Quantitative image analysis as a robust tool to assess effluent quality from an aerobic granular sludge system treating industrial wastewater

Joana G. Costa[1], Ana M.S. Paulo[2], Catarina L. Amorim[2], A. Luís Amaral[1,3,4], Paula M.L. Castro[2], Eugénio C. Ferreira[1], Daniela P. Mesquita[1*]

[1]CEB - Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[2]Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua Diogo Botelho 1327, 4169-005 Porto, Portugal

[3]Instituto Politécnico de Coimbra, ISEC, DEQB, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

[4]Instituto de Investigação Aplicada, Laboratório SiSus, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

*Corresponding Author: Daniela P. Mesquita, daniela@deb.uminho.pt

CEB - Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Abstract

Quantitative image analysis (QIA) is a simple and automated method for process monitoring, complementary to chemical analysis, that when coupled to mathematical modelling allows associating changes in the biomass to several operational parameters. The majority of the research regarding the use of QIA has been carried out using synthetic wastewater and applied to activated sludge systems, while there is still a lack of knowledge regarding the application of QIA in the monitoring of aerobic granular sludge (AGS) systems. In this work, chemical oxygen demand (COD), ammonium ($N\text{-}NH_4^+$), nitrite ($N\text{-}NO_2^-$), nitrate ($N\text{-}NO_3^-$), salinity (Cl^-), and total suspended solids (TSS) levels present in the effluent of an AGS system treating fish canning wastewater were successfully associated to QIA data, from both suspended and granular biomass fractions by partial least squares models. The correlation between physical-chemical parameters and QIA data allowed obtaining good assessment results for COD (R^2 of 0.94), $N\text{-}NH_4^+$ (R^2 of 0.98), $N\text{-}NO_2^-$ (R^2 of 0.96), $N\text{-}NO_3^-$ (R^2 of 0.95), Cl^- (R^2 of 0.98), and TSS (R^2 of 0.94). While the COD and $N\text{-}NO_2^-$ assessment models were mostly correlated to the granular fraction QIA data, the suspended fraction was highly relevant for $N\text{-}NH_4^+$ assessment. The $N\text{-}NO_3^-$, Cl^- and TSS assessment benefited from the use of both biomass fractions (suspended and granular) QIA data, indicating the importance of the balance between the suspended and granular fractions in AGS systems and its analysis. This study provides a complementary approach to assess effluent quality parameters which can improve wastewater treatment plants monitoring and control, with a more cost-effective and environmentally friendly procedure, while avoiding daily physical-chemical analysis.

Keywords: Aerobic granular sludge; Quantitative image analysis; Partial least squares; Salinity; Effluent quality parameters

1. Introduction

Aerobic granular sludge (AGS) is a compact and cost-effective wastewater treatment (WWT) system that has been replacing conventional activated sludge (AS) systems worldwide (Nancharaiah and Kumar, 2018). Biomass aggregation in the form of granules with good settling properties allows for a smaller process footprint. Also, the biomass growth forming spherical aggregates favours the existence of different redox microenvironments and, consequently, the simultaneous removal of organic matter and nutrients (nitrogen - N and

phosphorus - P) and also other pollutants in a single and compact system (Bengtsson et al., 2019; Nancharaiah and Kumar, 2018). In addition, different options are available to valorise the excess of produced sludge from the WWT, including the recovery of resources as exopolysaccharides (Adav et al., 2008; Nancharaiah and Kumar, 2018). Therefore, the use of AGS is being researched for the biological treatment of industrial wastewaters as dairy (Bumbac et al., 2015; Val del Río et al., 2012), fish canning (Val del Río et al., 2012), brewery (Corsino et al., 2017b), petrochemical (Caluwé et al., 2017), and textile (Lotito et al., 2014), and it has even been tested for heavy metals removal (Wei et al., 2017). The high salt content present in wastewaters produced by different industries, such as food-processing, leather, and petrochemical industries, often impairs the activity of the microbial communities within biological processes (Corsino et al., 2016; Pronk et al., 2014). However, AGS technology has already proven its ability to treat saline effluents (Paulo et al., 2021; Pronk et al., 2014; Wan et al., 2014). Nonetheless, there are specific drawbacks when treating industrial wastewater, mainly related to the structural stability of the biological granules and comprising physical modifications (Corsino et al., 2017a). The classic methods for process monitoring, based on physical-chemical methodologies, often solely detect system instabilities too late to intervene, delaying the implementation of control measures (Leal et al., 2020; Wagner et al., 2015). Quantitative image analysis (QIA) has become a very important tool to overcome the problems of the traditional methodologies, since it is a simple, quantitative and automated method that enables accurate monitoring of granular and suspended biomass and early detection of system disturbances (Amaral et al., 2013; Amaral and Ferreira, 2005; Grijspeerdt and Verstraete, 1997; Jenné et al., 2006; Leal et al., 2020; Mesquita et al., 2013, 2011). Moreover, characterizing the system based on the biomass dynamics, the key player of the process, by means of microscopy techniques coupled to QIA, provides important information about the biological process and may present, in this way, an advantage for assessment purposes, alongside mathematical modelling (Amaral et al., 2013; Mesquita et al., 2013). In fact, QIA is nowadays recognised as a valid monitoring tool to relate the sludge characteristics with its settling ability (Grijspeerdt and Verstraete, 1997), and to determine the aggregated biomass structure and filamentous bacteria contents (Amaral et al., 2013), quite important parameters in process performance monitoring. Fluctuations in the balance between granules and flocs, as well as in the biomass structure, which may affect the performance and stability of the system can be captured by QIA.

Similarly to other technological tools, QIA often leads to large amounts of gathered data, hindering its straightforward interpretation. In this way, chemometric techniques can be used as a valuable tool for data mining, optimization and development of predictive models in processes, and extraction of a maximum of information from experimental data (Grijspeerdt and Verstraete, 1997; Einax et al., 1997). A widely employed statistical technique is the partial least squares (PLS), used to assess different parameters and detect deviations (Einax et al., 1997), already successfully coupled to QIA to quantify intracellular storage polymers (Mesquita et al., 2013), estimate the sludge volume index (SVI) (Amaral and Ferreira, 2005; Mesquita et al., 2009) and total suspended solids (TSS) (Amaral and Ferreira, 2005), as well as estimate effluent chemical oxygen demand (COD), ammonia ($N\text{-}NH_4^+$), and nitrate ($N\text{-}NO_3^-$) (Mesquita et al., 2016) concentrations in AS systems. More recently, Leal and colleagues (Leal et al., 2020) used QIA-based methodology coupled to multilinear regression (MLR) to predict SVI, TSS and volatile suspended solids (VSS) of suspended and granular fractions of mature AGS. While QIA has been already applied to AS systems, with the majority of the research carried out using synthetic wastewater (Nancharaiah and Kumar, 2018), there is a lack of published studies relying on the application of QIA in monitoring AGS systems treating real wastewater.

The main objective of this study was to evaluate the feasibility of applying a QIA-based methodology, coupled to PLS, to assess effluent chemical quality in terms of COD, $N\text{-}NH_4^+$, nitrite ($N\text{-}NO_2^-$), $N\text{-}NO_3^-$, salinity (Cl^-), and TSS levels from an AGS system treating real fish canning industry wastewater, with variable composition. The treatment of wastewater with variable composition resulted in removal performance fluctuations, which increased the QIA data range and, hence, allowed to establish relationships between reactor biomass characterization and effluent chemical quality. Since the stability of AGS processes is related to the balance between suspended and granular fractions within the reactor (Wagner et al., 2015), the evaluation of both biomass fractions content was performed in order to establish the most appropriate data for assessing the effluent quality parameters. In this way, this methodology can improve the monitoring of the AGS systems and minimize the need of analytical measurements, often more costly, and prone to potentially pollutant byproducts release.

2. Materials and methods

2.1 Experimental Setup

A lab-scale sequencing batch reactor (SBR) with a working volume of 2.5 L was inoculated with granular sludge from Frielas WWTP (Lisbon, Portugal), and operated as described elsewhere (Paulo et al., 2021). Over almost 8 months (231 days), the reactor was fed with fish canning wastewater collected from an industrial facility. A total of 8 different batches of wastewater from a fish canning plant were collected and used for reactor feeding, after solids and fat removal by screening and coagulation/flotation processes performed at the plant. Brine wastewater was separated from the remaining wastewaters at the fish canning plant, decreasing the amount of salt in the final effluent.

2.2 Analytical methods

The reactor effluent was monitored for TSS, COD (soluble), $N\text{-}NH_4^+$, $N\text{-}NO_3^-$, $N\text{-}NO_2^-$ and Cl^-, used as an indirect measurement of the NaCl present in the wastewater (according to Paulo et al., 2021). Biomass was removed from the collected samples using syringe nylon membrane filters (0.45 µm pore-size), except the ones for TSS measurements. COD and TSS were analysed according to the standard methods (APHA, 1998). Photometric test kits (Spectroquant®, Merck Millipore) were used to determine the concentrations of $N\text{-}NH_4^+$, $N\text{-}NO_3^-$, $N\text{-}NO_2^-$, and Cl^- following the manufacturer's instructions.

2.3 Image acquisition, processing, and analysis

For the QIA analysis, and to obtain homogeneous and representative biomass samples, a duplicate 2 mL mixed liquor sample aliquot was collected from inside the reactor during the aeration phase, within complete mixture conditions. Then, the biomass samples were washed with phosphate-buffered saline (PBS) solution. Subsequently, biomass samples were incubated for 2 h at 4 ºC in a mixture of PBS and formaldehyde (4%) (0.25:1), washed with PBS solution and preserved in PBS and ethanol (96%) (1:1). To separate the granular and suspended biomass fractions a density separation technique was applied, taking into account the different settling times characterizing each fraction.

