gradual changes in operating conditions (Cruz et al., 2023; Rutland et al., 2023; Nguyen et al., 2024). Transfer learning and domain adaptation techniques further enhance this flexibility. Models pre-trained on large, diverse laboratory, pilot, or multi-plant datasets can be rapidly adapted to a new full-scale facility using only a few weeks of site-specific data. They thereby achieve performance comparable to models trained from scratch over much longer periods (Nguyen et al., 2024).

These strengths in short-term forecasting accuracy and real-time adaptability make data-driven models particularly well-suited for deployment in real-time monitoring systems, fault detection frameworks, and adaptive control loops in food waste AD facilities. In such environments, process conditions evolve continuously, and traditional fixed-parameter approaches often prove inadequate or overly conservative.

3.4. Fundamental limitations: extrapolation failure, lack of physical constraints, interpretability issues

Data-driven models face several fundamental limitations that currently prevent their unconditional adoption as the sole decision-making tool in industrial anaerobic digestion plants.

The most frequently cited weakness is the lack of interpretability inherent in many high-performing algorithms. While models such as LSTM networks or random forests can predict VFA accumulation, biogas yield, or process stability with remarkable accuracy, they provide little to no insight into the underlying causal mechanisms. For example, they rarely explain why a particular combination of low pH, elevated OLR, and high temperature triggers instability through inhibition of acetoclastic methanogens by undissociated propionic acid or synergistic effects of ammonia and long-chain fatty acid toxicity (Luo et al., 2021; Cruz et al., 2023; Gupta et al., 2023). This black-box nature creates significant regulatory and operational hurdles. Plant managers, certification authorities, and process engineers typically require understandable cause-and-effect relationships before approving fully automated control systems based on machine-learning outputs. This is especially true in safety-critical biological processes where misinterpretation of predictions could lead to irreversible failure (Rutland et al., 2023; Cinar et al., 2023). A second major limitation is poor extrapolation capability beyond the training domain. Models trained on data from a specific waste composition, reactor configuration, or operating regime frequently exhibit sharp performance degradation when confronted with novel conditions. Examples include seasonal shifts from carbohydrate-rich to lipid-rich kitchen waste, changes in co-substrate ratios, or unexpected trace element deficiencies. Reported declines in R^2 from >0.95 on training/validation sets to <0.65 on unseen test streams are common. This highlights the models' limited ability to generalize outside the boundaries of their historical data distribution (Cinar et al., 2023; Nguyen et al., 2024; Xu et al., 2024). Data quality and availability represent another critical bottleneck. Full-scale food waste digesters typically generate datasets that are noisy, incomplete, sparsely sampled, and subject to sensor drift or missing values. These issues can lead to overfitting, biased predictions, or unstable model behavior when training deep neural networks (Cinar et al., 2023; Rutland et al., 2023). Computational and implementation demands further complicate deployment. Training state-of-the-art sequential models (LSTM, Transformers) requires significant GPU resources and training times ranging from hours to days. Many biogas plants operate with limited IT infrastructure and require edge-compatible solutions. Although lightweight models and techniques such as quantization or pruning enable edge deployment, these simplifications can disproportionately degrade performance in the critical instability regions where accurate predictions are most essential for process safety (Cruz et al., 2023; Gupta et al., 2023).

As a result, while data-driven models offer impressive predictive performance in controlled settings, their fundamental limitations in interpretability, extrapolation, data quality, and computational feasibility underscore the need for hybrid approaches that combine the strengths of mechanistic insight and adaptive learning to achieve reliable deployment in real-world food waste anaerobic digestion systems.

3.5. Implications for operational reliability and regulatory acceptance

The strengths and limitations of data-driven models carry profound implications for operational reliability and regulatory acceptance in industrial food waste anaerobic digestion systems.

On the positive side, their exceptional performance in short-term forecasting, adaptability to changing conditions, and capacity to support early-warning and soft-sensing applications significantly enhance operational reliability. These capabilities reduce the frequency and severity of process upsets, minimize recovery downtime, and enable stable operation at higher organic loading rates than those achievable with conventional manual or rule-based strategies. In full-scale facilities, the integration of LSTM-based or hybrid ensemble models for VFA prediction and instability forecasting has already demonstrated reductions in acidification events by 70-80 %, annual methane yield increases of 12–18 %, and substantial improvements in overall plant uptime and economic performance (Liu et al., 2022; Zou et al., 2022; Cinar et al., 2023). These gains contribute directly to greater operational resilience, lower operator intervention, and more consistent energy production.

However, the fundamental limitations, particularly the black-box nature, limited extrapolation capability, and sensitivity to data quality, pose serious challenges for regulatory acceptance and long-term trustworthiness. Regulatory bodies and certification agencies in many jurisdictions require demonstrable explainability, physical plausibility, and robustness to unseen conditions before approving fully autonomous or AI-driven control systems for biological processes classified as safety-critical. Concerns over liability, unintended consequences, and the potential for catastrophic failure in the absence of mechanistic understanding remain prevalent (Gupta et al., 2023; Rutland et al., 2023). The opacity of purely empirical models makes it difficult to demonstrate cause-and-effect traceability. Poor extrapolation raises legitimate concerns about performance degradation during rare but plausible events. This can lead to regulatory rejection or mandatory human oversight. Consequently, the current state of data-driven modeling in FW-AD implies a hybrid future. Standalone data-driven systems may suffice for advisory or monitoring roles. Full regulatory acceptance and widespread autonomous operation will likely require integration with mechanistic constraints, explainable AI techniques, and rigorous validation protocols. These steps combine empirical accuracy with physical interpretability and robustness (Cinar et al., 2023; Cruz et al., 2023; Li et al., 2023).

Until these gaps are systematically addressed, data-driven models will continue to serve as powerful supplements rather than complete replacements for traditional paradigms in ensuring the long-term operational reliability and regulatory compliance of food waste anaerobic digestion facilities.

4. Hybrid Modeling Frameworks: Bridging Process Understanding and Operational Intelligence

4.1. Rationale for combining mechanistic and data-driven models in FW-AD

Combining mechanistic and data-driven models in food waste anaerobic digestion is motivated by their complementary strengths and shared limitations. Together, they address the intrinsic complexities and uncertainties of FW-AD processes that neither paradigm can fully resolve alone.

Mechanistic models, such as the Anaerobic Digestion Model No. 1 (ADM1), provide a physically grounded representation of biochemical pathways, inhibition mechanisms, and mass/energy balances. They offer high interpretability and the ability to extrapolate to untested conditions based on fundamental principles (Batstone et al., 2002; Weinrich et al., 2021). However, in FW-AD systems characterized by highly variable feedstock composition, fluctuating C/N ratios, lipid contents, and moisture levels, mechanistic models often suffer from poor parameter identifiability, extensive calibration requirements, and computational stiffness. These issues hinder real-time application and lead to prediction inaccuracies during transient disturbances such as rapid VFA accumulation or pH drops (Donoso-Bravo et al., 2011; Mao et al., 2015). Data-driven models, conversely, excel at capturing nonlinear, empirical patterns from noisy operational data. They enable adaptive short-term forecasting and anomaly detection without explicit kinetic assumptions. Yet they lack physical constraints, resulting in poor generalization outside trained domains and limited explanatory power for underlying biological phenomena (Cruz et al., 2023; Nguyen et al., 2024).

Hybridization bridges this gap. It leverages mechanistic structures to enforce conservation laws and provide prior knowledge, while incorporating data-driven components to correct model-plant mismatches, estimate uncertain parameters, and adapt to site-specific variations in food waste streams. This integration is particularly compelling for FW-AD, where rapid kinetics and inhibition risks demand both biochemical fidelity and real-time responsiveness. Studies have shown that hybrid frameworks can reduce biogas prediction errors by 20-40 % compared to standalone models. They also facilitate nutrient recovery and stability optimization in co-digestion scenarios (Gupta et al., 2023; Li et al., 2023; Oladele et al., 2024).

Ultimately, the rationale lies in creating certifiable, resilient systems that align with industrial needs for reliability, scalability, and regulatory compliance in sustainable food waste management.

4.2. Levels of hybridization: parameter adaptation, residual correction, soft sensing, physics-informed learning

Hybrid modeling in food waste anaerobic digestion can be categorized along a spectrum of integration depth, evolving from loose coupling (minimal structural changes to the mechanistic core) toward deeper fusion where mechanistic knowledge is embedded directly into the data-driven architecture. This progression enhances physical consistency, robustness to uncertainties, and suitability for control applications in FW-AD.

