Development of an Artificial Intelligence-Based Framework for Biogas Generation from a Micro Anaerobic Digestion Plant

Ikechukwu Offie[1], Farzad Piadeh[1], Kourosh Behzadian[1*], Luiza C. Campos[2], Rokiah Yaman[3]

[1] School of Computing and Engineering, University of West London, Ealing, London, W5 5RF, UK

[2] Civil, Environmental and Geomatic Engineering, University College London, Gower St, London WC1E6BT, UK

[3] Leap AD Micro Company, London, UK

* Corresponding author. Tel.: +44 (0) 20 8231 2466 E-mail address: kourosh.behzadian@uwl.ac.uk

Abstract

Despite the advantages of the Anaerobic Digestion (AD) technology for organic waste management, low system performance in biogas production negatively affects the wide spread of this technology. This paper develops a new artificial intelligence-based framework to predict and optimise the biogas generated from a micro-AD plant. The framework comprises some main steps including data collection and imputation, recurrent neural network/ Non-Linear Autoregressive Exogenous (NARX) model, shuffled frog leaping algorithm (SFLA) optimisation model and sensitivity analysis. The suggested framework was demonstrated by its application on a real micro-AD plant in London. The NARX model was developed for predicting yielded biogas based on the feeding data over preceding days in which their lag times were fine-tuned using the SFLA. The optimal daily feeding pattern to obtain maximum biogas generation was determined using the SFLA. The results show that the developed framework can improve the productivity of biogas in optimal operation strategy by 43% compared to business as usual and the average biogas produced can raise from 3.26 to 4.34 m^3/day. The optimal feeding pattern during a four-day cycle is to feed over the last two days and thereby reducing the operational costs related to the labour for feeding the plant in the first two days. The results of the sensitivity analysis show the optimised biogas generation is strongly influenced by the content of oats and catering waste as well as the optimal allocated day for adding feed to the main digester compared to other feed variables e.g., added water and soaked liner.

Keywords: Anaerobic digestion; Artificial intelligence framework; Biogas generation; Optimised operation strategy; Organic waste; Recurrent neural network.

1. Introduction

Over the years, the world has been subjected to unprecedented population growth, economic development, and rapid urbanisation. These series of development have given rise to a constant increase in organic waste generation globally. This constant increase in the generation of organic wastes has become a major source of concern globally following its negative impacts (Arun and Sivashanmugam, 2017). Organic wastes account mainly for approximately 105 billion tonnes of the total municipal solid waste generated on an annual basis globally (WBA, 2021). Lack of proper and efficient waste management strategies can lead to a series of environmental problems such as emerging pollution, ecosystem destruction, harm to human health and depletion of natural resources (Kumar *et al.*, 2021). The poor management of organic wastes also has the potential to contribute to climate change through the emission of greenhouse gases into the atmosphere (CIWEM, 2021). The effect of this has compelled nations and governments to invest more financial and material resources for the remediation of organic wastes in recent years (Wainaina *et al.*, 2020).

Presently, efforts are being made to revolutionise the waste management industry towards achieving sustainability and profitability (Abdallah *et al.*, 2020). This has led to the application of advanced recycling technologies such as anaerobic digestion (AD), composting and incineration amongst others in treating and managing wastes that having been identified to be better alternatives to landfill systems (Wainaina *et al.*, 2020). AD technology has been regarded as an established biological processing technique suitable for stabilising a plethora of organic solid wastes that also results in resource recovery of energy (i.e., methane biogas) and useful nutrients (i.e., organic fertilisers) (Wainaina *et al.*,2020). More specifically, the AD technology can deliver both de-fossilisation and decarbonisation, i.e., avoided GHG (greenhouse gases) emissions through converting organic wastes to (1) renewable energy thereby reducing the need for fossil fuel utilisation and (2) organic fertilisers reducing the need for chemical fertilisers (WBA., 2021). In addition, compared to other technologies such as incineration that may result in air pollution and GHG emissions, the ability of the AD technology for converting waste to useful energy and organic nutrients without causing any form of environmental pollution, i.e., avoided embodied energy/carbon and hence avoided GHG emissions, makes it a preferable option (Liang *et al*., 2022). The multi-faceted nature of the AD technology has rendered it as a highly ranked technique in the waste management industry and an excellent tool for the realisation of circular economy (WBA., 2021).

Evidently, the performance of the AD technology is mainly evaluated based on the biogas generation as the most valuable output which is the result of processes in four stages including hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Shahsavar *et al.*, 2021). Despite the plethora of advantages in the AD technology, its performance especially for biogas generation is heavily dependent upon the balanced mix of the waste and microbial groups and hence is highly sensitive to organic compounds and may result in process instability and failure (Cruz *et al.*, 2022). In addition, long residence time and low removal efficiency of organic compounds are other limitations that hinder the wide application and adoption of this technology to full potential (Xu *et al.*, 2021). All this can directly affect the efficiency of the biogas production. Therefore, modelling AD processes are of paramount importance and useful tool to first estimate and then optimise the AD performance (i.e., projection of biogas production and organic fertilisers). Although several conventional mathematical models (e.g., theoretical, analytical and statistical) are available, their application is limited due mainly to the complexity of their development, data demanding and challenges with model calibration (Cruz *et al.*, 2022). Hence, these models are widely used as useful tools for the AD planning and design such as AQUASIM, GRAINIT BIOGAS, ANESSA and ADM1 (Carlini *et al.*, 2020). However, the reliability of these models within the operation phase of AD plants is more challenging as the operation conditions of AD processes can be highly variable and rapid changes in control parameters are inevitable especially depending on waste composition (Cheela *et al.,* 2021). As a result, due to changes in various microbial species and the complex metabolic pathways, the above mathematical models are unable to properly estimate the model performance. However, data driven models such as Artificial Intelligence (AI) can be introduced as a good surrogate for process-based modelling that are dependent of complex physico-chemical processes. In other words, the AI-based models are developed based on historic data of the system variables and can be used for real-time operation of AD plants by using online data (Piadeh *et al.*, 2022).