For flocs and filaments' characterization, suspended biomass samples were analysed in triplicate (for representative reasons) through bright-field microscopy under a total magnification of 40x. For each slide, 10 µL of sample were deposited and 50 images were acquired in the upper, middle and bottom of the slide, resulting in a total of 150 images (3 x

50 images per slide). Images were acquired with a resolution of 1360x1024 pixels and 8-bit (256 grey levels) format, using the commercial software Cell^B (Olympus, Tokyo, Japan) on an Olympus BX51 microscope (Olympus, Shinjuku, Japan) coupled to an Olympus DP72 camera (Olympus, Tokyo, Japan). This resulted in more than 3750 aggregates per sample in average.

With respect to the granules' characterization, an Olympus SZ 40 stereoscope (Olympus, Shinjuku, Japan) was used. The stereoscope was coupled to a CCD Sony AVD D5CE (Sony, Tokyo, Japan) grey video camera, and to a Data Translation DT 3155 (Data Translation, Marlboro MA, USA) frame grabber. The images of the granules present in the 2 mL samples were acquired in a Petri dish under a total magnification of 15x. Images were acquired with a resolution of 768x576 pixels and 8-bit (256 grey levels) format. This resulted in more than 246 granules per sample in average. The resulting images were then processed by two dedicated programs in Matlab 9.2 (The Mathworks, Natick MA, USA) language, previously developed by Amaral, 2003 and further adapted thereafter. The main steps of each program are already well described in Leal et al. (2021), and can be found in the supplementary information. These programs return data of granules from the granular fraction and of flocs and filamentous bacteria from the suspended biomass fraction. These contents were individually characterized in terms of the most relevant size, morphological, and content variables (Table 1). The dynamic of the process related to degranulation and granulation phenomena throughout the reactor operation (Figure S1) provided the diversity of data needed for the establishment of the models. A more detailed definition of the variables can be found elsewhere (Amaral, 2003; Mesquita et al., 2009).

The selected variables have been proved to be useful for monitoring settling dysfunctions or previously associated to the effluent parameters in analysis, as COD removal and nitrification associated to granules disintegration (Luo et al., 2014; Paulo et al., 2021) or even the salinity and TSS associated with different descriptors from both fractions (Amaral and Ferreira, 2005; Corsino et al., 2018, 2017a; He et al., 2017; Leal et al., 2020; Li and Wang, 2008; Wan et al., 2014). This way, QIA data from both suspended and granular fractions were used to assess the different parameters. It should be noticed though that the granular or floccular biomass fractions were not considered, since the purpose of the present work was to assess the effluent quality parameters (comprising also the total TSS), using solely QIA data. Moreover,

179 the general dynamic between both fractions in AGS systems provided a wider range of QIA
180 results contributing to the prediction ability.

181 Furthermore, the flocs from the suspended biomass fraction were divided in 3 classes
182 according to their equivalent diameter (D_{eq}): small flocs (D_{eq} < 25 µm); intermediate flocs (25
183 µm < D_{eq} < 250 µm); and large flocs (D_{eq} ≥ 250 µm); while the granules from the granular
184 biomass fraction were divided in small (D_{eq} < 0.25 mm), intermediate (0.25 mm < D_{eq} < 2.5
185 mm), and large granules (D_{eq} ≥ 2.5 mm). The variables regarding morphology descriptors for
186 the small flocs class were not assessed since, at the employed magnification, the size of the
187 objects does not allow to distinguish with precision its shape. Also, filamentous bacteria were
188 only characterized in terms of content since the magnification used in the analysis of the
189 suspended fraction does not allow to distinguish different filamentous typologies.

190 Table 1–QIA variables collected from each biomass fraction

Main descriptor group	Parameter	Description		Formula	Fraction
Free filamentous bacteria contents	TL/Vol (mm/µL)	Total filaments length per volume		n.a. *	Suspended biomass
	Nb_{fil} /Vol (μL^{-1})	Number of filaments per volume		n.a.	
	Nb_{int}/Vol (μL^{-1})	Number of filaments intersections per volume		n.a.	
	TL/TA (mm/mm^2)	Total filaments length per total area of flocs		TL/TA=$\frac{TL}{TA}$	
Aggregates size	D_{eq} (µm)	Aggregate equivalent diameter	For small, intermediate and large classes separately	$D_{eq}=2F_{Cal}\sqrt{\frac{A}{\pi}}$	Suspended and granular biomass separately
	Per (µm)	Aggregate perimeter		P=$N_{Per} \times 1.1222 \times F_{Cal}$	
	Length (µm)	Aggregate length		Length=$F_{max} \times F_{Cal}$	
	Width (µm)	Aggregate width		Width=$F_{min} \times F_{Cal}$	
	Ar (μm^2)	Aggregate area		Ar=$N_{obj} \times F_{Cal}$	
	Vl (μm^3)	Granule volume		Vl = $\frac{\pi \times \text{Width}^2 \times \text{Length}}{6}$	
Aggregates morphology	FF	Aggregate form factor	For intermediate and large classes separately	FF=$\frac{P^2}{4\pi A}$	
	Conv	Aggregate convexity		Conv=$\frac{P_{Conv}}{P}$	
	Comp	Aggregate compactness		Comp =$\frac{\sqrt{\frac{4}{\pi}}A}{F_{max}}$	
	Round	Aggregate roundness		Round =$\frac{4\pi A}{P_{Conv}{}^2}$	
	Sol	Aggregate solidity		Sol =$\frac{A}{A_{Conv}}$	

	Ext	Aggregate extent		$\text{Ext}=\frac{A}{W_{BB}\times L_{BB}}$
	Ecc	Aggregate eccentricity		$\text{Ecc}=\frac{(4\pi^2)(M_{2X}-M_{2Y})^2+4M_{2XY}{}^2}{A^2}$
	Rob	Aggregate robustness		$\text{Rob}=\frac{2er_{obj}}{\sqrt{A}}$
	LrgC	Aggregate largest concavity		$\text{LrgC}=\frac{2er_{comp}}{\sqrt{A}}$
	RelArea	Flocs ratio between hole and object area		$\text{RelArea}=\frac{A_h}{A}$
Aggregates contents	Nb/Vol (flocs - μL^{-1}; granules - mL^{-1})	Total number of aggregates per volume	For classes ensemble, and small, intermediate and large classes separately	n.a.
	TA/Vol (flocs - $mm^2/\mu L$; granules - mm^2/mL)	Total area of aggregates per volume		n.a.
	TV/Vol (mm^3/mL)	Total volume of granules per volume		n.a.
	$\%Nb_i$	Number percentage of the aggregates *i* class	For small, intermediate and large classes separately	$\%\ Nb_i=\frac{\sum_{i=1}^{Nclass} Nb_i}{T_{Nb}}$
	$\%Ar_i$	Area percentage of the aggregates *i* class		$\%\ Ar_i=\frac{\sum_{i=1}^{Nclass} Ar_i}{TA}$
	$\%Vl_i$	Volume percentage of the granules *i* class		$\%\ Vl_i=\frac{\sum_{i=1}^{Nclass} Vl_i}{TV}$

n.a.: Not applicable; *$L=(N_{Thn}+N_{Int})\times 1.1222\times F_{Cal}$;

F_{Cal} is the calibration factor (µm per pixel); N_{Per} is the pixel sum of the aggregate boundary; P is the aggregate perimeter; F_{maz} is the aggregate maximum Feret Diameter; F_{min} is the aggregate minimum Feret Diameter; N_{obj} is the aggregate pixel sum; A is the aggregate area; P_{Conv} is the aggregate Convex Envelope perimeter; A_{Conv} is the aggregate Convex Envelope area; W_{BB} is the aggregate Bounding Box width; L_{BB} is the aggregate Bounding Box length; M_{2X} and M_{2Y} are the aggregate central second moments with respect to x-axis and y-axis respectively; M_{2XY} is the aggregate second horizontal and vertical order moment; er_{obj} is the number of erosions needed to delete the aggregate; er_{comp} is the number of erosions needed to delete the complement of the aggregate in relation to its convex envelope; A_h is the aggregate holes area; i subscript is each size class; Ar_i is area of each aggregate belonging to the i size class; Nb_i is number of each aggregate belonging to the i size class; T_{Nb} is the total number of aggregates; L is the filament length; N_{Thn} is the pixel sum of each thinned filament; N_{Int} is the number of filaments intersections; 1.1222 is used in order to homogenize the different angles of filaments

2.4 Multivariate statistical analysis

PLS is a linear regression method that performs a multivariate calibration of the dependent variable (Y) from a large independent dataset (X variables) by an iterative algorithm (Teppola et al., 1997; Wold et al., 2001). Prior to the development of the PLS model, the X dataset matrix was pre-processed using the standard normal variate (SNV) method in order to avoid biasing the weights determination of the input variables. A more detailed explanation of the PLS algorithm can be found in Teppola *et al.* and Woo *et al.* (Teppola et al., 1997; Woo et al., 2009).

