Data in Brief 38 (2021) 107323

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Low-temperature thermal hydrolysis of sludge prior to anaerobic digestion: Principal component analysis (PCA) of experimental data ☆

Mohamed N.A. Meshref[a,b], Seyed Mohammad Mirsoleimani Azizi[a], Wafa Dastyar[a], Rasha Maal-Bared[c], Bipro Ranjan Dhar[a,*]

[a] Civil and Environmental Engineering, University of Alberta, 116 Street NW, Edmonton, AB T6G 1H9, Canada
[b] Public Works Department, Faculty of Engineering, Ain Shams University, 1 El Sarayat St., Abbassia, Cairo, 11517, Egypt
[c] EPCOR Water Utilities, 2000–10423 101 Street NW, Edmonton, Alberta T5H 0E8, Canada

ARTICLE INFO

Article history:
Received 2 June 2021
Accepted 12 August 2021
Available online 25 August 2021

Keywords:
Principal component analysis (PCA)
Cluster analysis
Thermal hydrolysis process (THP)
Low-temperature thermal hydrolysis
Anaerobic digestion
Macromolecular compounds
Primary sludge fermentation
Methane yield

ABSTRACT

Here, we report data of the principal component analysis (PCA) assessment and clustering analysis related to low-temperature thermal hydrolysis process (THP) for enhancing the anaerobic digestion (AD) of sludge in wastewater treatment plants (WWTPs) with primary sludge fermentation (Azizi et al., 2021). The PCA was examined to pinpoint the influence of different THP schemes on the variations of macromolecular compounds solubilization after low-temperature THP and the relative performances in enhancing methane potential in AD. We established 2 experimental setups with a total of 18 treatment conditions (3 exposure times, 30, 60, and 90 min at three temperature levels 50, 70 and 90 °C) in comparison to the untreated control samples. Scheme-1 comprises the THP of a mixture of (1:1 vol ratio) fermented primary sludge (FPS) and thickened waste activated sludge (TWAS); while scheme-2 comprised the THP of TWAS only. The factors employed in the assessment of the PCA encom-

☆ Given her role as Section Editor of this journal, Rasha Maal-Bared had no involvement in the peer-review of this article and has no access to information regarding its peer-review.

DOI of original article: 10.1016/j.cej.2021.130485

* Corresponding author.
E-mail address: bipro@ualberta.ca (B.R. Dhar).

https://doi.org/10.1016/j.dib.2021.107323

passed the variations in the macromolecular compounds and other solubilization metrics. This included the variations in the levels of carbohydrates, lipids, proteins, and solubilization of chemical oxygen demand (COD) and volatile suspended solids (VSS). Furthermore, the evaluation considered the changes of volatile fatty acids (VFAs) and total ammonia nitrogen (TAN) with respect to time and temperature. The assessment of PCA classified the THP based on their differences and alterations that occurred after the treatment. The indices of the PCA assessments differed based on the factors of concern and the focus of each individual PCA assessment. In every individual PCA assessment, the respective contribution to the total variance in PCA analysis was calculated and manifested by the highest distribution of the principal components (PCs) axis PC1 and PC2. The differences in distributions of PCs after various PCA examinations can describe the relative influence of THP schemes and the most significant variables that can trigger major differences among THP conditions. The comparative differences demonstrated by PCA support the potential investigations of the efficiency of THPs conditions and their performance categories.

Specifications Table

Subject	Environmental Chemical Engineering
Specific subject area	Development and optimization of low-temperature THP for enhancing sludge anaerobic digestion
Type of data	Tables and Figures
How data were acquired	Laboratory experiments and statistical analysis
Data format	Experimental raw data and their analysis
Parameters for data collection	Chemical oxygen demand (COD), total suspended solids (TSS), volatile suspended solids (VSS), volatile fatty acids (VFAs), carbohydrates, proteins, lipids, total ammonia production (TAN), and biochemical methane potential (BMP).
Description of data collection	- Principal component analysis (PCA) and clustering analysis related to low thermal hydrolysis processes (THPs). - PCA data was processed based on the primary datasets (i.e., raw data) of the THPs experiments using Excel and Minitab 19. - Two THP schemes (Scheme-1: TWAS+FPS; Scheme −2: TWAS only) under different combinations of temperatures and exposure times (3 exposure times, 30, 60, and 90 min at three temperature levels 50, 70, and 90 °C) in addition to control samples. - Control and test samples were analyzed for COD, TSS, VSS, VFAs, carbohydrates, proteins, TAN. - Functional groups associated with various macromolecular compounds were analyzed with Fourier transform infrared (FTIR) spectroscopy. - BMP values of control and THP samples were analyzed with the batch test.
Data source location	University of Alberta, Civil and Environmental Engineering, Edmonton, AB, Canada 53.56123561773164 N, −113.41511692082224 E
Data accessibility	PCA data are provided within the article while raw data are provided in the supplementary file.
Related research article	Mohammad Mirsoleimani Azizi, S., W. Dastyar, M. N. A. Meshref, R. Maal-Bared & B. Ranjan Dhar (2021) Low-temperature thermal hydrolysis for anaerobic digestion facility in wastewater treatment plant with primary sludge fermentation. Chemical Engineering Journal, 130,485. https://doi.org/10.1016/j.cej.2021.130485