Several research works have studied the application of AI methods to the AD processes for modelling the relevant non-linear and complex relationships by focusing on optimising particle size of organic matters, organic loading rate, ratio of carbon to nitrogen (C/N), pH and temperature, and residence time (Zhang *et al.*, 2019). These research studies mainly followed three approaches: (1) using classification machine learning (ML) methods such as support vector machine, random forest (RF), K-nearest neighbourhood (KNN) to predict the corrected operation, (2) optimising parameters by particle swarm and genetic algorithm, and (3) employing various artificial neural networks (ANN) to predict control parameters (Cruz *et al.* 2022). To increase the rate of biogas yield, AI-based methods have been widely used in agricultural and industrial application (Kunatsa and Xia, 2022). However, to the best of our knowledge, few research works have presented an AI-based framework for developing operation strategies to improve the AD performance in producing biogas from the food waste generated in an urban area. More specifically, the KNN method employed by Wang *et al.* (2020) and RF used by Long *et al.* (2021) separately classify and find the regression between different operational control measurements and biogas generation. Tufaner and Demirci (2020) used simple ANN to predict biogas generation in a laboratory scale AD by using pH, alkalinity, organic load rate, chemical oxygen demand (COD) and total solid (TS) Park *et al.* (2021)

similarly used pH, alkalinity, COD removal and volatile solids as input variables for ANN to predict biogas yield. More recently, Pei *et al.* (2022) used data mining and ANN models to estimate biogas generation based on TS, C/N, pH and acid concentration. These efforts aimed at estimating the AD outputs especially biogas production based on the system variables especially pH, alkalinity, and effluent pollution. Although the development of smart and decision-making frameworks in waste management have recently attracted more attention by researchers (Shahsavar *et al.,* 2021, Shahsavar *et al.,* 2022), none of the previous works either developed a framework for the AD operation based on the ANN models or carried out proper investigations on the effect of different waste compositions and the water added to the AD on biogas yield. Furthermore, those previously developed models mainly used simple ML or ANN whereas the performance of the AD procedure may fit in better with simulation of time-series models that rely on earlier timesteps. This is particularly important because AD systems are operated continuously and are highly dependent on sequential and continuous input waste load (Yang *et al.,* 2022; Chozhavendhan *et al.,* 2023). This type of modelling can be envisaged through the application of a recurrent neural network (RNN) model for monitoring the performance of the AD processes (Offie *et al.*, 2022). Hence, this study aims to develop a new smart framework for optimal operation performance of micro-AD plants located in a residential area based on Recurrent Neural Network (RNN) and optimisation techniques. It is also aimed at determining the maximum volume of biogas that can be generated from the micro-AD plant. This framework is demonstrated by its application to historic data obtained from a real case of a micro AD plant in London, UK.

This paper is organised as follows: in section 2, the features of the micro-AD plant used as the pilot study as well as the description of the micro-AD site location will be clearly stated. The nature of data collected from the micro-AD plant and different techniques adopted for data imputation will be then presented. In addition, the type of artificial neutral networks (ANN)-

based model developed for monitoring and improving the efficiency of the micro-AD plant will also be presented and described in detail alongside the sensitivity and uncertainty analysis carried out to assess the performance of the developed ANN model. The results obtained from infilling the missing data and the ANN model development and testing will be presented and discussed in detail in section 3 followed by finally summarising key findings and remarking notes in section 4.

2. Methodology

This study presents a new AI-based framework for the simulation and optimisation of micro-AD plants based on data-driven models. Figure 1 shows the methodology in this study comprising three main steps as data collection/preparation, model development and performance assessment. These steps are commonly used for developing most data-driven environmental models (Piadeh *et al.*, 2022). The AI-based framework is mainly used as the core tool for estimating and optimising biogas generation based on the feed data collected over preceding days. All steps of the framework were carried out using MATLAB 2021b software which provides functions for estimation and optimisation of the system performance. These steps follow a series of procedures after collecting data from the micro-AD plant, which are described below with more details.

2.1 Data collection and preparation

This stage entails data collection and imputation for infilling missing data using some data-mining-based techniques, selection of relevant data for model development. The data in this study was collected from a micro-AD plant located in Camley Street Natural Park Central London, United Kingdom (UK) with the schematic diagram shown in Figure 2 (Walker *et al.*, 2017). The micro-AD plant in this site had a pre-feed tank consisting of a chopper mill, mixing pre-feed tank on load cells and a feed pump. It also had a main anaerobic digester containing an automated mechanical mixer and heater by an internal water heat exchanger. Other main components of the micro-AD plant as shown in the figure include the hydrogen sulphide scrubber filled with activated carbon pellets, floating gasometer for biogas storage, digestate sedimentation tank, digestate liquor storage tank. The micro-AD plant was monitored for a period of 310 days during which the operational parameters, biological stability, and energy requirements of the micro-AD plant were evaluated.

The data collected from the micro-AD plant include temperature, pH, volatile solids, total solids, feed into the main digester, feed composition into the pre-feed tank. The feed composition comprises apple, catering and coffee, coffee, digestate, green waste, oats, soaked peanuts and muesli, tea, tea leaves, tea bags, oil, soaked muesli, soaked liners, and catering. Other data collected are the water added to the pre-feed tank and the volume of biogas generation. The feed into either the pre-feed tank or the main digester was usually done every few days when both feed amounts and biogas volume in the storage were recorded. Hence, out of the monitoring period of 310 days of the micro-AD plant, there were days when no feed was added to either the pre-feed tank or the main digester and no recording of biogas generation while daily continuous data for both feed and biogas are necessary for developing a time-series ANN model that considers lag days. In addition, there were some days with missing output data (i.e., there was feed but there was no reading for biogas generation). This effect can hinder the model accuracy of the micro-AD plant especially for the prediction of the biogas volume generated. Hence, some data-mining techniques were first analysed in this study for estimating the missing data to determine the most suitable one for infilling the missing data. Note that missing data in this study refer to the absence of biogas readings in two types: (1) data samples with feed values available (input) but no reading for biogas generation (output); and (2) data samples with feed value equal to zero but no reading for biogas generation. Therefore, the entire dataset was first divided into two groups of data with feeding inclusive and data without feeding. Some data mining techniques were then tested to identify the relationship between the feed data and the generated biogas for data groups with feeding data. Out of those techniques, the best one was selected for infilling the missing data of the first type (i.e., data with feed values but no biogas values). The second type of missing data (i.e., data where feeding is zero and biogas is unavailable) were infilled based on the linear regression of the remaining total biogas data read. The data mining techniques explored here include Random Forest (RF), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Kriging, Feed Forward Neural Network (FFNN) and Linear Regression (LR).

Sequel to this, a sensitivity analysis was carried out for each of the operational feed variables to determine their correlation and impact on the volume of biogas generation. Based on the cross-correlation analysis of all input variables (demonstrated as Figure A1 and Table A1 in the Appendix), the daily feed into the digester, the water added to the digester showed significant correlation and corresponding impact on the biogas volume generated. In addition, out of various waste compositions, only oats, soaked liners, and catering were selected whereas other waste compositions were negligible as they had no significant correlation and hence no meaningful impact on the volume of the biogas generated. In addition, the volatile solids, total solids, pH, and temperature were measured but observed to be relatively constant during the operation and hence these parameters were also excluded from the analysis for estimating biogas generation.

2.2 Model development

To estimate biogas generation, a type of time-series RNN model known as Non-Linear Autoregressive Neural Network (NARX) was developed here with three hidden 10-neuron layers with the architecture shown in Figure 1. This model was developed based on the selected input variables of the micro-AD plant including the actual/estimated daily feed added to the main digester (X_1), the feed composition comprised of catering (X_2), oats (X_3), and soaked liners (X_4) added to the pre-feed tank (i.e. the top three highly correlated variables with biogas generation), the water added to the pre-feed tank (X_5), and the volume of biogas generation (Y). The model settings are as follows: Levenberg-Marquardt method used for training process; mean square error as the indicator to evaluate the model performance, and 6 epochs (iterations) adjusted for training failure. The database used for model development is divided into three parts as 70% for training, 15% for validation and 15% for test as a common practice (Eghbali *et al.* 2017). The trained model was then used to predict the biogas generation (Y_p) for the micro-AD plant in the case study.

