Estimation of effluent quality parameters from an activated sludge system using quantitative image analysis

Daniela P. Mesquita [a], A. Luís Amaral [a,b], Eugénio C. Ferreira [a,*]

[a] CEB – Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] Instituto Politécnico de Coimbra, ISEC, DEQB, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

HIGHLIGHTS

- Quantitative image analysis was used to evaluate an activated sludge system.
- COD, ammonia, and nitrate effluent concentrations were estimated by partial least squares.
- Using morphological and physiological data provided the best predicting abilities.
- Quantitative image analysis has the ability to be used in process monitoring.

ARTICLE INFO

ABSTRACT

The efficiency of an activated sludge system is generally evaluated by determining several key parameters related to organic matter removal, nitrification and/or denitrification processes. Off-line methods for the determination of these parameters are commonly labor, time consuming, and environmentally harmful. In contrast, quantitative image analysis (QIA) has been recognized as a prompt method for assessing activated sludge contents and structure. In the present study an activated sludge system was operated under different experimental conditions leading to a variety of operational data. Key parameters such as chemical oxygen demand (COD) and ammonium (N-NH$_4^+$), and nitrate (N-NO$_3^-$) concentrations, throughout the experimental periods, were measured by classical analytical techniques. QIA was further used for the microbial community characterization. Partial least squares (PLS) models were used to correlate QIA information and the aforementioned key parameters. It was found that the use of the morphological and physiological data allowed predicting, at some extent, the effluent COD, N-NH$_4^+$, and N-NO$_3^-$ concentrations based on chemometric techniques.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In environmental systems, processes and reactions are often dynamic, irreversible, and occurring in systems which are difficult to define, and unviable to describe using deterministic models [1]. Disturbances detection is an important task for process performance enhancement in wastewater treatment plants (WWTPs), due to strict regulations regarding effluent quality. Among the operational parameters most commonly used for assessing wastewater quality, the chemical oxygen demand (COD), as an indirect measure of the organic matter concentration, is of the utmost importance. Ammonium (NH$_4^+$) and nitrate (NO$_3^-$) concentrations are also widely assessed, reflecting the extent of nitrification/denitrification processes. However, traditional methodologies for determining these parameters are costly, labor and time-consuming, and may even present environmental risks associated to end products. Therefore, alternative methodologies that could predict these parameters, in a time and labor effective manner are of great interest.

In recent years, a great deal of attention has been given to different monitoring strategies to help clarifying the behavior of biological processes, such as fluorescent microscopy techniques and/or mathematical modeling. It has been previously found that Fluorescence in situ Hybridization (FISH) and epifluorescence microscopy (as an alternative to the use of confocal laser scanning microscopy) are among the best strategies to quantify nitrifying bacteria in activated sludge [2]. However, other microorganisms found in those systems also contribute to the behavior of the biological process. Recently, mathematical models were applied to

activated sludge systems and it has been deduced from many full-scale observations that improper sludge age, food/biomass ratio (fed and removed), and feed composition imbalance could result in bulking, foaming, and other malfunctions. These relations have been recently summarized in attempts to model a WWTP risk to suffer from solids separation problems [3–5]. Furthermore, another model was developed [6] showing the bacteria community predators significance on the performance of a nitrifying system. Indeed, these models have been shown to allow for the characterization of the effluent quality in activated sludge bioreactors. Furthermore, it is nowadays considered that biomass characterization provides important information about the biological processes leading to a number of different bulking events. In this sense, microscopy techniques coupled to quantitative image analysis (QIA), characterizing the system as a whole, may present an advantage for prediction purposes, alongside mathematical modeling.

In order to assess activated sludge operating parameters, QIA applications have been employed for aggregated and filamentous bacteria contents determination, biomass structure and physiology (Gram-positive/Gram-negative and viable/damaged bacteria ratios) assessment, and even intracellular storage polymers (PHA, glycogen, poly-phosphates) quantification [7–16]. All of the above demonstrate that image processing and analysis methodologies are able to provide valuable information about activated sludge systems, and are nowadays recognized as valid monitoring tools. Indeed, the possibility of accurately monitoring the filamentous bacteria contents (responsible for bulking phenomena), or to characterize and quantify the aggregated biomass contents, size stratification, and structure (related to pin point or viscous bulking phenomena), by this simple, quantitative, automated and near real time methodology, compares favorably to the conventional monitoring methods.

As recent advances in computer and instrumentation techniques lead to the collection of large amounts of data from different processes, the integration of mathematical and statistical methods is crucial. Furthermore, this increased number of monitored variables in WWTPs, due to the use of computerized measurement devices, hinders data interpretation. Therefore, a systematic method to handle and analyze data is needed to effectively extract relevant information for process monitoring and supervision [17]. In this sense, chemometric techniques are increasingly used for a wide variety of tasks, including data evaluation and interpretation, optimization and development of predictive models in processes, and extraction of a maximum of information from experimental data [1,18]. Partial least squares (PLS) is a widely used statistical technique able to model a given predicted variable and detect deviations [19]. Furthermore, PLS models have been recently used for the estimation of COD, nitrate (NO_3^-), total organic carbon (TOC), and mixed liquor suspended solids (MLSS) using ultraviolet–visible and near-infrared spectroscopy on wastewater treatment systems [20–22]. Despite the prior use of image analysis data, coupled to PLS models, to estimate sludge volume index (SVI) and MLSS [11,12,15], QIA has not yet been applied, to our knowledge, neither to estimate the COD removal ability nor nitrification/denitrification occurrence.

In this sense, the present work aims at assessing the performance of an activated sludge system in four different operational conditions. Microscopic images were obtained throughout each experiment and processed by a QIA procedure to obtain the image analysis data (QIA data) regarding the activated sludge biomass morphological and physiological characterization. PLS models were further applied by integrating QIA data alongside operational data. The final goal of this research was to establish the most appropriate data and methodology for the prediction of COD, ammonia (N-NH$_4^+$), and nitrate (N-NO$_3^-$) within the activated sludge system.

2. Materials and methods

2.1. Experimental setup

A lab-scale activated sludge system fed with synthetic wastewater containing acetate as the main carbon source was used. The reactor had a working volume of 17 L (aeration tank), followed by a 2.5 L clarifier. The system was equipped with two feeding pumps for influent dilution, air supply at the bottom allowing efficient agitation, and sensor apparatus such as a dissolved oxygen probe and a pH meter with associated control pump. Sludge recirculation from the clarifier to the aerated tank was also performed.

