Chemical Engineering Journal 331 (2018) 114–123

Contents lists available at ScienceDirect

Chemical Engineering Journal

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

On-line estimation of the aerobic phase length for partial nitrification processes in SBR based on features extraction and SVM classification

Francisco Jaramillo[a,*], Marcos Orchard[a], Carlos Muñoz[b], Christian Antileo[c], Doris Sáez[a], Pablo Espinoza[a]

[a] Department of Electrical Engineering, University of Chile, Av. Tupper 2007, Santiago, Chile
[b] Department of Electrical Engineering, University of La Frontera, Cas. 54-D, Temuco, Chile
[c] Department of Chemical Engineering, University of La Frontera, Cas. 54-D, Temuco, Chile

ARTICLE INFO

Keywords:
Feature extraction
Partial nitrification
SBR
SVM

ABSTRACT

We present a strategy for the on-line estimation of the aerobic reaction phase length for a partial nitrification process with pH and dissolved oxygen closed-loop control. To overcome existing drawbacks associated to partial nitrification (e.g., non-linearities and time-variant behaviors), our strategy is based on feature extraction over manipulated variables to identify interesting patterns associated to the end-point of nitrification. We use a support vector machine (SVM) classifier as a decision tool to determine the end-point of the aerobic phase. A database of lab-scale sequencing batch reactor (SBR) cycles selected from ten months of operation was used to train and test the proposed decision-making strategy. Results for all 533 SBR cycles showed 100% correct classifications. Most aerobic phase lengths in the analyzed database had a reduction time around 20 min, although time reductions greater than 60 min were also achieved.

1. Introduction

The biological nitrogen removal (BNR) process is an economically feasible alternative for the treatment of wastewater with high concentrations of nitrogen. BNR is divided into two sub-processes: nitrification and denitrification. Its implementation offers high conversion efficiency, reduced consumption of chemical products, and a low biomass generation [1].

The nitrification process, under aerobic conditions, transforms ammonium (NH_4^+) into nitrite (NO_2^-) by ammonia oxidizing bacteria (AOB). Subsequently nitrite is transformed into nitrate (NO_3^-) by means of nitrite oxidizing bacteria (NOB) [2]. In the denitrification process, under anoxic conditions and in presence of organic matter, heterotrophic bacteria (HB) convert the nitrate into nitrite, then nitric oxide (NO), nitrous oxide (N_2O), and finally into molecular nitrogen (N_2, released into the atmosphere) [3].

Nitrite is formed and consumed during the nitrification phase. Then it is formed again during denitrification, making nitrite oxidation redundant [4]. Partial nitrification then arises as an attractive alternative to optimize the process. To carry it out it is necessary to enhance the activity of the AOB and to selectively inhibit the activity of the NOB [5]. This can be achieved through the management of the pH, temperature, and dissolved oxygen (DO) [6]. After nitrite is accumulated, it is used to start denitrification [7] (Fig. 1). The partial nitrification approach may achieve savings in aeration costs in nitrification and savings in organic matter in denitrification [8]. Furthermore, surplus sludge production decreases.

Within the conventional systems used for BNR (and partial nitrification), the sequencing batch reactor (SBR) is highlighted due to its flexibility and low cost [9]. The SBR also presents the advantage of carrying out BNR in only one reactor, through the sequential development of aerobic (nitrification) and anoxic (denitrification) phases.

A SBR cycle is characterized by a series of phases: fill, react, settle and draw, each with a defined duration [10]. For BNR the reaction phase is critical, and it is divided into two phases: aerobic and anoxic. In the aerobic phase there is a permanent aeration and stir of the sludge inside the reactor that permits to carry out the nitrification. In the anoxic phase the aeration is finished, but without interrupting the stirring of the sludge (allowing to carry out the denitrification). Then comes the settle phase, when aeration and stir are completed so as the sludge settles at the bottom of the reactor.

A remarkable feature of SBR processes is its batch mode operation, which enables control of the reaction time (aeration/mixing) to be adjusted in response to the wastewater quality. This improves the

* Corresponding author.
E-mail addresses: francisco.jaramillo@ing.uchile.cl (F. Jaramillo), morchard@u.uchile.cl (M. Orchard), carlos.munoz@ufrontera.cl (C. Muñoz), christian.antileo@ufrontera.cl (C. Antileo), dsaez@ing.uchile.cl (D. Sáez), pablo.espinoza@ing.uchile.cl (P. Espinoza).

http://dx.doi.org/10.1016/j.cej.2017.07.185
Received 26 March 2017; Received in revised form 6 July 2017; Accepted 31 July 2017
Available online 02 August 2017
1385-8947/

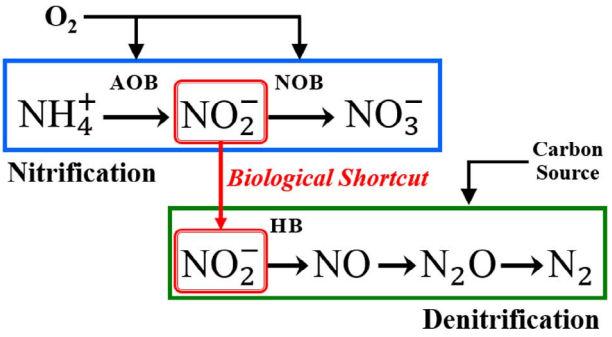

Fig. 1. Schematic view of partial nitrification process. Here a *biological shortcut* is produced, meaning that the nitrite formed in nitrification is then used in denitrification.

operational efficiency and energy savings of WWTPs, but it also brings more complex operation modes that require highly reliable automation methods [11].

For a normal design of the SBR, each phase has a prescribed duration regardless of the process dynamics and the nitrogen concentration in the wastewater influent. This may result in a highly inefficient operation in terms of energy consumption costs [12]. Therefore, there is a need of having advanced monitoring and control tools in order to adapt the duration of the aerobic/anoxic phase to the actual process requirements [9]. The accuracy required to estimate the duration of each phase is critical: underestimating it could lead to incomplete nitrogen removal, and thus effluent quality not meeting local discharge standards. In contrast, overestimating the phase length would decrease the WWTP capacity (volume of wastewater treated per day), increase operational costs (e.g., aeration associated [10]), and lead to NOB growth (which will limit a stable partial nitrification [13]).

An example of important monitoring tools are on-line nitrogen (ammonium and nitrate) sensors. But, their high price, maintenance costs, significant time delay, and complex operation constitute major disadvantages [14]. An alternative is the use of physico-chemical measurements (secondary variables), e.g., pH, DO, oxidation reduction potential (ORP), and oxygen uptake rate (OUR) as indirect indicators for the on-line monitoring and estimation of the nitrification and denitrification phases length [15]. These are on-line quantities which are readily available in industry. The use of secondary variables leads to a strategy called *bending points detection* [14] to estimate the aerobic/anoxic phase duration. The strategy consists in detecting relative changes, or bending points, in on-line profiles of pH, DO, ORP and OUR (which act as free variables as there is no closed-loop control on them). Once detected, these bending points define the end of nitrification/denitrification. In the short-term bending points detection leads to a reduction in aeration costs and reaction time [16]. In the long-term, an increase of AOB and a decrease of NOB populations [5] contribute to maintain a stable partial nitrification [2].