As the NARX model needs lag time specification in day (known as delay factor F_i), i.e., range of input variables for previous timesteps to use for estimation of biogas generation at one timestep ahead (Y_{t+1}) based on input data (decision variables), an optimisation method is used to find the optimal lag time for each decision variable, as a model tunning, to obtain the most accurate output data i.e., biogas generation. To this effect, the optimisation model was developed using shuffled frog leaping algorithm (SFLA). This is a memetic and nature-based algorithm with the ability to search in both local and global search space where each lag time represents one frog (Bui *et al.*, 2020). Here, each frog, i.e., decision variable, represents a lag time to find the minimum root mean square error (RMSE) and the highest Normalized Nash-Sutcliffe Efficiency (NNSE) in this optimisation approach. 4 trials for exploration and 4 trials for exploitation were set for each iteration of optimisation, and stopping criteria being set to an improvement of less than 1%. Each of the six decision variables (i.e., F_0-F_5 in Figure 1) is an integer value ranging between 0 and 10 due to the results of cross-correlation analysis on inputs, provided in Figure A1 in the appendix. Thus, this approach can be used to determine the delay factor (range of previous X_i data) for each input data/decision variable.

This algorithm was then used again to specify the required weights for the daily feeds added to the main digester, daily feed compositions and the water added to the pre-feed tank to maximise the output (i.e., maximum volume of biogas generation from the micro-AD plant) for each of the days in a cyclic period of feeds. Note that the cyclic period is based on the (lag time) delay factor specified in the first optimisation model. While stopping criteria and trials are set similarly, each NARX input for each day are selected as decision variables. To simulate the real operation and put a cap for the feeds/water added to the plant, constraints are defined based on the historic operation of the plant as follows: (1) maximum feed equals to 80 kg every 4 days. Note that 4 days is based on the cyclic period of 5 days (as four days input data and biogas generation in day 5) specified as a result of the optimisation model for the largest lag time (see the result section); (2) total weight of the feed and all pre-feeding compositions should be equalled during the optimisation; (3) added water is limited to 30% of the total feed weight, (4) all decision variables need to be either zero or positive values.

2.3 Performance assessment

Two metrics i.e., RMSE and NNSE are used in this study to evaluate the performance of the developed NARX model. The developed model is also evaluated using both the sensitivity analysis and uncertainty analysis. The sensitivity analysis is carried out to show the significance of each input variable for yielding the predicted output (biogas) by removing one input parameter and running the model afterwards with NNSE and RMSE both for observations. The uncertainty analysis on the other hand was done to show how the relative accuracy is changed when running the model with dataset reduction. When carrying out the sensitivity analysis, the optimal waste composition is taken into consideration where the impact of each waste composition is analysed and evaluated to determine the significant impact of each waste composition on the generation of biogas. Further analyses are also carried out on the feed data to further evaluate its importance in the performance of the developed model.

3. Results and discussion

3.1. Data imputation results

Figure 3 shows the performance of the data mining techniques applied for infilling the missing data. The performance of these techniques is shown for RMSE of the test data based on the cross-validation method in which all data samples participate in the evaluation of the test set (Eghbali *et al.,* 2017). The 6-fold cross-validation method was used in this study for the performance assessment of infilling the missing data. Based on the results presented, it is evident that the Kriging technique had the least range of fluctuations amongst other data mining techniques with an average RMSE value of 1.23 m^3/day compared to 1.25-2.25 m^3/day for the other techniques. As a result, the Kriging technique was selected and used to obtain the missing biogas values with complete feed values thereby giving rise to more accurate predictions of the biogas produced from the AD plant. This further confirms the effectiveness of the Kriging technique in infilling missing data where other previous studies have not highlighted dealing with missing data and usually have used simple techniques, especially linear regression (Pei *et al.*, 2022). This is specifically important because similar to many industrial and real practices, there was no daily measurement for the generated biogas and only 123 non-sequential data out of the 310 operation days were recorded. Hence, this infilling technique could provide acceptable results used to develop the model for accurate estimation and optimisation of the generated biogas volume.

3.2.Fine-tuning of lag times of input variables

Figure 4 shows the optimal number of lag times for each input variable is obtained after 8 trials. This figure also shows the obtained lag times in each iteration and their corresponding model performance metrics (i.e., RMSE and NNSE). As can be seen in Figure 4, the RMSE and NNSE were observed to decrease and increase gradually over the trial number, respectively. The results show that the optimal daily lag time data is 5 days for the added water (F_5), followed by

3 days for the feed added to the main digester (F_1), and then 1 day for other variables (i.e., catering, oat, soaked-liner, and biogas generation). This shows that adding water to the pre-feed tank can influence the content of the generated biogas in the main digester from the past 5 days whereas waste compositions can immediately impact on biogas generation from only one previous day. This also shows the yielded biogas can be dependent on the daily distribution of feeds and water. Furthermore, one day lag time in the waste composition indicates that the process of biogas generation is highly influenced by specific rate of waste compositions added to pre-digester, even if this rate is different from that of material added the digester.

Furthermore, the cross-correlation analysis, provided in Figure A1 in the appendix, shows the highest correlation coefficient between biogas generation and the feed adding to the main digester occurs in previous 5 days (used as initial for F_1 in Figure 4) whereas optimised lag time for feed is reduced to 3 days (trial 7 in Figure 4). Similarly, daily lag times for catering, oat and liners are reduced from 3 days in initial trial based on the cross-correlation analysis to only 1 day. However, the high correlation of 5^{th} to 3^{rd} days ago for the added water were initially ignored (number 2 in the initial row for F_5 in Figure 4 vs number 5 in the last row). This can be due to the impact of the combination of input variables on optimal lag times that is shown in the significant improvement of metrics, i.e., RMSE (decrease from 1.4 to 0.4) and NNSE (increase from 0.6 to around 0.9). Although most of the previously developed NARX models recommended using cross-correlation results for developing NARX model (Abdel daiem *et al.,* 2022), the difference between initial lag times and final lag times obtained from the SFLA method shows the added value of using optimisation models to fine tune these time-series models.

3.3. Performance assessment of biogas predictions

Figure 5a compares the biogas measurements with the corresponding estimations over the test period. From this figure, the disparity between the predicted biogas and the generated biogas was observed to be quite insignificant. Figure 5b also shows the performance assessment of the developed model where the scatter plot of predicted biogas versus corresponding measurements for one day lead time (i.e., one day ahead) for the three types of the feed. The RMSE values are observed to be 0.33 m^3/day for heavy weight feed (greater than 20 kg), 0.46 m^3/day for medium feed (within 10-20 kg) and 0.39 m^3/day for light weight feed (less than 10kg). While it can be deduced that the feed with weight greater than 20 kg had the least range of errors compared to other two feeds, the coefficient of variance (CV), indicates that the biogas estimation model is highly sensitive to the feed with lower weights compared to the feed with higher weights. However, the three RMSE values obtained which are relatively low indicating that the efficiency of the model developed is relatively high and hence reliable to be used as a surrogate model for estimation of biogas generation in the micro-AD plant. Furthermore, the coefficient of variance (CV) of 13% in this study (Figure 5b) can be compared to previously reported studies which are in a range between 31% by Wang *et al.*, (2020) and 23% by Long *et al.* (2021). This confirms the effectiveness of the developed model in predicting the biogas generated from the micro-AD plant. It also indicates that the developed NARX-ANN model is robust enough to be used for relatively high fluctuated biogas generation. Although the overall NNSE can be quite acceptable in this model but high variability of the measured biogas (3.26±1.21 reported in Table A1 in the appendix) may be considered as a drawback for the model in tracking biogas, particularly when there is a sudden change (e.g., droop at end days as shown in Figure 5a).