2.2. Experimental conditions

During the present work, four different experimental conditions (EC) were conducted by changing the organic loading rate (OLR) and the food to microorganism's ratio (F/M). The operated time of each EC was 75 (EC1), 85 (EC2), 118 (EC3), and 49 (EC4) days. The sludge residence time (SRT) was set to 20 d for the first experimental condition (EC1), 15 d for the second (EC2), 20 d for the third (EC3) and 6 d for the fourth (EC4). These conditions were defined based on the typical SRT range in conventional activated sludge systems [23]. During the EC1 a sharp decrease on the OLR and F/M from 2.7 to 0.67 kg_{COD} m^{-3} d^{-1} and 0.88 to 0.23 kg_{COD} kg_{MLSS}^{-1} d^{-1} was performed. EC2 was conducted decreasing gradually the OLR and F/M from 0.6 to 0.09 kg_{COD} m^{-3} d^{-1} and 0.16 to 0.07 kg_{COD} kg_{MLSS}^{-1} d^{-1}, respectively. For EC3 the OLR was gradually increased from 0.27 to 0.87 kg_{COD} m^{-3} d^{-1} and the F/M from 0.07 to 0.35 kg_{COD} kg_{MLSS}^{-1} d^{-1}. Finally, EC4 was performed with an OLR around 0.59 kg_{COD} m^{-3} d^{-1} and an F/M of 0.15 kg_{COD} kg_{MLSS}^{-1} d^{-1} with some minor fluctuations. These four strategies were conducted with the purpose of obtaining a variety of operational conditions, leading to different sludge morphological and physiological properties.

2.3. Synthetic wastewater composition

Beyond acetate and (NH$_4$)$_2$SO$_4$, the concentration of the other nutrients added in the synthetic feed is listed below (per L): 0.025 g MgSO$_4$·7H$_2$O; 0.044 g KH$_2$PO$_4$; 0.059 g K$_2$HPO$_4$·2H$_2$O; 0.03 g CaCl$_2$·2H$_2$O; 0.019 g FeCl$_3$·6H$_2$O; 0.105 g NaHCO$_3$, and 2 mL of a trace metals solution for biomass maintenance. The trace metals consisted of (g L^{-1}): H$_3$BO$_3$, 0.05; ZnCl$_2$, 0.05; CuI$_2$·H$_2$O, 0.04; MnCl$_2$, 0.02; (NH$_4$)$_6$Mo$_7$O$_{24}$·4H$_2$O, 0.055; AlCl$_3$, 0.05; NiCl·6H$_2$O, 0.11 (adapted from [24]).

2.4. Analytical procedures

The system was monitored for SVI and MLSS, COD, N-NH$_4^+$, and N-NO$_3^-$ concentrations. SVI was determined based on the sludge height variation monitored for 30 min and combined with MLSS results. MLSS were measured in accordance with the procedures described in Standard Methods [25]. Samples for COD, N-NH$_4^+$, and N-NO$_3^-$ were collected, centrifuged, and filtered, from the synthetic wastewater and aeration tank. COD was measured with Hach Lange COD cell tests (LCK 414 and LCK 514) on a spectrophotometer (Hach Lange DR 5000). N-NH$_4^+$ was determined according to Nessler's method [25]. N-NO$_3^-$ was determined by high-performance liquid chromatography (HPLC) according to Sarraguça et al. [20].

2.5. Staining procedures

The assessment of biomass composition (and physiology) was performed collecting daily fresh sludge samples from the reactor. The Live/Dead BacLight bacterial viability kit was used to differentiate viable and damaged bacteria (Molecular Probes, Eugene, Oregon, USA). This kit utilizes a mixture of SYTO 9 green-fluorescent and propidium iodide (PI) red-fluorescent nucleic acid stains. Viable bacteria are stained by SYTO 9 and damaged bacteria are stained by PI. The Live Baclight bacterial Gram stain allows classifying bacteria as Gram (+) or Gram (−) without the use of fixatives (Molecular Probes, Eugene, Oregon, USA). This kit utilizes a mixture of SYTO 9 green-fluorescent and hexidium iodide (HI) red-fluorescent nucleic acid stains. Gram (−) bacteria are stained by SYTO 9 and Gram (+) bacteria are stained by HI.

For the staining procedures a sterile solution of 0.85% NaCl was first prepared. For the staining solutions, 1.5 µL of each dye was diluted into 5 mL of the NaCl solution and the tube was wrapped with aluminum foil and protected from light to avoid photodegradation. A volume of 100 µL of undiluted biomass suspension from the reactor was mixed with 50 µL of staining solution and incubated in darkness for 15 min at room temperature. The bacteria population was then visualized through epifluorescence microscopy.

2.6. Image acquisition

Aggregated and filamentous biomass contents and structure (morphological data) were assessed by images acquired through bright-field microscopy, and biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria (physiological data) by epifluorescence microscopy from mixed liquor samples obtained in each of the four experiments. Daily samples were taken from the aerated tank for visualization and image acquisition using an Olympus BX51 microscope (Olympus, Tokyo, Japan) coupled with an Olympus DP71 camera (Olympus, Tokyo, Japan). A recalibrated micropipette with a sectioned tip at the end, with a large enough diameter to allow larger aggregates to flow, was used to deposit 10 µL of the sample on to the slide, and covered with a 20 mm × 20 mm cover slip. Three slides per sample were used for the morphological characterization (bright-field microscopy), resulting in a total of 150 images (3 × 50 images per slide) examined at 100× total magnification and acquired at 1360 × 1024 pixels in 8-bit format. For the physiological assessment (epifluorescence microscopy with 200× total magnification) two slides per sample were used, resulting in a total of 100 images (2 × 50 images per slide) of 1360 × 1024 pixels and 24-bit format. Images were acquired in the upper, middle and bottom of the slide in order to improve the representativeness of the microbial community. 24, 29, 18, and 31 samples were collected from EC1, EC2, EC3, and EC4, respectively, for microscopy image acquisition.

2.6.1. Image processing and analysis in bright-field images

Regarding biomass contents and structure determination (morphological data) a detailed description of the developed image processing and analysis programs is presented below, with the main steps summarized for pre-treatment, segmentation and debris elimination (image processing), and data determination (image analysis). The image processing and analysis programs were developed in Matlab 7.3 (The Mathworks, Natick, USA).

2.6.1.1. Pre-treatment. The image pre-processing stage relies on the enhancement of the original grayscale images by background determination and removal. In this stage, the original image is first divided by a previously acquired background image to minimize background light differences. Aggregates and filaments are further enhanced by using local histogram equalization in order to improve boundaries contrast.

2.6.1.2. Segmentation. This stage consists primarily in aggregates and filaments segmentation by simultaneous determination of aggregates boundary and core images. Following pre-treatment, a boundary grayscale image is determined by subtracting the eroded from the dilated pre-treated image, and segmented by a predefined 0.3 threshold level to the binary boundary image. The binary core image is obtained by segmenting the pre-treated image by a predefined 0.3 threshold level. The binary boundary and binary core images are then combined to form the binary aggregates image. Finally, a morphological opening (disk of radius 5) and reconstruction allows for the final aggregates binary image to be determined. For filaments segmentation, the aggregates binary image is used as a mask to remove the aggregates from the original grayscale image, next segmented by a 92.5% percentile based segmentation procedure. Next, a predefined 0.2 threshold level allows the determination of a filament marker image which is later used as a mask to reconstruct the filaments more accurately.