The bending points conventionally associated with the end of nitrification in the aerobic phase are: ammonia valley for pH, ORP plateau, and DO elbow [17,18]. For denitrification we have the nitrate knee for ORP and nitrate apex for pH [19]. Even though bending point detection is convenient for research and industry, it may be affected by operating conditions or measurement noise [18].

Bending points detection acting over no closed-loop control variables makes it difficult to complement with partial nitrification strategies, where pH and DO closed-loop control is often used. Recently, strategies that promote partial nitrification and can be complemented with bending points detection have appeared [20]. Their novelty reside on the on-line monitoring of the pH and DO manipulated variables in order to find patterns related to the end of the aerobic phase (nitrification).

The on-line monitoring of manipulated variables in SBRs is a relatively unexplored area. In [20], the DO in the reactor is controlled by a centrifugal blower equipped with a frequency converter. An strategy involving the application of moving slope change (MSC) to the blower frequency measurements and defining an MSC value equal to −1 as end-point criterion was designed and evaluated. In [13] the end of the aerobic phase is determined using information from the manipulated variable of pH (carbonate consumption) and DO (percentage of air valve opening). In this case the end of the aerobic phase was defined when the carbonate consumption keeps constant for 30 min, and the percentage of air valve opening reaches a value of 25%. In summary, these methods rely on finding characteristic behaviors in manipulated variables, and obtain satisfactory results. Nevertheless, it is worth noting that their detection capability of bending points may be affected on the medium to the long term, since the BNR process in SBRs is both time-variant and subject to multiple uncertainties. These change the controller tuning parameters, degrading the performance of the controller and changing the manipulated variables trends.

Our aim is to provide a new strategy to estimate the length of the aerobic phase in processes where partial nitrification is promoted through the use of closed-loop control of pH and DO in SBRs. To accomplish this, a strategy based on feature extraction and support vector machines (SVMs) is proposed to estimate the end-point of the aerobic phases while simultaneously overcoming the existing drawbacks related to bending point detection. To test/validate and further develop the proposed strategy, experimental data extracted from a lab-scale SBR is used.

The paper is organized as follows: In Section 2, the lab-scale reactor and experimental database are described. To develop the proposed methodology, feature extraction and SVMs are also introduced. In Section 3, performance results of the methodology are presented. In Section 4, a discussion of the obtained results is provided. Finally, in Section 5, we present the main conclusions extracted from this work.

2. Materials and methods

2.1. Lab-scale reactor and instrumentation

The lab-scale SBR used in this work has a volume of 2.4 litres. At the beginning of each cycle, a peristaltic pump (Masterflex easy-load 75-18-00) fills the wastewater. The DO, ORP, and pH/temperature were measured by three electrodes: WTW (Oxi 701, Germany), EUTECH (alpha-pH 2000, Singapore) and HACH (EC 310, USA), respectively. The temperature was regulated by means of an insulated water jacket and a thermostat (Julabo, Model EC, Germany). The reactor was homogenized by a mechanic mixer (HEIDOLPH, RSR 2050, Germany) at 360 rpm. The pH was regulated with a proportional-integral (PI) controller through the addition of pulses of sodium carbonate (0.125 mL of 0.8 M concentrated Na_2CO_3) by a diaphragmatic pump (LANG, type ELADOS EMP II, 41 L/h, Germany). Aeration was supplied by an aquarium aerator (COSMOS double type 1000, China) through two diffusers (Double COSMOS 1000, China) that distributed the inserted air inside the reactor. The DO was regulated by a PI controller using pulse width modulation (PWM) for pneumatic valve opening (Festo, 457, MSG-24DC, Germany), simulating an air valve opening (AVO) between 0 and 100%. At the end of each cycle a second peristaltic pump (Masterflex easy-load 75-18-00) drew the treated wastewater outside the reactor. The program KepServer acquired the on-line data from the described sensors and the manipulated variables: number of sodium carbonate pulses and %AVO. Furthermore, the system was automated by the use of a programmable logic controller (PLC) (Siemens, Simatic S7-200, CPU 214), operated with a PC/PLC interface programmed in MATLAB®. Communication between MATLAB® (client) and the KepServer (server) was effected by using DDE (dynamic data exchange protocol, Microsoft). The sampling time used for recording to the hard drive was 1 min for all the data. The plant scheme and its instrumentation are presented in Fig. 2.

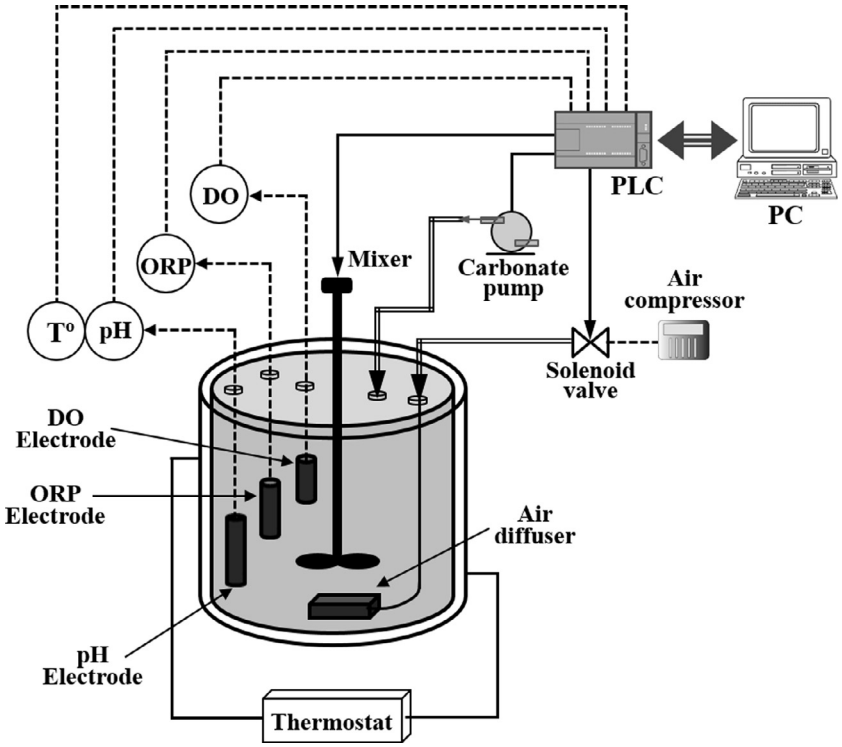

Fig. 2. Schematic diagram of the instrumentation (sensors, actuators, and control system) of the SBR.

2.1.1. SBR operation and issues with the aerobic phase length estimation

A complete process of nitrification and denitrification via nitrite (partial nitrification) was carried out through five phases: fill, aerobic reaction, anoxic reaction, settle, and draw. During the aerobic phase, partial nitrification was achieved by means of pH and DO control, where a set-point of 7.6 (and a few cases where the pH set-point was 8.5, depending of experimental criteria) for pH and 2.0 mg O_2/l for DO was established [13,8]. Every 15 min the aeration was interrupted for 33 s, in order to calculate the OUR (not considered for this work). The length of the aeration phase was determined using the strategy implemented by [13], where the bending points were obtained from the actuator profile of the pH (carbonate consumption) and DO (%AVO) control system. The denitrification process also follows [13].