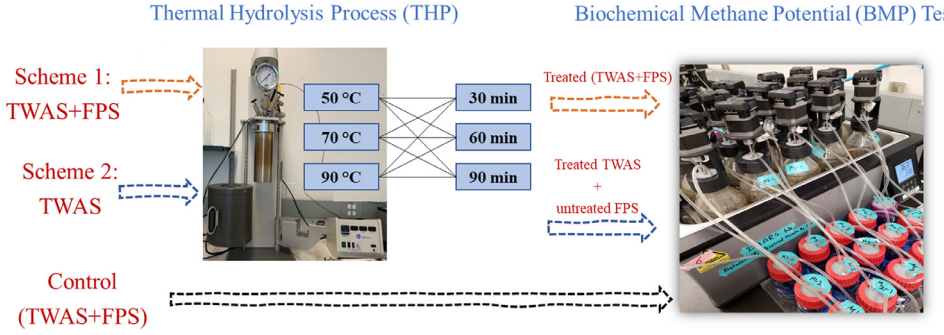

Fig. 1. Summary of the experimental schemes and setup.

Value of the Data

- PCA analysis was performed for a better assessment of the effectiveness, variations, and similarities of low-temperature THP under different temperatures and exposure times.
- The data is considered useful assistance and benefits operators of WWTPs and researchers regarding primary sludge fermentation to enhance sludge anaerobic digestion facility's process.
- With respect to exposure time, temperature, and relative performance metrics (organics solubilization and volatile solids reduction), the PCA models classified the data into various clusters and demonstrated the differences and correlations between various schemes and experimental conditions.
- PCA analysis can help to pave the road for future investigations of the efficiency of THPs conditions concerning the sludge characteristics and solubilization of macromolecular compounds. Overall, the data demonstrated the most significant variables favorable for PCA assessment to categorize various temperatures and exposure times in the low-temperature THP.

1. Data Description

PCA was performed to pinpoint the influence of the two low-temperature THP schemes (scheme-1; TWAS+FPS; scheme −2, TWAS only) on the variations of macromolecular compounds and volatile solids solubilization and the relative effects on anaerobic digestion. Fig. 1 depicts the schematic diagram of the two schemes. All raw THP experimental data associated with this article is available in supplementary materials.

Fig. 2. shows the PCA analysis of the FTIR peaks of the functional groups associated with the major macromolecular compounds as well as other solubilization metrics (e.g., COD, VSS, carbohydrates, and proteins). Based on the PCA analysis, most of the TWAS+FPS samples were clustered together in the bottom-right quadrant. In contrast, the TWAS samples were grouped in the right-top quadrant. Furthermore, the control samples were positioned in the left quadrants. Overall, the levels of carbohydrates, lipids, and proteins were considerably varied and altered after the THP compared to control (untreated samples). The variations in the PC1 direction contributed to 80.9% of the total variations, and this attribute to the broad separation of THP samples from the control in this direction. The negative scores of 50 °C, 30 min conditions in both schemes on PC1 and their relative positions in the same part with control emphasized less variation between those conditions and control.

Excluding the FTIR peaks, another investigation of PCA analysis was conducted with the focus only on the performance metrics of the solubilization (solubilization of COD, VSS, carbohydrates, and proteins) (Fig. 3). Generally, the pre-treated TWAS+FPS samples were highly scattered along