PLS was performed in Matlab 9.2 (The Mathworks, Natick MA, USA) using, solely, the QIA dataset (matrix X) to assess COD, $N\text{-}NH_4^+$, $N\text{-}NO_2^-$, $N\text{-}NO_3^-$, Cl^-, and TSS concentrations (matrix Y). The response matrix Y for COD, $N\text{-}NO_2^-$, $N\text{-}NO_3^-$, Cl^-, and TSS models totalled 22 samples and 21 samples for $N\text{-}NH_4^+$. Three studies were performed for each Y parameter assessment using different datasets regarding the obtained QIA results (matrix X): the QIA dataset of the suspended biomass fraction for PLS1; QIA dataset of the granular biomass fraction for PLS2; and the QIA dataset of the ensemble suspended and granular biomass fractions for PLS3. This resulted in a total of 53 variables in matrix X for PLS1, 57 variables for PLS2, and 110 variables for PLS3.

As in PLS is crucial to prevent overfitting problems (Wold et al., 2001), the maximum number of latent variables (LV) allowed for each model was set to half the number (n/2 components) of the training model observations (n). Cross-validation was used for selecting the optimal LV. The samples were randomly divided in a 67% training dataset to 33% validation dataset and repeated 5000 times to obtain a total number of 5000 randomly selected different datasets. This methodology for variables selection has been previously used and allowed reasonable assessment for the operational parameters MLSS (mixed liquor suspended solids) and SVI (Amaral et al., 2013). The correlation coefficient (R^2) between measured and assessed Y values was then determined for the overall set. To reduce the number of variables of matrix X, PLS was ran solely on the independent variables capable of influencing each parameter. The variable importance in the projection (VIP) was determined, allowing to identify the most relevant independent variables (X) for explaining matrix Y. A cross-correlation analysis (CC) between the variables with higher VIP scores was then performed, leading to the exclusion of one variable for each pair presenting a correlation factor above 0.9. In summary, for each predictive parameter the n uncorrelated variables with the higher VIP scores were selected.

The reduced dataset was further treated by PLS. Furthermore, the best model was selected considering the R^2 of the different datasets (training, validation and overall) and the root mean squared error of prediction (RMSEP) and residual prediction deviation (RPD) of the predictive model for the overall (training and validation) dataset.

3. Results and discussion

3.1 Variable reduction and VIP

In this work, partial least squares (PLS) was used to perform the analysis of the QIA data from suspended and granular biomass fractions and assess the reactor effluent quality in terms of COD, N-NH_4^+, N-NO_2^-, N-NO_3^-, Cl^-, and TSS concentrations. The most important variables regarding each parameter assessment (higher VIP values) are presented in Table 2 and Table S1. Investigating the treatment of a real wastewater with varied composition and diverse physical-chemical characterization, was found to be an important strategy to use in the PLS models. Besides the diversity of data, the possibility of testing the QIA methodology for WWT monitoring of a real effluent is highly relevant for its future application in full-scale WWTP using the AGS technology. Representative images of the granular and suspended fractions along reactor operation are presented in Figure S1.

Table 2–VIP values for the PLS regressions. Variables description is presented in Table 1, as well as the main descriptor group.

PLS	**1**		**2**		**3**	
Matrix y	**Variables used for modelling**	**VIP**	**Variables used for modelling**	**VIP**	**Variables used for modelling**	**VIP**
COD	%Nb_i_l	1.66	Vl_gs	1.63	Nb_{int}/Vol	1.36
	LrgC_l	1.42	Nb/Vol_gs	1.30	Nb/Vol_gs	1.23
	%Ar_i_s	1.22	Nb/Vol_g	1.20	Nb/Vol_g	1.15
	%Nb_i_s	1.04	Ext_gi	1.16	FF_gi	1.15
	TL/Vol	0.94	FF_gi	1.14	Width_s	1.09
	Nb/Vol_s	0.90	Per_gs	0.97	Ext_gi	1.09
	TL/TA	0.89	%Vl_gs	0.93	Per_gs	1.03
	LrgC_i	0.86	Rob_gi	0.90	Conv_gi	0.94
	Ecc_i	0.83	Conv_gl	0.88	%Nb_i_l	0.93
	Width_s	0.81	Comp_gi	0.80	Rob_gi	0.83
	Width_i	0.80	Conv_gi	0.78	Vl_gs	0.80
	TA/Vol_l	0.75	FF_gl	0.65	Comp_gi	0.77
	Conv_i	0.72	%Nb_i_gs	0.59	Conv_gl	0.74
	Nb_{int}/Vol	0.60	Rob_gl	0.39	FF_gl	0.57
N-NH_4^+	TL/TA	1.85	Comp_gi	1.28	TL/TA	1.42
	RelArea_l	1.08	Vl_gi	1.23	Ecc_gi	1.15

PLS	1		2		3	
Matrix y	Variables used for modelling	VIP	Variables used for modelling	VIP	Variables used for modelling	VIP
	Nb_{fil}/Vol	1.01	Vl_gs	1.19	Ecc_l	1.08
	Width_l	1.01	$\%Vl_i$_gl	1.18	Nb_{fil}/Vol	1.05
	Ecc_l	0.99	Length_gl	1.11	RelArea_l	1.02
	Ext_l	0.98	Conv_gi	1.08	Width_l	1.02
	Rob_l	0.98	Per_gs	1.04	LrgC_l	1.00
	Length_s	0.91	Ecc_gl	0.89	Rob_l	0.98
	Comp_i	0.90	FF_gl	0.83	$\%Ar_i$_l	0.97
	RelArea_i	0.85	Ar_gl	0.83	Conv_l	0.94
	FF_i	0.81	Conv_gl	0.82	Conv_gi	0.92
	Nb_{int}/Vol	0.81	Comp_gl	0.80	$\%Ar_i$_i	0.80
	$\%Nb_i$_i	0.77	Width_gs	0.79	D_{eq}_gs	0.74
	Length_i	0.42	$\%Nb_i$_gs	0.66	Per_gs	0.71
$N\text{-}NO_2^-$	Ar_i	1.36	$\%Nb_i$_gi	1.48	Ar_i	1.40
	Nb/Vol_l	1.36	Length_gl	1.31	Nb/Vol_gs	1.31
	$\%Nb_i$_i	1.31	LrgC_gl	1.15	$\%Nb_i$_gi	1.18
	Nb_{int}/Vol	1.20	Ecc_gl	1.09	D_{eq}_gs	1.17
	Rob_l	1.09	D_{eq}_gs	1.04	Conv_gl	1.08
	%Ar_l	1.06	Rob_gl	1.01	LrgC_gl	1.07
	TL/Vol	1.03	Nb/Vol_gl	1.01	$\%Ar_i$_gl	1.03
	TL/TA	0.86	Conv_gl	0.96	Nb/Vol_gl	1.01
	Ecc_l	0.79	Ext_gl	0.92	Rob_gl	0.97
	$\%Nb_i$_l	0.77	Vl_gl	0.88	FF_gl	0.94
	TA/Vol_l	0.73	FF_gl	0.87	$\%Nb_i$_l	0.74
	$\%Ar_i$_s	0.70	$\%Ar_i$_gl	0.76	Width_gs	0.66
	Nb/Vol_s	0.65	Comp_gl	0.64	RelArea_l	0.48
	Nb/Vol	0.62	Width_gl	0.38	Per_gs	0.34
$N\text{-}NO_3^-$	Width_s	1.96	Nb/Vol_g	1.45	Length_s	1.58
	TL/TA	1.62	Length_gs	1.40	TL/Vol	1.31
	TL/Vol	1.25	Vl_gs	1.16	Nb/Vol_gi	1.29
	$\%Nb_i$_l	1.19	TV/Vol_gl	1.15	Nb_{int}/Vol	1.09
	RelArea_i	0.99	$\%Nb_i$_gl	1.04	Conv_gl	1.02
	Nb/Vol_i	0.99	Conv_gl	1.03	$\%Vl_i$_gl	0.99
	Nb_{int}/Vol	0.75	Ext_gi	0.99	RelArea_i	0.93
	Nb/Vol_s	0.66	$\%Ar_i$_gl	0.95	Vl_gs	0.89
	Ar_l	0.64	Ecc_gl	0.83	TA/Vol_gi	0.85
	$\%Ar_i$_l	0.60	Comp_gi	0.78	TV/Vol_gl	0.85
	RelArea_l	0.60	LrgC_gl	0.75	$\%Nb_i$_gl	0.83
	Conv_l	0.50	Conv_gi	0.70	Conv_gi	0.68
	Ecc_i	0.37	FF_gi	0.69	Nb/Vol_gs	0.67
	$\%Ar_i$_i	0.29	D_{eq}_gs	0.63	FF_gi	0.39
Cl^-	D_{eq}_i	2.06	Ecc_gl	1.33	TA/Vol_gi	1.56
	$\%Nb_i$_l	1.19	Rob_gl	1.29	Ar_i	1.52
	Rob_l	1.19	$\%Vl_i$_gl	1.19	TL/TA	1.06
	Nb/Vol_s	1.02	Conv_gl	1.18	$\%Vl_i$_gl	1.04
	$\%Ar_i$_s	1.01	Comp_gl	1.14	TA/Vol_gs	0.92
	TL/Vol	0.92	Width_gl	1.07	$\%Nb_i$_gl	0.92
	$\%Nb_i$_s	0.87	$\%Nb_i$_gl	0.93	TV/Vol_gl	0.89
	Width_s	0.86	Vl_gl	0.85	$\%Nb_i$_l	0.86
	TA/Vol_l	0.82	TV/Vol_gl	0.84	Nb/Vol_gi	0.84