2.6.1.3. Debris elimination. Finally, a gyration radius based procedure is implemented in the aggregates final binary image to discard small filamentous-like debris by the use of a 1.2 cut-off value [26]. Residual aggregates (smaller than 3.5 µm in diameter) and debris elimination is performed by a 3rd order erosion and reconstruction operations and all aggregates cut off by the image boundaries are removed.

2.6.1.4. Data determination. From the aggregates final binary image, aggregates are classified according to their size, based on the equivalent diameter (D_{eq}), in small (sml) ($D_{eq} < 25$ µm), intermediate (int) (25 µm $< D_{eq} < 250$ µm), and large (lrg) aggregates ($D_{eq} > 250$ µm). For each size range the morphological data is determined according to Mesquita et al. [12]. From the filaments final binary image, filaments total length per volume is determined.

Fig. 1 shows illustrative bright-field images from each conducted experiment with corresponding aggregates and filaments binary images obtained from the image analysis procedure.

2.6.2. Image processing and analysis in epifluorescence images

With respect to the biomass composition (physiological assessment), two different staining kits were used from Molecular Probes (Eugene, Oregon, USA). Live/Dead® BacLight™ bacterial viability kit was used to differentiate viable and damaged bacteria, and Live Baclight™ bacterial Gram stain kit for Gram status identification. Two long pass filters were used in an epifluorescence microscope, one for determining SYTO 9 green emission dye with an excitation bandpass of 470–490 nm and emission cut-off at 516 nm (1st filter), and a second filter for determining PI or HI red emission dyes with an excitation bandpass of 530–550 nm and emission cut-off at 591 nm (2nd filter). A more detailed description of the developed image processing and analysis programs is presented below, with the main steps summarized in pre-treatment, segmentation, debris elimination, post-treatment and fluorescence-based intensity images determination (image processing) and data determination (image analysis). The image processing and analysis programs were developed in Matlab 7.3 (The Mathworks, Natick, USA).

2.6.2.1. Pre-treatment. The image pre-processing stage allows the enhancement of the working color channels (green channel at 516 nm cutoff and red channel at 591 nm cutoff) from the original RGB images by background determination and removal. In this stage, a block process methodology is applied to determine the 200 × 200 pixels minimum surrounding each pixel, and therefore
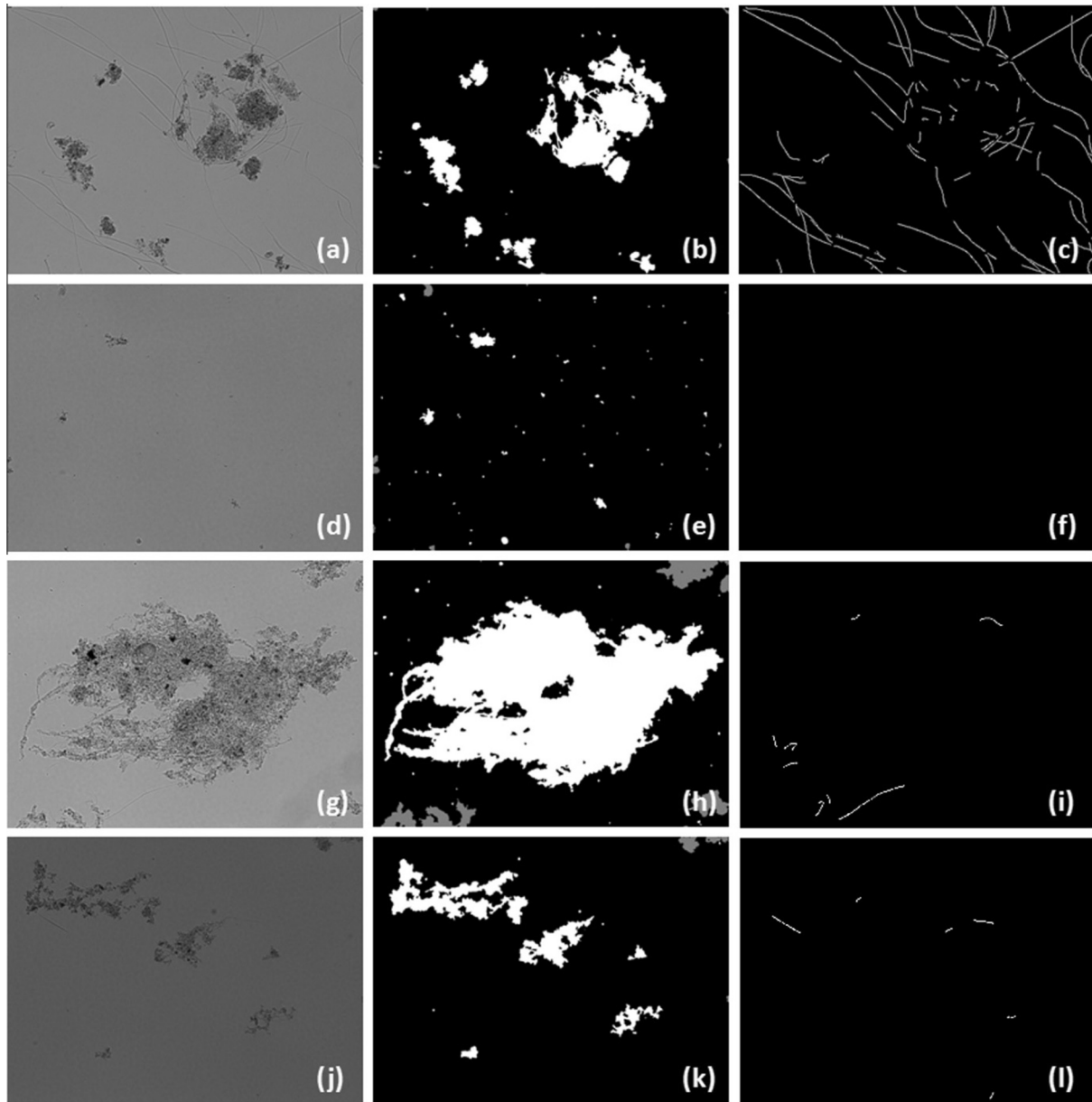
Fig. 1. (a) original grayscale image (EC1), (b) aggregates binary image, (c) filaments binary image, (d) original binary image (EC2), (e) aggregates binary image, (f) filaments binary image (g) original binary image (EC3), (h) aggregates binary image, (i) filaments binary image, (j) original grayscale image (EC4), (k) aggregates binary image, (l) filaments binary image.

compute the background images. The latter are then subtracted from the corresponding channel original images to minimize background intensity differences.

2.6.2.2. Segmentation. This stage consists primarily in the aggregates and filaments segmentation of the resulting pre-treatment images into binary images. Different predefined threshold levels are used for both the filaments and aggregates, as well as for each color channel. First the aggregates are segmented by predefined threshold levels, and then morphological opening (disk of radius 12) and reconstruction are applied for the elimination of small debris and attached filaments. Regarding filaments identification, after adaptive histogram equalization of the pre-treated images, filaments are segmented by predefined threshold levels.

2.6.2.3. Debris elimination. A morphological opening (disk of radius 3) is applied to the resulting binary images, and a gyration radius

based procedure is implemented to discard small filamentous-like debris by the use of a 1.2 cut-off value [26].

