Universidade do Minho
Escola de Engenharia

Daniela Patrícia Bernardino Mesquita

Image analysis and chemometric techniques as monitoring tools to characterize aggregated and filamentous organisms in activated sludge processes

Dezembro de 2010


QUADRO
DE REFERÊNCIA
ESTRATÉGICO
NACIONAL
PORTUGAL 2007.2013
UNIÃO EUROPEIA
Fundo Social Europeu
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR
Portugal

Universidade do Minho
Escola de Engenharia

Daniela Patrícia Bernardino Mesquita

Image analysis and chemometric techniques as monitoring tools to characterize aggregated and filamentous organisms in activated sludge processes

Programa Doutoral em Engenharia Química e Biológica

Trabalho efectuado sob a orientação do
Professor Doutor Eugénio Manuel de Faria Campos Ferreira
e do
Doutor António Luís Pereira do Amaral

Dezembro de 2010

AUTOR

Daniela Patrícia Bernardino Mesquita

E-mail: daniela@deb.uminho.pt

Telefone: +351 253 604 400

TÍTULO DA TESE

Image analysis and chemometric techniques as monitoring tools to characterize aggregated and filamentous organisms in activated sludge processes

ORIENTADORES

Professor Doutor Eugénio Manuel de Faria Campos Ferreira

Doutor António Luís Pereira do Amaral

ANO DE CONCLUSÃO

2010

PROGRAMA DOUTORAL EM ENGENHARIA QUÍMICA E BIOLÓGICA

Universidade do Minho, 14 de Dezembro de 2010

AGRADECIMENTOS

Gostaria de agradecer a todas pessoas que acompanharam a realização deste trabalho.

Em primeiro lugar agradeço aos meus orientadores, Professor Eugénio Ferreira e Doutor Luís Amaral pelo apoio incondicional, coordenação, amizade e encorajamento durante estes quatro anos. Obrigada pelas palavras de incentivo, por terem confiado em mim e por me transmitirem com entusiasmo as sugestões que contribuíram para a melhoria desta tese.

I want to express my gratitude to Doctor Marie-Noëlle Pons for receiving me during a short period at Nancy, and giving me the oportunity to perform some experiments. I would especially like to thank the people from Laboratoire des Sciences du Génie Chimique for the great atmosphere in the lab. Merci!

Agradeço a todos os amigos e colegas do Departamento de Engenharia Biológica e do laboratório, por me terem proporcionado um ambiente agradável e divertido e sobretudo pelas palavras de encorajamento. Aos colegas dos grupos BioProcess Systems Engineering group (BioPSEg) e Laboratório de Biotecnologia Ambiental (LBA) agradeço a colaboração, sugestões e incentivos. Obrigada a todos!

Agradeço à AGERE E.M. e à AGS pela colaboração e disponibilidade demonstrada durante o decorrer deste trabalho.

Finalmente, aos meus pais, ao Pedro e à minha família agradeço a confiança que sempre depositaram em mim, o apoio, a compreensão, o incentivo e ainda o carinho que foi extremamente importante durante todos estes anos. A vós dedico o meu esforço e empenho...

The work presented in this thesis was financially supported by a research grant SFRH/BD/32329/2006 from the Fundação para a Ciência e a Tecnologia (FCT) and Fundo Social Europeu (FSE).

SUMMARY

Activated sludge systems are commonly used in wastewater treatment plants (WWTPs). However, such systems are quite sensitive to sudden changes in operating conditions and a number of problems ranging from hydraulic and organic loads, nutrients and biomass disturbances is quite common. Furthermore, the solid-liquid separation is considered one of the critical stages of the process and quite dependent on the biomass structure and on the proliferation of specific microorganisms that can be assessed by microscopy techniques. For this reason, image analysis techniques have been increasingly used in the assessment of the structure and contents of activated sludge aggregated and filamentous biomass. Furthermore the potential of this technique is sought to be proportional to the array of different types of images fed to such system.

In this thesis eight urban full-scale wastewater treatment plants (WWTPs) were monitored by quantitative image analysis, and a lab-scale activated sludge system was used to simulate different system disturbances. Existing programs for the characterization of the biomass structure for images originating from both bright field and phase contrast were optimized during the course of this thesis, and new programs were developed for fluorescence microscopy images regarding the assessment of the biomass viability and Gram status. A comprehensive comparison between bright field and phase contrast acquisition methodologies was performed, alongside the use of different sample dilutions and microscopy magnifications, since many WWTPs operate with high biomass concentrations. Finally, chemometric techniques such as principal component analysis (PCA) and partial least squares (PLS) were applied to data from the biomass contents, structure, viability and Gram status, alongside reactor operational parameters. It was set for this thesis the main objective of developing and optimizing image analysis procedures to timely detect disturbances and predict ahead of time crucial parameters in an activated sludge system.

The fact that during phase contrast microscopy aggregates presented a halo surrounding them hindered this technique, with respect to both protruding filamentous biomass identification (when present in low contents) and aggregates size determination. Taking into account the overall assessment of the biomass structure and contents, and according to the advantages and disadvantages of both methods, bright field microscopy was considered as presenting the best overall results, with the exception of very high filamentous bacteria contents.

With respect to the dilution study (1:2 to 1:10), it was found that due to its different influence on several key biomass structure parameters, the use of a single correction factor for the overall results seems unfeasible. Furthermore, care should be taken in determining the best sample dilution and a methodology based on the aggregates recognition percentage is proposed. Regarding the magnification study (100x and 20x), the 100x magnification allowed for the detection of both small aggregates and filamentous biomass contents while still presenting a good representativeness of the total biomass contents.