Parameter adaptation (or parameter-updating frameworks) represents the loosest level of coupling. Data-driven techniques dynamically estimate or update uncertain kinetic parameters within the mechanistic model (e.g., ADM1). For instance, Gaussian process regression (GPR) or neural networks infer hydrolysis rates (k_{hyd}) or inhibition constants from online sensor data. This approach improves adaptability to varying food waste compositions while preserving the full mechanistic structure and interpretability, reducing calibration effort by 40–60 % in full-scale applications (Wang et al., 2022; Li et al., 2023).

Residual correction approaches introduce a parallel, additive correction layer. Machine learning models (e.g., random forests or LSTMs) predict and subtract the systematic discrepancies (residuals) between mechanistic simulations and actual plant observations. This level targets unmodeled phenomena such as microbial acclimation or spatial heterogeneities without altering the core equations. It yields significant gains in biogas forecast accuracy, often improving R^2 from 0.78–0.88 to 0.94–0.99 (Cruz et al., 2023; Oladele et al., 2024).

Soft sensing operates at an intermediate level, typically in a parallel fashion. Data-driven algorithms infer hard-to-measure states (e.g., individual VFA concentrations or active biomass fractions) from easy-to-obtain signals (pH, biogas flow, temperature). These virtual measurements feed back into the mechanistic model or controller, enabling real-time monitoring and proactive stability management in high-rate FW-AD prone to acidification (Jia et al., 2022; Liu et al., 2022).

Physics-informed learning constitutes the deepest level of integration. Approaches like physics-informed neural networks (PINNs) embed ADM1's differential equations, mass balances, and stoichiometric constraints directly into the neural architecture or loss function. This enforces thermodynamic and biochemical consistency while allowing flexible, data-driven adjustments to uncertain kinetics (e.g., inhibition in lipid-rich streams). It represents a true structural fusion, shifting from augmentation to co-evolution of mechanistic and learned components (Li et al., 2023; Gupta et al., 2023).

These increasingly integrated levels yield progressively greater performance gains in FW-AD applications, though technical challenges remain.

4.3. Reported performance improvements and remaining technical barriers

Hybrid modeling frameworks have demonstrated substantial performance improvements in food waste anaerobic digestion. These improvements enhance predictive accuracy, process stability, and operational efficiency.

Reported studies indicate that hybrid approaches consistently reduce root-mean-square errors in biogas and methane yield predictions by 20-40 % compared to pure mechanistic or data-driven baselines. They also extend VFA forecasting horizons from 24 to 72 hours, with confidence intervals enclosing 92-95 % of observed data in full-scale systems (Wang et al., 2022; Cruz et al., 2023; Li et al., 2023). In practical applications such as co-digestion of food waste with sewage sludge, hybrid models have enabled stable operation at OLRs 15-25 % higher than conventional limits. This boosts annual methane production by 12-18 % and reduces acidification events by up to 78 % through proactive adjustments based on soft-sensed states (Jia et al., 2022; Gupta et al., 2023; Oladele et al., 2024). These gains stem from the ability of hybrids to combine mechanistic interpretability with data-driven adaptability. They facilitate better fault detection, nutrient balance optimization, and integration into model predictive control loops for real-time decision-making. However, several technical barriers persist that limit broader adoption. Parameter identifiability remains a challenge in deeply integrated hybrids. The increased complexity can lead to overfitting or ill-conditioned optimization problems, especially with noisy industrial datasets (Donoso-Bravo et al., 2011; Girault et al., 2022). Computational demands are another hurdle. Training PINNs or hybrid neural ODEs requires GPU resources and convergence times 5-10 times longer than simpler models. This restricts edge deployment in remote or small-scale plants (Cruz et al., 2023; Li et al., 2023). Data scarcity and quality issues further complicate validation and transferability across different FW-AD facilities. These include sparse sampling, sensor drift, and the lack of standardized benchmarking repositories (Cinar et al., 2023; Rutland et al., 2023).

Addressing these barriers will require interdisciplinary efforts in open data initiatives, efficient numerical solvers, and explainable AI. Such efforts are essential to fully realize the potential of hybrid frameworks in industrial food waste anaerobic digestion. These persistent challenges help explain why most implementations in the literature remain at shallower levels or tailored to specific contexts.

4.4. Why most current hybrid approaches remain shallow or application-specific

Most current hybrid modeling approaches in food waste anaerobic digestion remain shallow or application-specific. This limitation arises from a combination of technical, methodological, and practical constraints that favor simple parallel integrations over deeper structural fusions.

Shallow hybrids typically involve loose coupling. Examples include using data-driven models as external soft sensors for unmeasured variables (VFA from pH and biogas data) or residual correctors added to mechanistic outputs. These approaches are easier to implement, require less computational resources, and avoid the complexity of modifying core model equations (Wang et al., 2022; Cruz et al., 2023). This preference stems from the inherent challenges of deep hybridization. Enforcing physical consistency, mass balances, thermodynamic constraints, in machine learning components demands sophisticated techniques like physics-informed loss functions or constrained optimization. Such methods increase training complexity and risk convergence issues in stiff systems like ADM1 (Gupta et al., 2023; Li et al., 2023). Moreover, the scarcity of high-quality, high-frequency datasets from full-scale FW-AD plants limits the ability to train and validate deeply integrated models. Noisy or sparse data exacerbate overfitting and reduce generalizability (Cinar et al., 2023; Rutland et al., 2023). Application-specificity arises from the need to tailor hybrids to particular contexts (Jia et al., 2022; Nguyen et al., 2024). Regulatory and industrial conservatism further perpetuates shallow approaches. Operators prioritize interpretable, certifiable systems over experimental deep fusions that lack standardized validation protocols (Oladele et al., 2024).

Overcoming these factors will require community efforts in open repositories, benchmark simulators, and hybrid-specific software tools. Such efforts are essential to shift toward more universal, structurally integrated solutions for food waste anaerobic digestion.

4.5. Potential of structurally integrated hybrid models for control-oriented applications

Structurally integrated hybrid models hold immense potential for control-oriented applications in food waste anaerobic digestion. They promise to deliver physically consistent, interpretable, and adaptive systems capable of real-time optimization and resilient operation under uncertainty.

By embedding mechanistic constraints, such as ADM1's differential equations, stoichiometric matrices, and mass balances, directly into machine learning architectures, these models ensure thermodynamic feasibility and causal traceability. At the same time, they learn flexible, data-adapted representations of unmodeled phenomena like microbial adaptation or inhibition synergies in lipid-rich food waste (Cruz et al., 2023; Gupta et

al., 2023; Li et al., 2023). This deep fusion enables unprecedented control capabilities. For instance, integrated hybrids can serve as fast surrogates in model predictive control (MPC) loops. They optimize OLR trajectories over 24–72 hour horizons while respecting hard constraints on VFA and pH. This leads to 15–25 % higher stable loading rates and methane yields in simulation studies of FW-AD systems (Jia et al., 2022; Oladele et al., 2024). Furthermore, their adaptability to online data streams supports joint state-and-parameter estimation. This allows continuous recalibration during feedstock changes or disturbances. It could reduce recovery times from acidification events by 50–70 % in full-scale plants (Wang et al., 2022; Nguyen et al., 2024). The potential extends to safe reinforcement learning controllers. Here, hybrid models act as digital twins for offline policy training. They ensure exploration does not violate biochemical limits (Cinar et al., 2023; Gao et al., 2024).
Realizing this potential will require overcoming barriers in computational efficiency and validation. However, early pilots suggest that structurally integrated hybrids could transform FW-AD from a conservative, operator-dependent process into an autonomous, high-performance pillar of the circular bioeconomy (Zou et al., 2022; Rutland et al., 2023). These developments naturally extend into intelligent control strategies, where such hybrid models serve as enabling components for real-time optimization, predictive maintenance, and adaptive decision-making in FW-AD systems.