2.6.2.4. Post-treatment. In order to prevent a given filament to be identified as both red and green, and thus characterized simultaneously as Gram negative and positive (or viable and damaged), superimposed filaments in both binary channel images have to be correctly assigned. This is achieved by combining the final filaments binary images, from both channels, and determining the intensity of the superimposed filaments, in each channel. The filament is then attributed to the filaments binary channel image in which it presented the highest intensity value and, subsequently, deleted from the other.

2.6.2.5. Fluorescence-based intensity images determination. The last step of the image processing program consists in determining the fluorescence-based intensity images of the aggregated and

filamentous bacteria in both channels. For that purpose, the aggregates and filaments final binary images, in each channel, are used as masks to respectively determine the aggregates and filaments intensity images.

2.6.2.6. Data determination. Gram negative/positive and viable/damaged aggregated and filamentous biomass are characterized both in terms of projected area (or length) and in mean intensity value. The ratios between Gram positive/negative and viable/damaged bacteria are next determined. A more detailed description of the physiological data can be found in Mesquita et al. [14].

Fig. 2 shows illustrative epifluorescence images from Live/Dead and Gram staining with corresponding binary and intensity images provided by the image analysis procedure. Table 1 presents the overall dataset used in this work.

2.7. Partial least squares (PLS) analysis

In partial least squares (PLS) regression, the decomposition of **X** and **Y** is carried out iteratively. By exchanging information between the two blocks in each step, the latent variables (LVs) of the **X**-space are rotated so that the predictive power of the **X**-space with regard to the **Y**-space is enhanced [19]. In PLS it is critical to determine the optimal number of LVs and cross-validation (CV) is a reliable way to test the predictive significance of each PLS regression. Part of the training dataset is kept out of the model, predicted by the model and finally compared with the actual values (using CV). A more detailed explanation of PLS algorithm could be found elsewhere [1,27–29]. However, during the present study, it was found that the number of LVs depicted by this methodology did not allow an acceptable training model. Thus, the number of LVs was selected by correlating the predicted and the

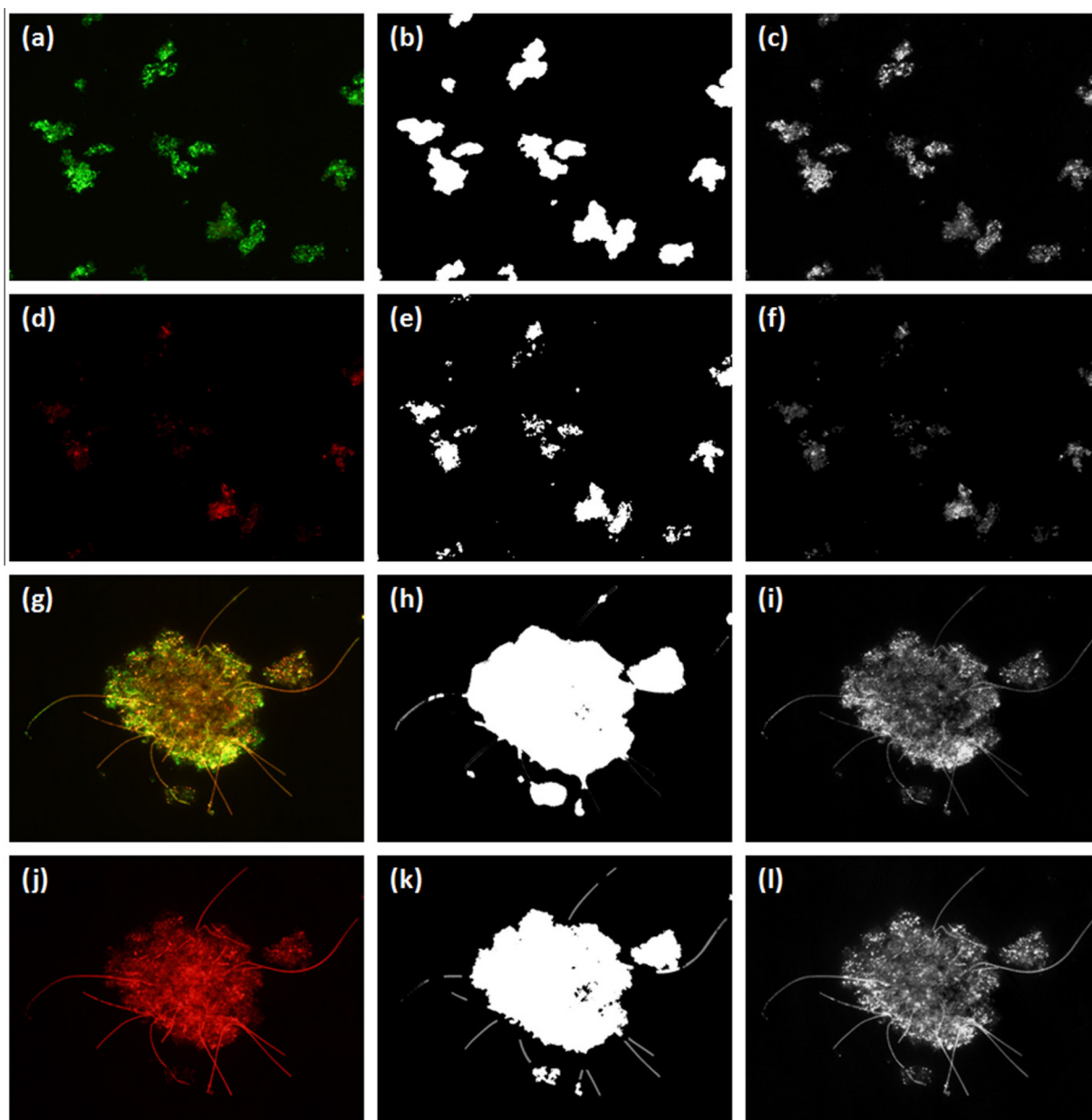**Fig. 2.** Images from the Live/Dead staining – 1st filter: (a) original, (b) area detection image, (c) intensity image; and from the 2nd filter: (d) original image, (e) area detection image, (f) intensity image. Images from the Gram staining – green filter: (g) original, (h) area detection image, (i) intensity image; and from the red filter: (j) original image, (k) area detection image, (l) intensity image.

Table 1
Morphological and physiological data included in the dataset used in the PLS analyses.