The relationship between the biomass structure and contents and sludge settling ability was extensively studied in eight full-scale wastewater treatment plants in both normal and filamentous bulking conditions. Upon the determination of the uncertainty range regarding the effectiveness of classifying bulking/non-bulking situations by sludge volume index (SVI)

prediction, a satisfactory quantification was obtained inside this range, whereas outside the uncertainty area the classification was quite accurate. A further parameter reduction procedure coupled to a PLS analysis allowed the establishment of a 10 independent variables procedure for SVI estimation, after the identification of the most important variables regarding SVI prediction. The analysis of the variables with the highest importance in the projection demonstrated the importance of the free filamentous bacteria contents descriptor group on the SVI prediction ability.

The study conducted in a lab-scale activated sludge system, allowed the establishment of different disorders that can occur in WWTPs, namely filamentous bulking, zoogleal bulking, and pinpoint flocs formation, alongside normal operating conditions. Data was collected from the biomass contents, structure, viability and Gram status apart from traditional operating parameters. A study was then conducted on the effect of the disturbances regarding the total suspended solids (TSS) and the sludge volume index (SVI), which allowed to correctly classifying the collected samples into the studied conditions. It was also found that the SVI correlated well with the filamentous bacteria contents and the TSS with the total aggregates area. A PCA was performed and permitted the identification of the samples clusters belonging to each of the above conditions. Through the use of a multivariate partial least squares in the collected image analysis data, parameters such as SVI, chemical oxygen demand (COD), nitrate ($N\text{-}NO_3^-$) and TSS inside the reactor were estimated both globally and for each disorder. A previous identification of the different disturbances proved to be crucial for improving the assessment of each parameter, and more than satisfactory correlations were obtained between the measured and the predicted values, with the exception of COD. The further use of the operational parameters presented no significant advantages for the SVI and TSS estimation, but was invaluable for COD and $N\text{-}NO_3^-$ assessment.

SUMÁRIO

Os sistemas convencionais de lamas activadas são habitualmente utilizados em estações de tratamento de águas residuais (ETARs). No entanto, estes sistemas são bastante sensíveis a alterações bruscas nas condições de operação podendo ser afectados por um elevado número de problemas como variações súbitas da carga orgânica, e da concentração de nutrientes e biomassa. Sabe-se também que a separação sólido-líquido, uma das etapas críticas do processo se encontra dependente da estrutura da biomassa e da proliferação de microrganismos específicos que podem ser avaliados através de microscopia. Por este motivo, tem existido um aumento considerável no uso de técnicas de análise de imagem na determinação da estrutura e conteúdo da biomassa agregada e filamentosa. O potencial de utilização desta tecnologia pode ser considerado como proporcional ao diferente tipo de imagens que pode ser tratado.

Nesta tese foram monitorizadas por análise de imagem oito estações de tratamento de águas residuais (ETARs) reais, tendo sido também utilizado um reactor laboratorial de lamas activadas para simular várias perturbações. Foram ainda optimizados programas prévios de processamento e análise de imagens para a caracterização da estrutura da biomassa para imagens resultantes de campo claro e de contraste de fase e desenvolvidos novos programas para a determinação da viabilidade e carácter Gram da biomassa. Foi efectuada uma comparação exaustiva entre as metodologias de aquisição de campo claro e de contraste de fase assim como do recurso a diferentes diluições e ampliações microscópicas, uma vez que muitas ETARs operam com elevadas concentrações de biomassa. Finalmente, foram aplicadas técnicas quimiométricas como a análise de componentes principais (PCA) e o método dos mínimos quadrados parciais (PLS) aos dados relativos à estrutura, conteúdo, viabilidade e carácter Gram da biomassa, assim como aos parâmetros operacionais do reactor. Assim, o principal objectivo desta tese consistiu no desenvolvimento e optimização de metodologias de análise de imagem para uma atempada detecção de perturbações e na previsão dos vários parâmetros considerados fundamentais num sistema de lamas activadas.

O facto de os agregados apresentarem um halo quando observados por microscopia de contraste de fase, prejudicou a detecção quer da biomassa filamentosa protuberante (quando presente em baixo conteúdo), quer a determinação do tamanho dos agregados formados. Tendo em conta a avaliação global da estrutura e conteúdo em biomassa, e de acordo com as vantagens e desvantagens de ambos os métodos, foi considerado que a microscopia de campo claro apresentou melhores resultados globais, com excepção de biomassa com elevado conteúdo em bactérias filamentosas.

No estudo das diluições (1:2 a 1:10), verificou-se que devido à sua influência diferenciada em vários parâmetros-chave da estrutura da biomassa, é inviável a utilização de um único factor de correcção nos resultados globais. Uma vez que a escolha da melhor diluição deve ser cuidadosamente determinada, foi proposta uma metodologia para esse fim baseada na percentagem de reconhecimento dos agregados. Relativamente ao estudo das ampliações (100x e 20x) que melhor define a biomassa agregada e filamentosa, concluiu-se que a ampliação de 100x permitiu uma elevada representatividade do conteúdo total da biomassa assim como da detecção dos agregados de menores dimensões e da biomassa filamentosa.

A relação entre a estrutura e conteúdo da biomassa e a sedimentabilidade das lamas activadas foi extensamente estudada em oito estações de tratamento de águas residuais, quer em funcionamento normal quer em *bulking* filamentoso. Foi também determinada a zona de incerteza no tocante à correcta classificação de situações de *bulking* filamentoso pela previsão do índice volumétrico de lamas (IVL). Verificou-se que, mesmo dentro desta zona de incerteza se obtiveram resultados aceitáveis, e fora da mesma os resultados puderam ser considerados como bastante exactos. Após um procedimento de redução de parâmetros pela análise dos seus factores de projecção no PLS, este permitiu o estabelecimento de uma metodologia englobando 10 parâmetros estruturais independentes para a previsão do IVL. A análise das variáveis com maior factor de projecção permitiu ainda demonstrar a importância do conteúdo em bactérias filamentosas na previsão do IVL.