Figures 5c and 5d present further sensitivity analysis for the impact of the percentage of the data used and each of the feed type, respectively, on the accuracy of the developed model.

Figure 5c shows the prediction accuracy of the model development in both metrics (NNSE and RMSE) is changed with a relatively similar and linear trend for the percentage of the data used. As the limited dataset (only 310 days) was used for all steps of training, validation and testing, the developed model would still be highly dependent on the volume of dataset. This would also confirm a relatively low coefficient of cross-correlation analysis between the input variables and the generated biogas, as illustrated in Figure A1 in the appendix. The sensitivity analysis presented in Figure 5d is also conducted by removing one decision variable and running model for one step ahead. From the sensitivity analysis presented, the catering and oats compositions demonstrate the highest impact on the prediction accuracy compared to other input variables. This also indicates that both the oat and catering compositions have the most significant impact on the biogas generation in the micro-AD plant compared to the other input variables. On the other hand, the liner-soaked composition has the least impact on the prediction accuracy and hence minimum impact on the biogas generation in the waste composition.

3.4.Optimal feeding pattern and maximum potential biogas generation

The SFLA optimisation method is used to specify the optimal feeding pattern to obtain maximum biogas generation. Based on the optimal number of lag times (days) for each variable obtained in Figure 4, the optimisation method is arranged for 18 decision variables as follows: four variables for feed added to the main digester at days t-3, t-2, t-1 and t; six variables for the water added to pre-feed tank at days t-5, t-4, t-3, t-2, t-1 and t; eight variables for each of the three waste composition types (i.e., catering, oat, and liner) and biogas generation at days t-1 and t. The objective value is to maximise the biogas generation at day t+1. The SFLA is run and the results of the optimal feeding pattern within the these setting days for each variable along with optimal biogas generation are shown in Table 1. Note that as the values for the added water are zero for all days except day t-1, the optimisation model was run for smaller number of days for the added water and the results showed that maximum biogas is generated when a four-day cycle is considered for feeding pattern. Hence, the optimal decision variables in Table 1 are shown for 4 days between t-3 and t. The estimation of maximum biogas generation at day t+1 is 4.52 m^3/day based on the optimal decision variables at the preceding days between days t and t-3. This is the maximum possible biogas generation from this micro-AD plant based on the optimal values of decision variables. The analysis of the optimal decision variables shows that the entire feed for a four-day cycle only needs to be added on the last day (i.e., 80kg on day t). Out of this 80kg feed, catering added to the pre-feed tank needs to be 60kg with a distribution of 55kg at day t-1 and 5kg at day t; the remaining waste added to the pre-feed tank is 20kg oat added only at day t-1; there is no need for adding any liner; and finally, 15kg water added to the pre-feed tank is needed only at day t-1. Accordingly, the biogas generation of the following day (i.e., day t+2) is estimated 4.23 m^3/day based on the input (decision) variables of the preceding four days (i.e., between days t-2 and t+1). Table 1 shows the summary of the estimated biogas generation but more details of biogas generation for these four days (i.e., between days t+1 and t+4) are given in Table A2 in the appendix. This decrease in biogas generation is because the daily distribution of input especially feed, catering and water is different from the optimal values obtained above. Similarly, estimation of biogas generation in the following days (i.e., day t+3 and t+4) was observed to decrease further. On the other hand, if the same amounts of feed, catering and water are added every 4 days, the estimated biogas generation is repeated every 4 days. In other words, the estimation of biogas generation after day 4 is observed to be repeated as the same for the input variables with the same feeding pattern. This indicates the volume of biogas generation and feeding pattern can be repeated every four days. As there is no feeding in the first two days, this can also be economically beneficial for the operation of the micro-AD plant which can be mainly operated by local communities with minimum labour (i.e., most of the feeding is arranged for one day every four days) to achieve the maximum efficiency of biogas generation.

Figure 6 shows the sensitivity analysis for biogas generation at day t+1 based on percentage of variables over the past three days i.e., days t-2, t-1, and t for all decision variables. Figure 6a shows the impact of the "feed to the main digester" on the biogas generated where different percentages of the feed data for days t and t-1 are shown in horizontal axes and feed data for day t-2 are shown as graphs with an interval of 10%. As can be seen, the maximum biogas generation (4.52 m^3/day) can only occur when AD is fed only on the last day (day t). In addition, any redistribution of feeding shows a decrease in the biogas generation which can be translated as relative sensitivity of the model to the daily distribution of feed. For example, when feeding the AD plant at day t-1 instead of day t, the biogas generation dropped from 4.5 to around 2.5 m^3/day (see blue circle in left-top and right-bottom). The results also show that model is highly sensitive to the amount of feed in last two days (i.e., day t and t-1) and feed ratio for day t-2 is less important (See blue circles are varied more than other lines indicating the model is sensitive to day t and day t-1).

Figure 6b shows the impact of the water added on the volume of biogas generation with 20% intervals. Compared to feed distribution in Figure 6a, the daily distribution of added water has a relatively low impact on the biogas generation, that slightly changes between 4.3 to 4.5 m^3/day. Figure 6c presents the impact of the three different composition variables on the volume of biogas generation. It shows the catering composition added to the pre-feed tank results in the maximum biogas generation compared to other variables. This implies that the catering composition has a higher influence on biogas generation than the other composition variables. Following this, the oat composition also generates a high volume of biogas as shown in Figure 6c. This also indicates that the oat composition has a strong influence on the volume of biogas generation. This is also in line with the sensitivity analysis presented in Figure 6d that shows the model is more sensitive to adding liner rather than daily distribution of composition. For example, while daily distribution of oats and catering has no significant impact on the generated biogas, it is highly sensitive to the amount of added liner in Figure 6e.