	Variable	Name
Predictor **(X)**	*Morphological data (from bright-field microscopy)*	
	Deq[a]	Equivalent diameter
	FF[b]	Form factor
	Conv[b]	Convexity
	Comp[b]	Compactness
	Round[b]	Roundness
	Solid[b]	Solidity
	Ext[b]	Extent
	Ecc [b]	Eccentricity
	Rob[b]	Robustness
	LrgC[b]	Largest concavity
	RelArea[b]	Ratio between hole and object area
	% Nb[a]	Aggregates number percentage
	% Area[a]	Aggregates area percentage
	Nb/Vol[a]	Aggregates number per volume
	TA/Vol	Total aggregate area per volume
	TL/Vol	Total filament length per volume
	Ln(TL/TA)	Logarithm of the ratio between total filament length and total aggregate area
	TL/MLSS	Total filament length per mixed liquor suspended solids
	MLSS/TA	Mixed liquor suspended solids per total aggregates area
	Physiological data (from epifluorescence microscopy)	
	G_AA_R	Gram-positive aggregated bacteria area
	G_AA_G	Gram-negative aggregated bacteria area
	G_AF_R	Gram-positive filaments area
	G_AF_G	Gram-negative filaments area
	LD_AA_R	Damaged aggregated bacteria area
	LD_AA_G	Viable aggregated bacteria area
	LD _AF_R	Damaged filaments area
	GR_AF_G	Viable filaments area
Response **(Y)**	COD_{out}	Chemical oxygen demand in the effluent
	$N\text{-}NH_{4\,out}^{+}$	Ammonium nitrogen concentration in the effluent
	$N\text{-}NO_{3\,out}^{-}$	Nitrate nitrogen concentration in the effluent

[a] Determined for small (Deq < 25 μm), intermediate (25 μm < Deq < 250 μm), and large (Deq > 250 μm) aggregates.
[b] Determined for intermediate (25 μm < Deq < 250 μm), and large (Deq > 250 μm) aggregates.

measured COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ concentrations, comprehending both training and validation data, expressed by the mean squared error. Additionally, and in order to obtain random training and validation datasets, up to 2000 different training and validation datasets from the original dataset were tested. The root mean squared error of prediction (RMSEP) and the correlation coefficient (R^2) between measured and model-generated **Y** values was then determined for the training set. Matrix (**X**) was always preprocessed using the standard normal variate (SNV) method to remove undesirable variations. Matrix (**Y**) containing the measured effluent COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ values was used to test the model performance.

Matlab™ 7.3 (The Mathworks, Natick, MA) was used for predicting COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ by PLS with a total of 102 observations from the four experiments. For each **Y** parameter prediction, two different studies, resulting in two PLS models, were performed. The first PLS model (PLS1) modeled all the different conditions as a whole, whereas the second study (PLS2) modeled each individual condition separately and then combined the ensemble of the individually obtained results for each EC. All PLS models were randomly divided into a training set (67% of the observations) to calibrate the model, and a validation set (33% of the observations) to validate the model. Prior to the PLS analysis, a data reduction step took place in which the PLS **X** matrix variables were chosen by a principal component analysis (PCA) procedure with $n/2$ components (n being the number of the training model observations). Accordingly, the final PLS **X** matrix variables for COD, $N\text{-}NH_4^+$, and

$N\text{-}NO_3^-$ prediction, presented the highest correlations (maximum 0.1 weighted average distance between the principal components values) with COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ respectively. Whenever two selected variables presented a high correlation between themselves (maximum 0.05 weighted average distance between the principal components values), the variable of the pair presenting the lower correlation with COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ respectively, was removed. Therefore, for each PLS model, in both for PLS1 and PLS2, within a total 79 variables provided by QIA, just 22, 19, and 17 variables were used for COD, $N\text{-}NH_4^+$, and $N\text{-}NO_3^-$ prediction, respectively.

3. Results and discussion

Figs. 3 and 4 displays the behavior of the operational data during the four experimental conditions studied in the activated sludge system.

Regarding EC1 (Fig. 3a–c), the SVI range was always higher than 150 mL g^{-1} (Fig. 3a), the considered limit for bulking conditions, according to Jenkins et al. [30], with the exception of the last three monitoring days. The experiment started with a very high influent COD (COD$_{in}$ of 3900 mg L^{-1}), decreasing after 15 days, influencing the OLR and the F/M ratio (Fig. 3c). With this strategy, COD$_{out}$, $N\text{-}NH_{4out}^+$ and $N\text{-}NO_{3out}^-$ concentrations were clearly affected as well as the activated sludge structure. Due to the high COD and $N\text{-}NH_4^+$ concentrations in the influent (COD$_{in}$ and $N\text{-}NH_{4in}^+$), the highest effluent COD and $N\text{-}NH_4^+$ concentrations (COD$_{out}$ and $N\text{-}NH_{4out}^+$) were obtained in this EC1 (Fig. 3b). At the very end of EC1, the $N\text{-}NH_{4out}^+$ increased, indicating that the performance of the biological system was affected as well as the nitrifying bacteria community. Also, the highest $N\text{-}NO_{3out}^-$ formation was achieved in this experiment, although presenting a decreasing trend with time after 40 days of experiment, which can be related to the $N\text{-}NH_{4out}^+$ increase in that period. It was also found that, throughout EC1, high contents of filamentous bacteria were present (Fig. 1a) influencing the operational parameters at the aeration tank, such as the SVI, and implying filamentous bulking as reported by Amaral et al. [15].

During EC2, SVI values were always lower than 150 mL g^{-1} (Fig. 3d). For this experiment the strategy encompassed the gradual decrease of the COD$_{in}$ (from 1000 to 140 mg L^{-1}) (Fig. 3d) influencing the OLR and the F/M ratio (Fig. 3f). In spite of the COD$_{out}$ fluctuation, the final COD concentration was always lower than the COD$_{out}$ from EC1. Also a constant decrease was found in the $N\text{-}NH_{4in}^+$ and $N\text{-}NH_{4out}^+$ concentration (Fig. 3e), alongside a steady formation of $N\text{-}NO_3^-$, although smaller than in EC1. MLSS in the reactor gradually decreased to 0.8 g L^{-1} at the end of EC2 (Fig. 3f), indicating a loss of biomass, mostly due to the low density properties of small flocs. QIA results established the presence of small flocs (Fig. 1d) with an extremely low number of filamentous bacteria (floc backbone) which influenced the correct balance between floc-forming and filamentous bacteria, resulting in a pinpoint flocs phenomenon [15].