O estudo realizado na instalação laboratorial de lamas activadas, permitiu o estabelecimento de diferentes perturbações que podem ocorrer em ETARs, nomeadamente *bulking* filamentoso, *bulking* zoogleal e formação de flocos *pinpoint*, para além de condições normais de funcionamento. Para além dos parâmetros tradicionais de operação, foi determinada a estrutura, conteúdo, viabilidade e carácter Gram da biomassa. Foi seguidamente conduzido um estudo para a determinação do efeito de cada uma das perturbações ao nível dos sólidos suspensos totais (SST) e do IVL, que permitiu classificar as amostras recolhidas nas respectivas condições de operação. Foi ainda determinado que o IVL se correlacionou com o conteúdo em bactérias filamentosas e os SST com a área total de agregados. A aplicação de PCA aos dados obtidos permitiu a identificação de aglomerados correspondentes a cada uma das condições descritas anteriormente. Através da utilização da técnica de mínimos quadrados parciais nos dados obtidos por análise de imagem, parâmetros como IVL, carência química de oxigénio (CQO), nitrato ($N\text{-}NO_3^-$) e SST no interior do reactor foram estimados globalmente e para cada perturbação. A prévia identificação das perturbações ocorridas mostrou-se crucial para a melhoria da avaliação de cada um dos parâmetros, tendo sido obtidas correlações mais do que aceitáveis entre os valores medidos e previstos, com excepção para a CQO. A utilização adicional dos parâmetros operacionais não apresentou vantagens significativas na previsão dos valores de IVL e SST, mas apresentou grande utilidade para a previsão da CQO e $N\text{-}NO_3^-$.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

% $Area_{int}$	Area percentage of intermediate aggregates
% $Area_{lrg}$	Area percentage of larger aggregates
% $Area_{sml}$	Area percentage of small aggregates
$Area_{ovr}$	Aggregates average area
% Nb	Aggregates number percentage
Comp	Compactness
Conv	Convexity
$Conv_{int}$	Convexity of intermediate aggregates (mean value)
$Conv_{lrg}$	Convexity of large aggregates (mean value)
$Conv_{ovr}$	Aggregates average Convexity
Deq	Equivalent Diameter
Deq_{int}	Equivalent diameter of intermediate aggregates (mean value)
Deq_{lrg}	Equivalent diameter of larger aggregates (mean value)
Deq_{ovr}	Aggregates average equivalent diameter
Ecc	Eccentricity
Ecc_{int}	Eccentricity of intermediate aggregates (mean value)
Ecc_{lrg}	Eccentricity of large aggregates (mean value)
Ecc_{ovr}	Aggregates average Eccentricity
Ext	Extent
Fcal	Calibration Factor
FF	Form Factor
FL	Filament Length
FL_{avg}	Filaments average length
$FL_{avg}/Area_{avg}$	Filaments average length per aggregates average area
FL_{avg}/Deq_{avg}	Filaments average length per aggregates average equivalent diameter
G_AG/AR	Gram-negative and Gram-positive filaments area ratio
LD_AG/AR	Viable and damaged filaments area ratio
G_AA_G/AA_R	Ratio between gram-negative and gram-positive aggregates area
G_AA_R; G_AA_G	Gram-positive aggregated bacteria area; Gram-negative aggregated bacteria area
G_AF_G/AF_R	Ratio between gram-negative and gram-positive filaments area
G_AF_R; G_AF_G	Gram-positive filaments area; Gram-negative filaments area
G_INA_G/INA_R	Ratio between Gram-negative and Gram-positive aggregates intensity
G_INA_R; G_INA_G	Gram-positive aggregated bacteria intensity; Gram-negative aggregated bacteria intensity
G_INF_G/INF_R	Ratio between Gram-negative and Gram-positive filaments intensity
G_INF_R; G_INF_G	Gram-positive filaments intensity; Gram-negative filaments intensity

LD _AF_G/AF_R	Ratio between viable and damaged filaments area
LD _AF_R; GR_AF_G	Damaged filaments area; Viable filaments area
LD _INA_G/INA_R	Ratio between viable and damaged aggregates intensity
LD _INA_R, LD_INA_G	Damaged aggregated bacteria intensity; Viable aggregated bacteria intensity
LD _INF_G/INF_R	Ratio between viable and damaged filaments intensity
LD _INF_R; LD _INF_G	Damaged filaments intensity; Viable filaments intensity
LD_AA_G/AA_R	Ratio between viable and damaged aggregates area
LD_AA_R; LD_AA_G	Damaged aggregated bacteria area; Viable aggregated bacteria area
L_{int}	Length of intermediate aggregates (mean value)
L_{lrg}	Length of larger aggregates (mean value)
L_{ovr}	Aggregates average length
LrgC	Largest concavity
Nb_{int}/Vol	Number of intermediate aggregates per volume
Nb_{lrg}/Vol	Number of larger aggregates per volume
Nb_{sml}/Vol	Number of small aggregates per volume
Nb_{tot}/Vol	Total number of aggregates per volume
P_{int}	Perimeter of intermediate aggregates (mean value)
P_{lrg}	Perimeter of larger aggregates (mean value)
P_{ovr}	Aggregates average perimeter
RelArea	Ratio between hole and object area
Rob	Robustness
Round	Roundness
$Round_{int}$	Roundness of intermediate aggregates (mean value)
$Round_{lrg}$	Roundness of large aggregates (mean value)
$Round_{ovr}$	Aggregates average roundness
Solid	Solidity
$Solid_{int}$	Solidity of intermediate aggregates (mean value)
$Solid_{lrg}$	Solidity of large aggregates (mean value)
$Solid_{ovr}$	Aggregates average solidity
TA/Vol	Total Aggregate Area per volume
TL/TA	Total Filament Length per Total Aggregate Area
TL/TSS	Total Filament Length per Total Suspended Solids
TL/Vol	Total Filament Length per volume
TRA/Vol	Total area of recognized aggregates per unit volume