PLS	1		2		3	
Matrix y	Variables used for modelling	VIP	Variables used for modelling	VIP	Variables used for modelling	VIP
TSS	Nb_{int}/Vol	0.74	Per_gs	0.80	Conv_gl	0.83
	TL/TA	0.68	Vl_gs	0.78	Ecc_l	0.82
	Ecc_i	0.61	TA/Vol_gi	0.78	Nb_{fil}/Vol	0.78
	Per_l	0.51	LrgC_gl	0.78	Rob_l	0.76
	Nb/Vol_l	0.48	Ext_gl	0.76	Vl_gs	0.74
	Width_i	1.81	$\%Ar_i$_gs	1.53	Length_gs	1.80
	TL/TA	1.54	Per_gs	1.27	TL/TA	1.38
	Conv_i	1.31	Rob_gi	1.17	TV/Vol_g	1.35
	TA/Vol_i	1.08	Width_gs	1.15	TA/Vol_g	1.30
	$\%Nb_i$_s	1.01	Nb/Vol_gi	1.11	Nb/Vol_gi	1.07
	Ecc_i	0.97	$\%Nb_i$_gs	1.03	Nb_{int}/Vol	0.83
	Rob_l	0.87	$\%Vl_i$_gi	0.94	$\%Ar_i$_l	0.79
	Length_l	0.86	Vl_gs	0.86	TL/Vol	0.75
	$\%Nb_i$_l	0.75	Ecc_gi	0.85	RelArea_l	0.69
	RelArea_l	0.69	Nb/Vol_gl	0.83	$\%Nb_i$_gi	0.69
	FF_l	0.53	TA/Vol_g	0.79	Rob_i	0.65
	Nb_{fil}/Vol	0.52	Ext_gl	0.78	Rob_gi	0.64
	Conv_l	0.51	Width_gl	0.65	Vl_gs	0.61
	RelArea_i	0.42	TV/Vol_g	0.58	Ext_gi	0.41

_l – large flocs; _i – intermediate flocs; _s – small flocs; _gl – large granules; _gi – intermediate granules; _gs – small granules

3.1.1 COD assessment

In what concerns to COD assessment (Table 2), 3 variables related to the filamentous bacteria were considered for the PLS1 model, however presenting a VIP below 1. On the other hand, 4 variables stand out (VIP above 1) related to the small and large flocs size and large flocs morphological descriptors. In PLS2, the main variables for modelling the COD were related with the granules content (total number of granules and small granules), size of small granules and morphology of intermediate granules. It is also possible to observe that all variables used for the COD assessment in PLS3 were common to PLS1 and PLS2. The main variables used for this model are related to the content of filamentous bacteria, small and total granules content, intermediate granules morphology, and, at last, small flocs and granules size. It should be stressed that in PLS3 there is a clear prevalence of the granular fraction variables (11 out of 14 variables) when compared to the variables related to the suspended fraction.

The importance of the granular fraction for the assessment of COD may be related to the influence of this parameter in the granulation and disintegration processes (Nancharaiah and Kumar, 2018), since these depend on the availability of substrate for bacterial growth and EPS production required for aggregation. A higher organic strength wastewater resulted in a

lower COD removal, leading to granules instability and consequent disintegration (Paulo et al., 2021). This possibility is supported by the fact that the most important descriptors are related to the number and size of small granules, and the morphology of intermediate granules. Moreover, the high COD content in the wastewater may have promoted filamentous bacteria proliferation, thus affecting the granules' stability (Paulo et al., 2021). In accordance, the number of filaments intersections fluctuates (Figure S2), contributing to the COD assessment. The close relationship between COD in the final effluent and biomass dynamics, best captured using data on smaller sized granules, is possibly due to the fact that both are directly dependent on the COD present in the wastewater. Wastewater strength will affect not only the balance of the microbial community and granules stability, but also the COD removal performance. Therefore, although COD correlates better with QIA data on the granular fraction (PLS2), the quantification of filamentous bacteria also contributes for assessing COD in the outlet.

3.1.2 N-NH_4^+ assessment

Relatively to the N-NH_4^+ assessment, 4 variables stand out in the PLS1 model, being 2 of them related to the filamentous bacteria content and the remaining to the morphology and size of large flocs. For PLS2, the main variables for modelling the N-NH_4^+ are morphological descriptors of intermediate granules and size descriptors of the three different classes. Regarding PLS3, 10 out of 14 variables are related to the suspended fraction, in particular to the filamentous bacteria content, large flocs morphological and size descriptors, and intermediate granules morphology. Between the different PLS, 8 common variables arise (out of a total of 14): the filamentous bacteria content, large flocs morphology and size, the morphology of intermediate granules and size of small granules. Furthermore, the highest VIP variable, for both PLS1 and PLS3, was related to the total length of filaments per aggregates total area, revealing its importance for the assessment of N-NH_4^+.

In fact, previous studies showed that the nitrification efficiency can be related to the disintegration of the granules (Luo et al., 2014), also associated to the increase of filamentous bacteria when the COD to nitrogen (COD/TN) ratio in the feeding wastewater decreases, or during periods of higher organic and nitrogen loading rates (OLR and NLR) (Paulo et al., 2021). In this work, it was possible to observe a high OLR and NLR along operation (Figure 1). These conditions could have allowed for the growth of suspended biomass with heterotrophic

bacteria, that outcompete for oxygen, required for nitrification, and also play a role on the destabilization of the granular biomass (Paulo et al., 2021; Winkler et al., 2018). Despite the nitrification process can be affected by the disintegration phenomena, the system performance was able to recover, since the formation of micro-granules proved to be beneficial for the nitrification process (Luo et al., 2014; Paulo et al., 2021).

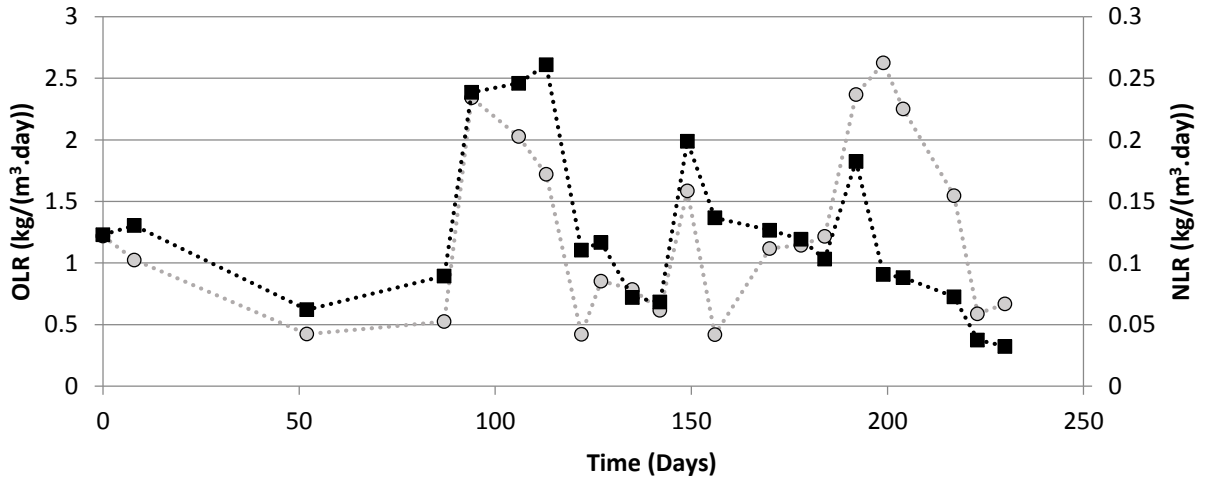


Figure 1 - Variation of OLR ( ) and NLR ( ) in the feeding wastewater

3.1.3 N-NO_2^- and N-NO_3^- assessment

In what concerns the N-NO_2^- assessment, the PLS1 model (Table 2) also encompasses the use of variables reflecting the content of filamentous bacteria. The remaining variables stress out mostly the large flocs content, size and morphological descriptors and intermediate flocs size. For the PLS2, most of the main variables concern to large granules and are, in its majority, morphological descriptors. The variables influencing this model also comprise other granules size related variables and large granules content. Regarding PLS3 there is a prevalence of the granular fraction variables (11 out 14 variables) opposite to the absence of descriptors related to the filamentous bacteria. However, it should be noticed that the parameter with higher VIP, for both PLS1 and PLS3, are related to the suspended fraction, namely the average area of intermediate flocs, revealing its importance for the assessment of N-NO_2^-. The remaining main variables are mostly related to the large granules morphology, size of the different granules classes and content descriptors of small and large granules. Furthermore, the variables used in PLS3 largely coincide with the ones used in PLS1 and PLS2 (10 out of 14).

The main variables considered in PLS1 for the $N\text{-}NO_3^-$ assessment include the filamentous bacteria content and size descriptors for small and large flocs. Similar to the PLS1, the PLS2 variables are mostly related to small and large granules size, although presenting also granules content (total number and large granules) and large granules morphological descriptors. Moreover, the variables used for the PLS3 model are mostly associated to the granules content, size and morphology (10 out of 14 variables). However, variables linked to the filamentous bacteria contents were also found important for this model. Furthermore, 9 out of the 14 variables used in PLS3 coincide with the ones used in PLS1 and PLS2.

The growth of filamentous (Figure 2) and other heterotrophic bacteria that compete for oxygen and nutrients with slow-growing bacteria, can lead to a change in granules microbial composition (Paulo et al., 2021). This effect can be more pronounced in larger granules, due to oxygen and nutrients diffusion limitation, promoting their disintegration, which can result in biomass washout of slow-growing bacteria as is the case of ammonium-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB). However, when compared to AOB, NOB presented a higher sensibility to operational changes, namely to increased carbon and free ammonia concentrations, possibly associated to lower oxygen availability (Paulo et al., 2021). Changes in these parameters might have induced changes in the suspended and granular biomass (Figure S1). Although QIA data from the granular fraction, more specifically the contents and size descriptors of large granules related variables, could provide the most important information for both $N\text{-}NO_2^-$ and $N\text{-}NO_3^-$ assessment, NOB activity was also found to be related with the $N\text{-}NH_4^+$ amount present in the process (Paulo et al., 2021).