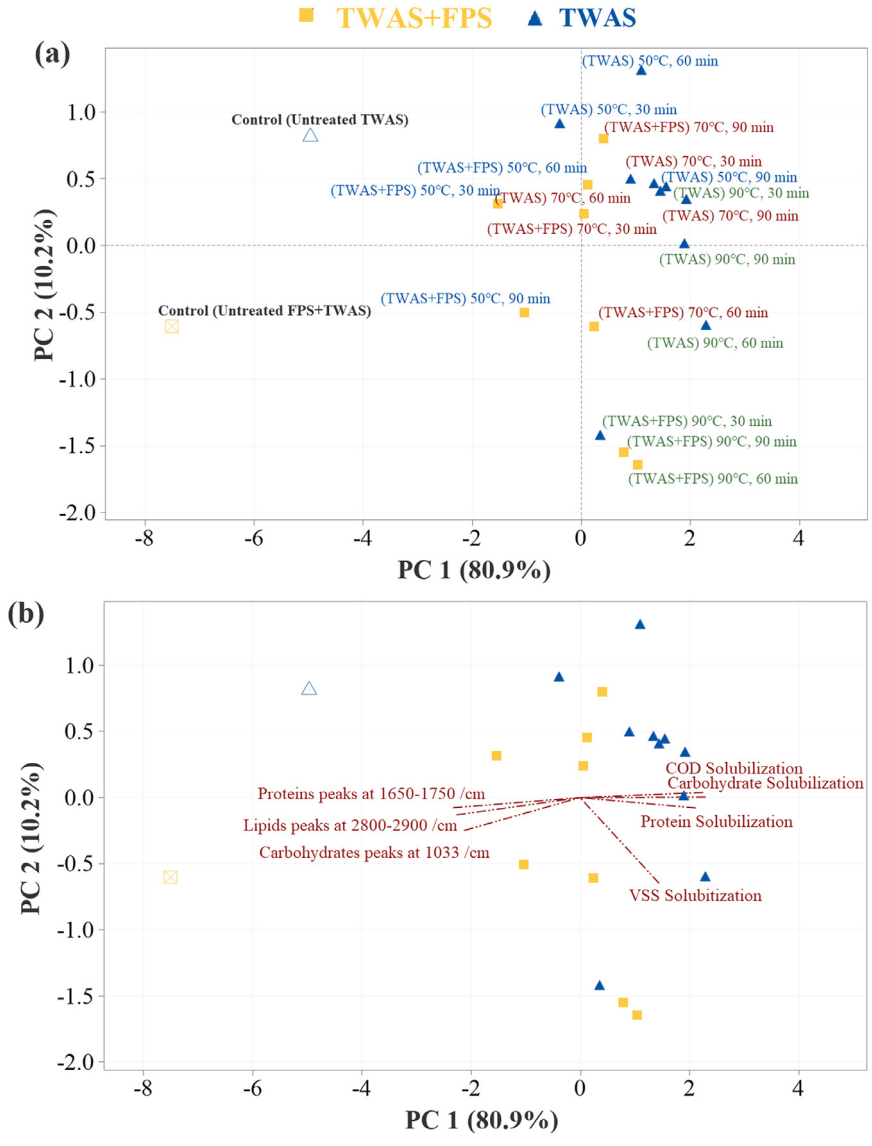

Fig. 2. (a) Score plot for PCA analysis of all pretreated samples, PC 1 (80.9%) and PC 2 (10.2%), (b) Biplot and loading plot of PC1 and PC2 with project lines of all pretreated samples where sample loadings are represented as vectors radiating from the origin. Sample scores are indicated by symbols (according to each scheme), samples that are chemically similar will plot near to each other (clustered together), samples are color-coded by substrate source. Raw data file is available in supplementary materials.

both axes ($-3.3 < |x| < -0.4$ vs $1.1 < |y| < -2.1$). In contrast, the TWAS samples in scheme-2 were mostly clustered in the right-top quadrant with a significant grouping near to origin ($0.1 < |x| < 2.1$ vs $1.8 < |y| < -1.1$) and less scattering along both axes. Generally, at 90 °C at different exposure times 30–90 min, three pre-treated TWAS+FPS samples (90 °C and 30–90 min) were clustered together (left bottom of the plot), showing the high-level similarity of these conditions in terms of VSS solubilization.