ABBREVIATIONS

ANN	Artificial Neural Networks
AS	Activated Sludge System
BF	Bright Field
BMP	Windows Bitmap format
CLSM	Confocal Laser Scanning Microscopy
COD	Chemical Oxygen Demand
DA	Discriminant Analysis
DO	Dissolved Oxygen

DT	Decision Trees
EGSB	Expanded Granular Sludge Bed Reactor
EPS	Extracellular Polymeric Substances
FB	Filamentous Bulking
FISH	Fluorescence *in situ* hybridization
GIF	CompuServe Bitmap format
HI	Hexidium iodide
HPLC	High Performance Liquid Chromatography
JPEG	Joint Photographers Expert Group format
MLSS	Mixed liquid suspended solids
MPEG	Moving Pictures Expert Group format
NC	Normal Conditions
NH_4^+	Ammonium
NO_2^-	Nitrite
NO_3^-	Nitrate
OLR	Organic Loading Rate
PC	Phase Contrast
PCA	Principal Component Analysis
PHB	Poly-β-hydroxybutyrate
PI	Propidium Iodide
PLS	Partial Least Squares
PP	Pinpoint flocs
RGB	Color space based in the Red, Green and Blue channels
SRT	Solid Retention Time
SVI	Sludge Volume Index
TIFF	Tagged Image File Format
TPE-LSM	Photon Excitation Laser Scanning Microscopy
TSS	Total Suspended Solids
UASB	Upflow anaerobic sludge bed reactor
VIP	Variable importance in the projection
WWTP	Wastewater Treatment Plant
ZB	Zoogleal Bulking

Abstract

This chapter introduces the fundamentals of wastewater treatment based on activated sludge processes. Image acquisition and the field of image analysis applications are discussed regarding implemented procedures to improve activated sludge characterization. Finally, the aim and thesis outline are also explained.

1. CHAPTER 1 – CONTEXT, AIM AND OUTLINE

1.1. CONTEXT

The emergence of new technologies and products over the last decades, promoted not only progress but several environmental problems. Thus, environmental concerns associated with sustainable development, led to the appearance of restrictive legislation, limiting the levels of contaminants discharge in watercourses. For example, effluents generated in several industrial operations (washing, rinsing, raw materials transportation, and cleaning equipment) are closely associated to water contamination, when an appropriate treatment is not taking into account. Many effluents are usually composed by organic substances, nutrients, inorganic and also aromatic compounds, and a feasible treatment is required.

The appearance of wastewater treatment processes to solve water quality issues led to the development of activated sludge aerobic processes, as a significant biological process in wastewater treatment plants (WWTPs). Also, in the past few years, with the increase of contaminants complexity, anaerobic digestion technology emerged to improve wastewater treatment. Both are biological processes, and as consequence, several microbiological concerns are always associated.

Activated sludge contains a complex ecosystem composed of different types of microorganisms, such as: filamentous and floc-forming bacteria, protozoa and metazoa. Sludge flocculation and settling are crucial to the effective operation of an activated sludge system (Liu and Fang, 2003). However, due to the complex nature of the microbial communities, imbalances between the different types of bacteria take place and disturb the WWTP with large economical and environmental consequences. Given the massive growth, in the past decades, of optical microscopy methodologies for visual inspection in a wide variety of fields, a number

of wastewater treatment applications have been studied. Using this technique, microorganism recognition or even aggregates characterization were performed. Aggregates size and density are two of the factors that most influence the performance of floc processes, significantly affecting activated sludge settling ability (Schuler and Jang, 2007a,b; Jones and Schuler, 2010).

Filamentous bacteria have been traditionally identified by their morphology (Eikelboom, 2000; Jenkins *et al.*, 2003). Nevertheless, in filamentous bacteria recognition, due to the similarity between some microorganisms many uncertainties have arisen, and therefore some staining techniques were developed.

For microorganism identification, classical Gram and Neisser staining are referred as the most significant techniques (Jenkins *et al.*, 2003). More recently, in the fluorescence field, novel nucleic-acid binding stains were found to detect microorganism status (such as Gram and viability) based on cell walls integrity (Ziglio *et al.*, 2002; Foster *et al.*, 2002; Invitrogen Molecular Probes, 2004; Hao *et al.*, 2009). Overcoming the limitations of bright field microscopy, fluorescent *in situ* hybridization (FISH) has been used, in past years, to determine the contents of specific microorganisms in environmental samples (Wagner *et al.*, 1994, 1996; de los Reyes *et al.*, 1997; Erhart *et al.*, 1997; Schuppler *et al.*, 1998; Davenport *et al.*, 2000; Kanagawa et al, 2000; Wagner *et al.*, 2002; Lopez *et al.*, 2005; Hug *et al.*, 2005).

Nowadays, technological evolution and advances in digital imaging allow for a fast and efficient classification and quantification of microorganisms. With the increase of computers availability, quantitative image analysis has become a routine, principally through grayscale imaging. Thus, image analysis is, at present, a well-established technique when combined to optical microscopy.

In the last years, several authors have already reported the use of automated image analysis in wastewater treatment processes, and an increasing number of activated sludge systems have been characterized for aggregates and filaments recognition (Dagot *et al.*, 2001; da Motta *et al.*, 2001a,b; da Motta *et al.*, 2002; Cenens *et al.*, 2002a,b; Amaral, 2003; Contreras *et al.*, 2004; Casellas *et al.*, 2004; Amaral and Ferreira, 2005; Banadda *et al.*, 2005; Smets *et al.*, 2006; Jenné *et al.*, 2005, 2006, 2007; Arelli *et al.*, 2009). It is known that protozoa and metazoa play an important role in activated sludge systems by controlling bacteria growth, floc size and acting as good indicators of treatment. Already, image analysis procedures have been used for the recognition of protozoa and metazoa in WWTPs by semi-automated image analysis (Ginoris *et al.*, 2007a,b,c; Amaral *et al.*, 2004, 2008). Furthermore, in the field of WWTP anaerobic digestion, granular sludge processes were recently studied using automated image analysis for the detection of granulation time, biomass activity recovery, organic loading disturbances, detergent shock loads, and organic solvents shock loads (Araya-Kroff *et al.*, 2004; Abreu *et al.*, 2007; Costa *et al.*, 2007, 2009a,b).