Recorded data including pH, DO, and the manipulated variables of the pH/DO control from the aerobic phase are shown in Fig. 3. The Figure shows a case where the aerobic phase finished after carbonate pulses kept constant for 30 min (to ensure the end of nitrification) and %AVO reached a 25% [13].

However, there were SBR cycles where a sharp drop (below 20%) in the %AVO was observed, but disturbances in the controller performance due to sensor uncertainties or the time-variant and non-linear nature of the nitrification process did not let the carbonate consumption be kept constant. This means the strategy developed in [13] loses accuracy for the aerobic phase length estimation. Fig. 4 shows an example of an inaccurately aerobic phase length estimation (approximately, a delay of one hour from the end of the aerobic phase).

2.1.2. Experimental database

Our database consists of 533 completed nitrification and denitrification processes at a lab-scale SBR. Only the nitrification phase is considered to develop the methodology. The database corresponds to ten months of reactor operation; additional parameters for aerobic phases are shown in Table 1.

2.2. Feature extraction

Feature extraction involves a pre-processing of sensors measurements, to obtain parameters or indicators that reveal if an interesting pattern emerges. The extracted features represent the original data in a compressed form, under the assumption that the relevant structure in the original data lies in a lower dimensional space [21]. Feature extraction begins from segmentation of the registered signals (or on-line measurements) on separating time windows [22]. Then, every selected feature is computed for each segment. In practice, depending on the characteristics of the process or the registered signals, features are extracted from different domains (time, frequency and time–frequency), as detailed below:

- Time domain methods: They are based in extracting features directly from samples of the measured signal. Examples include statistical moments (minimum, maximum, expected value, variance, skewness, kurtosis, etc.) [23,24], and dynamics features (overshoot, rise time, settling time) [25].
- Frequency domain methods: They are based in extracting features from the signal in the frequency-domain. The fast Fourier Transform (FFT) is an efficient method for converting time domain signals into frequency domain signals. Examples include Fourier coefficients, frequency bands, energy of a frequency band, locations of the frequency peaks, maximum power spectral density (PSD), and mean PSD [26,27].
- Time-Frequency domain methods: They are based in extracting features from the temporal evolution of the signal in the frequency domain. Signal spectrograms and Wavelet transform [28,29] are commonly used methods.

2.2.1. Feature extraction over the manipulated variables

In this experiment, both pH and DO (manipulated variables) exhibit interesting patterns. For pH, the carbonate consumption always decreases along the aerobic phase. DO always shows oscillations after the aeration interruption (33 s, every 15 min). Taking this into account, two features were chosen for this work:

i) Pulses Rate (P_R): The quantity of carbonate pulses injected to the reactor within a defined time window. The expression for Pulse Rate

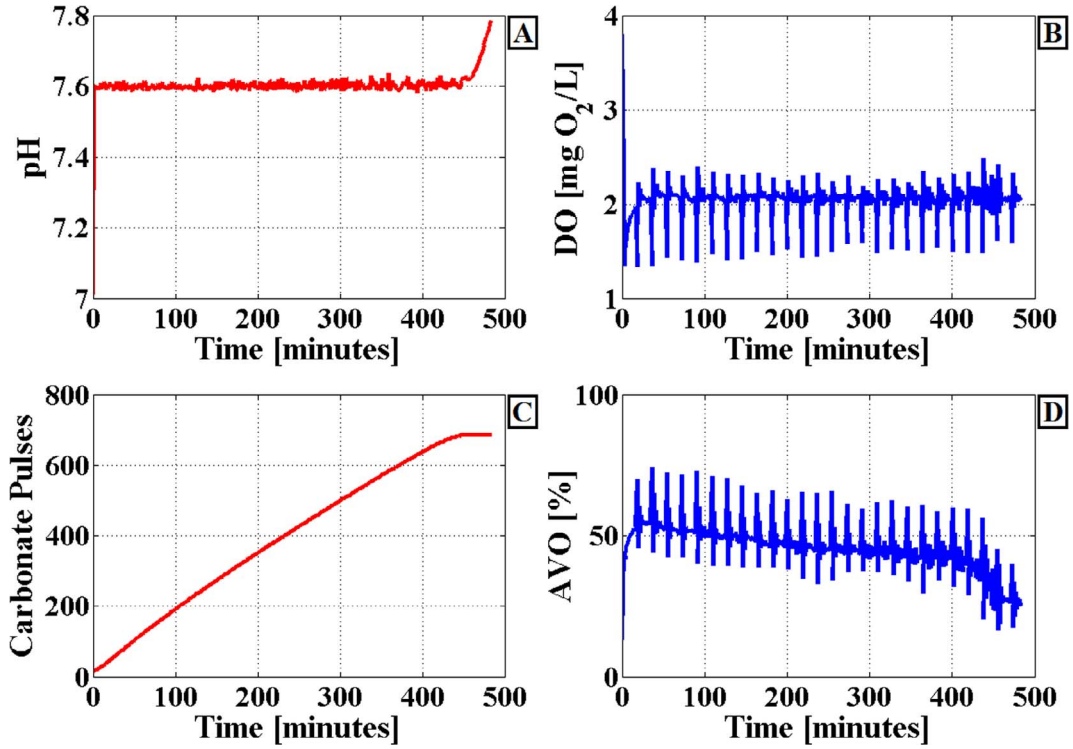

Fig. 3. Data sensors pH (A), DO (B) and their manipulated variables carbonate pulses (C) y %AVO (D). The aerobic phase is finished when carbonate pulses kept constant for 30 min and %AVO reached 25%.

is shown in Eq. 1.

$$P_R = \sum \text{Carbonate Pulses}_{[\text{Time Window}]} \quad (1)$$

ii) Power of the signal %AVO ($P_{\%AVO}$): The average of the empirical %AVO PSD, computed over a time window of n samples. $P_{\%AVO}$ is computed as in Eq. 2, where $X_{\%AVO}(\omega)$ is the %AVO signal in the frequency domain, and n also corresponds to the total number of its discrete frequency components.