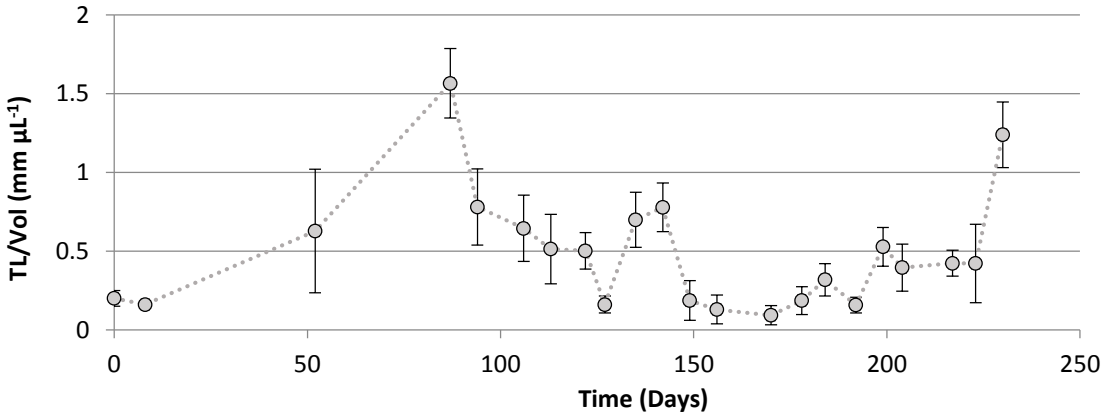


Figure 2 - Variation of the total filaments length per volume during reactor operation

3.1.4 Cl^- assessment

In the Cl^- assessment, the main variables in PLS1 were mostly size descriptors from the different flocs classes, small flocs content and large flocs morphological descriptors. Regarding PLS2, the key variables concern mostly to morphological descriptors and then to size descriptors of large granules. In PLS3, it was possible to observe a somewhat balanced use of variables from both suspended and granular fractions (8 out of 14 variables regard the granular fraction). Despite 8 out of 14 variables in PLS3 coinciding with PLS1 and PLS2, there is a shift in the typology of the granular fraction descriptors, which were more related with content of intermediate granules and size of large granules, instead of morphological descriptors. Moreover, the filamentous bacteria content and intermediate flocs size were also considered for the final PLS3 model.

Some authors have already described that the high salinity in wastewater affects the granules compactness and robustness (Li and Wang, 2008; Wan et al., 2014), as well as their stability, leading to shifts in their size and structure regularity and a consequent increase of the flocculent sludge due to the granules breakage (Corsino et al., 2018, 2017a). Moreover, salinity can inhibit the growth of filamentous bacteria and influence the size of small flocs (Corsino et al., 2018; He et al., 2017). These observations are in accordance with the results obtained in this study, where it is possible to observe the influence of the different descriptors of the granules in PLS2, namely the eccentricity, robustness, convexity, compactness, width, and volume percentage, and the importance of the size and contents of the suspended and granular biomass for the assessment ability of the PLS3 model, possibly related to the granules breakage. It should be noticed that the salt variations did not cause any detrimental effect towards the performance of biological removal processes (Paulo et al., 2021). However QIA coupled to PLS allowed the analysis of changes in biomass and to establish a relationship between Cl^- and other parameters. This indicates that QIA methodology might be very useful for investigating the effect of a specific compound on the AGS process, adding information which is not retrieved through performance results.

3.1.5 TSS assessment

Finally, for the TSS assessment, PLS1 used mostly the intermediate and small flocs size and intermediate flocs morphological descriptors. However, the filamentous bacteria content was

also considered. As in PLS1, PLS2 considers mostly size descriptors, specifically of small granules, and intermediate granules morphological and content descriptors. Regarding the model with the overall variables, the TSS assessment is somewhat balanced between the suspended and granular fractions (8 out of 14 variables concern to the granular fraction), including, as in PLS1, two variables regarding the filamentous bacteria contents. The variables with higher VIP for this model are related to the granular fraction, namely size of small granules and intermediate and total granules content descriptors, comprising the filamentous bacteria contents, as well. It is possible to observe that half of the variables are common to the different PLS used for assessing TSS, including the total and intermediate granules contents, small granule size, intermediate granule and large flocs morphology and the filamentous bacteria contents.

Previous studies assessing the TSS also indicated the importance of the filamentous bacteria and flocs content (total filaments length per total area of flocs and total area of aggregates per volume) and flocs morphology (convexity) in aerobic sludge (Amaral and Ferreira, 2005). Moreover, in an AGS system, the importance of flocs content (total area of aggregates per volume) and morphology (eccentricity) regarding the suspended biomass fraction, and granules content (volume percentage of large granules and total volume of intermediate granules) and morphology (robustness of intermediate granules), regarding the granular fraction, was previously reported (Leal et al., 2020). This is in accordance with the results of the present study where the filamentous bacteria content (total filaments length per total area of flocs) presents the second highest VIP value for the PLS1 and PLS3 models, the flocs morphology (eccentricity) is included in the final variables used in the PLS1 model and the granules content (total area and volume) and granules morphology (robustness of intermediate granules) are included in both PLS2 and PLS3.

In summary, the performed analyses indicated the most important descriptors to assess different effluent quality parameters associated to AGS process performance, while treating a real wastewater. Besides, it was also possible to understand which fraction gave the most information about the predictive parameter, when performing the PLS using QIA data from the ensemble suspended and granular fractions. While the VIP from the granular fraction QIA data was higher for COD and $N\text{-}NO_2^-$ assessment, the $N\text{-}NO_3^-$, Cl^- and TSS assessment benefited from the use of the suspended fraction data for modelling as well. On the other

412 hand, the suspended fraction gave the most valuable information for the N-NH_4^+ assessment.
413 Overall, filamentous bacteria content, which has been already used as indicator of process
414 stability (Amaral et al., 2013), has contributed for assessing different parameters, namely N-
415 NH_4^+, N-NO_3^-, Cl^- and TSS.