However, color imaging attracts more and more attention mainly due to the following reasons: (1) color images can provide more information than gray level images; (2) the power of personal computers is increasing rapidly, now able to speedily process color images (Cheng *et al.*, 2001). The use of color scales as a substitute for brightness values allows seeing small changes locally, and identifying the same brightness values globally in an image (Russ, 2002).

By now, quantitative image analysis procedures were used in environmental samples for color images, where filamentous bacteria based on traditional Gram staining methods and storage of poly-β-hydroxybutyrate (PHB) in filamentous bacteria by Sudan Black staining were detected (Pandolfi and Pons, 2004; Pandolfi *et al.*, 2007). Also, specialized software was recently developed for samples

acquired in traditional fluorescence microscopy (Hug *et al.*, 2005; Zhou *et al.*, 2007), confocal laser scanning microscopy (CLSM) (Daims *et al.*, 2006; Daims and Wagner, 2007), and the combination of both with photon excitation laser scanning microscopy (TPE-LSM) (Lopez *et al.*, 2005).

The large amount of data in wastewater treatment regarding plant performance, operational analytical and physical data, and biomass morphological characterization, requests the use of mathematical and statistical methods. This could be an excellent opportunity for the quantitative description of experimental results and effects extracting essential information in wastewater treatment using chemometric techniques. For example, principal component analysis (PCA) is a common technique for finding patterns in data of high dimension, highlighting their similarities and differences. The other main advantage of PCA is that once found these patterns in the data, data compression is possible by reducing the number of dimensions, without much loss of information. Partial Least Squares (PLS) is particularly useful to predict a set of dependent variables from a large set of variables.

PCA was previously used by Costa *et al.* (2009c, 2010) to study the effects by toxics in anaerobic granular sludge and to distinguish organic loading from toxic disturbances, combining image analysis data with reactor operational parameters. In activated sludge systems, PCA was applied to settleability measurements and image analysis parameters, finding a good correspondence between the model output and the sludge volume index (SVI) measurement. This work demonstrated how image analysis can act as a sensing instrument for sludge settleability, by combining the most relevant aggregated and filamentous biomass characteristics (Jenné *et al.*, 2006). Also, PLS was previously used by Amaral and Ferreira (2005), revealing a strong relationship between the operational parameters and image analysis information in filamentous bulking conditions.

In conclusion, image analysis applications especially in wastewater treatment is becoming a widespread tool, attracting gradually more attention due to the opportunities offered by microscopes and computers evolution.

1.2. RESEARCH AIM

This thesis reports several studies performed in lab-scale and full-scale activated sludge systems, to improve the evaluation of microbial aggregates and filamentous bacteria through quantitative image analysis. Studies were conducted based on three microscopic acquisition techniques: bright field, phase contrast, and epifluorescence microscopy. Relationships between image analysis and operational parameters datasets were sought, as well as viability and gram staining assessment by epifluorescence microscopy. Chemometric techniques, such as principal component analysis (PCA) and partial least squares (PLS) were applied for the identification of different settling problems and to predict sludge settling ability and other parameters related to the biological process (total suspended solids, chemical oxygen demand, and nitrate concentrations).

1.3. OUTLINE OF THIS THESIS

Figure 1.1 shows the structure of this dissertation. The current CHAPTER 1 provides the basic concepts about the techniques used for this research. Essentially, all the procedures applied and developed are presented as a subject of this thesis.

A description of activated sludge systems, microscopy, image analysis, and multivariate statistical techniques is presented in CHAPTER 2.

In CHAPTER 3 a comparison between bright field and phase contrast microscopy routines is performed. Microbial aggregates and protruding filaments recognition from activated sludge systems of several WWTP are compared. Image analysis parameters are determined for bright field and for phase contrast. Statistical analysis is applied and the best acquisition method for activated sludge characterization is selected.

Biomass content is an important issue in wastewater treatment plants. When activated sludge systems operate with high biomass contents, samples must be diluted for accurate sludge characterization. Furthermore, the selection of the best image acquisition magnification is directly related to the amount of biomass screened. In CHAPTER 4 the best dilution and magnification for a given activated sludge system is investigated by quantitative image analysis.

A comparison between activated sludge systems from eight WWTP is discussed in CHAPTER 5. During a combined period of two years, an image analysis routine was used providing the necessary data for monitoring the biological system. Results from protruding filamentous bacteria and microbial aggregates recognition were further used for sludge volume index (SVI) prediction during normal and bulking conditions.

The work presented in CHAPTER 6 uses an automated image analysis method to characterize activated sludge structure, focusing on the prediction ability of sludge settling properties, in both normal conditions and filamentous bulking periods. In this sense, the collected images were treated in order to characterize the aggregated and filamentous bacteria, thus originating the different morphological parameters, further fed into a PLS regression for the sludge volume index (SVI) model determination.

The determination of disturbances within an activated sludge system is discussed in CHAPTER 7. Lab-scale experiments were performed using an activated sludge system for normal operating conditions, filamentous bulking, zoogleal bulking and pinpoint flocs formation. Additional to morphological grayscale imaging, a new routine using epifluorescence microscopy is presented to assess viability and gram bacterial status.