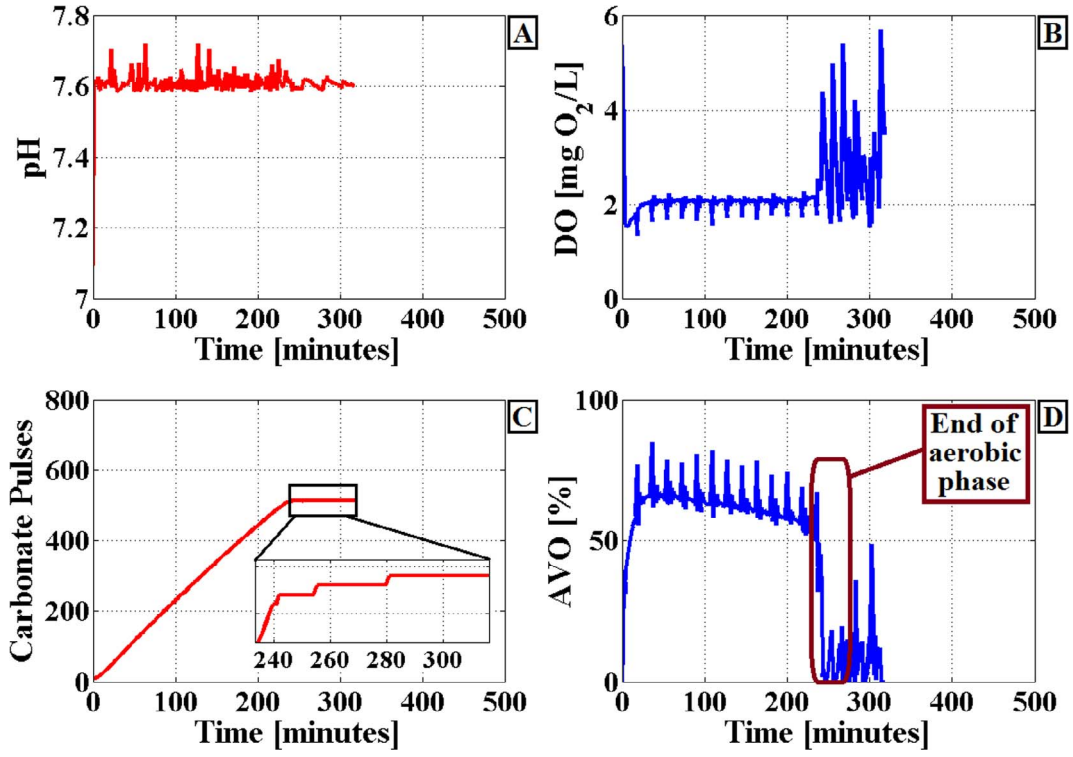

Fig. 4. Cycle where disturbances in the controller performance did not let the carbonate pulses (C) be kept constant, despite having a sharp drop below 20% in the %AVO (D), resulting in an inaccurately aerobic phase length estimation.

Table 1
Database additional parameters for aerobic phases.

Variable	Operational conditions
Initial TAN (mg N/L)	479 ± 63
C/N (g COD/g L)	0
pH	7.6 (294 cycles) 8.5 (239 cycles)
DO (mg O_2/L)	2.0

$$P_{\%AVO} = \frac{1}{n}\sum |X_{\%AVO}(\omega)|^2 \quad (2)$$

The "length" (n) of the time window was chosen to maximise the correlation among the extracted features and the ammonium degradation. According to this, we have selected a time window length of 11 min (sampling time: 1 min), which is always initiated at the moment when the aeration is interrupted (as stated above, these interruptions occur every 15 min). An example of the manipulated variables, the defined time window, and the extracted features are shown in Fig. 5.

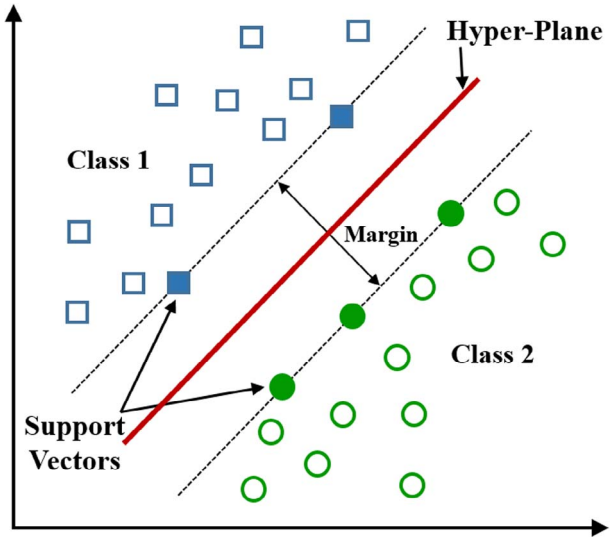

Fig. 6. Binary classification through the hyper-plane calculated by SVM. Adapted from [36].

2.3. Support vector machines

Classification is one of the main application domains of machine learning algorithms. The aim of classification is to build a model (or a decision rule), based on training data, to predict (or assign a class label to) independent data samples for which the class label is unknown [30]. When there are only two classes, the problem is known as binary classification, otherwise it becomes a multi-class classification.

Among machine-learning-based classifiers, SVMs excel for its robustness, efficiency, and generalization performance [31]. Additionally, SVMs have been used for applications related to regression, novelty detection, data mining, computer vision, and bioinformatics [32–35]. It was originally designed for two-class classification, although several approaches extend SVMs to multi-class problems. Its mathematical background is explained in [32]. The aim of SVMs is to minimize classification error, by finding the optimal hyper-plane that maximizes the margin to the closest data points [36]. These closest data points used to define the margin to the separating hyper-plane are called support vectors. An example for binary classification when the training data is linearly separable (mutually exclusive) is shown in Fig. 6.

SVMs can also be used in non-linear classification when the training data are not linearly separable. In that case, and through the use of kernel functions (e.g., linear, polynomial, Gaussian, sigmoidal), the training data is mapped onto a higher dimensional space where a linear

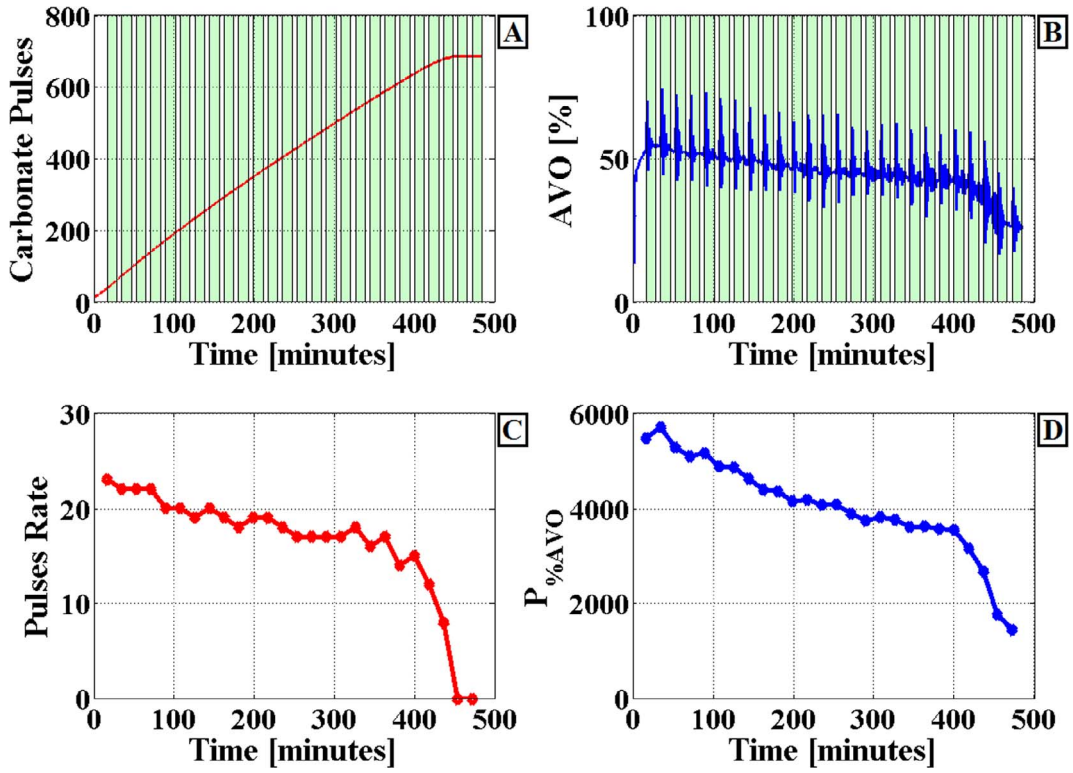

Fig. 5. The pulses rate feature (C) extracted from carbonate pulses (A), and $P_{\%AVO}$ feature (D) extracted from %AVO (B). In green, over the manipulated variables plots (A-B), the time windows used to compute the features are observed.