When applying optimal operation strategy, it is crucial to understand the significance of the distribution of composition variables added to the pre-feed tank over the cyclic period. Hence, the impact of distribution of the optimal amount of the pre-feed composition variables on biogas generation is further analysed in Figure 6d-e for three individual variables i.e., catering, oats, and liner separately. More specifically, Figure 6d considers distribution of optimal value of the catering variable at day t where other variables are fixed here as 20kg for oat, zero for liner and optimum condition of accumulative rate of catering is 60kg (i.e., 55kg for day t-1 and 5kg for day t). As can be seen in Figure 6d, the biogas generation slightly decreases for other distribution rates down to 4.515 m^3. This indicates that while the AD plant is highly dependent on catering (as shown in the sensitivity figures), its distribution between the day t and t-1 has no significant impact on biogas generation. Although there is a drop in the volume of biogas generation as the share of oat on day t increases compared to the optimum value, the drop in biogas volume has no significant impact on biogas generation. In Figure 6e, the share of liner in day t (%) is observed where each line corresponds to a percentage of liner in total waste. The horizontal axis shows how this percentage is distributed between day t and day t-1. Hence, for 20% of the liner, the available data is for 0-5-10-15 and 20%. From Figure 6e, it can be observed that the increasing the liner results in a decrease in the biogas volume as the liner has low impact on biogas generation compared to other waste types.

Finally, the strategy for generating optimised biogas generation is compared with best feeding events and entire test period (47 days). As it can be seen, the biogas generation in all three best identified feeding events (as shown in Figure 7a) is relatively similar and a uniform increase with a maximum weekly volume of 26.14 m^3. However, the maximum volume of biogas generation for the optimised operation is 29.97 m^3 i.e., an improvement rate of 15%. In Figure 7b, the average daily biogas generated over the first 40 days is observed to 3.26 m^3/day. The generated biogas decreases from days 40-46 as the average volume of biogas generation between days 41-46 is observed to be 1.48 m^3 /day. Similarly, the generated biogas for the entire test period for the measured event shows an increase up until day 40 when it experienced a slight decrease in the generated biogas volume. It then increases the next day with a maximum volume of 139.51 m^3. On the other hand, the generated biogas in the optimised operation can uniformly increase with a steeper slope and achieve up to a maximum of 199.46 m^3 i.e., a significant improvement of 43% for biogas generation compared to the business-as-usual operation. The proposed model has better performance in both short- and long-term operation, i.e., 7 days and 47 days which longer period results in more biogas enhancement. This indicates the potential benefit of developing an optimised strategy for the operation of the micro-AD plant that results in maximum biogas generation and hence the improvement of overall performance and productivity of the micro-AD plant.

4. Conclusions

This paper developed a three-step AI-based framework for estimating and optimising biogas generation from the real-world micro-AD plant. The first step entailed data collection and imputation for infilling the missing data by using several methods, in which the Kriging technique outperformed conventional techniques, particularly KNN, SVM, LR and FFNN.

A NARX model was developed in the second step for estimation of biogas generation of the following day based on variables of feed and waste composition added to the main digester and the pre-feed tank for the preceding days which were initially determined based on cross-correlation of time-series of biogas generation and the above variables. The SFLA optimisation model was then used to fine tune the number of lag times of the input layer variables. The result of fine-tuning the lag times showed the yielded biogas is highly sensitive to waste composition (catering, oat and liner-soaked) up to only one previous day, whereas this is up to three previous days for the feed and five previous days for the added water.

The SFLA optimisation model was used again to determine the optimal daily feeding pattern that generate maximum biogas in the micro-AD plant. The results show that the optimal feeding pattern can increase the biogas generation up to 15% and 43% for a 7-day and 45-day period, respectively with an average RMSE of 0.39 m^3/day. The optimal daily feeding pattern is obtained for a four-day cycle in which water, catering and oat are added to the pre-feed tank within the last two days (mainly on day 3) and the feed is added to the main digester in day 4 and hence the no feeding is needed within the first two days. This can also reduce the operational costs related to the labour for feeding the plant.

The sensitivity analysis of the developed model shows that biogas generation is strongly influenced by the oats and catering content compared to other feed types. In addition, any change to the optimal daily distribution of the feed added to the main digester is much more sensitive to biogas generation compared to other three feed types (i.e., added water, oat, and catering). Hence, to obtain high biogas generation, it is recommended that the micro-AD plant is fed in one day and allowed to rest for three days compared with gradual feeding every day. Furthermore, adding liner to the plant can significantly reduce the volume of biogas generation.

Although the results in this study show significant performance improvement of the micro-AD plant in generating biogas, the developed framework needs to be further tested and verified in other AD plants with longer analysis periods to show the efficacy of the developed approach. In addition, further studies need to be carried out for improving the model ability in tracking sudden change of feed as a common challenge for the AD operation in practice. Further analysis and data modelling can be performed to alleviate other challenges in the AD technology such as long residence time.

Declaration of competing interests

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

Acknowledgements

This work is supported by the Knowledge Exchange (KE) Seed Fund allocated to the fifth author (industrial partner) and the Fellowship allocated to the third author. The authors wish to acknowledge the KE seed fund supported by the University of West London and the Fellowship supported by the Royal Academy of Engineering under the Leverhulme Trust Research Fellowships scheme. The authors also wish to thank the Diego Vega from LEAP Ltd and Dr Davide Poggio from the University of Sheffield for their great support to provide and analyse the data collected from the case study. The authors also wish to thank the editor and the three anonymous reviewers for making constructive comments which substantially improved the quality of the paper.

References

Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., Mahfood B. (2020). Artificial intelligence applications in solid waste management: A systematic research review. *Waste Management*, 109, 231-246.

Abdel daiem, M., Hatata, A., Said, N. (2022). Modeling and optimization of semi-continuous anaerobic co-digestion of activated sludge and wheat straw using Nonlinear Autoregressive Exogenous neural network and seagull algorithm, *Energy*, 241, 122939.

Ankun Xu, H. C. (2021). Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review. *Waste Management 124*, 385-402.

Antoniou, N., Monlau., F., Sambistu., C., Ficara., E., Barakat., A., Zabaniotou., A. (2019). Contribution to Circular Economy options of mixed agricultural wastes management: Coupling anaerobic digestion with gasification for enhanced energy and material recovery. *Journal of Cleaner Production 209,* 505-514.

Arun, C., Sivashanmugam, P. (2017). Study on optimization of process parameters for enhancing the multi-hydrolytic enzyme activity in garbage enzyme produced from preconsumer organic waste. *Bioresource Technology 226*, 200-210.Association, W. B. (2021). *Biogas: Pathways to 2030.* Retrieved from worldbiogasassociation.org

Carlini, M., Castellucci, S., Mennuni, A., Selli, S. (2020). Simulation of anaerobic digestion processes: Validation of a novel software tool ADM1-based with AQUASIM. *Energy Reports*, 6(6), pp. 102-115.

Chartered Institute of Water and Environmental Management (CIWEM). (2021). *Policy Position Statement on Waste and Resource Management*. Available at www.ciwem.org [retrieved 02-11-2022].

Cheela, V., Ranjan, V., Goel, S., John, Dubey, B. (2021). Pathways to sustainable waste management in Indian Smart Cities. *Journal of Urban Management*, 10(4), pp. 419-429.

Chozhavendhan, S., Karthigadevi, G., Bharathiraja, B., Kumar, R., Abo, L., S. Prabhu, Balachandar, R., Jayakumar, M. (2023). Current and prognostic overview on the strategic exploitation of anaerobic digestion and digestate: A review, *Environmental Research*, 216(2), 114526.

