Biochemical Engineering Journal 53 (2010) 38–43

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

A chemometric tool to monitor high-rate anaerobic granular sludge reactors during load and toxic disturbances

J.C. Costa, M.M. Alves, E.C. Ferreira*

IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

ARTICLE INFO

Article history:
Received 1 October 2008
Received in revised form
27 November 2008
Accepted 12 December 2008

Keywords:
Anaerobic digestion
Control
Organic loading disturbance
Principal Component Analysis
Quantitative image analysis
Toxic shock load

ABSTRACT

Knowing that wide fluctuations in flow rate and presence of toxic compounds can damage the high efficiency of high-rate anaerobic granular sludge reactors, the use of Principal Component Analysis (PCA) to detect organic and toxic disturbances was tested. As earlier these disturbances are detected, more accurate would be the corrective actions, and less damage will be caused to the microorganisms involved in the process. The PCA determined a latent variable, combining a weighted sum of operational, physiological, and morphological data, which showed high sensitivity to recognize the operational problems occurred when four organic loading disturbances (OLDs) and three toxic shock loads (TSLs) were applied to Expanded Granular Sludge Bed (EGSB) reactors. The high loadings/weights linked with the morphological parameters, specially the aggregates size distribution (>0.1, >1), obtained using quantitative image analysis techniques, demonstrate the usefulness of monitor the anaerobic granular sludge structural changes. The application of PCA chemometric tool to dataset gathering information from all disturbances allowed the differentiation between organic loading and toxic shock disturbances, as well as the main effects caused by each class of disturbance.

1. Introduction

The development of high-rate reactors, based in anaerobic granular sludge, was the key feature that allowed for a great increase in the use of anaerobic technology for the treatment of a growing variety of industrial wastewaters [1]. Anaerobic granules are particulate biofilms, formed spontaneously by self-immobilization of anaerobic bacteria in the absence of a support material [1]. Hence, each granule is a functional unit comprising all the different microorganisms necessary for methanogenic degradation of organic matter. Consequently, uncoupling the hydraulic retention time (HRT) from the solids retention time allowed the application of high organic loading rates (OLRs), and therefore the use of compact and economical wastewater treatment plants. However, these systems are designed with reference to a nominal operating condition, in which the OLR is assumed to be constant in time. Also, some compounds can have inhibitory or toxic effects to the microbial populations, such as detergents and solvents. These facts, coupled with the long start-up periods imply the need to monitor the anaerobic granular sludge stability in order to achieve an appropriate control and sustainability of the process.

The recognition of parameters that could be used for monitoring the process is important to efficient control of those processes. It is equally feasible to obtain values of parameters measured in solid, liquid or gaseous phases. However, parameters involved in reactors control had been limited to indicators of the liquid and the gaseous phases [2], due to difficulties in obtain and inaccuracy associated with morphological parameters.

With the rapid development of instrumental methods the amount of diverse data generated in an environmental process monitoring and/or control is increasingly drastically [3–5]. This advance guide analysts and researchers to gathering further more multivariate data. Concurrently, with computer science and technology developments, apply computers and advanced statistical and mathematical methods to analyse this data became easier.

In this framework, image analysis techniques appear as a promising tool to provide quantitative parameters of the solid phase evolution. Chemometrics-based techniques, such as Principal Component Analysis (PCA), can be useful to detect groups, trends, correlations, and outliers in datasets gathering vast amounts of information. This method allows identifying patterns in data, and expressing them in order to highlight their similarities and differences. PCA is a projection method for analyse data and reduce it from an n-dimensional space to few latent/hidden variables, while keeping information on its variability.

Chemometric tools have been proved to be able to monitor wastewater treatment reactors [6–8]. Also, multivariate statistical analysis has been used together with image analysis techniques to pattern recognition, such as discriminant analysis, neural networks, and decision trees [9]. The relationships between morpholog-

* Corresponding author. Tel.: +351 253604407; fax: +351 253678986.
E-mail address: ecferreira@deb.uminho.pt (E.C. Ferreira).

1369-703X/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.bej.2008.12.006

Table 1
Organic loading disturbances (OLDs) and toxic shock loads (TSLs) conditions.

Disturbance	OLD1	OLD2	OLD3	OLD4	TSL1	TSL2	TSL3
Ethanol (gCOD/L)	5	1.5	15	15	1.5	1.5	1.5
HRT (h)	8	2.5	8	8	8	8	8
Toxic	–	–	–	–	Detergent	Detergent	Solvent
[Toxic] (mg/L)	–	–	–	–	1.6	3.1	40
Exposure phase (h)	72	72	72	384	56	222	222
Recovery phase (d)	7	7	7	7	14	12	7

ical parameters and biomass properties in aerobic wastewater treatment processes were also assessed by partial least squares regression [10] and PCA [11].