416

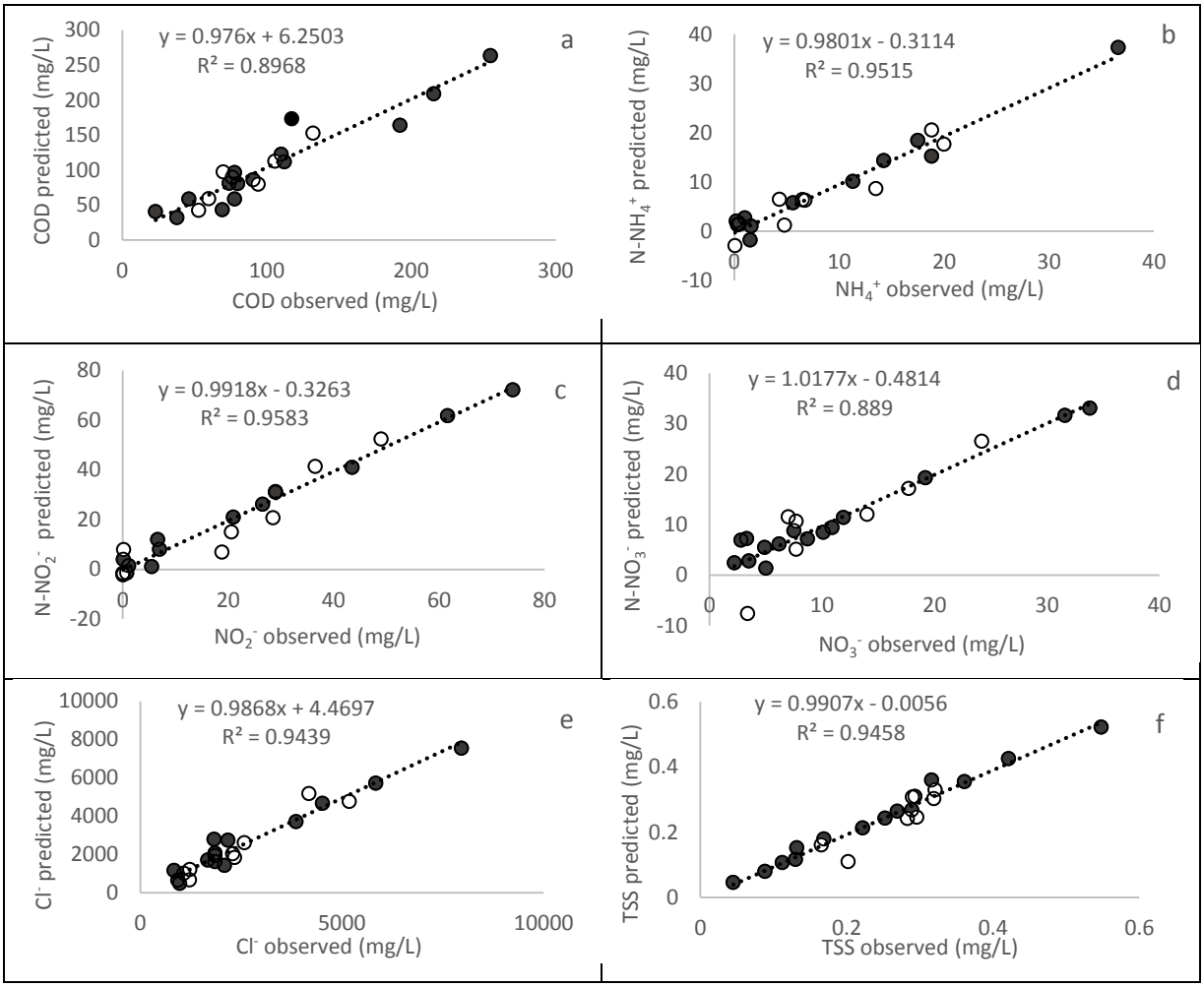
417 **3.2 Assessment ability**

418 The PLS results (cumulative fraction of the Y explained by the components R2Y(cum)), LVs,
419 linear regression equations and respective regression coefficients (R^2), RMSEP and RPD, as
420 well as the limits of applicability of the models (minimum and maximum) and the average
421 values for each studied parameter are depicted in Table 3. The regression analysis presented
422 reasonable to good assessment abilities (overall R^2 values above 0.80) for all the models
423 (Figure 3 a-f, Figure 4 a-f and Figure 5 a-f). However, R^2 values below or around 0.80 were
424 obtained for the PLS1 validation set in COD, N-NO_3^-, and TSS assessment (0.80, 0.79, and 0.80,
425 respectively) and for PLS2 models in N-NH_4^+ and TSS assessment (0.80 and 0.75, respectively).
426 This is in accordance with the VIP results, since the COD assessment model is mostly described
427 by the granular fraction, the N-NH_4^+ is mostly described by the suspended fraction and the N-
428 NO_3^- and TSS need variables from both fractions for its assessment when considering the
429 overall QIA data. Moreover, it should be noticed that for PLS3 (ensemble suspended and
430 granular biomass QIA data) the lowest correlation value achieved was of 0.88 for the
431 validation set when assessing N-NO_3^-, and 0.89 for the COD and N-NO_2^- assessments.
432 Considering all assessed parameters, QIA data from both suspended and granular biomass
433 might lead to better prediction results.

434

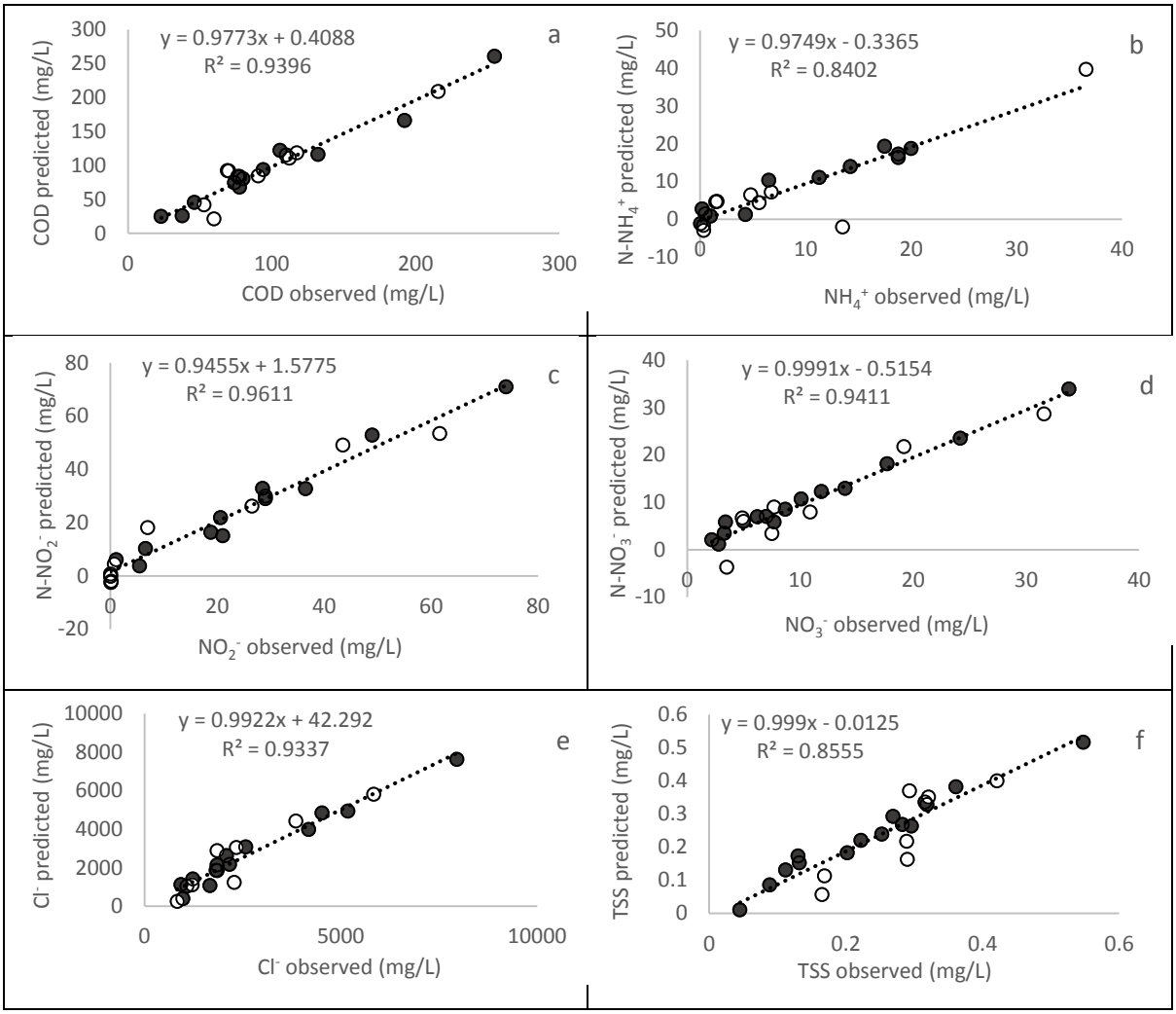
Table 3 – PLS prediction results for COD, N-NH_4^+, N-NO_2^-, N-NO_3^-, Cl^-, and TSS; root mean squared error of prediction (RMSEP) and residual prediction deviation (RPD) of the predictive model for the overall dataset for PLS1, PLS2 and PLS3; and minimum, maximum, average, and standard deviation of the observed parameters

Matrix Y	PLS	R2Y (cum)	LV	Training set		Validation set		Overall set		RMSEP (mg L^{-1})	RPD	Observed parameter (mg L^{-1})			
				Slope	R^2	Slope	R^2	Slope	R^2			Min	Max	Avg	SD
COD	**1**	0.88	7	0.95	0.95	1.42	0.80	0.98	0.90	19	2.95	23	255	99	56
	2	0.92	7	0.96	0.96	1.04	0.91	0.98	0.94	14	3.99				
	3	0.89	6	0.95	0.95	1.00	0.89	0.96	0.93	15	3.76				
N-NH_4^+	**1**	0.93	6	0.98	0.98	1.03	0.90	0.98	0.95	2.1	4.38	0.06	36.6	8.8	9.3
	2	0.82	6	0.93	0.93	1.03	0.80	0.97	0.84	4.0	2.33				
	3	0.96	7	0.93	0.99	0.99	0.96	0.97	0.98	1.3	7.20				
N-NO_2^-	**1**	0.95	7	0.99	0.99	1.01	0.86	0.99	0.96	4.4	4.80	0.04	74	20.9	21.2
	2	0.91	7	0.98	0.98	0.90	0.94	0.95	0.96	4.2	5.03				
	3	0.92	7	0.98	0.98	1.00	0.89	0.98	0.94	5.5	3.84				
N-NO_3^-	**1**	0.87	6	0.96	0.96	1.28	0.79	1.02	0.89	3.2	2.77	2.2	33.8	11.1	8.8
	2	0.94	7	0.99	0.99	1.05	0.89	1.00	0.94	2.3	3.89				
	3	0.94	6	0.98	0.98	1.02	0.88	1.00	0.95	2.1	4.19				
Cl^-	**1**	0.93	7	0.96	0.96	1.12	0.91	0.99	0.94	437	4.14	836	7960	2652	1809
	2	0.93	7	0.97	0.97	1.07	0.86	0.99	0.93	479	3.78				
	3	0.95	6	0.98	0.98	0.96	0.95	0.98	0.98	285	6.34				
TSS	**1**	0.94	7	0.98	0.98	1.21	0.80	0.99	0.95	0.03	4.04	0.05	0.55	0.25	0.11
	2	0.85	7	0.96	0.96	1.33	0.75	1.00	0.86	0.05	2.35				
	3	0.92	6	0.98	0.98	1.01	0.93	0.98	0.94	0.03	3.70				



440 Figure 3 – Relationship between the predicted and observed COD (a), N-NH4+ (b), N-NO2- (c), N-NO3-
441 (d), Cl- (e), and TSS (f) for the PLS regression with suspended fraction QIA data (PLS1). ●– Training set
442 ○– Validation set

443



444 Figure 4 - Relationship between the predicted and observed COD (a), N-NH4+ (b), N-NO2- (c), N-NO3-
445 (d), Cl- (e), and TSS (f) for the PLS regression with granular fraction QIA data (PLS2). ●– Training set ○
446 – Validation set