Cruz, I., Chuenchart, W., Long, F., Surendra, K., Andrade, L., Bilal, M., Liu, H., Figueiredo, R., Khanal, S., Ferreira, L. (2022). Application of machine learning in anaerobic digestion: Perspectives and challenges, *Bioresource Technology*, 345, 126433.

Eghbali, A.H., Behzadian, K., Hooshyaripor, F., Farmani, R. and Duncan, A.P. (2017). Improving prediction of dam failure peak outflow using neuroevolutionary combined with K-means clustering. *Journal of Hydrologic Engineering, 22(6).*

Kumar, M., Dutta, S., You, S., Luo, G., Zhang, S., Show, P., Sawarkar, A., Singh, L., Tsang, D. (2021). A critical review on biochar for enhancing biogas production from anaerobic digestion of food waste and sludge, *Journal of Cleaner Production*, 305, 127143.

Kunatsa, T., Xia, X. (2022). A review on anaerobic digestion with focus on the role of biomass co-digestion, modelling and optimisation on biogas production and enhancement, *Bioresource Technology*, 344(B), 126311.

Long, F., Wang, L., Cai, W., Lesnik, K., Liu, H. (2021). Predicting the performance of anaerobic digestion using machine learning algorithms and genomic data, *Water Research*, 199, 117182. doi.org/10.1016/j.watres.2021.117182

Offie, I., Piadeh, F., Behzadian, K., Alani, A., Yaman, R. and Campus, L., (2022). Real-time monitoring of decentralised Anaerobic Digestion using Artificial Intelligence-based framework, *International Conference on Resource Sustainability (icRS)*, Virtual conference. http://repository.uwl.ac.uk/id/eprint/9368/1/iCRS%20Paper.pdf [retrieved 02-11-2022].

Park, J., Jun, H., Heo, T. (2021). Retraining prior state performances of anaerobic digestion improves prediction accuracy of methane yield in various machine learning models, *Applied Energy*, 298, 117250.

Piadeh, F., Behzadian, K. and Alani, A., (2022). A critical review of real-time modelling of flood forecasting in urban drainage systems. Journal of Hydrology, p.127476.

Pei, Z., Liu, S., Jing, Z., Zhang, Y., Wang, J., Liu, J., Wang, Y., Guo, W., Li, Y., Feng, L., Zhou, H., Li, G., Han, Y., Liu, D., Pan, J. (2022). Understanding of the interrelationship between methane production and microorganisms in high-solid anaerobic co-digestion using microbial analysis and machine learning, *Journal of Cleaner Production*. 373, 133848.

Shahsavar, M.M., Akrami, M., Gheibi, M., Kavianpour, B., Fathollahi-Fard, A.M. and Behzadian, K., (2021). Constructing a smart framework for supplying the biogas energy in green buildings using an integration of response surface methodology, artificial intelligence and petri net modelling. *Energy Conversion and Management*, 248, p.114794.

Shahsavar, M.M., Akrami, M., Kian, Z., Gheibi, M., Fathollahi-Fard, A.M., Hajiaghaei-Keshteli, M. and Behzadian, K., (2022). Bio-recovery of municipal plastic waste management based on an integrated decision-making framework. *Journal of Industrial and Engineering Chemistry*, 108, pp.215-234.

Tufaner, F., Demirci, Y. (2020). Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. *Clean Technology Environmental Policy*, 22, pp. 713–724.

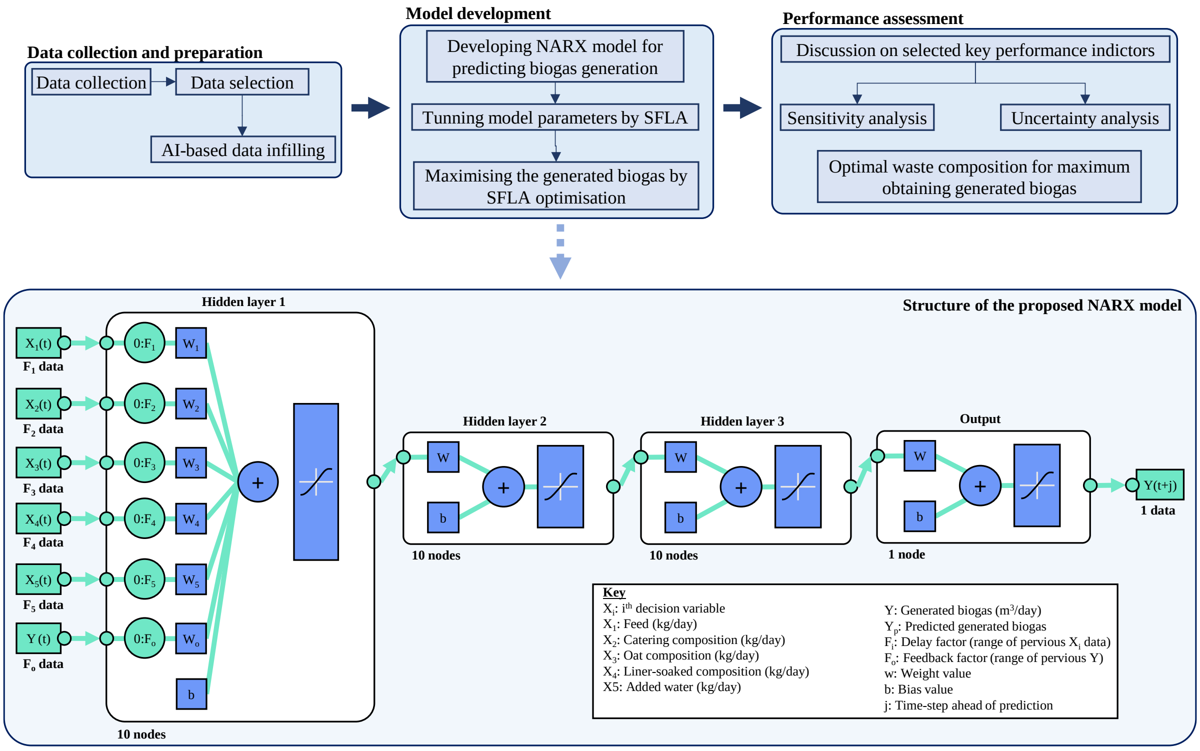
Yang, W., Li, S., Qv, M., Dai, D., Liu, D., Wang, W., Tang, C., Zhu, L. (2022). Microalgal cultivation for the upgraded biogas by removing CO2, coupled with the treatment of slurry from anaerobic digestion: A review, *Bioresource Technology*, 364, 128118.

Walker., M., Theaker., H., Yaman., R. Poggio., D., Nimmo., W., Bywater., A., Pourkashanian., M. (2017). Assessment of Micro-Scale Anaerobic Digestion for Management of Urban Organic Waste: A Case Study in London, UK. Journal of Waste Management Journal of Waste Management 2017.01.036.

Wang., L., Long, F., Liao. W., Liu, H., (2020). Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. *Bioresource. Technology*. 298, 122495.