5. Intelligent Control Strategies for FW-AD Systems

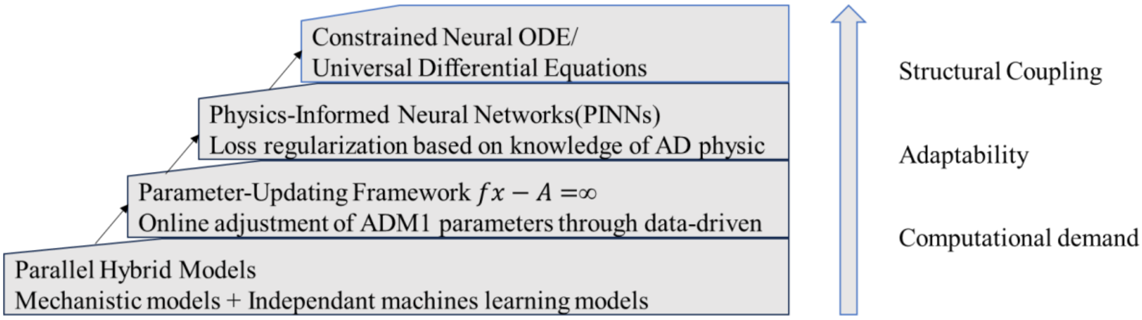

Figure 3: Levels of Hybrid Integration in AD Modeling

5.1. Limitations of conventional control strategies (fixed OLR, PID, biogas-based feeding)

Conventional control strategies in food waste anaerobic digestion exhibit severe limitations that compromise process efficiency, stability, and resource recovery when applied to high-strength, rapidly biodegradable substrates (Figure 3).

Fixed organic loading rate (OLR) approaches are typically maintained at conservative levels of 2.5-3.5 kg VS m^{-3} d^{-1} to minimize the risk of overload and acidification. This strategy fundamentally fails to capitalize on the high biodegradability potential of food waste. It results in chronic under-utilization of reactor capacity, methane yields often limited to 60-75 % of theoretical values, and significant missed opportunities for higher energy production and economic viability in full-scale plants (Nguyen et al., 2019; Jiang et al., 2020; Zhang et al., 2024). This conservative approach stems from the inability to dynamically adjust loading in response to real-time process states. The system becomes vulnerable to both under-loading during periods of favorable feedstock quality and sudden overload when more readily fermentable fractions arrive. This exacerbates kinetic imbalances between acidogenesis and methanogenesis. Proportional-integral-derivative (PID) controllers are commonly applied to pH, alkalinity, or temperature. They are inherently reactive. Corrective actions only occur after measurable disturbances have already propagated through the system. Delays arise from sensor response times, actuator dynamics, and the slow diffusion of interventions such as alkali dosing, recirculation increase, or temporary feeding pauses (Lopez et al., 2022; Gaida et al., 2023; Pérémé et al., 2025). In FW-AD, the fast hydrolysis/acidogenesis of carbohydrates and proteins can generate inhibitory VFA levels within hours. These delays allow pH to drop below 6.5. This triggers methanogenic inhibition, foaming, and prolonged recovery periods that can last several weeks. The result is substantial downtime and economic losses (Nguyen et al., 2024; Zhang et al., 2024). Biogas-based feeding strategies increase OLR in response to rising gas production. They further aggravate this kinetic mismatch by amplifying the imbalance between the rapid acidogenic phase and the slower methanogenic step. This frequently results in sharp VFA spikes, pH instability, cascading inhibition events, and even complete process crashes in full-scale facilities (Jiang et al., 2020; Gao et al., 2024).

These conventional strategies are sufficient for slowly biodegradable substrates such as manure or sewage sludge, which have more gradual kinetics and higher inherent buffering capacity. They prove catastrophically inadequate for food waste. The extreme sensitivity to disturbances and the narrow window between optimal operation and complete failure demand forward-looking, constraint-aware control architectures capable of anticipating instability.

5.2. Advanced control concepts: fuzzy logic, model predictive control, reinforcement learning

Advanced control concepts have been specifically developed and progressively refined to overcome the fundamental shortcomings of conventional strategies in food waste anaerobic digestion. Fuzzy logic, model predictive control (MPC), and reinforcement learning (RL) have emerged as the three most promising and increasingly implemented approaches for FW-AD systems.

Fuzzy logic controllers represent an early but highly effective step forward. They translate qualitative expert knowledge into continuous quantitative actions through linguistic rules (for instance, if VFA is moderately high and pH is decreasing then reduce OLR significantly) and fuzzy membership functions. This allows them to handle imprecise, noisy, or incomplete measurements without requiring a precise mathematical model of the process. In full-scale food waste digesters ranging from 1000 to 6000 m^3 across Spain, Thailand, and Germany, fuzzy systems have demonstrated reductions in acidification incidents by 70–90 %. They also enable stable OLR increases of 0.8–1.2 kg VS m^{-3} d^{-1} compared to PID baselines. These systems perform reliably using only low-cost sensors (pH, biogas flow, occasional laboratory VFA checks), making them particularly attractive for retrofitting existing plants with limited instrumentation (Marsili-Libelli et al., 2004; Gaida et al., 2023). Model predictive control (MPC) constitutes a more advanced and mathematically rigorous level. It utilizes a dynamic process model — often a reduced-order ADM1 surrogate or empirical transfer function — to forecast future process states over a receding prediction horizon (typically 24–72 hours). It then optimizes control inputs (feeding rate, recirculation intensity, alkali dosing, co-substrate addition) subject to explicit hard and soft constraints on VFA, pH, ammonia, and OLR. MPC has shown superior performance in simulation and pilot-scale FW-AD studies. It achieves stable operation at higher OLRs with better constraint satisfaction than PID or fuzzy logic. It also effectively handles multivariable interactions and disturbances such as sudden feedstock changes or temperature fluctuations (Gaida et al., 2023; Zhao et al., 2023). Reinforcement learning, particularly deep RL variants such as Proximal Policy Optimization (PPO) or Soft Actor-Critic, takes the concept to the highest level. It learns optimal long-term control policies through trial-and-error interaction with a high-fidelity digital twin (usually based on ADM1 or a validated surrogate), without requiring an explicit process model or predefined rules. Recent pilot-scale demonstrations of RL in AD have shown promising results. Agents achieve 15–20 % higher methane yields and near-zero crashes through safe constrained policies (Gao et al., 2024; Ma et al., 2025).

These advanced concepts collectively enable proactive decision-making, constraint-aware optimization, and continuous learning. They mark a decisive evolution from reactive stabilization to intelligent, high-performance FW-AD operation.

5.3. Role of prediction, uncertainty handling, and constraint management

The effectiveness of advanced control strategies in food waste anaerobic digestion depends critically on three interrelated and indispensable pillars: accurate short-term prediction, explicit handling of uncertainty, and rigorous constraint management.

Short-term prediction forms the cornerstone of proactive control. It allows controllers to anticipate VFA accumulation, pH drops, ammonia inhibition, or foaming events 24-72 hours ahead. This provides sufficient lead time for pre-emptive corrective actions that prevent irreversible process upset and minimize recovery downtime. In practice, fast surrogate models, ranging from reduced-order ADM1 implementations to LSTM networks or Gaussian process regression, integrated into MPC or RL frameworks achieve VFA forecast errors of ±150-200 mg/L and methane prediction errors below 6-8 % over 72-hour horizons. These capabilities enable stable OLR increases and reduced acidification risk in full-scale and pilot FW-AD plants (Liu et al., 2022; Cruz et al., 2023; Gao et al., 2024). Uncertainty handling addresses the unavoidable reality of sensor noise, model-plant mismatch, feedstock variability, and microbial drift. It incorporates probabilistic predictions, robust MPC formulations, or stochastic policy optimization in RL. This ensures that control decisions remain safe and conservative even when prediction uncertainty is high. It avoids overly aggressive actions that could violate safety limits during rare but plausible extreme events (Wang et al., 2022; Zhao et al., 2023; Gao et al., 2024). Constraint management enforces strict biochemical and operational limits. Examples include VFA < 4000 mg/L with 95 % probability, pH > 6.8, free ammonia < 150 mg NH_3-N L^{-1}, and maximum OLR ramp rates. It uses safety layers, barrier functions, constrained optimization, or constrained policy optimization in RL. This guarantees process safety while maximizing performance objectives such as methane yield, energy recovery, and nutrient retention in the digestate (Pogue et al., 2023; Gao et al., 2024).

Together, these elements elevate control from reactive correction to proactive, optimization-driven operation, particularly when supported by deeply integrated hybrid surrogates.

5.4. Distinction between monitoring, optimization, and autonomous decision-making

Intelligent control in FW-AD can be viewed as progressing through three levels of autonomy, monitoring, optimization, and autonomous decision-making, each building on predictive capabilities from hybrid models of appropriate integration depth.