447

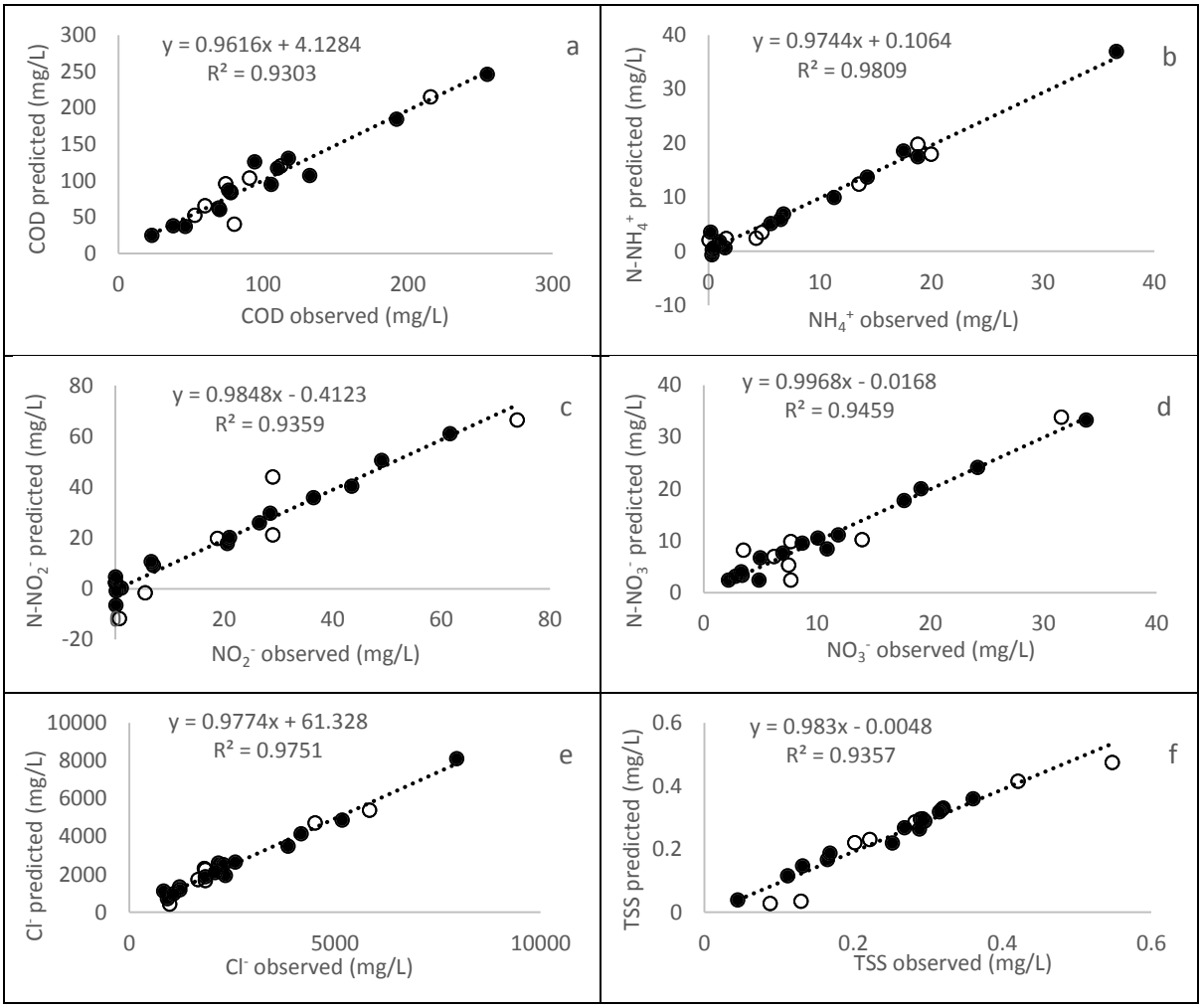


Figure 5 - Relationship between the predicted and observed COD (a), N-NH4+ (b), N-NO2- (c), N-NO3- (d), Cl- (e) and TSS (f) for the PLS regression with suspended and granular fractions QIA data (PLS3). ● – Training set ○– Validation set

When analysing the slopes of the PLS models, it is possible to observe values close to 1 for all PLS3 and PLS2 models, with the exception, in the later, of the validation set for the TSS assessment (1.33). In what concerns PLS1, the slopes for the validation set were further apart from 1 for the assessment of the majority of the parameters (slopes of 1.42, 1.28, 1.12, and 1.21 for COD, $N\text{-}NO_3^-$, Cl^-, and TSS assessment, respectively). However, even when assessing these parameters, with exception of COD and $N\text{-}NO_3^-$, the RMSEP values are less than a third of the standard deviation of the observed values, i.e. RPD above 3, indicating a good agreement (Cozzolino et al., 2004). According to this definition, good previsions were obtained for: COD and $N\text{-}NO_3^-$ with PLS2 and PLS3 models; $N\text{-}NH_4^+$ and TSS with PLS1 and PLS3 models; and for $N\text{-}NO_2^-$ and Cl^- with all the models.

Furthermore, the PLS2 presented the highest RPD for the COD and $N\text{-}NO_2^-$ assessment, while the PLS3 was the best model for $N\text{-}NH_4^+$, $N\text{-}NO_3^-$, and Cl^- assessment according to this factor. In the case of COD and $N\text{-}NO_2^-$ assessments, these are in accordance to the above VIP analysis for PLS3 (11 out of 14 variables used for the model respect to the granules, including the ones with higher VIP). In the VIP analysis for PLS3 of the Cl^- assessment it is possible to observe that 8 out of 14 variables used for the model respect to the granules, which indicates a more equal distribution of the use of the QIA data from both fractions as well as an equal distribution in the variables with higher VIP. In contrast, the $N\text{-}NO_3^-$ assessment by PLS3 resulted from the use of 10 out of 14 final variables from the granules' QIA data. However, in this case, the variables with higher VIP included both fractions.

In the case of $N\text{-}NH_4^+$, 10 out of 14 variables used for the PLS3 model respect to the suspended fraction, including the ones with higher VIP. However, it is the ensemble of the QIA data from both fractions that resulted in the best model for its assessment. It should also be stressed that the COD assessment is highly dependent of the granular fraction in this system and that a high COD leads to the overgrowth of filamentous bacteria, which interfere with granular stability and inhibits the nitrification process competing for the oxygen needed for this process (Paulo et al., 2021). Hence, the granular fraction can contribute for the assessment ability of the $N\text{-}NH_4^+$ model giving an indication of the flow of inhibition through the different biological processes.

Considering only the QIA data from the suspended biomass fraction (PLS1) the results presented the worst assessment abilities when compared to the remaining models, with the exception of TSS. TSS prediction presented better R^2 results using this model due to the direct relationship of TSS values with the suspended fraction present in the effluent. However, the high slope and low correlation value for the validation set are drawbacks for this model, which is overcome in the PLS3 model, having both a RMSEP of approximately 0.03 mg L^{-1} and a standard deviation of 0.11 mg L^{-1}. And although the PLS1 model presented better results than PLS2, when considering the model with the ensemble of the QIA data, 8 out of 14 variables respect to the granular fraction. This model may be related to the biomass dynamic and correlation between suspended and granular fractions.

The study of each variable assessment model, using individual or ensemble biomass fractions QIA data, allowed to determine which fractions and parameters provided the most valuable information in each case, or if the use of data from one of the fractions undermines the model by noise introduction or suppression of important variables on the step of dataset reduction.

This way, it is demonstrated the importance of the QIA data when analysing the performance of the system. While physical-chemical parameters do not always allow to find links in the system dynamics, the QIA data can detect small biomass changes. These changes enable the monitoring and differentiation between process phases while also elucidating on upcoming performance events, not foreseen by the physical-chemical analysis.

4. Conclusion

AGS systems can be disturbed by changes in wastewater composition, leading to instability. This study proved that QIA was useful for the assessment of the main effluent physical-chemical quality parameters. QIA data, coupled to PLS, allowed to achieve good assessment abilities for the effluent COD, $N\text{-}NH_4^+$, $N\text{-}NO_2^-$, $N\text{-}NO_3^-$, Cl^-, and TSS. While COD and $N\text{-}NO_2^-$ assessment benefited with the use of the granular fraction, $N\text{-}NH_4^+$, $N\text{-}NO_3^-$, Cl^- and TSS assessment benefited with the use of both fractions, revealing the importance of monitoring both suspended and granular fractions in AGS systems. And although the granular sludge is the main component of an AGS system, this study showed that the suspended biomass fraction was also useful for process monitoring.

The use of a real industrial wastewater in this study, besides being an important data source, showed that QIA and chemometric techniques can be considered promising tools in quality assessment of AGS performance for WWT monitoring and control at industrial level. Once the model is optimized, QIA can assess different operational parameters using a few samples. This can be a very useful tool for regular process monitoring and control, with a more environmentally friendly procedure, while avoiding daily physical-chemical analysis, enabling a decrease of analysis costs and hazardous waste generation. Nevertheless, the future use of this methodology requires more in-depth studies in full-scale WWTP to assess its overall applicability.

Acknowledgements

The authors thank the Portuguese Foundation for Science and Technology (FCT) under the scope of the strategic funding to the research units CEB (UIDB/04469/2020) and CBQF (UIDB/50016/2020) and the project AGeNT - PTDC/BTA-BTA/31264/2017 (POCI-01-0145-FEDER-031264). The authors wish to thank the company Águas do Tejo Atlântico, S.A. for

supplying the granules. Daniela P. Mesquita and Catarina L. Amorim thank FCT for funding through program DL 57/2016 – Norma transitória.