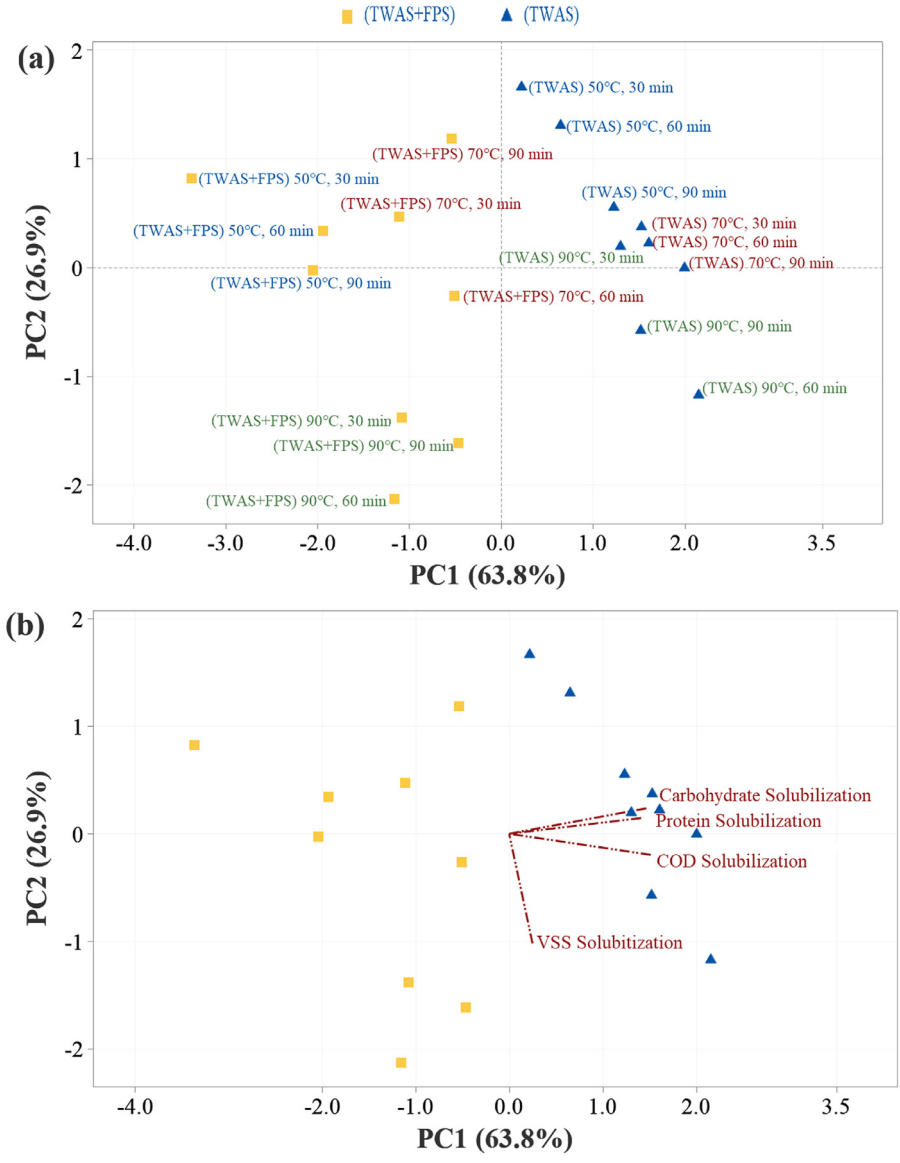

Fig. 3. (a) Score plot for PCA analysis of all pretreated samples, PC 1 (63.8%) and PC 2 (26.9%), (b) Biplot and loading plot of PC1 and PC2 with project lines of all pretreated samples where sample loadings are represented as vectors radiating from the origin. Sample scores are indicated by symbols (according to each scheme), samples that are chemically similar will plot near to each other (clustered together), samples are color-coded by substrate source. Raw data file is available in supplementary materials.

Most of the scheme-1 (TWAS+FPS) samples at 50 °C (30–60 min) and 70 °C (30 and 90 min) were grouped together in the left-top quadrant, while the remainder samples 90 °C (30–90 min) were categorized in the left-bottom quadrant. Generally, the pre-treated TWAS+FPS samples were highly scattered and separated along both axes, which indicates that the solubilization effectiveness was increasingly altered along temperature and exposure time. Nevertheless, the

Table 1
Eigenanalysis of the correlation matrix.

Eigenvalue	2.5505	1.0759	0.2985	0.0751
Proportion	0.638	0.269	0.075	0.019
Cumulative	0.638	0.907	0.981	1.000

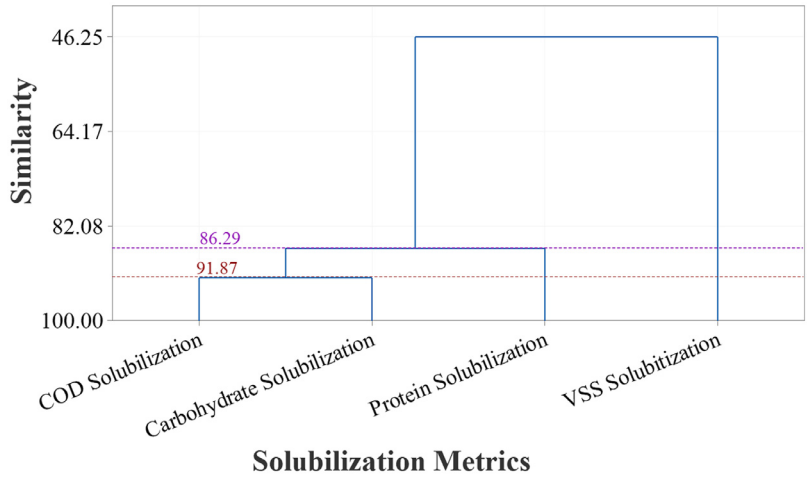

Fig. 4. Dendrogram of the cluster analysis for solubilization metrics. Raw data file is available in supplementary materials.

treated TWAS+FPS samples at 90 °C and 30–90 min were clustered together, revealing a high similarity in solubilization at 90 °C regardless of the time increase. For clustering of scheme-2, the TWAS samples were mostly clustered in the right-top quadrant with a significant grouping near to origin and minimal separation in the direction of the PC1 axis. Except for 70 °C (30–90 min) and 90 °C and 30 min, the TWAS treated samples were marginally separated and deviated in the two quadrants. Generally, the total variance in PCA analysis was manifested by total variation percentages of 90.7% (*i.e.*, 63.8% in the direction of PC 1 and 26.9% in the direction of PC 2) and loading vectors radiating towards the bottom and top-right quadrants (Fig. 3).

Table 1 illustrates the eigenanalysis of the correlation matrix and the respective distribution portion in each axis. The highest significant conditions in terms of solubilization efficacy shifted the loading vectors towards their clusters. For instance: the maximum protein and carbohydrate solubilization and its loading vector were shifted towards 70 and 90 °C and 30–90 min for TWAS. In contrast, nearly all 50 °C, 30–90 min conditions for both TWAS+FPS and TWAS samples attained minimal solubilization.

The PCA validated a positive correlation between the solubilization of COD, proteins, and carbohydrates compared to VSS (Fig. 3). In the same sense, the cluster analysis of factors showed 2 clusters at a similarity level of 86.29 in which: (i) cluster 1 encompassed protein solubilization, carbohydrate solubilization, COD solubilization, and (ii) cluster 2 encompassed VSS solubilization (Fig. 4). Considering a higher level of similarity at 91.87, 3 clusters were revealed: VSS only, protein only, and a combined cluster for the solubilization of COD and carbohydrates together.

Table 2 illustrates the amalgamation steps for the clustering analysis of data based on the solubilization variables and metrics. To establish the variations and the differences in the pretreated samples after THP with regard to several factors, such as time, temperature, VFA increase, and solubilization of VSS, carbohydrates, proteins, COD, and methane yield amongst, other pa-

Table 2
Amalgamation Steps for the clustering analysis of data based on the Solubilization variables and metrics.