Large SVI values were obtained for EC3 (reaching a maximum of 483 mL g^{-1}), indicating the presence of bulking conditions, although much lower than the ones obtained for filamentous bulking (EC1), and with a floc morphology pointing to the presence of viscous bulking (Fig. 1g) [15]. The experiment started with a COD influent of 400 mg L^{-1}, with a two-step increase after 30 and 90 days, respectively (Fig. 4a), influencing the OLR and the F/M ratio (Fig. 4c). The MLSS was slightly affected throughout the experiment whereas the COD$_{out}$ in the reactor was approximately constant and lower (Fig. 4b) than in EC1. The influent and reactor $N\text{-}NH_{4in}^+$ concentrations increased after 90 days, influencing the $N\text{-}NH_{4out}^+$ at the end of the experiment. Regarding the $N\text{-}NO_{3out}^-$, it is clear that with this strategy, the nitrification process was not conducted in the biological system.
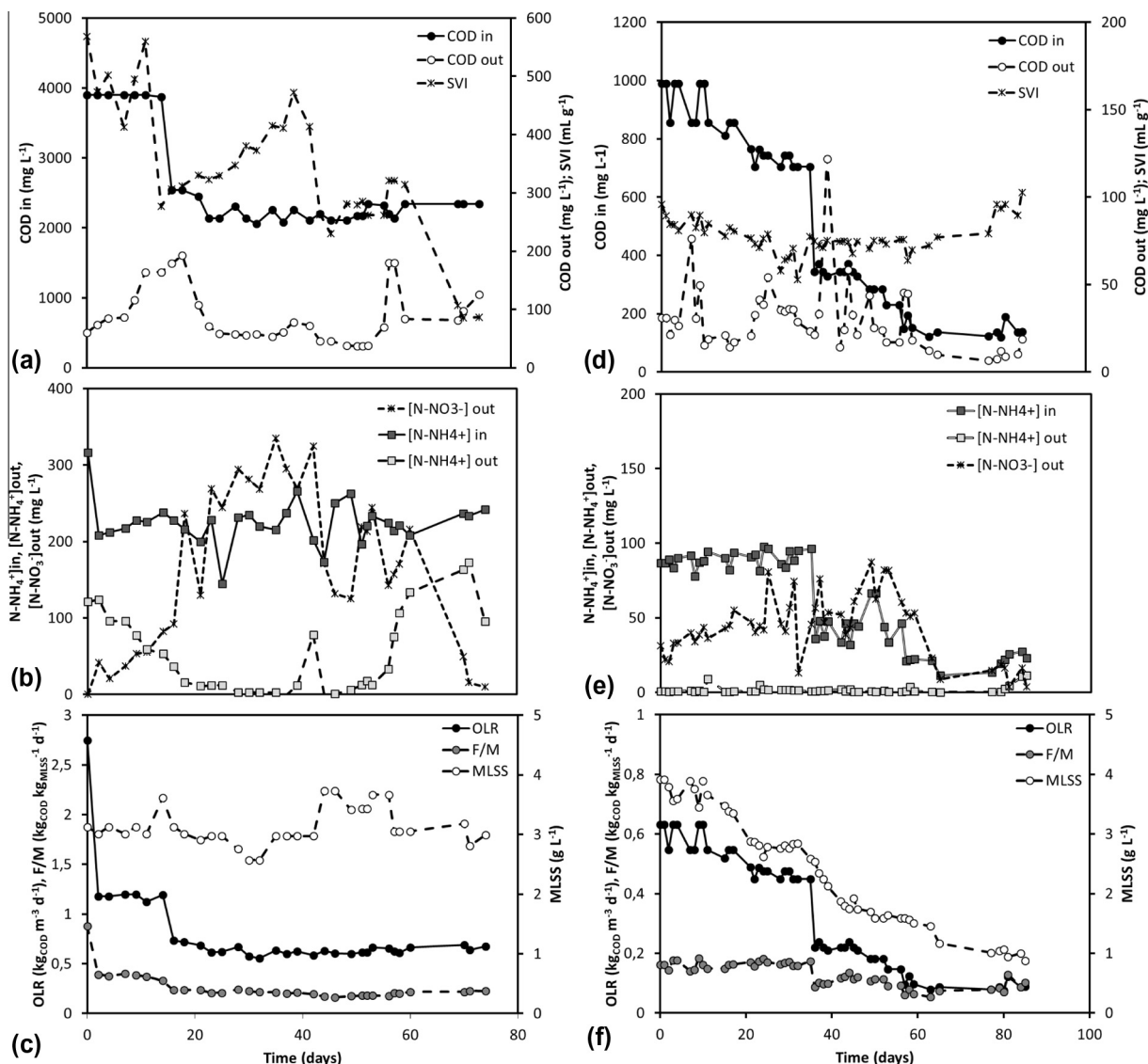
Fig. 3. OLR, F/M, MLSS, COD$_{in}$, COD$_{out}$, N-NH$_4^+$$_{in}$, N-NH$_4^+$$_{out}$, and N-NO$_3^-$$_{out}$ (operational data) behavior for EC1 (a–c) and for EC2 (d–f).

Regarding EC4, SVI values were always lower than 150 mL g^{-1}, indicating good settling properties (Fig. 4d). The MLSS attained a plateau value around 3–4 g L^{-1} (Fig. 4f). COD$_{in}$ was maintained constant during the experiment; however, COD$_{out}$ was quite higher than the ones obtained during EC3. At the same time, N-NH$_4^+$$_{in}$, remained approximately constant (Fig. 4e) and there was practically no formation of N-NO$_3^-$ (with a slight increase at the end of the experimental period). The QIA results revealed normal floc structures (Fig. 1j) with a good balance between floc-forming and filamentous bacteria [14,15].

The possibility of using image analysis data to predict the effluent COD, N-NH$_4^+$, and N-NO$_3^-$ through PLS analysis was then sought. The collected data from the aggregated and filamentous biomass characterization (morphological data), biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria (physiological data), and operational parameters was used for that purpose (Table 1). Two different studies, resulting in two PLS models each for COD$_{out}$, N-NH$_4^+$$_{out}$, and N-NO$_3^-$$_{out}$, were performed. The first PLS model (PLS1) modeled all the different conditions as a whole, whereas the second study (PLS2) modeled each individual condition separately and then combined the ensemble of the individually obtained results for each EC.