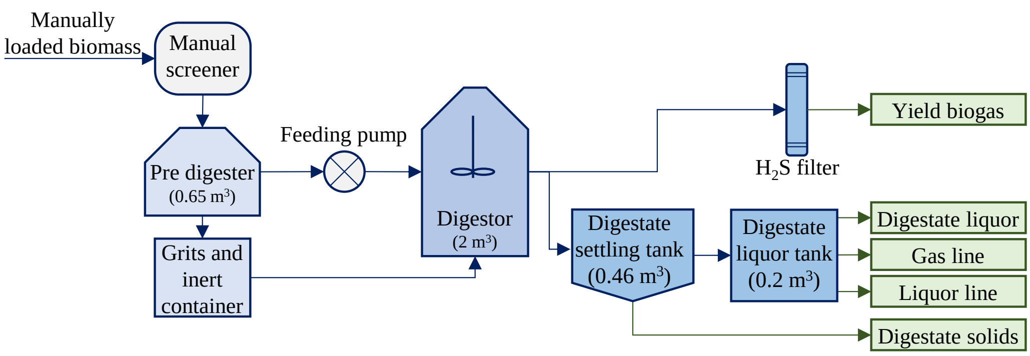
World Biogas Association (WBA). (2021) *Biogas: Pathways to 2030.* Retrieved from *worldbiogasassociation.org*

Zhang, L., Loh, K., Zhang, J. (2019). Enhanced biogas production from anaerobic digestion of solid organic wastes: Current status and prospects, *Bioresource Technology Reports*, 5, pp. 280-296.



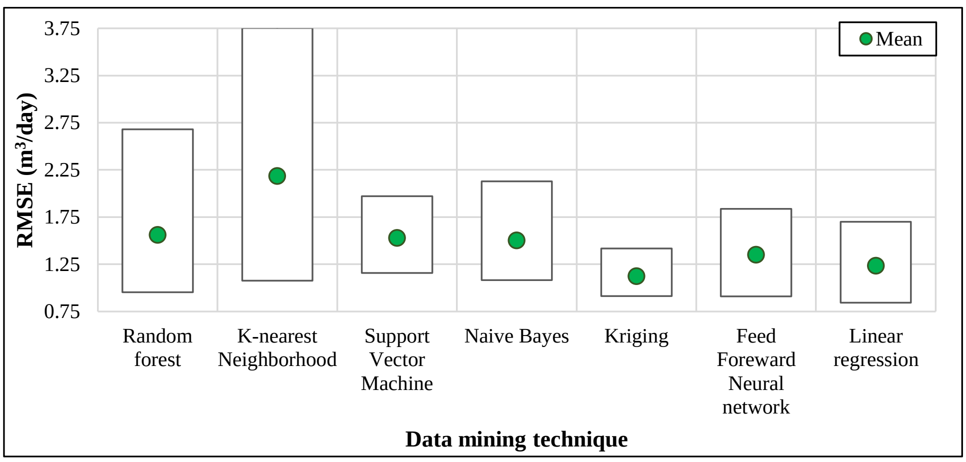
571

572 **Figure 1. AI-based framework for the operation of the micro anaerobic digestion plant**



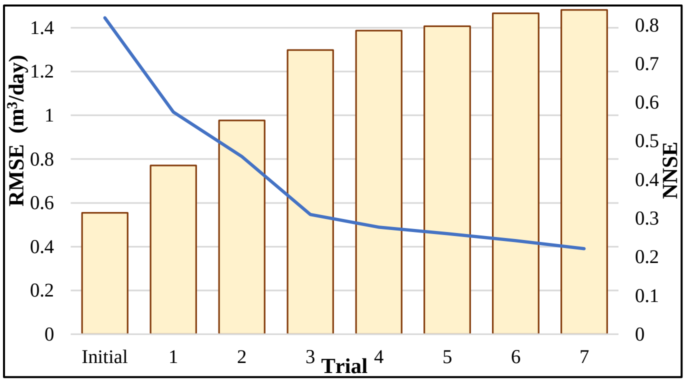
573

574 **Figure 2. Schematic diagram of the micro-AD plant used in this study**



575

576 **Figure 3. Performance of data mining techniques applied for infilling the missing data**



Trial	Optimised number of lag times (days)					
	F_1	F_2	F_3	F_4	F_5	F_o
Initial	5	3	3	3	2	1
1	6	1	2	2	2	1
2	4	2	1	1	3	2
3	7	1	2	1	3	7
4	10	2	1	1	3	11
5	3	1	1	1	5	5
6	3	1	1	1	5	3
7	**3**	**1**	**1**	**1**	**5**	**1**

F_1: Feed F_2: Catering composition F_3: Oat composition F_4: Liner-soaked composition F_5: Added water F_o: Generated Biogas
NNSE RSME

Figure 4. The trend of the SFLA method to specify optimal number of lag times for input variables

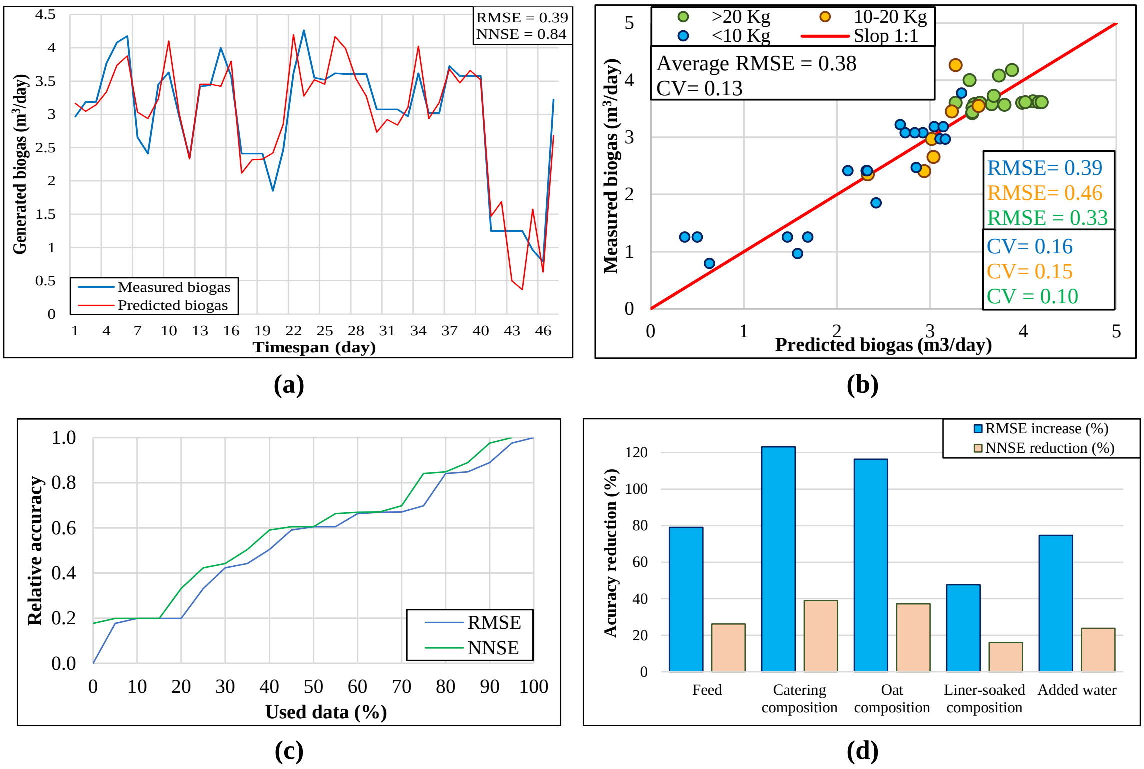


Figure 5. (a) scatter plot of predicted biogas vs corresponding measurements for 1 day ahead, (b) comparison of observations with estimations, (c) relative accuracy based on the percentage of dataset used for model development, (d) Impact of feed compositions in the pre-feed tank on the biogas generation