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	3	91.8717	0.16257	1	4	1	2
2	2	86.2905	0.27419	1	3	1	3
3	1	46.2500	1.07500	1	2	1	4

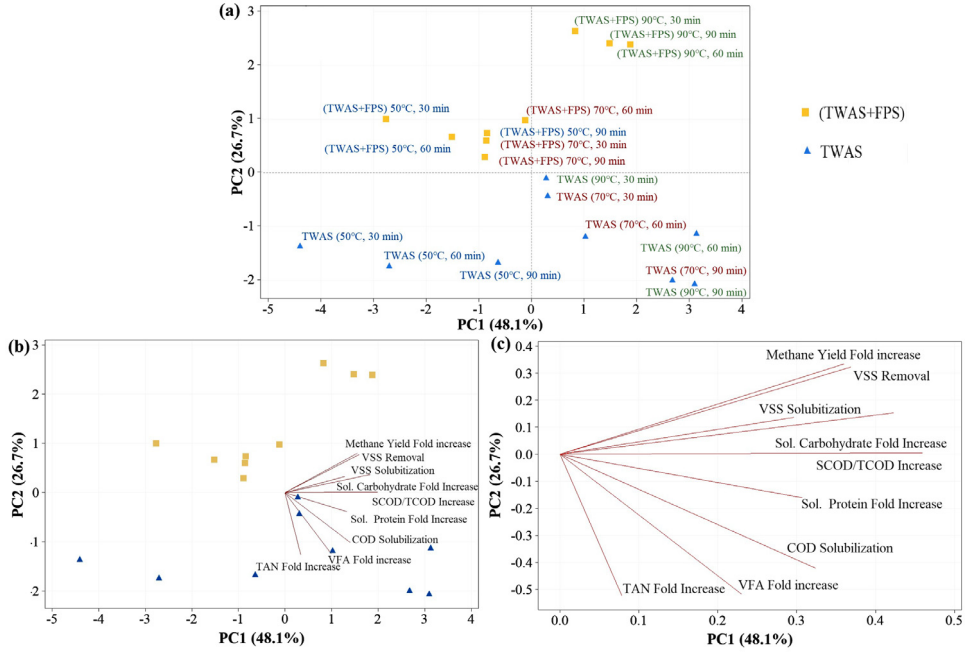

Fig. 5. (a) Score plot for PCA analysis of all pretreated samples, PC 1 (48.1%) and PC 2 (26.7%), (b) Biplot with project lines of all pretreated samples, PC 1 (48.1%) and PC 2 (26.7%), and (c) Loading plot where sample loadings are represented as vectors radiating from the origin. Sample scores are indicated by symbols (according to each scheme), samples that are chemically similar will plot near to each other (clustered together), samples are color-coded by substrate source. Raw data file is available in supplementary materials.

rameters, we re-examined PCA analysis for data as one entire group of samples then dividing them into two subgroups based on the characteristics of sludge (TWAS+FPS and TWAS only).

As one entire group, Fig. 5 depicts the score, biplot, and loading plots established for all pre-treated samples in scheme-1 (TWAS+ FPS) and scheme-2 (TWAS). Considering the loading vectors of the 9 parameters, the total variations along axis 1 (highest variations) then axis 2 (as the second most variations) contributed to 74.8% of the total variance emphasizing distinctive differences in all samples. All pre-treated samples were noticeably separated in PC axis 1 (48.1%) and PC axis 2 (26.7%). The distinctly scattered samples in various parts of the graph can be explained according to the correlation matrix that suggests a third direction accounting for 12.1% of the total variance of the analyzed data (Table 3). Considering the variations of the analyzed data in three components, the total variations can be computed and increased to 86.8%. Table 3 demonstrates the eigenanalysis of the correlation matrix and the respective distribution portion in each axis.

The highest contribution of methane production with scheme-1 (TWAS+FPS) at 90 °C can be clearly distinguished with the clustering of the samples closely in the top-right of the graph.

Table 3
Eigenanalysis of the correlation matrix.

Eigenvalue	4.3245	2.4006	1.0887	0.6436	0.2654	0.1479	0.0891	0.0378	0.0025
Proportion	0.481	0.267	0.121	0.072	0.029	0.016	0.010	0.004	0.000
Cumulative	0.481	0.747	0.868	0.940	0.969	0.986	0.996	1.000	1.000

Similarly, in the bottom-left of the graph and middle- left, the lowest temperature (50 °C) in two schemes were closely clustered in which the TAN increase was predominant. Furthermore, COD solubilization, VFA increase, and soluble proteins have a positive relationship with TWAS samples, particularly at (70 and 90 °C) samples since their position was in the same quadrant (bottom-right). Both SCOD/TCOD and soluble carbohydrates also had positive linkages with the samples in the two schemes, in particular, 70 and 90 °C and pre-treatment time 30–90 min as they were positioned in the middle part of the plot near to origin.

Considering all samples belong to one group in Fig. 5, the respective contribution to the total variance in PCA analysis was manifested by 75%. However, these observations lead us to re-examine the analysis by splitting the pre-treated samples into two subgroups based on substrate origin. These investigations were clearly verified in Figs. 6 and 7 in which the PC1 and PC2 distributions were considerably increased in TWAS+FPS (scheme-1) and TWAS (scheme-2) to (65 and 16%, the total variance of 81%, Fig. 6) and (73.8 and 12.3%, the total variance of 86.1%, Fig. 7) respectively.