PLS1 results for the effluent COD, N-NH$_4^+$, and N-NO$_3^-$ prediction are depicted in Table 2. This model was obtained using solely the selected variables for the ensemble of the conditions (22 for COD$_{out}$, 19 for N-NH$_4^+$$_{out}$, and 17 for N-NO$_3^-$$_{out}$). In all cases, both the regression coefficients and regression equations correspond to the linear correlations found by the least squares method. Furthermore, the presented regression analyses and coefficients were obtained for the joint predicted and measured values from the overall (training and validation) datasets and the RMSEP was also determined. The regression analysis between PLS1 predicted and measured COD$_{out}$, N-NH$_4^+$$_{out}$, and N-NO$_3^-$$_{out}$ presented reasonable prediction abilities, with the exception of the effluent COD, where the RMSEP was 14.81 mg L^{-1} and the correlation coefficient was 0.7. At this point, it is important to notice that by combining the four experimental conditions, a high variability on the COD$_{out}$, N-NH$_4^+$$_{out}$, and N-NO$_3^-$$_{out}$ was obtained. This was due to the different OLR used in each experimental condition, leading to a diverse feed stream composition in terms of carbon and/or nitrogen. In spite of the highest correlation coefficient (R^2 of 0.93) obtained for N-NO$_3^-$$_{out}$, a very high RMSEP of 19.19 mg L^{-1} was achieved. This was due to the fact that the global analysis was strongly dependent of the condition presenting the largest N-NO$_3^-$ range (EC1). Regarding PLS1, the lowest
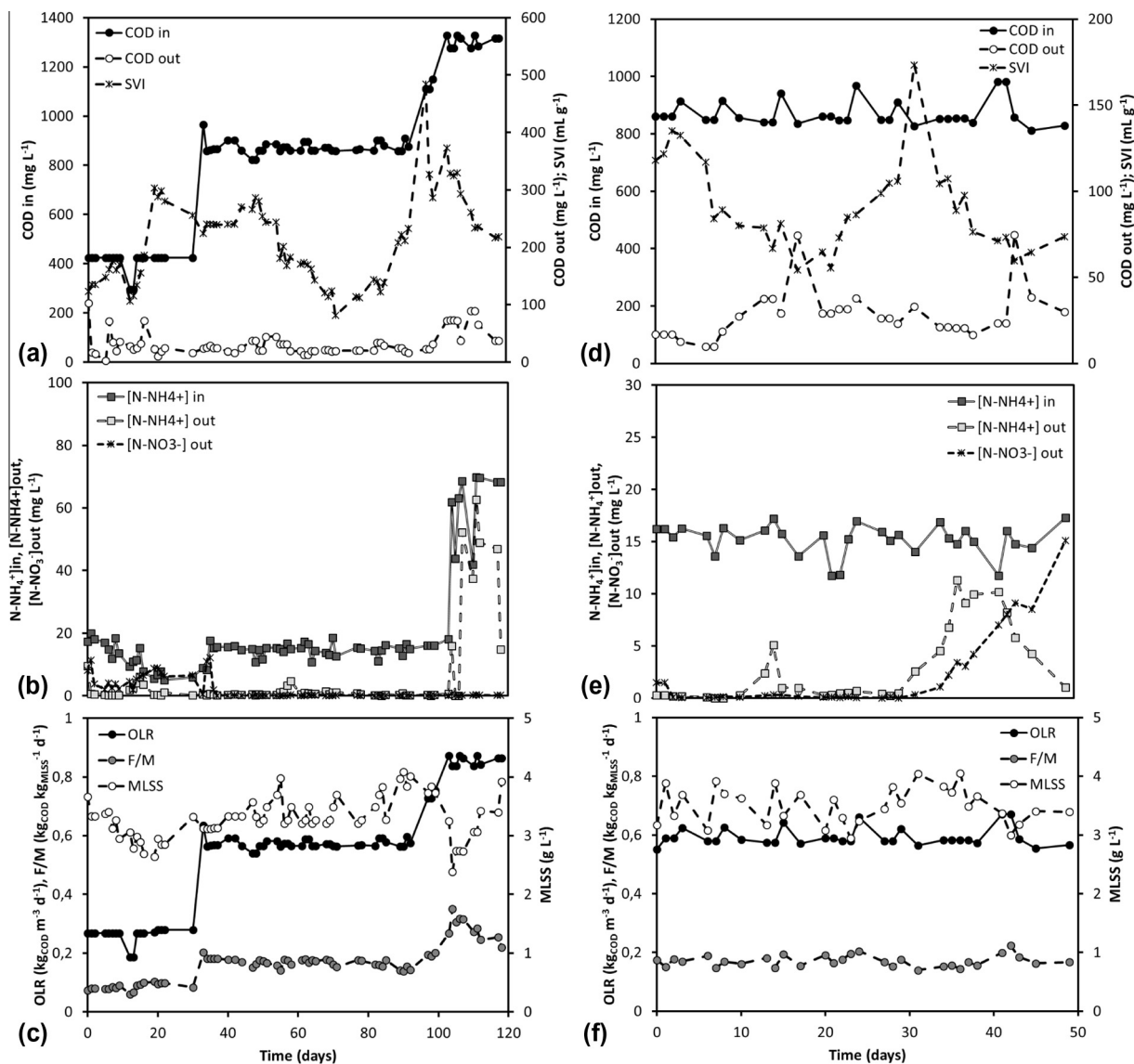
Fig. 4. OLR, F/M, MLSS, COD_{in}, COD_{out}, $N-NH_{4\,in}^{+}$, $N-NH_{4\,out}^{+}$, and $N-NO_{3\,out}^{-}$ (operational data) behavior for EC3 (a–c) and for EC4 (d–f).

Table 2
PLS prediction results for the effluent COD, $N-NH_4^+$, and $N-NO_3^-$ for the ensemble of the data obtained by analyzing the four experimental conditions as a whole (PLS1), and for the ensemble of the individually obtained results for each EC (PLS2).

		PLS1	PLS2
COD_{out} $(mg\ L^{-1})$	RMSEP $(mg\ L^{-1})$	14.81	12.48
	R^2 overall (training + validation) set	0.70	0.81
	Linear regression (training + validation set)	$y = 0.98x$	$y = 0.97x$
$N-NH_{4\,out}^{+}$ $(mg\ L^{-1})$	RMSEP $(mg\ L^{-1})$	4.61	5.44
	R^2 overall (training + validation) set	0.91	0.89
	Linear regression (training + validation set)	$y = 0.95x$	$y = 1.00x$
$N-NO_{3\,out}^{-}$ $(mg\ L^{-1})$	RMSEP $(mg\ L^{-1})$	19.19	9.61
	R^2 overall (training + validation) set	0.93	0.98
	Linear regression (training + validation set)	$y = 0.97x$	$y = 0.97x$

RMSEP (4.61 mg L^{-1}) and an acceptable correlation coefficient (R^2 of 0.91) was obtained for $N-NH_{4out}^{+}$.

PLS2 results for the effluent COD, $N-NH_4^+$, and $N-NO_3^-$ prediction are also represented in Table 2. This model was obtained using solely the selected variables for each of the individual conditions (22 for COD_{out}, 19 for $N-NH_{4out}^{+}$, and 17 for $N-NO_{3out}^{-}$). A clear reduction of the RMSEP (12.48 mg L^{-1}) was obtained regarding the COD_{out} prediction ability, increasing also the correlation coefficient (R^2 of 0.81). It was also observed higher prediction ability for $N-NO_{3out}^{-}$, presenting a higher correlation coefficient (R^2 of 0.98) and a lower RMSEP of 9.61 when compared to the PLS1 model. Regarding the effluent $N-NH_4^+$, slightly higher RMSEP (5.44 mg L^{-1}) and lower correlation coefficient (R^2 of 0.89) were achieved, and therefore using each EC individually for the PLS analysis could not be proven to be the best strategy in this case. However, and given the similitude of the obtained results for PLS1 and PLS2 for the $N-NH_{4out}^{+}$ prediction, both strategies may be considered in equal terms to that effect.

Overall, the obtained results emphasize the usefulness of dividing the data into the four experimental conditions for individual PLS analysis over a global PLS analysis. It is clear that better prediction abilities could be found by PLS2 methodology and mainly for

the effluent COD and N-NO$_3^-$. This fact is related to morphological and physiological conditions affecting each experimental condition resulting in different factors influencing the key parameters for each experimental condition. As a result, modeling each experimental condition individually is able to take into account the different dependencies in much more detail than a single PLS analysis.