Credit author statement

Joana Costa: Investigation, Software, Formal analysis, Writing – Original draft preparation. **Ana M.S. Paulo**: Investigation, Formal analysis, Writing – Reviewing & Editing. **Catarina L. Amorim**: Writing – Reviewing & Editing. **A. Luís Amaral:** Writing – Reviewing & Editing. **Paula M.L. Castro**: Supervision, Writing – Reviewing & Editing, Funding raising. **Eugénio C. Ferreira**: Supervision, Writing – Reviewing & Editing, Funding raising. **Daniela Mesquita**: Conceptualization, Supervision, Writing - Reviewing & Editing.

Declaration of competing interest

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

References

Adav, S.S., Lee, D., Show, K., Tay, J., 2008. Aerobic granular sludge: Recent advances. Biotechnol. Adv. 26, 411–423. https://doi.org/10.1016/j.biotechadv.2008.05.002

Amaral, A.L., 2003. Image analysis in biotechnological processes: Applications to wastewater treatment. Universidade do Minho.

Amaral, A.L., Ferreira, E.C., 2005. Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. Anal. Chim. Acta 544, 246–253. https://doi.org/10.1016/j.aca.2004.12.061

Amaral, A.L., Mesquita, D.P., Ferreira, E.C., 2013. Automatic identification of activated sludge disturbances and assessment of operational parameters. Chemosphere 91, 705–710. https://doi.org/10.1016/j.chemosphere.2012.12.066

APHA, 1998. Standard Methods for the Examination of Water and Wastewater. American Public Health Association, Washington DC.

Bengtsson, S., Blois, M. De, Wilén, B., Gustavsson, D., 2019. A comparison of aerobic granular sludge with conventional and compact biological treatment technologies. Environ. Technol. 40, 2769–2778. https://doi.org/10.1080/09593330.2018.1452985

Bumbac, C., Ionescu, I.A., Tiron, O., Badescu, V.R., 2015. Continuous flow aerobic granular sludge reactor for dairy wastewater treatment. Water Sci. Technol. 71, 440–445. https://doi.org/10.2166/wst.2015.007

561 Caluwé, M., Dobbeleers, T., D'aes, J., Miele, S., Akkermans, V., Daens, D., Geuens, L.,
562 Kiekens, F., Blust, R., Dries, J., 2017. Formation of aerobic granular sludge during the
563 treatment of petrochemical wastewater. Bioresour. Technol. 238, 559–567.

564 Corsino, S.F., Capodici, M., Morici, C., Torregrossa, M., Viviani, G., 2016. Simultaneous
565 nitritation-denitritation for the treatment of high-strength nitrogen in hypersaline
566 wastewater by aerobic granular sludge. Water Res. 88, 329–336.
567 https://doi.org/10.1016/j.watres.2015.10.041

568 Corsino, S.F., Capodici, M., Torregrossa, M., Viviani, G., 2018. A comprehensive comparison
569 between halophilic granular and flocculent sludge in withstanding short and long-term
570 salinity fluctuations. J. Water Process Eng. 22, 265–275.
571 https://doi.org/10.1016/j.jwpe.2018.02.013

572 Corsino, S.F., Capodici, M., Torregrossa, M., Viviani, G., 2017a. Physical properties and
573 Extracellular Polymeric Substances pattern of aerobic granular sludge treating
574 hypersaline wastewater. Bioresour. Technol. 229, 152–159.
575 https://doi.org/10.1016/j.biortech.2017.01.024

576 Corsino, S.F., di Biase, A., Devlin, T.R., Munz, G., Torregrossa, M., Oleszkiewicz, J.A., 2017b.
577 Effect of extended famine conditions on aerobic granular sludge stability in the
578 treatment of brewery wastewater. Bioresour. Technol. 226, 150–157.
579 https://doi.org/10.1016/j.biortech.2016.12.026

580 Cozzolino, D., Kwiatkowski, M.J., Parker, M., Cynkar, W.U., Dambergs, R.G., Gishen, M.,
581 Herderich, M.J., 2004. Prediction of phenolic compounds in red wine fermentations by
582 visible and near infrared spectroscopy. Anal. Chim. Acta 513, 73–80.
583 https://doi.org/10.1016/j.aca.2003.08.066

584 Einax, J.W., Zwanziger, H.W., Geiss, S., 1997. Chemometrics in environmental analysis, in:
585 Chemometrics in Environmental Analysis. VCH, Weinheim, Germany, pp. 747–833.

586 Grijspeerdt, K., Verstraete, W., 1997. Image analysis to estimate the settleability and
587 concentration of activades sludge. Water Res. 31, 1126–1134.

588 He, H., Chen, Y., Li, X., Cheng, Y., Yang, C., Zeng, G., 2017. Influence of salinity on
589 microorganisms in activated sludge processes: A review. Int. Biodeterior. Biodegrad.
590 119, 520–527. https://doi.org/10.1016/j.ibiod.2016.10.007

591 Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., Impe, J.F. Van,
592 2006. Use of image analysis for sludge characterisation: studying the relation between
593 floc shape and sludge settleability. Water Sci. Technol. 54, 167–174.
594 https://doi.org/10.2166/wst.2006.384

595 Leal, C., Val del Río, A., Ferreira, E.C., Mesquita, D., Amaral, A.L., 2021. Validation of a
596 quantitative image analysis methodology for the assessment of the morphology and
597 structure of aerobic granular sludge in the presence of pharmaceutically active
598 compounds. Environ. Technol. Innov. 23, 101639.
599 https://doi.org/10.1016/j.eti.2021.101639

600 Leal, C., Val del Río, A., Mesquita, D.P., Amaral, A.L., Castro, P.M.L., Ferreira, E.C., 2020.
601 Sludge volume index and suspended solids estimation of mature aerobic granular
602 sludge by quantitative image analysis and chemometric tools. Sep. Purif. Technol. 234.

603 Li, Z.H., Wang, X.C., 2008. Effects of salinity on the morphological characteristics of aerobic

granules. Water Sci. Technol. 58, 2421–2426. https://doi.org/10.2166/wst.2008.838

Lotito, A.M., De Sanctis, M., Di Iaconi, C., Bergna, G., 2014. Textile wastewater treatment: Aerobic granular sludge vs activated sludge systems. Water Res. 54, 337–346. https://doi.org/10.1016/j.watres.2014.01.055

Luo, J., Hao, T., Wei, L., Mackey, H.R., Lin, Z., Chen, G., 2014. Impact of influent COD/N ratio on disintegration of aerobic granular sludge. Water Res. 62, 127–135. https://doi.org/10.1016/j.watres.2014.05.037

Mesquita, D.P., Amaral, A.L., Ferreira, E.C., 2016. Estimation of effluent quality parameters from an activated sludge system using quantitative image analysis. Chem. Eng. J. 285, 349–357. https://doi.org/10.1016/j.cej.2015.09.110

Mesquita, D.P., Amaral, A.L., Ferreira, E.C., 2011. Characterization of activated sludge abnormalities by image analysis and chemometric techniques. Anal. Chim. Acta 705, 235–242. https://doi.org/10.1016/j.aca.2011.05.050

Mesquita, D.P., Dias, O., Dias, A.M.A., Amaral, A.L., Ferreira, E.C., 2009. Correlation between sludge settling ability and image analysis information using partial least squares. Anal. Chim. Acta 642, 94–101. https://doi.org/10.1016/j.aca.2009.03.023

Mesquita, D.P., Leal, C., Cunha, J.R., Oehmen, A., Amaral, A.L., Reis, M.A.M., Ferreira, E.C., 2013. Prediction of intracellular storage polymers using quantitative image analysis in enhanced biological phosphorus removal systems. Anal. Chim. Acta 770, 36–44. https://doi.org/10.1016/j.aca.2013.02.002

Nancharaiah, Y. V, Kumar, G.K., 2018. Aerobic granular sludge technology: Mechanisms of granulation and biotechnological applications. Bioresour. Technol. 247, 1128–1143.

Paulo, A.M.S., Amorim, C.L., Costa, J., Mesquita, D.P., Ferreira, E.C., Castro, P.M.L., 2021. Long-term stability of a non-adapted aerobic granular sludge process treating fish canning wastewater associated to EPS producers in the core microbiome. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.144007

Pronk, M., Bassin, J.P., Kreuk, M.K. De, Kleerebezem, R., Van Loosdrecht, M.C.M., 2014. Evaluating the main and side effects of high salinity on aerobic granular sludge. Appl. Microbiol. Biotechnol. 98, 1339–1348. https://doi.org/10.1007/s00253-013-4912-z

Teppola, P., Mujunen, S.-P., Minkkinen, P., 1997. Partial least squares modeling of an activated sludge plant: A case study. Chemom. Intell. Lab. Syst. 38, 197–208.

Val del Río, A., Figueroa, M., Arrojo, B., Mosquera-corral, A., Campos, J.L., García-Torriello, G., Méndez, R., 2012. Aerobic granular SBR systems applied to the treatment of industrial effluents. J. Environ. Manage. 95, S88–S92. https://doi.org/10.1016/j.jenvman.2011.03.019

Wagner, J., Gregory, D., Manguin, V., Helena, R., Morgenroth, E., Derlon, N., 2015. Effect of particulate organic substrate on aerobic granulation and operating conditions of sequencing batch reactors. Water Res. 85, 158–166. https://doi.org/10.1016/j.watres.2015.08.030

Wan, C., Yang, X., Lee, D., Liu, X., Sun, S., Chen, C., 2014. Partial nitrification of wastewaters with high NaCl concentrations by aerobic granules in continuous-flow reactor. Bioresour. Technol. 152, 1–6.