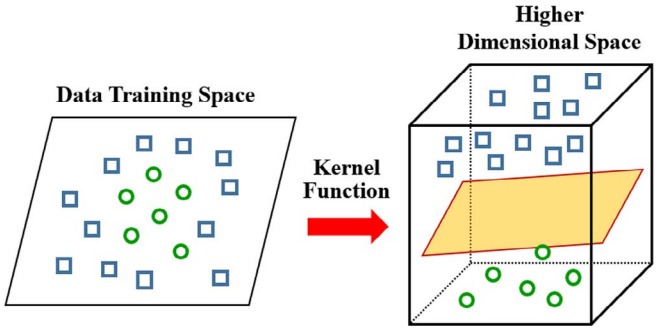

Fig. 7. Non-linear classification using kernel functions that allows to separate the classes in a higher dimensional space.

classification is possible. This is equivalent to non-linear classification in the original space [37]. An example of the separating hyper-plane in a higher dimension for binary classification is shown in Fig. 7.

In practice, the SVM training involves the selection of the kernel function and the tuning of two hyper-parameters: BoxConstraint and KernelScale [38]. On the one hand, the BoxConstraint hyper-parameter controls the trade-off between the margin width and training errors [34]. On the other hand, the KernelScale hyper-parameter (which depends on the selected kernel function) controls the flexibility of the SVM classifier in fitting the data [33].

2.3.1. SVM classifier development: training and testing

We re-defined the estimation of the aerobic phase length (or bending point detection) as a binary classification problem, where one class is the degradation ammonium state, and the other is the complete ammonium degradation state. Hence in this case, the strategy consisted in training and testing an SVM classifier to find the better decision boundary between the two classes.

The following steps are taken to develop the SVM classifier (summarized in Fig. 8). Each step is detailed as follows:

i) *Database Pre-processing*: All the faulty SBR cycles were eliminated, the features were extracted and the information of each class was assigned ("ammonium degradation state" and "complete ammonium degradation state"). Fig. 9 shows an example of the extracted features in the pulses rate - power of the signal %AVO plane from four different aerobic phases (cycles 182, 285, 293, 354). The blue dots represent the features of the ammonium degradation class, while the red dots represent the features of the complete ammonium degradation class. Two elements are important about the features projected in Fig. 9. The first is the trends followed by the features of each phase: every time a new aerobic phase started, its features appear far from the origin, but at the end the features appear close to it. The second relates to the values of the features when the complete ammonium state is reached. The pulses rate feature values were 0 (the minimum value that can be obtained by this feature); the $P_{\%AVO}$ values were different, showing a range between 0 to 3500 approximately. These two elements show important characteristics, useful in the selection and design of the classifier.

ii) *Training Data*: It was defined as the 80% of the database, and it were used to train the SVM classifier to find the decision boundary. In Fig. 10 all the training set is plotted in the features plane, where the two classes are represented in well defined, but not linearly separable areas.

iii) *Testing Data*: The remaining 20% of the database, used to test the previously trained SVM classifier and evaluate its classification accuracy [35].

iv) *SVM Training*: An SVM with Gaussian kernel was chosen, since it has a better classification accuracy than other kernel functions [39]. The Matlab function fitcsvm was used to train this SVM classifier according to the procedure shown in [38], where the *BoxConstraint* and *KernelScale* hyper-parameters are tuned to maximize the classification based on the classification error retained from cross-validation [39].

V) *SVM Testing*: In this step the classifier performance was determined using the testing data.

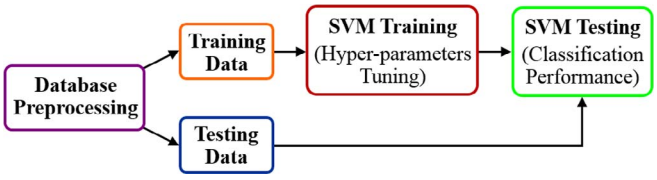

Fig. 8. Steps for training and testing the SVM classifier.

2.3.2. Performance evaluation

The performance evaluation for the SWM classifier was based on the concepts of Confusion Matrix, (%) Accuracy, and (%) Error. In Table 2 a confusion matrix for a two-class classification problem is illustrated. The matrix terms were adapted to the aerobic phase length estimation problem as follow:

- TP (True Positive): Ammonium degradation correctly classified.
- TN (True Negative): End of phase correctly classified.
- FP (False Positive): Ammonium degradation incorrectly classified.
- FN (False Negative): End of phase incorrectly classified.

The classifier (%) Accuracy and (%) Error are defined in Eq. 3 and 4, respectively.

$$(\%)\ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$(\%)\ \text{Error} = \frac{FP + FN}{TP + TN + FP + FN} \tag{4}$$

Due to the fact of having imbalance data (both classes are not equally represented in the database), with an imbalance ratio of approximately 9:1 of ammonium degradation class over complete degradation class, two useful performance measures are also included: Sensitivity (Eq. 5) and Specificity (Eq. 6)

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

In this way, the classifier performance becomes a compromise between these two additional performance measures, since an increase in sensitivity usually involves a decrease in specificity and vice versa [40].

In this work (and in the field of biological nitrogen removal), the importance of obtaining FN or FP is not the same, because an FN implies to keep the aerobic phase although it has already finished. Otherwise, an FP implies to finish the aerobic phase although the ammonium in the reactor has not been completely removed. Therefore, specificity is more important and is strictly necessary that its value be equal to 1 (FP = 0).

2.4. Real-time aerobic phase length estimation

The real-time aerobic phase length estimation flow is illustrated in Fig. 11. Once the aerobic phase has been initialized, on-line measurements of the manipulated variables must be acquired in order to generate feature extraction over these signals. At the next step, the previously trained classifier will determine if the extracted feature corresponds either to the ammonium degradation class or the complete degradation class. For the first case the system will return to the on-line

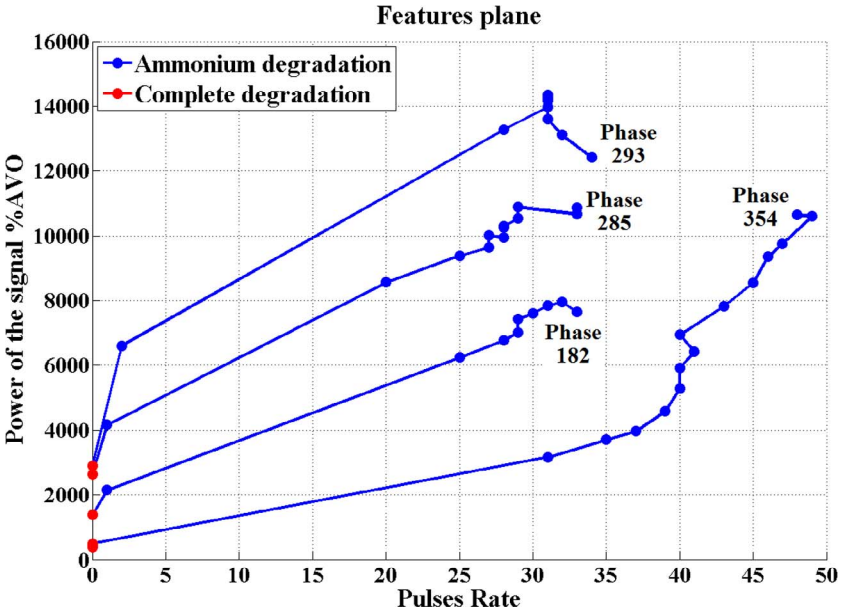

Fig. 9. Example of the features trends in the Pulses Rate – $P_{\%AVO}$ plane. The blue dots represent the features of the ammonium degradation class, and the red dots represent the features of the complete ammonium degradation class. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measurements stage, while for the second the system will end the aerobic phase.