4. Conclusions

Using PLS models, key parameters of an activated sludge system such as effluent COD, N-NH$_4^+$, and N-NO$_3^-$ were estimated. Regarding the PLS analyses performed, it was concluded that combining the biomass morphological and physiological data allowed, quite reasonably, predicting the effluent COD, N-NH$_4^+$, and N-NO$_3^-$. In conclusion, these results sustain the use of the current QIA procedures in quality assessment of activated sludge systems, and namely on the plant performance, thus emphasizing their ability to be used in process monitoring and control. On the other hand, it should be noted that the implementation of the proposed methodology in real WWTP is subjected to some constraints and that the effluent and operational (bulking/non-bulking, diurnal, and seasonal) variations in time should be taken into account. Indeed, the general acceptance of the proposed methodology for widespread plant performance monitoring still lacks more studies in a larger set of real WWTP operating in a broad diversity of conditions.

References

[1] J.W. Einax, H.W. Zwanziger, S. Geiss, Chemometrics in Environmental Analysis, VCH, Weinheim, Germany, 1997.
[2] R. Manser, K. Muche, W. Gujer, H. Siegrist, A rapid method to quantify nitrifiers in activated sludge, Water Res. 39 (2005) 1585–1593.
[3] J. Comas, I. Rodríguez-Roda, K.V. Gernaey, C. Rosen, U. Jeppsson, M. Poch, Risk assessment modelling of microbiology-related solids separation problems in activated sludge systems, Environ. Model. Software 23 (2008) 1250–1261.
[4] L. Belanche, J.J. Valdés, J. Comas, I.R. Roda, M. Poch, Prediction of the bulking phenomenon in wastewater treatment plants, Artif. Intell. Eng. 14 (2000) 307–317.
[5] X. Flores-Alsina, J. Comas, I. Rodriguez-Roda, K.V. Gernaey, C. Rosen, Including the effects of filamentous bulking sludge during the simulation of wastewater treatment plants using a risk assessment model, Water Res. 43 (2009) 4527–4538.
[6] M.S. Moussa, C.M. Hooijmans, H.J. Lubberding, H.J. Gijzen, M.C.M. van Loosdrecht, Modelling nitrification, heterotrophic growth and predation in activated sludge, Water Res. 39 (2005) 5080–5098.
[7] E. Liwarska-Bizukojc, Application of image analysis techniques in activated sludge wastewater Treatment processes, Biotechnol. Lett. 27 (2005) 1427–1433.
[8] E. Liwarska-Bizukojc, M. Bizukojc, O. Andrzejczak, Validation of a new image analysis procedure for quantifying filamentous bacteria in activated sludge, Water Sci. Technol. 70 (2014) 955–963.
[9] M. da Motta, M.N. Pons, N. Roche, H. Vivier, Characterisation of activated sludge by automated image analysis, Biochem. Eng. J. 9 (2001) 165–173.
[10] R. Jenné, E.N. Banadda, I.Y. Smets, A. Bamelis, L. Verdickt, J.F. Van Impe, Activated sludge image analysis system: monitoring settleability and effluent clarity, Water Sci. Technol. 52 (2005) 193–199.
[11] A.L. Amaral, E.C. Ferreira, Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression, Anal. Chim. Acta 544 (2005) 246–253.
[12] D.P. Mesquita, O. Dias, A.M.A. Dias, A.L. Amaral, E.C. Ferreira, Correlation between sludge settleability and image analysis information using partial least squares, Anal. Chim. Acta 642 (2009) 94–101.
[13] D.P. Mesquita, A.L. Amaral, E.C. Ferreira, Characterization of activated sludge abnormalities by image analysis and chemometric techniques, Anal. Chim. Acta 705 (2011) 235–242.
[14] D.P. Mesquita, A.L. Amaral, E.C. Ferreira, Identifying different types of bulking in an activated sludge system through quantitative image analysis, Chemosphere 85 (2011) 643–652.
[15] A.L. Amaral, D.P. Mesquita, E.C. Ferreira, Automatic identification of activated sludge disturbances and assessment of operational parameters, Chemosphere 91 (2013) 705–710.
[16] D.P. Mesquita, C. Leal, J.R. Cunha, A. Oehmen, A.L. Amaral, M.A.M. Reis, E.C. Ferreira, Prediction of intracellular storage polymers using quantitative image analysis in enhanced biological phosphorus removal systems, Anal. Chim. Acta 770 (2013) 36–44.
[17] D.S. Lee, P.A. Vanrolleghem, Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis, Biotechnol. Bioeng. 82 (2003) 489–497.
[18] C.E. Miller, Chemometrics in process analytical chemistry, in: A. Bakeev (Ed.), Process Analytical Technology, Blackwell Publishing, Oxford, UK, 2005.
[19] C. Rosen, A chemometric approach to process monitoring and control with applications to wastewater treatment operation (Ph.D thesis), Lund Institute of Technology, Sweden, 2001.
[20] M.C. Sarraguça, A. Paulo, M.M. Alves, A.M.A. Dias, J.A. Lopes, E.C. Ferreira, Quantitative monitoring of an activated sludge reactor using on-line UV–visible and near-infrared spectroscopy, Anal. Bioanal. Chem. 395 (2009) 1159–1166.
[21] A.M.A. Dias, I. Moita, R. Páscoa, M.M. Alves, J.A. Lopes, E.C. Ferreira, Activated sludge process monitoring through in situ near-infrared spectral analysis, Water Sci. Technol. 57 (10) (2008) 1643–1650.
[22] N.D. Lourenço, F. Paixão, M.H. Pinheiro, A. Sousa, Use of spectra in the visible and near-mid-ultraviolet range with principal component analysis and partial least squares processing for monitoring of suspended solids in municipal wastewater treatment plants, Appl. Spectrosc. 64 (2010) 1061–1067.
[23] G. Tchobanoglous, F.L. Burton, H.D. Stensel, Metcalf and Eddy Wastewater engineering – treatment and reuse, fourth ed., McGraw-Hill, New York, 2003.
[24] B.Y.P. Moy, J.H. Tay, S.K. Toh, T.L. Tay, High organic loading influences the physical characteristics of aerobic sludge granules, Lett. Appl. Microbiol. 34 (2002) 407–412.
[25] APHA, AWWA, WPCF, Standard Methods for the Examination of Water and Wastewater, American Public Health Association, Washington, DC, 1989.
[26] M.N. Pons, H. Vivier, Biomass quantification by image analysis, Adv. Biochem. Eng./Biotechnol. 66 (2000) 133–184.
[27] P. Teppola, S.P. Mujunen, P. Minkkinen, Partial least squares modeling of an activated sludge plant: a case study, Chemometr. Intell. Lab. 38 (1997) 197–208.
[28] S.H. Woo, C.O. Jeon, Y.S. Yun, H. Choi, C.S. Lee, D.S. Lee, On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant, J. Hazard. Mater. 161 (2009) 538–544.
[29] K.P. Singh, N. Basant, A. Malik, G. Jain, Modeling the performance of "up-flow anaerobic sludge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches – a case study, Anal. Chim. Acta 658 (2010) 1–11.
[30] D. Jenkins, M.G. Richard, G. Daigger, Manual on the Causes and Control of Activated Sludge Bulking, Foaming and other Solids Separation Problems, Lewis Publishing, Boca Raton, FL, 2003.