In CHAPTER 8, lab-scale disturbances are sought and identified using principal component analysis (PCA), and operational parameters are predicted through the use of partial least squares (PLS) on the image analysis data.

CHAPTER 9 contains the most relevant conclusions as well as some perspectives for further work.

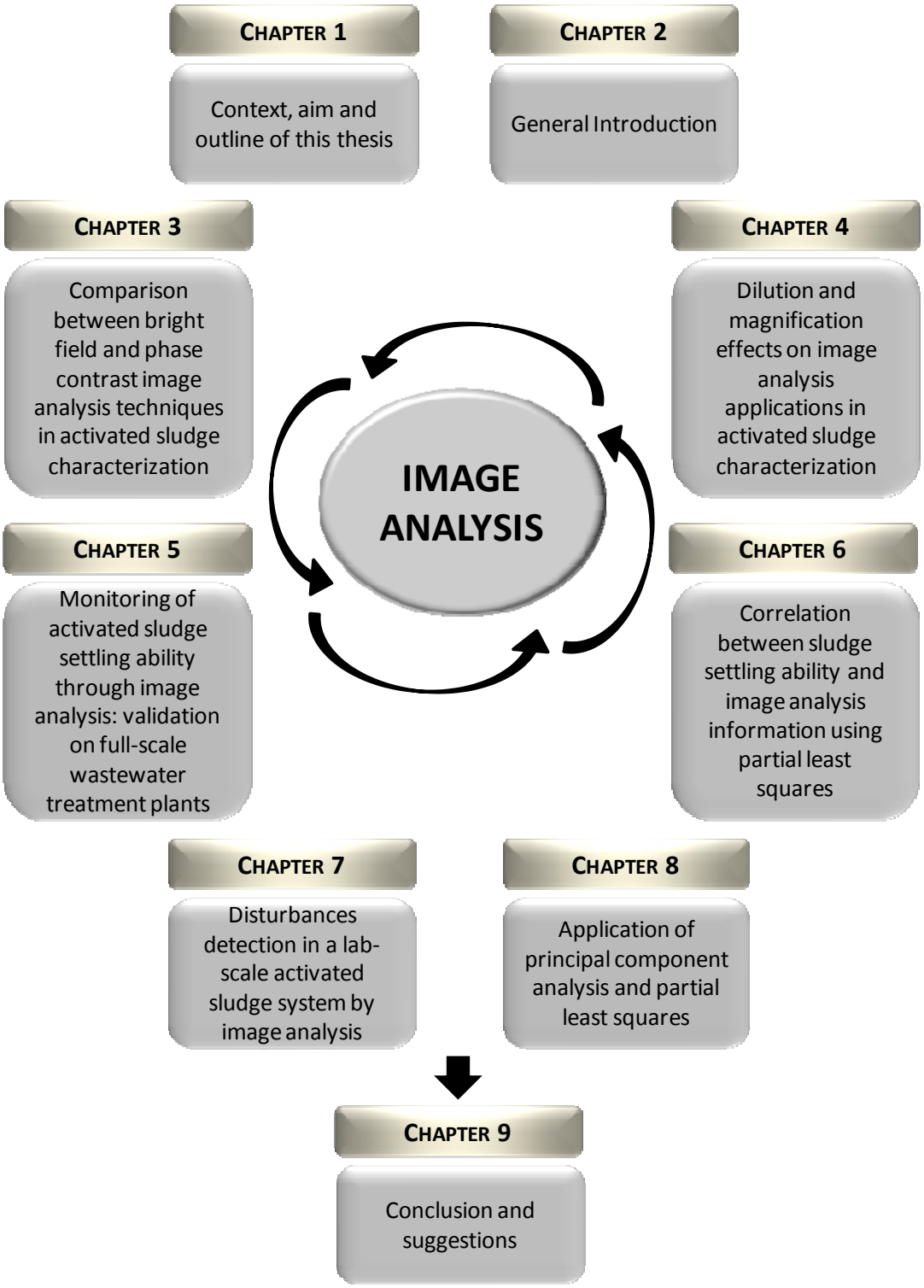

Figure 1.1. Thesis structure.

1.4. REFERENCES

Abreu, A.A., Costa, J.C., Araya-Kroff, P., Ferreira, E.C., and Alves M.M. (2007) Quantitative image analysis as a diagnostic tool for identifying structural changes during a revival process of anaerobic granular sludge. *Water Research* **41**, 1473-1480.

Amaral, A.L. (2003) *Image analysis in biotechnological processes: applications to wastewater treatment.* PhD Thesis, University of Minho, Braga, Portugal (http://hdl.handle.net/1822/4506)

Amaral, A.L., da Motta, M., Pons, M.N., Vivier, H., Roche, N., Mota, M., and Ferreira, E.C. (2004) Survey of Protozoa and Metazoa populations in wastewater treatment plants by image analysis and discriminant analysis. *Environmetrics* **15**, 381-390.

Amaral, A.L. and Ferreira E.C. (2005) Activated sludge monitoring of wastewater treatment plant using image analysis and partial least squares. *Analytica Chimica Acta* **544**, 246-253.

Amaral, A.L., Ginoris, Y.P., Nicolau, A., Coelho, M.A.Z., and Ferreira E.C. (2008) Stalked protozoa identification by image analysis and multivariable statistical techniques. *Analytical and Bioanalytical Chemistry* **391** (4), 1321-1325.

Araya-Kroff, P., Amaral, A.L., Neves, L., Ferreira, E.C., Pons, M.N., Mota M., and Alves M.M. (2004) Development of image analysis techniques as a tool to detect and quantify morphological changes in anaerobic sludge: I. Application to granulation process. *Biotechnology and Bioengineering* **87** (2), 184-193.

Arelli, A., Luccarini, L., and Madoni, P. (2009) Application of image analysis in activated sludge to evaluate correlations between settleability and features of flocs and filamentous species. *Water Science and Technology* **59** (10) 2029-2036.