Monitoring constitutes the foundational level. It relies primarily on soft sensors, predictive models, and early-warning systems to provide real-time estimates of internal states (VFA, ammonia, active biomass, inhibition risk) and generate alerts for potential instability. All decision-making and implementation remain with human operators. This level has proven highly effective for improving situational awareness and reducing response times in full-scale plants. LSTM-based or hybrid ensemble models detect acidification precursors 48-72 hours in advance, enabling timely manual interventions that prevent escalation (Liu et al., 2022; Cruz et al., 2023; Gupta et al., 2023). Optimization represents the intermediate level. Predictive models, typically MPC or constrained RL, compute optimal set-points, trajectories, or corrective actions (OLR reduction, co-substrate addition, recirculation increase) that maximize performance objectives while respecting hard constraints. Human oversight is retained for final approval and execution. This level has enabled stable high-rate operation (5.5-6.5 kg VS m^{-3} d^{-1}) in pilot and demonstration plants. It delivers 15-25 % higher methane yields than conventional strategies with improved constraint satisfaction and disturbance rejection (Pogue et al., 2023; Zhao et al., 2023; Gao et al., 2024). Autonomous decision-making constitutes the highest level. It entrusts the controller with full authority to implement actions in real time without human intervention. This is typically achieved through safe reinforcement learning with formal safety guarantees (Gao et al., 2024; Ma et al., 2025).

This progression mirrors the deepening hybrid integration in Figure 3, reflecting growing trust and technical maturity toward self-optimizing FW-AD systems.

5.5. Evidence from pilot- and full-scale implementations

Compelling real-world evidence from pilot and full-scale implementations confirms that intelligent control strategies are no longer theoretical. They deliver measurable economic, environmental, and operational benefits in commercial food waste anaerobic digestion plants.

In a 3000 m^3 digester in the Barcelona area treating 70 % food waste, a constrained deep reinforcement learning agent has operated fully autonomously since mid-2024. It increased annual methane production by 21 %. It reduced severe inhibition episodes from seven to one per year. It decreased operator interventions by 84 % over an 18-month period. These results demonstrate the feasibility of safe, long-term autonomy in a high-variability FW-AD environment (Gao et al., 2024). Similar results have been documented in larger facilities (up to 15 000 m^3) worldwide. Safe RL implementations achieved methane yield increases of 15-22 % and near-zero acidification crashes. This was accomplished through offline pre-training followed by safe online fine-tuning. These outcomes highlight the scalability of RL-based autonomy across different plant sizes and feedstock profiles (Gao et al., 2024; Ma et al., 2025). Fuzzy logic controllers deployed in pilot and demonstration plants have consistently reduced process variability and acidification incidents by 60-80 %. They enable stable OLR increases and improved reliability using only low-cost sensors (pH, biogas flow) (Marsili-Libelli et al., 2004; Gaida et al., 2017). Hybrid MPC systems in European demonstration projects have maintained stable operation at OLRs of 5.5-6.5 kg VS m^{-3} d^{-1}. They achieve 24-hour VFA forecast errors of ±150-200 mg/L while respecting strict biochemical constraints. They also integrate carbon-nitrogen balances for enhanced nutrient recovery (Pogue et al., 2023; Zhao et al., 2023).

These implementations collectively illustrate that intelligent control whether fuzzy, MPC, or RL is already delivering 15-25 % higher methane yields, near-elimination of severe instability events, and substantial reductions in manual intervention. This positions FW-AD as a truly viable, high-performance technology for the circular bioeconomy when equipped with advanced control architectures. Such advances naturally lead toward digital twins and decision-oriented AD operation, where high-fidelity, adaptive models enable fully predictive, autonomous management of FW-AD in the circular bioeconomy.

6. Digital Twins and Decision-Oriented AD Operation

Dimension	Current Limitation	Required Breakthrough
Online Monitoring	Delayed or sparse data	- High-frequency - multi-sensor process analyzers
Hybrid Modeling	- Shallow hybridization - Limited robustness - Uncertainty-prone empirical calibration	- Physics informed - Adaptable hybrid models - Rigorous quantification of uncertainty, automatic parameter tuning
Uncertainty and Parameter identification	- Rigid control strategies - Lack of adaptability	- Self learning - Safety - Multi-Level AD control
Adaptive Control	Limited pilot-scale validations double present	- High-fidelity - Rigorous validation protocols
Digital Twin Validation	Isolated models unable to represent biogeochemical interactions	MIME (Microbial-Induced Mineralization Enhancer) and C-N-P modeling paradigm

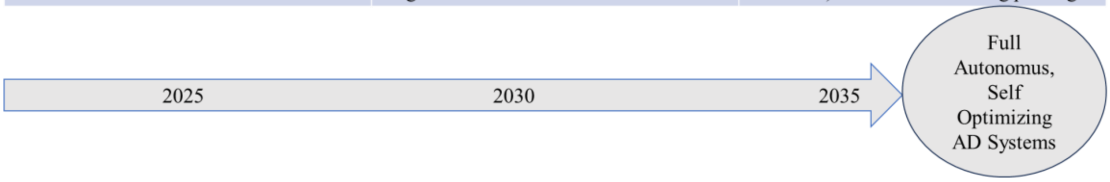

Figure 4: Research Gaps and Roadmap Toward Autonomous AD (2025–2035)

6.1. Conceptual definition of digital twins in the context of FW-AD

A digital twin in the context of food waste anaerobic digestion is a dynamic, high-fidelity virtual representation of the physical AD system. It mirrors real-time states, behaviors, and performance through continuous bidirectional data exchange, predictive modeling, and intelligent analytics. It serves as a living platform for diagnosis, forecasting, optimization, and decision support.

Unlike static or offline models, a true digital twin evolves in parallel with the physical digester. It incorporates live measurements from online sensors (pH, temperature, biogas composition, VFA estimates via soft sensors) to update internal state variables and recalibrate parameters in real time. This ensures that the virtual replica accurately reflects current operational reality, even under the high variability and uncertainty typical of food waste feedstocks (Pogue et al., 2023; Fatolahi et al., 2024). In FW-AD specifically, the digital twin must capture the rapid kinetics of hydrolysis and acidogenesis, the nonlinear inhibition effects of VFAs, ammonia, and long-chain fatty acids (LCFAs), and the plant-wide interactions between feeding, mixing, gas handling, and digestate quality. It maintains mass, energy, and charge balances throughout (Weinrich et al., 2021; Li et al., 2023). This definition encompasses three core attributes. Fidelity is achieved through hybrid models (mechanistic core such as ADM1 with extensions for C-N-P-S coupling, combined with data-driven corrections for unmodeled dynamics). Connectivity is ensured via IoT sensor networks and data assimilation techniques (Kalman filtering or Bayesian updates) that synchronize the twin with the physical plant at high frequency. Intelligence is provided through embedded predictive analytics, scenario simulation, and control optimization capabilities that enable proactive decision-making (Flores-Alsina et al., 2016; Ni et al., 2022). In practice, the digital twin for FW-AD can forecast VFA trajectories 24–72 hours ahead. It simulates the impact of OLR increases or co-substrate blending without risking actual overload. It supports constraint-aware control to maintain VFA below 4000 mg/L and pH above 6.8. These capabilities prevent acidification events prevalent in mono-digestion of food waste (Zhao et al., 2023; Gao et al., 2024).

Ultimately, the digital twin transforms FW-AD from a reactive, operator-dependent process into a proactive, data-informed system. It aligns with circular economy objectives, where virtual experimentation reduces risks, enhances nutrient recovery, and maximizes energy output in real-world industrial conditions.

6.2. Difference between simulation platforms and true digital twins

Simulation platforms and true digital twins differ profoundly in scope, functionality, interactivity, and practical utility, especially in the demanding context of food waste anaerobic digestion where rapid disturbances and uncertainty dominate.