3. Results

3.1. SVM training and testing results

The BoxConstraint and KernelScale hyper-parameters of the SVM classifier with Gaussian kernel were tuned in the training phase, and their values are shown in Table 3. Using these hyper-parameters, an optimum decision boundary of the SVM classifier was found to discriminate between the two classes of the aerobic phase length estimation problem. In Fig. 12, the decision boundary (in green) and the two classes can be visualized.

The first step to evaluate the classification performance is through the application of the classifier over the training database, the confusion matrix, and the previously defined performance measures. The results of these performance hyper-parameters are shown in Table 4.

To prevent the overfitting of the classifier (poor capacity of classifying different data with respect to the training database), the trained SVM classifier was tested over the testing database, the results of its performance are given in Table 5.

Table 2
Two-class classification confusion matrix.

	Classified Ammonium degradation	Classified End of phase
Actual Ammonium degradation	TP	FP
Actual End of phase	FN	TN

3.2. Improvements regarding to the accuracy of the aerobic phase length estimation

The sum of all aerobic phase lengths of the database and the new aerobic phase lengths updated using the proposed methodology were

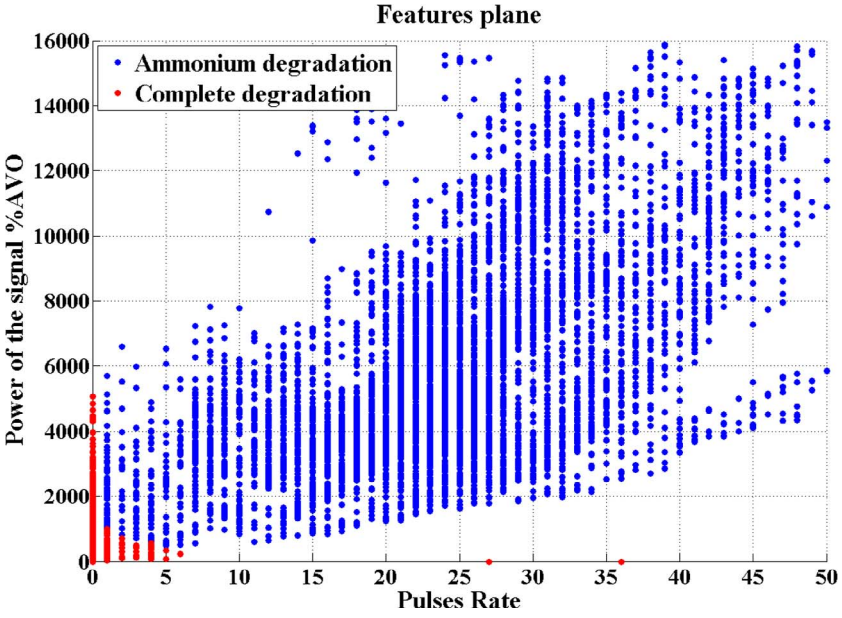

Fig. 10. Training set in the features plane. The ammonium degradation class (blue dots) and the complete degradation class (red dots) form two well defined, but not linearly separable areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

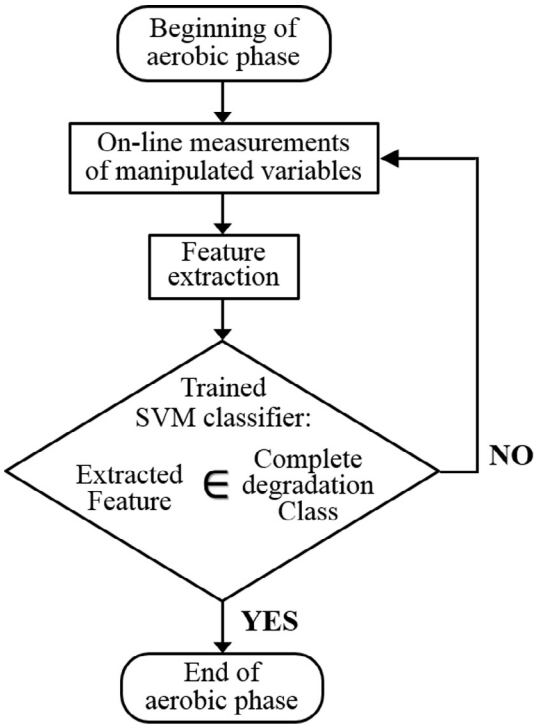

Fig. 11. Flow chart of the proposed methodology for its implementation to real-time aerobic phase length estimation.

Table 3
SVM tuned hyper-parameters.

KernelScale	BoxConstraint
176.14	33656

computed. The total times were 3041.74 h and 2812.84 h, respectively. This corresponds to a 7.52% reduction in total operation time, highlighting the use of advanced monitoring tools which ensure robustness and accuracy on the aerobic phase length estimation.

To better understand the reduction of operation times, a histogram of time reduction for each aerobic phase is shown in Fig. 13. Most of the aerobic phases had a reduction of around 20 min, however there are also phases with time reduction greater than 60 min (or even 120 min).

Table 4
Confusion matrix of the classifier over the training database.

	Classified Ammonium degradation	Classified End of phase
Actual Ammonium degradation	6496	0
Actual End of phase	0	729
Correct classifications: 7225		Accuracy: 100%
Misclassifications: 0		Error: 0%
Sensitivity (%) = 100		Specificity (%) = 100

Table 5
Confusion matrix of the classifier over the testing database.

	Classified Ammonium degradation	Classified End of phase
Actual Ammonium degradation	1624	0
Actual End of phase	0	182
Correct classifications: 1806		Accuracy: 100%
Misclassifications: 0		Error: 0%
Sensitivity (%) = 100		Specificity (%) = 100

4. Discussion

We use a SVM classifiers to generate a decision rule between two classes. This is achieved by searching the best hyper-plane that maximizes the margin to the closest features. A correct feature extraction/selection considerably improves the performance of the classifier, because it allows to have a greater separability between classes. In this sense, we have the advantage of incorporating more information (features) than other strategies that search for patterns directly in the sensor signal (e.g., ammonia valley, ORP plateau, DO elbow), which present constraints for operating parameters (e.g., low ammonia loads, excess of aeration, and temperature) [18].