Banadda, E.N., Smets, I.Y., Jenné, R., and Van Impe, J.F. (2005) Predicting the onset of filamentous bulking in biological wastewater treatment systems by exploiting image analysis information. *Bioprocess and Biosystems Engineering* **27**, 339-348.

Casellas, M., Dagot, C., Pons, M.N., Guibaud, G., Tixier, N., and Baudu, M. (2004) Characterization of structural state of flocculent microorganisms in relation to the purificatory performances of sequencing batch reactors. *Biochemical Engineering Journal* **21**, 171-181.

Cenens, C., Jenné, R., and Van Impe, J.F. (2002a) Evaluation of different shape parameters to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **45** (4-5), 85-91.

Cenens, C., Van Beurden, K.P., Jenné, R., and Van Impe, J.F. (2002b) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated slude images. *Water Science and Technology* **46** (1-2), 381-387.

Cheng, H.D., Jiang, X.H., Sun, Y., and Wang, J. (2001) Color image segmentation: advances and prospects. *Pattern Recognition* **34**, 2259-2281.

Contreras, E.M., Giannuzzi, L., and Zaritzky, N.E. (2004) Use of image analysis in the study of competition between filamentous and non-filamentous bacteria. *Water Research* **38**, 2621-2630.

Costa, J.C., Abreu, A.A., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for monitoring structural changes of anaerobic granular sludge during detergent shock loads. *Biotechnology and Bioengineering* **98** (1), 60-68.

Costa, J.C., Moita, I., Ferreira, E.C., and Alves, M.M. (2009a) Morphology and physiology of anaerobic granular sludge exposed to an organic solvent. *Journal of Hazardous Materials* **167** (1-3), 393-398.

Costa, J.C., Moita, I., Abreu, A.A., Ferreira, E.C., and Alves, M.M. (2009b) Advanced monitoring of high rate anaerobic reactors through quantitative image analysis of granular sludge and multivariate statistical analysis. *Biotechnology and Bioengineering* **102** (2), 445-456.

Costa, J.C., Alves, M.M., and Ferreira, E.C. (2009c) Principal component analysis and quantitative image analysis to predict effects of toxics in anaerobic granular sludge. *Bioresource Technology* **100**, 1180-1185.

Costa, J.C., Alves, M.M., and Ferreira, E.C. (2010) A chemometric tool to monitor high-rate anaerobic granular sludge reactors during load and toxic disturbances. *Biochemical Engineering Journal* **53** (1), 38-43.

da Motta, M., Pons, M.N., and Roche, N. (2001a) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* **43** (7), 91-96.

da Motta, M., Pons, M.N., Roche, N., and Vivier, H. (2001b) Characterization of activated sludge by automated image analysis. *Biochemical Engineering Journal* **9**, 165-173.

da Motta, M., Pons, M.N., and Roche, N. (2002) Study of filamentous bacteria by image analysis and relation with settleability. *Water Science and Technology* **46** (1-2), 363-369.

Dagot, C., Pons, M.N., Casellas, M., Guibaud, G., Dollet, P., and Baudu, M. (2001) Use of image analysis and rheological studies for the control of settleability of filamentous bacteria: application in SBR reactor. *Water Science and Technology* **43** (3), 27-33.

Daims, H., Lucker, S., and Wagner, M. (2006) Daime, a novel image analysis program for microbial ecology and biofilm research. *Environmental Microbiology* **8** (2), 200-213.

Daims, H. and Wagner, M. (2007) Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. *Applied Microbiology and Biotechnology* **75**, 237-248.

Davenport, R.J., Curtis, T.P., Goodfellow, M., Stainsby, F.M., and Bingley, M. (2000) Quantitative use of fluorescent in situ hybridization to examine relationships between mycolic acid-containing actinomycetes and foaming in activated sludge plants. *Applied and Environmental Microbiology* **66** (3), 1158-1166.

de los Reyes, F.L., Ritter, W., and Raskin, L. (1997) Group-specific small-subunit rRNA hybridization probes to characterise filamentous foaming in activated sludge systems. *Applied and Environmental Microbiology* **63**, 1107-1117.

Eikelboom, D.H. (2000) *Process control of activated sludge plants by microscopic investigation*. First edition, IWA Publishing, London.

Erhart, R., Bradford, D., Seviour, R.J., Amann, R., and Blackall, L.L. (1997) Development and use of fluorescent in situ hybridization probes for the detection and identification of *''Microthrix parvicella''* in activated sludge. *Systematic and Applied Microbiology* **20**, 310-318.

Forster, S., Snape, J.R., Lappin-Scott, H.M., and Porter, J. (2002) Simultaneous Fluorescent Gram Staining and Activity Assessment of Activated Sludge Bacteria. *Applied and Environmental Microbiology* **68** (10), 4772-4779.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira E.C. (2007a) Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system. *Water Research* **41**, 2581-2589.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007b) Raw data pre-processing in the protozoa and metazoa identification by image analysis and multivariate statistical techniques. *Journal of Chemometrics* **21**, 156-164.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007c) Recognition of protozoa and metazoa using image analysis tools, discriminant analysis, neural networks and decision trees. *Analytica Chimica Acta* **595**, 160-169.

Hao, X., Wang, Q., Zhang, X., Cao, Y., and van Loosdrecht, M. (2009) Experimental evaluation of decrease in bacterial activity due to cell death and activity decay in activated sludge. *Water Research* **43**, 3604-3612.

Hug, T., Gujer, W., Siegrist, H. (2005) Rapid quantification of bacteria in activated sludge using fluorescence in situ hybridization and epifluorescence microscopy. *Water Research* **39**, 3837-3848.