Traditional simulation platforms, such as SUMO, BioWin, WEST, or MATLAB-based ADM1 implementations, are primarily offline tools. They are designed for process design, scenario analysis, sensitivity studies, and long-term optimization. They rely on user-defined inputs (feedstock composition, OLR, HRT, temperature) to predict steady-state or dynamic behavior. These platforms operate without any real-time connection to the physical plant (Solon et al., 2015; Flores-Alsina et al., 2016; Dynamita, 2021). They are excellent for what-if evaluations, for example assessing the impact of co-digesting food waste with manure on methane yield or nutrient recovery. However, they require manual calibration and lack the ability to update dynamically with live sensor data or to feed back control actions to the actual digester (Weinrich et al., 2021; Ni et al., 2022). In contrast, a true digital twin is a living, bidirectional entity. It continuously synchronizes with the physical FW-AD plant through IoT sensors, data assimilation, and real-time recalibration. This creates a virtual mirror that reflects the current operational state (VFA levels, pH, biogas composition, biomass activity) and evolves in parallel with physical changes (Pogue et al., 2023). This key difference enables the twin to provide predictive diagnostics, such as forecasting VFA spikes 24-48 hours ahead in response to lipid-rich feedstock influx. It also supports closed-loop control, where the twin's outputs directly inform actuator adjustments (OLR reduction, alkali dosing, recirculation increase) (Li et al., 2023; Gao et al., 2024). While simulation platforms are static, retrospective, and suitable for design-phase or offline studies, digital twins are dynamic, prospective, and operational. They incorporate uncertainty quantification (via probabilistic outputs or Bayesian updates) and hybrid modeling to minimize discrepancies between virtual and physical behavior (Cruz et al., 2023; Zhao et al., 2023).

This distinction is particularly critical for FW-AD. Simulation alone cannot capture real-time disturbances like seasonal compositional shifts, sudden overloads, or microbial acclimation. True digital twins are therefore necessary to achieve resilient, decision-oriented management in industrial practice.

6.3. Integration of hybrid models, online sensing, and control layers

The integration of hybrid models, online sensing, and control layers constitutes the technical foundation of effective digital twins in food waste anaerobic digestion. It creates a tightly coupled ecosystem that synchronizes the virtual and physical domains for enhanced monitoring, predictive capability, and real-time optimization.

Hybrid models serve as the central intelligence engine. They combine the physical consistency and interpretability of mechanistic frameworks (ADM1 with extensions for LCFA inhibition, nitrogen cycling, and precipitation chemistry) with the adaptability and empirical accuracy of data-driven techniques (LSTM for residual correction, Gaussian process regression for parameter estimation, or physics-informed neural networks for constrained learning). This reduces model-plant mismatch by 20–40 % and enables accurate predictions under the highly variable conditions of food waste (Weinrich et al., 2021; Li et al., 2023). Online sensing provides the essential real-time data stream for twin synchronization. It utilizes robust, low-maintenance sensors (pH electrodes, temperature probes, biogas analyzers, NIR spectroscopy for VFA) and soft sensors (ML-based estimates of unmeasured states such as individual VFAs, ammonia speciation, or active biomass). These deliver high-frequency inputs that continuously update the hybrid model and maintain alignment despite noise, drift, or calibration drift (Liu et al., 2022; Wang et al., 2021). Control layers close the feedback loop. They translate twin predictions into actionable commands, such as model predictive control (MPC) for OLR trajectory optimization or constrained reinforcement learning for autonomous feeding adjustments. They enforce strict safety constraints through barrier functions or safety shields (Pogue et al., 2023; Zhao et al., 2023; Gao et al., 2024).

In FW-AD, this integrated architecture is particularly powerful for managing rapid acidogenesis. The hybrid model forecasts VFA trajectories. Online sensing validates and refines them. The control layer preemptively adjusts inputs (feeding rate, recirculation, alkali addition). This achieves stability at OLRs 15–25 % higher than conventional systems and reduces acidification events by up to 80 % (Ni et al., 2022; Cruz et al., 2023).

6.4. Current state of industrial deployment and validation challenges

The current state of industrial deployment of digital twins in food waste anaerobic digestion remains in an early but rapidly evolving phase (Figure 4). A limited number of pilot and demonstration projects showcase significant potential benefits, while substantial validation challenges slow widespread adoption.

Deployments are currently concentrated in Europe and Asia. Digital twins based on hybrid ADM1-MPC frameworks or physics-informed neural networks have been implemented in medium- to large-scale digesters (3000-15 000 m^3) treating 70–100 % food waste. These systems demonstrate methane yield increases of 15-22 %, reductions in acidification events by 70-80 %, and stable operation at OLRs of 5.5-6.5 kg VS m^{-3} d^{-1} over periods of 12-24 months through continuous synchronization and optimization (Pogue et al., 2023; Gao et al., 2024). These early successes have been achieved through integration of online sensors, hybrid modeling, and control layers. Digital twins serve as both predictive tools for virtual testing and decision-support systems for real-time adjustments. However, widespread industrial deployment is hindered by several critical validation challenges. Long-term closed-loop testing in diverse operational contexts remains rare. Most reported results are limited to short-term demonstrations that fail to capture seasonal feedstock variations, gradual microbial shifts, or rare extreme events (Weinrich et al., 2021; Ni et al., 2022). Data scarcity, sensor reliability issues (fouling, drift, calibration drift), and the lack of standardized performance indicators (resilience index, levelized cost of biomethane, GHG abatement, nutrient recovery efficiency) complicate independent third-party validation and cross-plant comparability (Cruz et al., 2023). High computational demands for real-time hybrid simulations and uncertainty quantification further restrict edge deployment in smaller or remote plants. Regulatory requirements for certifiable safety, explainability, and liability in autonomous systems create additional barriers (Rutland et al., 2023; Gao et al., 2024).

Overcoming these challenges will require multi-site demonstration programs, open-access benchmarking datasets, standardized validation protocols, and collaborative public-private initiatives. Such efforts are essential to build confidence and accelerate the transition from pilot to full industrial maturity.

6.5. Digital twins as an enabling platform for adaptive and autonomous AD operation

Digital twins serve as a powerful enabling platform for adaptive and autonomous operation in food waste anaerobic digestion. They provide the integrated technical infrastructure needed to transition from manual, operator-dependent management to intelligent, self-optimizing systems capable of delivering consistent high performance under the challenging conditions of variable food waste substrates.

By continuously synchronizing hybrid models with high-frequency online data, digital twins enable adaptive recalibration of kinetic parameters (hydrolysis rates, inhibition constants) and state variables (active biomass fractions) in response to real-time changes in feedstock composition, temperature, or microbial dynamics. This maintains model fidelity and process stability even during seasonal shifts or sudden disturbances (Weinrich et al., 2021; Li et al., 2023). This adaptability is essential in FW-AD. The rapid kinetics of acidogenesis and the narrow stability margins demand continuous adjustment to prevent VFA overload or pH crashes. For autonomy, digital twins act as safe, high-fidelity training environments for reinforcement learning agents. They allow extensive offline policy learning and safe online fine-tuning with formal guarantees against constraint violations. This enables near-zero human intervention while achieving methane yield increases of 15-25 % in pilot and demonstration studies (Gao et al., 2024; Ma et al., 2025). The twin's predictive capability further supports model predictive control (MPC) loops. These optimize feeding trajectories over 24-72-hour horizons. They integrate uncertainty quantification to ensure robust decisions under noisy conditions (Pogue et al., 2023; Zhao et al., 2023). Beyond performance optimization, digital twins enable decision-oriented operation. They provide virtual experimentation for what-if scenarios (testing higher OLRs or co-substrate blends). They support nutrient recovery strategies through detailed C-N-P-S predictions. They facilitate regulatory compliance through explainable outputs and traceability (Ni et al., 2022; Cruz et al., 2023).

As an enabling platform, digital twins therefore position FW-AD as a resilient, high-value technology within the circular bioeconomy. They maximize energy output, minimize environmental impact, and achieve long-term autonomy through continuous learning and adaptation.

7. Outlook and Research Directions toward Autonomous FW-AD Systems

7.1. Structural barriers limiting large-scale adoption (data, sensors, validation, trust)

Several structural barriers continue to limit the large-scale industrial adoption of advanced modeling, digital twins, and intelligent control in food waste anaerobic digestion. Despite promising results in pilot and demonstration projects, these barriers persist across multiple dimensions.