As illustrated in Fig. 6 for TWAS+FPS; similar clusters were noticed for 90 °C samples at all exposure times. Additionally, soluble protein and carbohydrates and ammonia production depicted a strong positive relationship with 30 min samples of 70 °C and 60–90 min samples of 50 °C treatment. This attribute to their places in the same quadrant (top-right) in the graph. For the TWAS+ FPS samples, the higher temperature and longer pre-treatment duration positively impacted various parameters such as removal of volatile solids and organics hydrolysis. In contrast, for the TWAS+FPS samples, a minimal impact of low temperature and time (50 °C and 30 min) can be remarked on the parameters (VSS solubilization, methane, VFA, and TAN).

On examining the relationships between exposure times and temperatures in TWAS samples (Fig. 7), factors such as COD solubilization, soluble carbohydrates and proteins, VFAs production, methane yields, were positively correlated with increasing temperatures and exposure times, particularly 70 and 90 °C (60–90 min). The PC1 alone explains (73.8%) of the total variations, which are mainly dominated by the changes in the temperature of treatment and positive scores of methane yield, COD solubilization, soluble carbohydrates and protein, and VSS reduction. However, the major alterations in the orientation of the methane production (*i.e.*, was towards the top in the PC 2 direction in Fig. 5) to the right direction in Fig. 7 led to the increase in the variations of PC1. Overall, the observations of PCA analysis relevant to the relative performance of conditions (based on substrate, time, and temperature) appraise the distinction and potential interactions amongst all parameters in addition to the variability of pre-treatment conditions in terms of exposure time, temperature, and relative performance metrics (solubilization and organics hydrolysis, and solids reduction).

2. Experimental Design and Methodology

Detailed descriptions for all experiments and analyses can be found in the original research paper [1]. All raw datasets associated with this article are available in supplementary materials.

2.1. Sampling

The inoculum and sludge (TWAS and FPS) used in this experiment and data collection were obtained from the Gold Bar WWTP (Edmonton, Alberta, Canada). All samples were stored at 4 °C

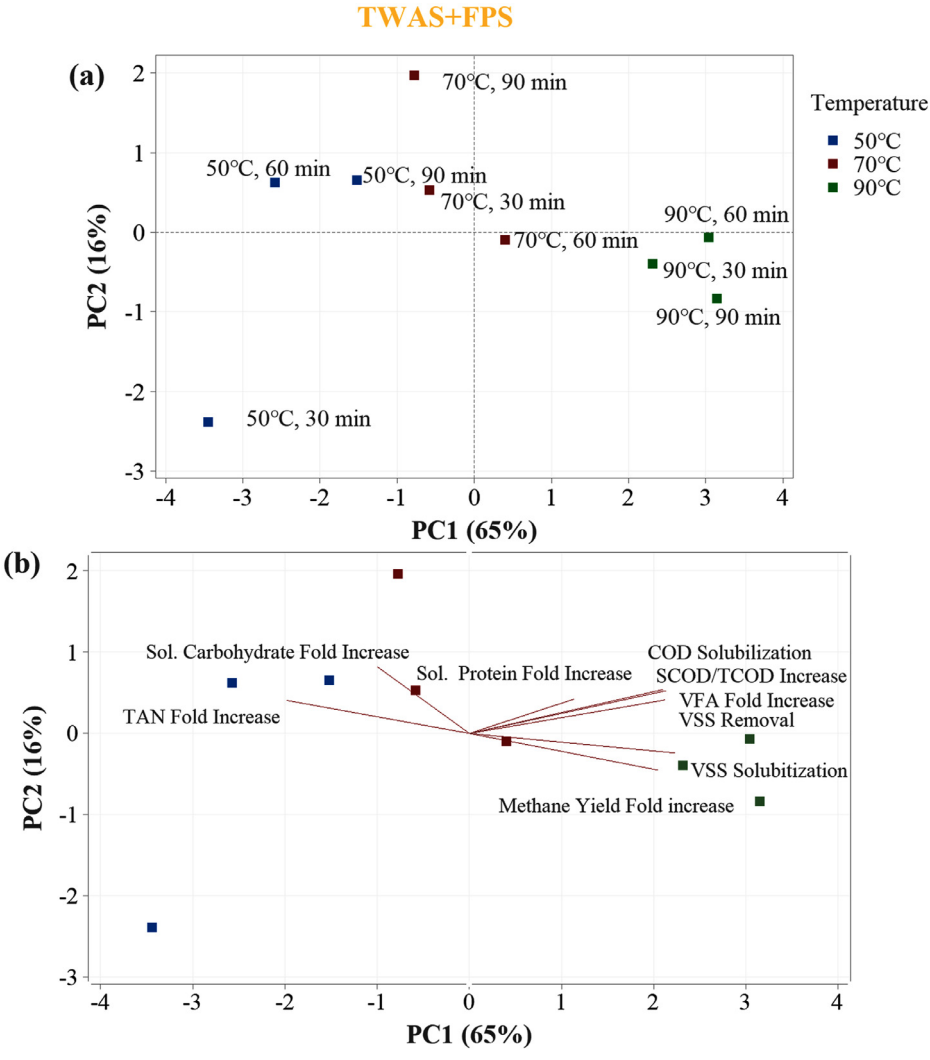

Fig. 6. (a) Score plot for PCA analysis of TWAS+FPS pretreated samples, PC 1 (65%) and PC 2 (16%), and (b) Biplot with project lines where sample loadings are represented as vectors radiating from the origin, PC 1 (65%) and PC 2 (16%). Sample scores are indicated by symbols (according to temperature and time), samples that are chemically similar will plot near to each other (clustered together), samples are color-coded by temperature. Raw data file is available in supplementary materials.

right after collection. A summary of the characteristics of inoculum and sludge can be found elsewhere [1].