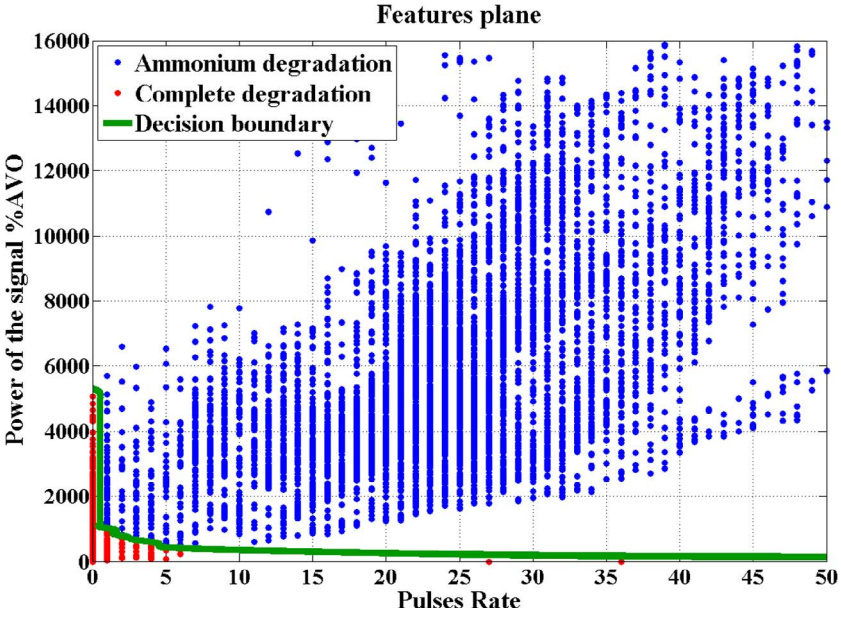

Fig. 12. In green, the optimum decision boundary to discriminate between the ammonium degradation class and the complete degradation class. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

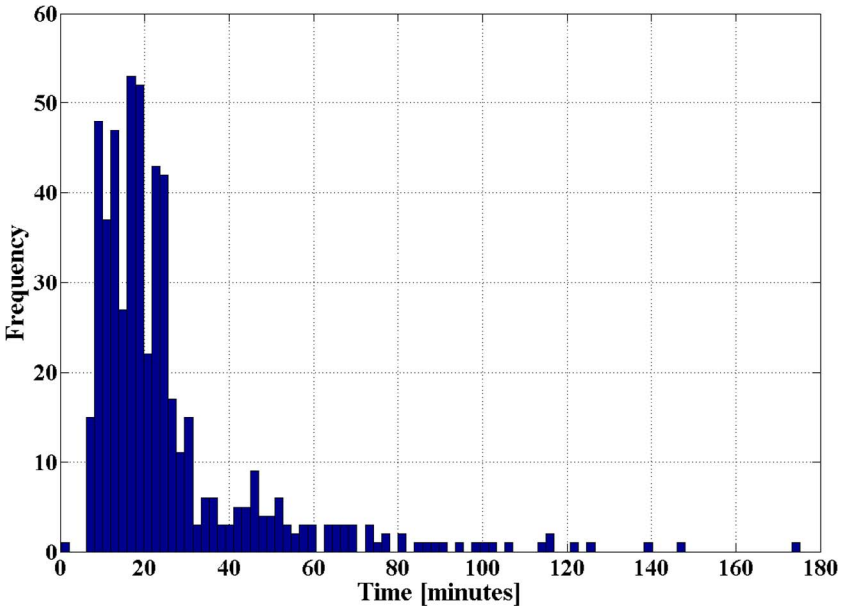

Fig. 13. Histogram which represents the differences between each phase length in the database and its respective phase length estimated with the proposed methodology.

4.1. SVM performance

The results in Table 4 are important, because the SVM classifier obtained a performance with 100% accuracy. All the classes were correctly classified, 6496 cases of ammonium degradation, and 729 cases of complete ammonium degradation. In other words, a correct aerobic phase length estimation was obtained for all the SBR cycles in the training database. This is an important first result for this methodology, as it was possible to find the optimal decision boundary, which acts as a bending point detector. In the case of sensitivity and specificity, a performance of 100% were obtained.

The results in Table 5 show that a 100% accuracy, 0% of error, 100% of sensitivity and 100% of specificity were obtained using the testing dataset, thereby validating the proposed methodology to estimate the aerobic phase length. Generally, in most cases the classifier performance decreases when unknown data is tested. If the performance does not reach a required standard, it is necessary to apply improvements in the followed methodology. Examples of improvements that increase the separability between classes include re-training using the same training dataset, re-training using a different training dataset, changing the classifier structure, and/or adding new features. But in our case, the proposed methodology was able to obtain a 100% accuracy, in spite of operating drawbacks such as changes in the set-point of the pH controller (from 7.5 to 8.5), or deterioration along the batch cycles of the closed-loop DO controller performance.

Imbalanced data tends to overfit the majority class in the training dataset, increasing incorrect classifications for the minority class samples [40]. Even when our database has an imbalance ratio of 9:1, the classification performance always reached a 100% of accuracy, 100% of sensitivity, and 100% of specificity for both training and testing stages. These remarkable results support that the feature extraction/selection and SVM training were adequate to the aerobic phase length estimation problem re-defined as a binary classification.

Furthermore, it is important to mention that the features were selected based on the characteristics of the original signals. For different signal characteristics (e.g., sampling time, behavior), or kind of signals (e.g., ORP, or different manipulated variables), other features may be appropriate to reach the best separability between the classes in the feature space. This implies that our proposed strategy could be used both in nitrification–denitrification and partial nitrification processes.

4.2. Improvements of the aerobic phase length estimation

Results show that the proposed methodology does not require a fixed time window to ensure the end of nitrification (30 min as it was proposed by [13]). Furthermore, it shows evidence of being able to overcome complex cases where the achieved time reduction were greater than 30 min.

4.3. Future work considerations

We propose to study the influence of our strategy over partial nitrification performance in SBRs on the long term. We also look to implement this strategy in systems (nitrification–denitrification or partial nitrification processes) where other kind of variables must be monitored to detect the aerobic phase end-point.

5. Conclusions

A novel strategy of aerobic phase length estimation based on feature extraction and SVM classifiers over a BNR process was presented. Both extracted features, Pulses Rate and Power of the signal %AVO, showed to be fundamental to obtain representative patterns of the aerobic phase end-point. This despite changes in operating condition of the reactor and non-linear/time-variant characteristics of the nitrification processes. From a point of view of binary classification, the selected features always allowed to have separability between the classes defined in this work. According to the classification results, a 100% of accuracy were obtained for both the training database and the testing database, validating the selected features and SVM training performance. The use of the proposed strategy, in comparison with previous methodologies, allows for a total reduction in aerobic phase lengths of approximately 7.52% (corresponding to 9.54 days). Most aerobic phase lengths in the analyzed database had a reduction time around 20 min, but time reductions greater than 60 min were also achieved.

Acknowledgements

The work of Francisco Jaramillo has been supported by a Ph.D scholarship from CONICYT, Chile. CONICYT-PCHA/Doctorado Nacional/ 2014-21140201. Francisco Jaramillo and Marcos Orchard also acknowledge the support from the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008 and

FONDECYT 1170044.

References

[1] K.M. Boaventura, N. Roqueiro, M.A.Z. Coelho, O.Q.F. Araújo, State observers for a biological wastewater nitrogen removal process in a sequential batch reactor, Bioresour. Technol. 79 (2001) 1–14.

[2] J. Guo, Y. Peng, S. Wang, Y. Zheng, H. Huang, S. Ge, Effective and robust partial nitrification to nitrite by real-time aeration duration control in an SBR treating domestic wastewater, Process Biochem. 44 (2009) 979–985.