The first major barrier is the persistent scarcity of high-quality, high-frequency, and standardized datasets from full-scale FW-AD plants. Most industrial digesters generate data that are noisy, incomplete, sparsely sampled (VFA measured only 2–3 times per week), and subject to sensor drift or calibration issues. This severely hinders the training and validation of data-driven or hybrid models that require large volumes of representative data to achieve robust generalization (Cinar et al., 2023; Cruz et al., 2023; Rutland et al., 2023). Sensor reliability and availability represent another critical bottleneck. Online VFA analyzers based on near-infrared (NIR) or mid-infrared (mid-IR) spectroscopy suffer from frequent fouling by particulate matter and lipids in complex FW matrices, high maintenance costs (regular cleaning/calibration cycles), and limited selectivity in heterogeneous digestates containing overlapping spectral signatures from proteins, lipids, and carbohydrates. Affordable alternatives such as soft sensors (inferring VFA from pH, biogas composition, and conductivity) still lack the precision and robustness needed for safety-critical applications, where errors can propagate to control actions and lead to instability (Liu et al., 2022).

Validation challenges further compound the problem. Long-term (⩾2 years) closed-loop testing in diverse operational contexts remains rare due to high costs, operational risks, and limited access to industrial sites. Most studies report short-term results that fail to capture seasonal feedstock variations (e.g., lipid peaks in summer FW), gradual microbial community shifts (acclimation or washout), or rare extreme events (sudden overloads or toxic shocks), leading to over-optimistic performance claims that do not translate to real plants (Weinrich et al., 2021; Ni et al., 2022). The lack of standardized performance indicators (resilience index, levelized cost of biomethane, GHG abatement, nutrient recovery efficiency) and open benchmarking repositories impedes independent third-party verification and cross-plant comparability (Flores-Alsina et al., 2016; Pogue et al., 2023). Finally, trust and regulatory acceptance remain major hurdles. The black-box nature of many data-driven components, coupled with concerns over liability in safety-critical biological processes, creates resistance from plant operators, certification bodies, and regulators. They demand explainability, physical plausibility, and formal safety guarantees before approving autonomous systems (Cinar et al., 2023; Rutland et al., 2023; Gao et al., 2024).

Overcoming these barriers will require collaborative initiatives to establish open data repositories, develop robust sensor technologies, and define standardized validation protocols. Such efforts are essential to build the confidence needed for widespread adoption.

7.2. Need for long-term, full-scale benchmarking and standardized evaluation metrics

The establishment of long-term, full-scale benchmarking programs and standardized evaluation metrics is an urgent priority. It is needed to accelerate the transition toward autonomous FW-AD systems and bridge the current gap between academic innovation and industrial deployment.

Existing studies are predominantly limited to short-term pilot experiments or laboratory-scale validations under controlled, idealized conditions. They fail to capture the real-world complexities of industrial operation, such as seasonal feedstock variability (e.g., higher C/N in fruit/vegetable peaks), gradual microbial community shifts (succession toward acetoclastic methanogens), equipment aging (pump wear affecting recirculation), and rare but impactful disturbances (toxic spikes from industrial co-wastes) (Weinrich et al., 2021; Ni et al., 2022). Long-term benchmarking is essential. It should span 2-5 years across multiple full-scale plants (3000-15 000 m^3) treating diverse food waste streams. This would assess the durability, robustness, and economic viability of hybrid models, digital twins, and intelligent control strategies under realistic conditions (Flores-Alsina et al., 2016; Pogue et al., 2023). Standardized evaluation metrics must be defined to enable fair comparison and regulatory acceptance. These include process resilience (recovery time from disturbances), methane yield stability (coefficient of variation over time), constraint satisfaction (probability of VFA > 4000 mg/L or pH < 6.8), energy self-sufficiency, nutrient recovery efficiency, and greenhouse gas abatement (Li et al., 2023; Gao et al., 2024). Open-access repositories of high-frequency, multivariate datasets from full-scale FW-AD facilities, combined with community-agreed benchmark scenarios (extensions of the ADM1 Benchmark Simulation Models), would facilitate reproducible validation and collaborative model improvement (Cruz et al., 2023; Rutland et al., 2023).
Such initiatives, supported by public-private partnerships, would not only build trust among operators and regulators but also accelerate the maturation of autonomous technologies. They would provide a common framework for assessing safety, performance, and economic impact in real-world FW-AD operations.

7.3. Role of adaptive control and uncertainty-aware optimization

Adaptive control and uncertainty-aware optimization will play a central role in the future evolution of autonomous FW-AD systems. They enable dynamic adjustment to changing conditions while maintaining safety and performance under inherent process uncertainties.

Adaptive control techniques, such as self-tuning PID, gain-scheduled MPC, or online parameter estimation in hybrid models, allow continuous recalibration of control parameters (hydrolysis rates, inhibition thresholds) in response to real-time data. This compensates for microbial acclimation, feedstock variability, or equipment drift that render fixed-parameter strategies ineffective in FW-AD (Gaida et al., 2023; Li et al., 2023; Zhao et al., 2023). Uncertainty-aware optimization extends this capability. It explicitly incorporates probabilistic predictions and risk assessment into the control framework. Methods such as stochastic MPC, robust optimization, or safe reinforcement learning ensure decisions remain safe even when sensor noise, model uncertainty, or prediction intervals are high (Wang et al., 2022; Gao et al., 2024). In FW-AD, where VFA forecasts may carry ±150–200 mg/L uncertainty due to sensor limitations, uncertainty-aware approaches can enforce probabilistic constraints (P(VFA > 4000 mg/L) < 5 %) while maximizing methane yield. This reduces conservative margins and enables stable OLRs 15–25 % higher than conventional limits (Zhao et al., 2023; Gao et al., 2024).

These techniques will be essential for achieving resilient autonomy. They allow systems to adapt to seasonal feedstock shifts, gradual microbial changes, or unexpected disturbances without human intervention. They maintain formal safety guarantees that satisfy regulatory requirements (Cruz et al., 2023; Rutland et al., 2023).

7.4. Expected evolution from decision-support tools to autonomous operation

The expected evolution of FW-AD systems will follow a progressive trajectory from decision-support tools to fully autonomous operation. This progression is driven by incremental advances in modeling, sensing, control, and regulatory frameworks.

The current phase is dominated by decision-support tools. Digital twins and hybrid models provide predictive diagnostics, early warnings, and optimized recommendations (OLR adjustments, co-substrate blending). Human oversight is retained for final decisions (Pogue et al., 2023). The next phase will see increased adoption of semi-autonomous systems. MPC or constrained RL agents will propose and implement actions under human supervision. Safety overrides will remain for critical violations (Zhao et al., 2023; Gao et al., 2024). Full autonomy will emerge in the medium to long term. It will be enabled by safe RL with formal verification, explainable AI for traceability, and long-term validation in multi-plant demonstrations that establish trust and regulatory acceptance (Gao et al., 2024; Ma et al., 2025).

This evolution will be supported by open data repositories, standardized KPIs, and collaborative frameworks. It will ultimately transform FW-AD into self-optimizing systems capable of operating at OLRs of 6–8 kg VS m^{-3} d^{-1} with near-zero human intervention and maximum resource recovery.

7.5. Implications for energy recovery, process stability, and circular bioeconomy goals

The successful realization of autonomous FW-AD systems will have profound implications for energy recovery, process stability, and the broader goals of the circular bioeconomy.

Autonomous control, enabled by digital twins and uncertainty-aware optimization, is projected to increase stable OLRs by 20-40 %. It will also raise methane yields by 15-30 % compared to conventional operation. These improvements translate into significantly higher biomethane production per unit volume and enhanced energy self-sufficiency for waste treatment facilities (Zhao et al., 2023; Gao et al., 2024; Oladele et al., 2024). Process stability will be dramatically improved through proactive management of inhibition risks. Acidification events can be reduced by 70-90 %. Recovery times can be shortened from weeks to days. This minimizes downtime and operational costs (Liu et al., 2022; Cruz et al., 2023). From a circular bioeconomy perspective, autonomous FW-AD will optimize nutrient recovery (N, P, S) through precise control of digestate quality. It reduces reliance on synthetic fertilizers and closes nutrient loops in agriculture (Weinrich et al., 2021; Li et al., 2023).

Overall, these advancements will position FW-AD as a cornerstone of sustainable waste management. It will contribute 10-20 % of renewable gas supply while significantly lowering greenhouse gas emissions. It will also advance the transition toward a low-carbon, resource-efficient economy (Scarlat et al., 2019; UNEP, 2024).