2.2. THP and BMP experiments

A 2 L batch hydrothermal reactor (Parr 4848, Max. temperature: 350 °C, Max. pressure: 1900 psi, Parr Instrument Company, Moline, IL, USA) was used for performing the low-temperature THPs of both schemes. The sludge temperature was monitored and maintained to

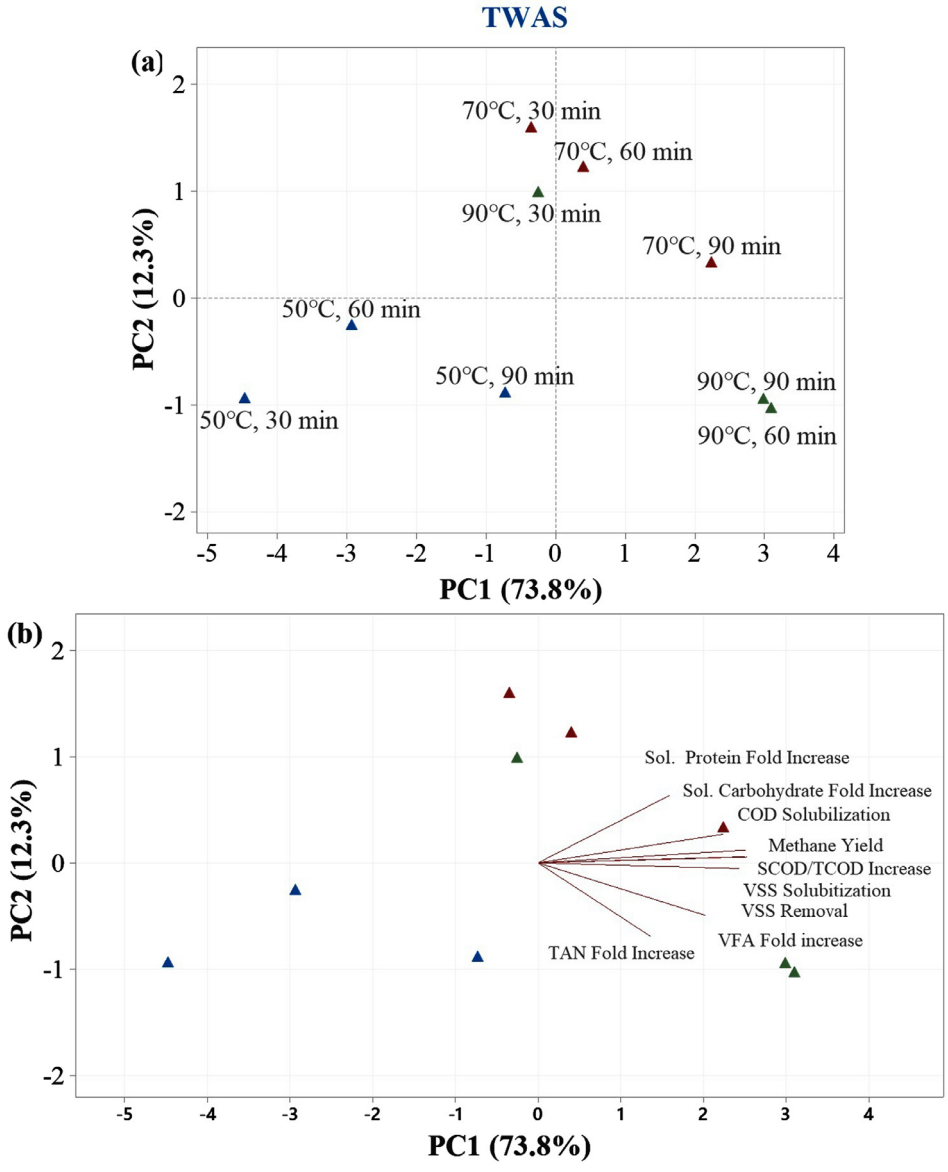

Fig. 7. (a) Score plot for PCA analysis of TWAS pretreated samples, PC 1 (73.8%) and PC 2 (12.3%), and (b) Biplot with project lines where sample loadings are represented as vectors radiating from the origin, PC 1 (73.8%) and PC 2 (12.3%). Sample scores are indicated by symbols (according to temperature and time), samples that are chemically similar will plot near to each other (clustered together), samples are color-coded by temperature. Raw data file is available in supplementary materials.

the desired and designed values (50, 70, and 90 °C) as well as targeted exposure times (30, 60, and 90 min). As mentioned in our parallel report [1], prior to the BMP tests, the THP experiments were conducted in two schemes (Fig. 1). In scheme-1, before the BMP test, THP was applied to the mixture of TWAS and FPS (volume ratio 1:1). In scheme-2, THP was performed on the TWAS only, and then it was mixed with the raw FPS before the BMP tests.