[3] D. Rodŕguez, N. Pino, G. Peñuela, Monitoring the removal of nitrogen by applying a nitrification-denitrification process in a Sequencing Batch Reactor (SBR), Bioresour. Technol. 102 (2011) 2316–2321.

[4] C. Antileo, A. Werner, G. Ciudad, C. Muñoz, C. Bornhardt, D. Jeison, H. Urrutia, Novel operational strategy for partial nitrification to nitrite in a sequencing batch rotating disk reactor, Bioche. Eng. J. 32 (2006) 69–78.

[5] J. Guo, Y. Peng, H. Huang, S. Wang, S. Ge, J. Zhang, Z. Wang, Short- and long-term effects of temperature on partial nitrification in a sequencing batch reactor treating domestic wastewater, J. Hazard. Mater. 179 (2010) 471–479.

[6] V. Pambrun, E. Paul, M. Spérandio, Control and modelling of partial nitrification of effluents with high ammonia concentrations in sequencing batch reactor, Chem. Eng. Process. 47 (2008) 323–329.

[7] B. Sinha, A. Annachhatre, Partial nitrification – Operational parameters and microorganisms involved, Rev. Environ. Sci. Biotechnol. 6 (2007) 285–313.

[8] G. Ciudad, R. González, C. Bornhardt, C. Antileo, Modes of operation and pH control as enhancement factors for partial nitrification with oxygen transport limitation, Water Res. 41 (2007) 4621–4629.

[9] S. Marsili-Libelli, A. Spagni, R. Susini, Intelligent monitoring system for long-term control of sequencing batch reactors, Water Sci. Technol. 57 (2008) 431–438.

[10] P. Wilderer, R. Irvine, M. Goronszy, Sequencing Batch Reactor Technology, IWA Publishing, London, UK, 2001.

[11] Q. Yang, S. Gu, Y. Peng, S. Wang, X. Liu, Progress in the development of control strategies for the SBR process, Clean – Soil Air Water 38 (2010) 732–749.

[12] N. Zhu, W. Chen, X. Zheng, Effectiveness of ph as control parameter for nitrogen removal via snd in a sequencing batch reactor, in: 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009, pp. 1–5.

[13] C. Antileo, H. Medina, C. Bornhardt, C. Muñoz, F. Jaramillo, J. Proal, Actuators monitoring system for real-time control of nitrification-denitrification via nitrite on long term operation, Chem. Eng. J. 223 (2013) 467–478.

[14] M.V. Ruano, J. Ribes, A. Seco, J. Ferrer, An advanced control strategy for biological nutrient removal in continuous systems based on pH and ORP sensors, Chem. Eng. J. 183 (2012) 212–221.

[15] S.G. Won, C.S. Ra, Biological nitrogen removal with a real-time control strategy using moving slope changes of pH(mV)- and ORP-time profiles, Water Res. 45 (2011) 171–178.

[16] S. Puig, L. Corominas, M.T. Vives, M.D. Balaguer, J. Colprim, J. Colomer, Development and implementation of a real-time control system for nitrogen removal using OUR and ORP as end points, Ind. Eng. Chem. Res. 44 (2005) 3367–3373.

[17] P.T. Martín de la Vega, E. Martínez de Salazar, M.A. Jaramillo, J. Cros, New contributions to the ORP & DO time profile characterization to improve biological nutrient removal, Bioresour. Technol. 114 (2012) 160–167.

[18] K.M. Poo, J.H. Im, B.H. Jun, J.R. Kim, L.S. Hwang, K.S. Choi, C.W. Kim, Full-cyclic control strategy of SBR for nitrogen removal in strong wastewater using common sensors, Water Sci. Technol. 53 (2006) 151–160.

[19] W. Zeng, Y. Peng, S. Wang, C. Peng, Process control of an alternating aerobic-anoxic sequencing batch reactor for nitrogen removal via nitrite, Chem. Eng. Technol. 31 (2008) 582–587.

[20] S. Gu, S. Wang, Q. Yang, P. Yang, Y. Peng, Start up partial nitrification at low temperature with a real-time control strategy based on blower frequency and pH, Bioresour. Technol. 112 (2012) 34–41.

[21] G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, Wiley, 2006.

[22] S. Arik, T. Huang, W.K. Lai, Q. Liu, Neural Information Processing: 22nd International Conference, ICONIP 2015, November 9–12, 2015, Proceedings, number parte 4 in Lecture Notes in Computer Science, Springer International Publishing, 2015.

[23] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mech. Syst. Signal Process. 17 (2003) 317–328.

[24] J. Zarei, Induction motors bearing fault detection using pattern recognition techniques, Expert Syst. Appl. 39 (2012) 68–73.

[25] A. Bleakie, D. Djurdjanovic, Feature extraction, condition monitoring, and fault modeling in semiconductor manufacturing systems, Comput. Ind. 64 (2013) 203–213.

[26] E. Sejdić, I. Djurović, J. Jiang, Time–frequency feature representation using energy concentration: an overview of recent advances, Digital Signal Process. 19 (2009) 153–183.

[27] O. Smart, H. Firpi, G. Vachtsevanos, Genetic programming of conventional features to detect seizure precursors, Eng. Appl. Artificial Intelligence 20 (2007) 1070–1085.

[28] D. Cvetkovic, E. Übeyli, I. Cosic, Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study, Digital Signal Process. 18 (2008) 861–874.

[29] Z. Wang, S. Bian, M. Lei, C. Zhao, Y. Liu, Z. Zhao, Feature extraction and classification of load dynamic characteristics based on lifting wavelet packet transform in power system load modeling, Int. J. Electr. Power Energy Syst. 62 (2014) 353–363.

[30] D. Ruan, G. Chen, E.E. Kerre, G. Wets, Intelligent Data Mining: Techniques and Applications, Studies in Computational Intelligence, Springer, Berlin Heidelberg, 2005.

[31] V.N. Mandhala, V. Sujatha, B.R. Devi, Scene classification using support vector machines, in: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, 2014, pp. 1807–1810.

[32] S. Abe, Support Vector Machines for Pattern Classification, Advances in Computer Vision and Pattern Recognition, Springer, London, 2010.

[33] A. Ben-Hur, C.S. Ong, S. Sonnenburg, B. Schölkopf, G. Rätsch, Support vector machines and kernels for computational biology, PLoS Comput. Biol. 4 (2008) 1–10.

[34] C.M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.

[35] P. Mather, B. Tso, Classification Methods for Remotely Sensed Data, second ed., CRC Press, 2009.

[36] F. Chen, B. Tang, R. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, Measurement 46 (2013) 220–232.

[37] A. Widodo, B.-S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (2007) 2560–2574.

[38] Mathworks, Statistics and Machine Learning Toolbox User's Guide, The MathWorks Inc, 2015.

[39] K.H. Hui, M.H. Lim, M.S. Leong, S.M. Al-Obaidi, Dempster-Shafer evidence theory for multi-bearing faults diagnosis, Eng. Appl. Artif. Intell. 57 (2017) 160–170.

[40] R. Muscat, M. Mahfouf, A. Zughrat, Y.Y. Yang, S. Thornton, A.V. Khondabi, S. Sortanos, Hierarchical fuzzy support vector machine (SVM) for rail data classification, IFAC Proceedings Volumes 47 (2014) 10652–10657.