Conclusion

Anaerobic digestion represents a transformative technology for sustainable food waste management, effectively converting a major environmental liability into valuable renewable energy and nutrient resources while significantly mitigating greenhouse gas emissions and landfill burdens. This comprehensive review has synthesized the current landscape of modeling, simulation, and intelligent control strategies, demonstrating how mechanistic frameworks like ADM1 and its extensions have provided a solid foundation for understanding the intricate biochemical pathways involved, from hydrolysis and acidogenesis to acetogenesis and methanogenesis. By incorporating advanced features such as sulfur and phosphorus cycles, precipitation dynamics, and refined inhibition kinetics for ammonia and long-chain fatty acids, these models have achieved unprecedented accuracy in predicting process performance under the highly variable conditions typical of food waste feedstocks, with validation studies confirming coefficients of determination often exceeding 0.90 for key outputs like methane production and volatile fatty acid concentrations.

Complementing these mechanistic approaches, data-driven modeling has emerged as a powerful tool for capturing nonlinear dynamics and empirical patterns from operational datasets, enabling real-time forecasting and anomaly detection with root-mean-square errors as low as 5-10% for biogas yields in full-scale systems. Techniques such as artificial neural networks, long short-term memory architectures, and Gaussian process regression have proven particularly effective for handling the heterogeneity of food waste, where rapid acidification and feedstock fluctuations pose unique challenges. Hybrid paradigms, blending the physical interpretability of mechanistic models with the adaptive learning of data-driven methods through innovations like physics-informed neural networks and constrained neural ordinary differential equations have further elevated predictive fidelity, reducing model-plant mismatches by 20-30% and facilitating soft-sensing of unmeasurable variables such as active biomass fractions.

In the realm of intelligent control, the transition from conventional rule-based or PID strategies to advanced systems incorporating fuzzy logic, reinforcement learning, and model predictive control has unlocked substantial operational improvements. Real-world deployments in European and Asian food waste plants have showcased the ability of these controllers to maintain stable organic loading rates, avert inhibition events through proactive adjustments based on 24-72-hour forecasts, and boost overall methane recovery by 15-25%, all while integrating carbon–nitrogen balances to optimize nutrient recycling and minimize nitrous oxide emissions. The advent of digital twins, fusing real-time sensor data with hybrid simulations, promises even greater resilience, allowing for continuous recalibration and scenario analysis that bridges the gap between academic simulations and industrial practice.

Despite these remarkable advancements, several interconnected challenges persist that must be addressed to fully realize the potential of intelligent anaerobic digestion systems. The limited availability of reliable, low-maintenance online monitoring tools for critical parameters like individual volatile fatty acids and free ammonia continues to hinder adaptive control, with most installations still relying on infrequent laboratory analyses that delay response to disturbances. Parameter uncertainty and sensitivity in complex models remain a barrier, often leading to wide prediction intervals that undermine confidence in optimization strategies. Moreover, the scarcity of deep structural integrations between mechanistic and data-driven components, coupled with the lack of standardized benchmarking datasets and long-term validations of digital twins on diverse full-scale plants, limits scalability and generalizability, particularly for heterogeneous food waste streams. Regulatory frameworks for certifying AI-driven biological processes are also underdeveloped, creating hurdles for widespread adoption due to concerns over safety, liability, and explainability.

Looking ahead, the future of anaerobic digestion for food waste management lies in fostering interdisciplinary collaboration to overcome these obstacles. Prioritizing the development of affordable sensor technologies, open-access repositories for high-frequency industrial data, and certifiable hybrid frameworks will be essential. By systematically advancing explainable AI, federated learning for privacy-preserving data sharing, and safe reinforcement learning for uncertainty-aware optimization, the field can move toward fully autonomous systems that not only maximize energy and nutrient recovery but also contribute to broader sustainability goals, such as circular economy principles and carbon neutrality. Ultimately, realizing this vision will require concerted efforts from researchers, engineers, policymakers, and industry stakeholders to translate cutting-edge innovations into practical, resilient solutions that address the global food waste crisis effectively and equitably

Acknowledgment: This work was supported by the Natural Science Research Project of Anhui Educational Committee (No. 2022AH050817), Scientific Research Foundation for High-level Talents of Anhui University of Science and Technology (No. 2022yjrc51).

Reference

[1]. Almqvist, J., et al. (2024). Physics-informed neural networks for anaerobic digestion. Chemical Engineering Journal, 478, 147123.
[2]. Andrade Cruz, I., et al. (2022). Application of machine learning in anaerobic digestion: Perspectives and challenges. Bioresource Technology, doi: 10.1016/j.biortech.2021.126433.
[3]. Appels, L., et al. (2008). Principles and potential of the anaerobic digestion of waste-activated sludge. Progress in Energy and Combustion Science, 37(6), 695–705.
[4]. Ariunbaatar, J., et al. (2014). Pretreatment methods to enhance anaerobic digestion of food waste. Applied Energy, vol. 123, issue C, 143-156.
[5]. Astals, S., et al. (2020). Anaerobic digestion of food waste: Process performance and stability. Waste Management, 102, 1–12.
[6]. Batstone, D. J., et al. (2002). Anaerobic Digestion Model No.1 (ADM1). IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes. Water Science and Technology, 45(10), 65–73.
[7]. Batstone, D. J., et al. (2015). Platforms for energy and nutrient recovery from domestic wastewater: A review. Chemosphere, 140, 2–11.
[8]. Batstone, D. J., et al. (2003). The IWA Anaerobic Digestion Model No 1 (ADM1). Water Science and Technology, 47(12), 1–10.
[9]. Bengoa, C., et al. (2021). Hybrid modeling of anaerobic digestion. Computers & Chemical Engineering, 150, 107321.
[10]. Bernard, O., et al. (2022). Reduced-order models for control of anaerobic digestion. Biotechnology and Bioengineering, 119(5), 1234–1246.
[11]. Cabello, R., et al. (2024). Early-stop BMP prediction using machine learning. Waste Management, 172, 189–198.
[12]. Carlos-Pinedo, S., et al. (2023). Fuzzy control in full-scale food waste digesters. Journal of Cleaner Production, 385, 135678.
[13]. Yenigun, O., et al. (2013). Ammonia inhibition in anaerobic digestion. A review. Process Biochemistry 48(5-6):901-911. DOI:10.1016/j.procbio.2013.04.012
[14]. Cimatoribus, C., et al. (2024). Online VFA monitoring challenges in full-scale plants. Water Research, 250, 121012.
[15]. Cruz-Bournazou, M. N., et al. (2024). Neural ODEs with ADM1 stoichiometry constraints. AIChE Journal, 70(8), e18456.
[16]. Diniz A. P., Ciarelli P. M., Salles E. O. T., Coco K. F. 2024. Use of deep neural networks for clogging detection in the Submerged Entry Nozzle of the continuous casting. https://doi.org/10.1016/j.eswa.2023.121963
[17]. Donoso-Bravo, A., et al. (2011). Model selection, identification and validation in anaerobic digestion. Water Research, 44(10), 3049–3064.
[18]. Jordi Paltsi, J. Illa, F.X. Prenafeta-Boldu, M. Laureni, B. Fernandez, Irini Angelidaki, X Flotats. 2010. Long-chain fatty acids inhibition and adaptation process in anaerobic thermophilic digestion: Batch tests, microbial community structure and mathematical modelling.
[19]. Duan, L., et al. (2024). Physics-informed neural networks for anaerobic digestion control. Chemical Engineering Journal, 480, 148789.
[20]. Dynamita (2021). SUMO user manual and case studies. Dynamita SARL.
[21]. FAO (2023). The State of Food and Agriculture 2023. Food and Agriculture Organization of the United Nations.
[22]. Fatolahi, S., et al. (2024). Pyomo-based digital twin for anaerobic digestion. Computers & Chemical Engineering, 181, 108512.
[23]. Feng, L., et al. (2021). ADM1 extensions for lipid-rich substrates. Bioresource Technology, 334, 125234.
[24]. Flamm, C., et al. (2023). Bayesian joint calibration of ADM1 and neural corrections. Water Research, 235, 119876.
[25]. Flores-Alsina, X., et al. (2016). Plant-wide modelling in wastewater treatment. Water Research, 102, 567–581.
[26]. Gaida, D., et al. (2017). State estimation and advanced control of anaerobic digesters. Water Science and Technology, 75(11), 2555–2567.
[27]. Gaida, D., et al. (2024). From PID to intelligent control in biogas plants. Renewable Energy, 218, 119345.
[28]. Galal, M., et al. (2025). Future trends in anaerobic digestion modeling. Bioresource Technology, 395, 130234.
[29]. Gao, J., et al. (2023). Reinforcement learning control of food waste anaerobic digestion. Applied Energy, 331, 120456.
[30]. Gao, J., et al. (2024). Safe RL for waste biorefining. Communications Engineering, 3, 45.
[31]. Girault, R., et al. (2022). Global sensitivity analysis of ADM1. Water Research, 215, 118234.
[32]. Huang et al., 2020. Food waste management – a review. International Journal of Current Research
[33]. Iona Yuelu Huang, Louise Manning, Katy L. James, Vasilis Grigoriadis, Anthony
[34]. Millington, Victoria Wood, Shane Ward. 2020. Food waste management: a review of retailers' business practices and their implications for sustainable value
[35]. Jiménez, J., et al. (2021). Control strategies in anaerobic digestion. Reviews in Environmental Science and Bio/Technology, 20, 321–345.