Total 18 conditions of THP (3 exposure times, 30- 90 min at three temperature levels 50, 70, and 90 °C) in addition to control samples were examined. A total of fifty-four (500 mL) anaerobic glass bioreactors (ISES-Canada, Vaughan, ON, Canada); the working volume of 350 mL, and a 150 mL headspace were used in the BMP tests. Triplicate bioreactors of each condition were operated and maintained for 45 days at mesophilic conditions (37 ± 2 °C) using water baths supplied from General Purpose Water Bath, Digital, 20 L, PolyScience, Illinois, USA. Gas bags connected with a CO_2 sequestration unit were used to collect and monitor the produced methane from each bioreactor [2]. The daily measurement of methane volume was performed daily using a frictionless glass syringe. Further details can be found elsewhere [1,3].

2.3. Analytical methods

TAN, TCOD, and SCOD concentrations were analyzed with HACH reagent kits (HACH, Loveland, CO, USA) in which the SCOD and TAN samples were prepared using 0.45 µm membrane filters. For TSS and VSS concentrations, they were determined using standard methods [4]. An ion chromatograph (Dionex^{TM} ICS-2100, Sunnyvale, USA) equipped with an electrochemical detector and microbore AS19, 2 mm column was employed for determining the concentration of major VFAs. Both total and soluble carbohydrates were measured using the modified phenol-sulfuric acid method and glucose as a standard [5]. Similarly, the total and soluble proteins were determined using the TKN kit (HACH, Loveland, CO, USA).

Standard methods were employed to determine concentrations of TSS and VSS [4]. FTIR spectroscopy analysis was conducted with an FTIR Perkin-Elmer 2000 spectrophotometer for all THP samples and control at infrared spectra measured over the range of 4000–400 cm^{-1}. Further details can be found elsewhere [1].

2.4. Principal component analysis

The results of COD, carbohydrates, proteins, lipids, VSS, TAN, and methane yields were examined and evaluated using PCA and clustering analysis in order to highlight the relative efficacy, performances and correlations between the two THP schemes. All quantification results and concentrations of the above-mentioned parameters were employed for the PCA assessment and cluster analysis in Minitab 19. To offer best clustering results and extract the significant variables with the highest original variance and indices that can cause the major differences among samples, the PCA analysis was evaluated based on PCA model consistency and samples with high statistical value (*i.e.*, outlier values were not considered and removed) [6,7]. The aim of PCA evaluation was to provide a statistical comparison [8,9] between the two schemes and to determine the significant differences between the variations in characteristics of pretreated samples at various THP conditions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Mohamed N.A. Meshref: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Investigation, Writing – original draft; **Seyed Mohammad Mirsoleimani Azizi:** Conceptualization, Methodology, Writing – review & editing; **Wafa Dastyar:** Methodology, Writing – review & editing; **Rasha Maal-Bared:** Project administration, Writing – review & editing;

Bipro Ranjan Dhar: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Acknowledgments

The fund of this research work was aided by the Collaborative Research and Development Grant (CRD) from the Natural Sciences and Engineering Research Council of Canada (NSERC) and EPCOR Water Services. The John R. Evans Leaders Fund from the Canada Foundation for Innovation (CFI) supported the acquisition of a hydrothermal reactor used for lab experiments. Special thanks go to Dr. Yangang (Rick) Feng, Mr. Abdul Mohammed, and Ms. Bing Lin from EPCOR for their great support in this research project.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107323.

References

[1] S. Azizi, M. Mirsoleimani, W. Dastyar, M.N.A. Meshref, R. Maal-Bared, B.R. Dhar, Low-temperature thermal hydrolysis for anaerobic digestion facility in wastewater treatment plant with primary sludge fermentation, Chem. Eng. J. (2021) 130485.
[2] S. Barua, B.S. Zakaria, B.R. Dhar, Enhanced methanogenic co-degradation of propionate and butyrate by anaerobic microbiome enriched on conductive carbon fibers, Bioresour. Technol. 266 (2018) 259–266.
[3] P. Zhou, M.N.A. Meshref, B.R. Dhar, Optimization of thermal hydrolysis process for enhancing anaerobic digestion in a wastewater treatment plant with existing primary sludge fermentation, Bioresour. Technol. 321 (2020) 124498.
[4] A.P.H. Association, A.W.W. Association, W.P.C. Federation, W.E. Federation, Standard Methods For the Examination of Water and Wastewater, American Public Health Association., 1915.
[5] M. Dubois, K.A. Gilles, J.K. Hamilton, P.A. t Rebers, F. Smith, Colorimetric method for determination of sugars and related substances, Anal. Chem. 28 (1956) 350–356.
[6] I.T. Jolliffe, Principal Component Analysis, 2nd ed., Springer-Verlag, New York, 2002.
[7] L. Lin, B. Chowdhury, B.S. Zakaria, B.R. Dhar, Temperature-dependent (20–55 C) electrocatalytic characteristics during ethanol/propionate degradation by methanogenic communities grown on conductive carbon fibers, Chem. Eng. J. 391 (2020) 123566.
[8] B.K. Biswal, H. Huang, J. Dai, G.H. Chen, D. Wu, Impact of low-thermal pretreatment on physicochemical properties of saline waste activated sludge, hydrolysis of organics and methane yield in anaerobic digestion, Bioresour. Technol. 297 (2020) 122423.
[9] J. Liu, J. Zheng, J. Zhang, D. Yu, Y. Wei, The performance evaluation and kinetics response of advanced anaerobic digestion for sewage sludge under different SRT during semi-continuous operation, Bioresour. Technol. 308 (2020) 123239.