The objective of this work was to apply the chemometric technique PCA in order to recognize fluctuations and respective effects in high-rate anaerobic granular sludge reactors performance caused by organic loading and toxic disturbances. The role of quantitative morphological parameters in the potential fault detection was also assessed.

2. Materials and methods

2.1. Datasets

Four organic loading disturbances (OLDs) [12] were applied to an Expanded Granular Sludge Bed (EGSB) reactor fed with 1.5 gCOD-ethanol/L and HRT of 8 h, in steady-state conditions. Also, three toxic shock loads (TSL) were applied in EGSB reactors operating in similar conditions [13,14]. Summary of the disturbances applied are presented in Table 1.

Three programmes previously developed [15] were used as the final step of a procedure [13] to obtain quantitative morphological information from anaerobic granular sludge.

Three datasets were created gathering morphological, physiological, and reactors performance information. Datasets 1 and 2 included observations of OLD and TSL, respectively. The objective consisted in examine the sensitivity of the latent variables to recognize the disturbances. Dataset 3 encompassed all observations to study the differentiation of the OLD from the TSL, and respective effects. The variables used during the experiments are defined in Table 2.

2.2. Principal Component Analysis

PCA aims at finding and interpreting hidden complex, and possibly causally determined, relationships between features in a dataset. Correlating features are converted to the so-called factors which are themselves noncorrelated [16].

SIMCA-P (Umetrics AB) software package was used to perform the PCA. The first step of the analysis consists in the pre-treatment of data by standardization of the variables, i.e. guarantee that each individual variable has about the same range, avoiding that some variables would be more important than others because of scale effects. During this work each variable was autoscaled so that each variable has mean zero and unit standard deviation.

Subsequently, the software iteratively computes one Principal Component (PC) at a time, comprising a score vector t_a and a loading vector p_a. The score vectors contain information on how the samples relate to each other. Otherwise, the loading vectors define the reduced dimension space and contain information on how the variables relate to each other. Usually, few PCs (2 or 3) can express most of the variability in the dataset when there is a high degree of correlation among data.

The criterion used to determine the model dimensionality (number of significant components) was cross-validation (CV). Part of data is kept out of the model development, and then are predicted by the model and compared with the actual values. The prediction error sum of squares (PRESSs) is the squared differences between observed and predicted values for the data kept out of the model fitting. This procedure is repeated several times until data element has been kept out once and only once. Therefore, the final PRESS has contributions from all data. For every dimension, SIMCA computes the overall PRESS/SS, where SS is the residual sum of squares of the previous dimension. A component is considered significant if PRESS/SS is statistically smaller than 1.0.

3. Results and discussion

3.1. Summary of the disturbances effects

Since the main objective of this study was the rapid detection of potential problems, emphasis should be given to

Table 2
Loadings/weights associated with the first (t[1]) and second (t[2]) latent variable in organic loading disturbances (OLD) and toxic shock loads (TSL).

Variable	OLD		TSL		Name
	t[1]	t[2]	t[1]	t[2]	
OLR	−**0.40**	0.06	0.12	−0.20	Organic loading rate
Csd	–	–	−**0.34**	−0.07	Toxic (detergent or solvent) concentration
Eff	**0.40**	0.02	−0.20	**0.30**	COD removal efficiency
pH	**0.32**	−0.12	0.00	**0.38**	pH
VSS	−**0.40**	−0.12	−0.05	−0.28	Effluent volatile suspended solids
<0.1	−0.12	−0.29	−0.20	**0.30**	% of Aggregates projected area with $D_{eq} < 0.1$ mm
>0.1	−**0.35**	0.19	−**0.43**	−0.09	% of Aggregates projected area with $0.1 \leq D_{eq}$ (mm) < 1
>1	**0.36**	−0.16	**0.44**	0.07	% of Aggregates projected area with $D_{eq} \geq 1$ mm
SAA	**0.33**	0.13	−0.26	**0.30**	Specific acetoclastic activity
SHMA	0.04	−**0.41**	0.18	−0.04	Specific hydrogenotrophic methanogenic activity
LfA	0.13	**0.55**	0.24	**0.38**	Total filaments length/total aggregates projected area
VSS/TA	0.16	−0.18	0.17	−**0.33**	VSS/total aggregates projected area
TL/VSS	0.05	**0.55**	0.13	**0.44**	Total filaments length/volatile suspended solids
vsed	–	–	**0.46**	0.01	Settling velocity

Bold values are marked the loadings with higher influence in each score (higher than 0.30).