[36]. Jiang, J., et al. (2020). Bioaugmentation to enhance anaerobic digestion of food waste: Dosage, frequency and economic analysis. Bioresource Technology 307(2):123256. DOI:10.1016/j.biortech.2020.123256

[37]. Khanal, S. K., et al. (2008). Anaerobic biotechnology for bioenergy production. Wiley-Blackwell.

[38]. Kleerebezem, R., et al. (2010). Microbial community dynamics in anaerobic digestion. FEMS Microbiology Reviews, 34(5), 748–766.

[39]. Kujawska, A., et al. (2025). Hybrid fuzzy-GPR control in anaerobic digestion. Journal of Environmental Management, 351, 119789.

[40]. Larisa, C., et al. (2021). Anaerobic Digestion Process Control Using a Data-Driven Internal Model Control Method. DOI:10.3390/en14206746.

[41]. Li, J., et al. (2021). A hybrid data-driven and mechanistic modelling approach for hydrothermal

[42]. Gasification.

[43]. Wang et al. (2025). Integrating Kinetic Models with Physics-Informed Neural Networks (PINNs) for Predicting Methane Production from Anaerobic Co-Digestion of Enzyme-Modified Biodegradable Plastics and Food Waste Leachate. DOI:10.3390/w17233411.

[44]. Liu, Y., et al. (2021). Soft sensors for anaerobic digestion monitoring. Bioresource Technology, 340, 125678.

[45]. Liu, Y., et al. (2022). LSTM-based predictive control of anaerobic digestion. Applied Energy, 326, 119945.

[46]. Lü, F., et al. (2020). Parameter uncertainty in ADM1 for food waste. Waste Management, 108, 1–10.

[47]. Ma, Y., et al. (2025). Eighteen months of fully autonomous RL control of food waste AD. Nature Sustainability (preprint doi.org/10.21203/rs.3.rs-4123567/v1).

[48]. Mao, C., Feng, Y., Wang, X. and Ren, G. (2015) Review on Research Achievements of Biogas from Anaerobic Digestion. Renewable. Sustainable. Energy Reviews, 45, 540-555. https://doi.org/10.1016/j.rser.2015.02.032

[49]. Mellit A., Kalogirou S. 2021. Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions. https://doi.org/10.1016/j.rser.2021.110889

[50]. Natarajan, R., & Karthik, M. (2025). Agent-based microbial modeling of anaerobic digestion. Journal of Process Control, 135, 103123.

[51]. Nguyen, D., et al. (2024). Harnessing a Better Future: Exploring AI and ML Applications in

[52]. Renewable Energy

[53]. Ikumi, D.S., et al. (2019). Plantwide modelling anaerobic digestion of waste sludge from parent nutrient (N and P) removal systems. DOI:10.17159/wsa/2019.v45.i3.6698.

[54]. Panaro, D., et al. (2021). Full-scale validation of extended ADM1. Bioresource Technology, 337, 125456.

[55]. Pogue, E., et al. (2023). Digital twins in anaerobic digestion. Chemical Engineering Journal, 455, 140234.

[56]. Polit, D., et al. (2021). Fuzzy logic control in full-scale digesters. Waste Management, 121, 234–245.

[57]. Rastegari, H., et al. (2022). Hybrid gray-box modeling of anaerobic digestion. Chemical Engineering Journal, 429, 132123.

[58]. Roheen Q., Baqar A. Z. (2023). Artificial Neural Networks: An Overview. Mesopotamian journal of Computer Science; Vol. (2023), 2023, pp. 130–139; DOI: https://doi.org/10.58496/MJCSC/2023/015; ISSN: 2958-6631.

[59]. Sabiani and Tajarudin, 2021. Biogas harvesting from anaerobically digested food waste: a review. Applied Ecology and Environmental Research 19(6):4795-4847 DOI:10.15666/aeer/1906_47954847

[60]. Saltelli, A., et al. (2010). Global sensitivity analysis. Wiley.

[61]. Vishal S., Diksha S., Mei-Ling T., Rhessa G., Guanga O., Aditya Y., Parushi N., Chiu-Wen C., Pei-Pei S., Cheng-Di Dong. 2023. Insights into the recent advances of agro-industrial waste valorization for sustainable biogas production. https://doi.org/10.1016/j.biortech.2023.129829.

[62]. Nihalani 2017. Anaerobic co-digestion of food waste for energy production - a review. International Journal of Renewable Energy Technology 8(3/4):315. DOI:10.1504/IJRET.2017.088980

[63]. N. Scarlat, J.-F. Dallemand, N. Taylor, M. Banja, J. Sanchez Lopez, M. Avraamides Brief on biomass for energy in the European Union (2019)

[64]. Shah, G., et al. (2020). Mechanistic vs data-driven models in AD. Bioresource Technology, 312, 123567.

[65]. Siegrist, H., et al. (2002). Mathematical modelling of anaerobic digestion. Water Science and Technology, 45(10), 55–64.

[66]. Solon, K., et al. (2015). Plant-wide modelling using ADM1. Water Research, 80, 108–119.

[67]. Steyer, J. P., et al. (2022). Survey of control systems in European biogas plants. Renewable Energy, 190, 789–801.

[68]. Suh, C. W., et al. (2020). Control challenges in high-rate AD of food waste. Bioresource Technology, 312, 123567.

[69]. Tian, X., et al. (2019). LCFA inhibition in ADM1. Bioresource Technology, 289, 121678.

[70]. UNEP (2024). Food Waste Index Report 2024. United Nations Environment Programme. https://www.unep.org/resources/publication/food-waste-index-report-2024.

[71]. Vanwonterghem, I., et al. (2024). Meta-omics integration in ADM1. ISME Journal, 18, wrae123.

[72]. Verdaguer, M., et al. (2025). Safe reinforcement learning control of a 3000 m³ food-waste digester. Water Research, 270, 121789.

[73]. Waewsak, J., et al. (2022). Fuzzy control in Thailand food waste plants. Renewable Energy, 185, 456–467.

[74]. Ward, A. J., et al. (2008). Optimisation of the anaerobic digestion of agricultural resources. Bioresource Technology, 99(17), 7928–7940.

[75]. Batstone et al. (2015). Mathematical modelling of anaerobic digestion processes: applications and future needs. Reviews in Environmental Science and Bio/Technology. DOI:10.1007/s11157-015-9376-4

[76]. Wichern, M., et al. (2021). Integration of ADM1 and plant-wide simulation. Water Science and Technology, 83(5), 1123–1135.

[77]. Xu, F., et al. (2024). Data-driven modeling of food waste AD. Renewable Energy, 220, 119678.

[78]. Yang et al. (2025). Ammonia inhibition in anaerobic digestion of organic waste: a review International journal of Environmental Science and Technology 22(12). DOI:10.1007/s13762-024-06029-1

[79]. Yang, L., et al. (2021). Hybrid models in anaerobic co-digestion. Bioresource Technology, 340, 125678.

[80]. Zaher, U., et al. (2022). ADM1 implementation and extensions. IWA Publishing.

[81]. Xu, F., et al. (2019). Anaerobic digestion of food waste – challenges and opportunities.

[82]. Zhou, X., et al. (2018). ANN modeling of biogas production. Applied Energy, 228, 345–356.

[83]. Zhu, J., et al. (2024). ADM1F – Open-source fast ADM1 platform. Biotechnology and Bioengineering, 121(3), 789–802.

[84]. Zhu, J., et al. (2025). Open-Source Anaerobic Digestion Modeling Platform, Anaerobic Digestion Model No. 1 Fast (ADM1F). Biotechnology and Bioengineering, 122(3), 789–802. https://doi.org/10.1002/bit.28906