Table 1. Optimum condition for the operation of the micro-AD plant for maximum biogas generation

Parameter	Days							
	Predictors (input data)				Predictions			
	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
Feed	0	0	0	80				
Biogas	-	-	4.11	4.08	4.52	4.23	4.11	4.08
Catering	-	-	55	5				
Oat	-	-	20	0				
Liner	-	-	0	0				
Water	0	0	15	0				

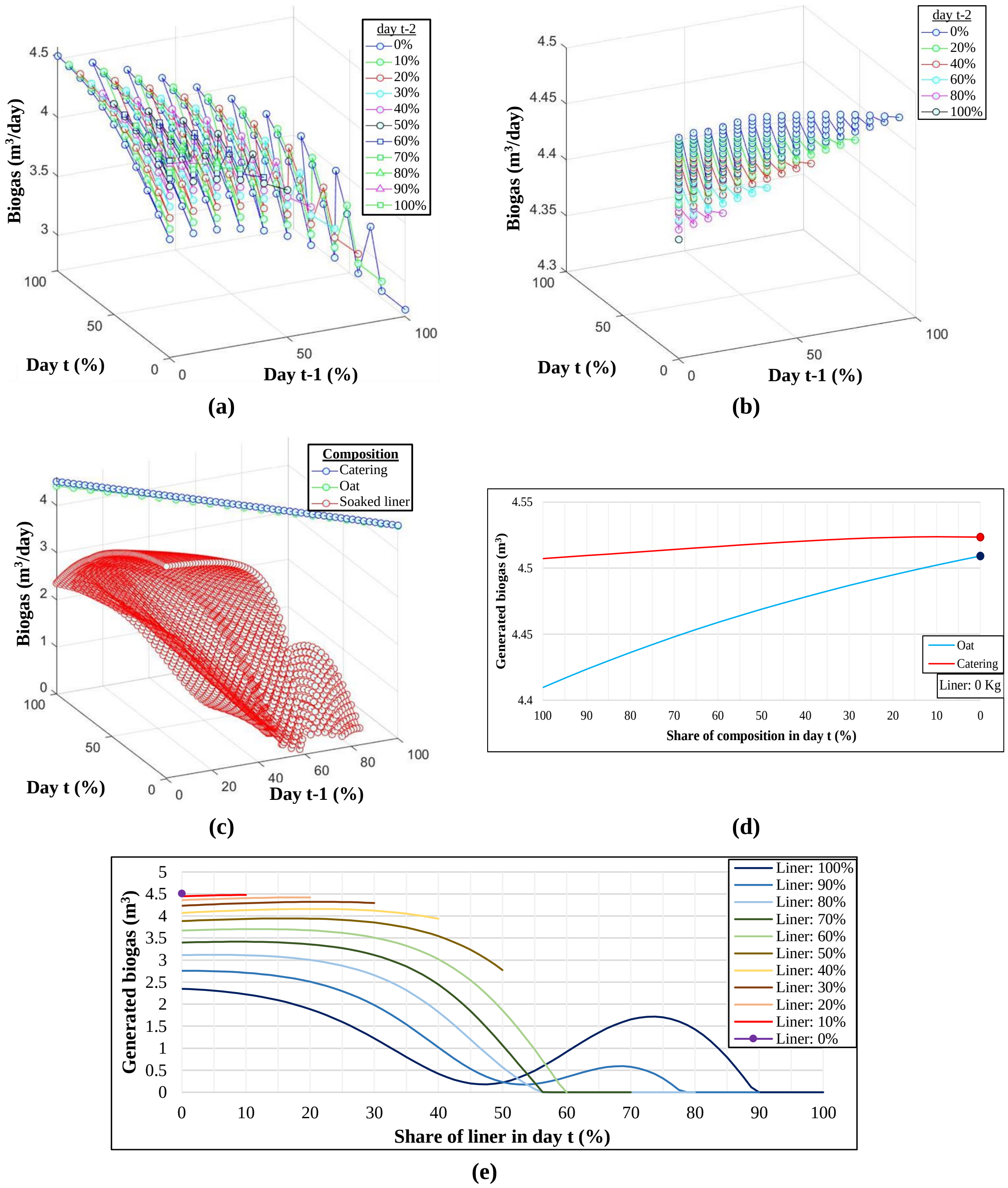
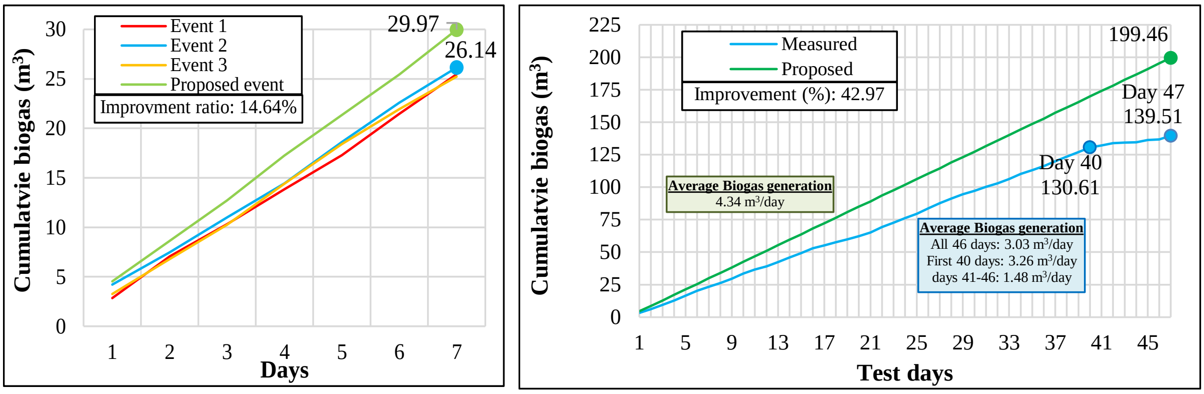


Figure 6. Sensitivity analysis of biogas generation for different decision variables: (a) feed, (b) water, (c) waste composition; and the impact of the distribution of the optimal values of the pre-feed composition variables on biogas generation for (d) catering and oat, and (e) liner



(a) (b)

586 **Figure 7. Comparison between the best feeding events and the proposed (optimised) operation for**
587 **operation in (a) 7 days and (b) 47 days**