Invitrogen Molecular Probes (2004) LIVE/DEAD *BacLight™ Bacterial Viability Kits. Manuals and Product Inserts*. http://probes.invitrogen.com/media/pis/mp07007.pdf.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jenné, R., Banadda E.N., Smets I.Y., Bamelis A., Verdickt L., and Van Impe, J.F. (2005) Activated sludge image analysis system: monitoring settleability and effluent clarity. *Water Science and Technology* **52** (10-11), 193-199.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of filamentous bulking problems: developing an image analysis system for sludge composition monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Jones, P.A. and Schuler, A.J. (2010) Seasonal variability of biomass density and activated sludge settleability in full scale wastewater treatment systems. *Chemical Engineering Journal* **164**, 16-22.

Kanagawa, T., Kamagata, Y., Aruga, S., Kohno, T., Horn, M., and Wagner, M. (2000) Phylogenetic analysis of and oligonucleotide probe development for Eikelboom Type 021N filamentous bacteria isolated from bulking activated sludge. *Applied and Environmental Microbiology* **66**, 5043-5052.

Liu, Y. and Fang, H.P. (2003) Influence of extracellular polymeric substances on flocculation, settling and dewatering of activated sludge. *Critical Reviews in Environmental Science and Technology* **33** (3), 237-273.

Lopez, C., Pons, M.N., and Morgenroth, E. (2005) Evaluation of microscopic techniques (epifluorescence microscopy, CLSM, TPE-LSM) as a basis for the quantitative image analysis of activated sludge. *Water Research* **39**, 456-468.

Pandolfi, D. and Pons, M.N. (2004) Gram-staining characterization of activated sludge filamentous bacteria by automated colour analysis. *Biotechnology Letters* **26**, 1841-1846.

Pandolfi, D., Pons, M.N., and da Motta, M. (2007) Characterization of PHB storage in activated sludge extended filamentous bacteria by automated colour image analysis. *Biotechnology Letters* **29**, 1263-1269.

Russ, J.C. (2002) *The Image Processing Handbook*. Fourth edition, CRC Press, Boca Raton.

Schuppler, M., Wagner, M., Schön, G., and Göbel, U. (1998) *In situ* identification of nocardioform actinomycetes in activated sludge using fluorescent rRNA-targeted oligonucleotide probes. *Microbiology* **144**, 249-259.

Schuler, A.J. and Jang, H. (2007a) Causes of variable biomass density and its effects on settling in full scale biological wastewater treatment systems. *Environmental Science and Technology* **41** (5), 1675-1681.

Schuler, A.J. and Jang, H. (2007b) Density effects on activated sludge zone settling velocities. *Water Research* **41** (8), 1814-1822.

Smets, I.Y., Banadda, E.N., Deurinck, J., Renders, N., Jenné, R., and Van Impe, J.F. (2006) Dynamic modelling of filamentous bulking in lab-scale activated sludge processes. *Journal of Process Control* **16**, 313-319.

Wagner, M., Amann, R., Kämpfer, P., Assmus, B., Hartmann, A., Hutzler, P., Springer, N., and Schleifer, K.H. (1994) Identification and *in situ* detection of gram-negative filamentous bacteria in activated sludge. *Systematic and Applied Microbiology* **17**, 405-417.

Wagner, M., Rath, G., Koops, H.P., Flood, J., and Amann, R. (1996) *In situ* analysis of nitrifying bacteria in sewage treatment. *Water Science and Technology* **34** (1-2), 237-244.

Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., and Daims, H. (2002) Microbial community composition and function in wastewater treatment plants. *Antonie van Leeuwenhoek* **81**, 665-680.

Zhou, A., Pons, M.N., Raskin, L., and Zilles, J.L. (2007) Automated image analysis for quantitative fluorescence *in situ* hybridization with environmental samples. *Applied and Environmental Microbiology* **73** (9), 2956-2962.

Ziglio, G., Andreottola, G., Barbesti, S., Boschetti, G., Bruni, L., Foladoria, P., and Villa, R. (2002) Assessment of activated sludge viability with flow cytometry. *Water Research* **36** (2), 460-468.

Abstract

In this chapter, a general introduction of the main subjects discussed in this thesis is presented. The relevance of biological processes in wastewater treatment, mainly the use of activated sludge systems is given. Activated sludge microbiology is discussed, as well as microscopy identification techniques, used to monitor the biological process. A review of image analysis applications in activated sludge systems is given. Finally, it is also described the usefulness of chemometric tools, such as principal component analysis and partial least squares.

2. CHAPTER 2 – GENERAL INTRODUCTION

2.1. BIOLOGICAL WASTEWATER TREATMENT

In the early 20th century, research was basically developed in filtration processes, and biological processes were limited to septic tanks. Septic tanks have lost some of its popularity due to a restriction, but were replaced by new anaerobic processes. Physical and chemical processes were definitely the most popular technologies at time, including sedimentation, chemical treatment, and dilution. However, none of these processes offer a high quality treatment. In this context, emerged a new biological process called *activated sludge* capable of producing a clear effluent with high quality (Metcalf and Eddy, 2003).

Nowadays, biological processes are extensively used in wastewater treatment plants. The use of bacteria and other microorganisms to remove contaminants is increasingly applied in sewage and industrial wastewater treatment. After the emergence and field application of biological processes with higher efficiencies, environmental laws establishing more strict values regarding the discharge effluent parameters were imposed. Furthermore, the biological process monitoring can be accomplished through the microbial community observation. The microbial community assessment by microscopy inspection is a useful methodology for a fast diagnosis of the problem. Timely predict the problem occurrence is the highest benefit.

2.2. ACTIVATED SLUDGE SYSTEMS

Classically, an activated sludge system is a continuous biological process widely used in wastewater treatment plants, which includes metabolic conversion of the contaminants. This system is composed of an aerated tank where the biochemical

(pollutants removal) phase takes place and of a clarifier where the physical (settling) phase occurs. In the aerated tank, microbial organisms (globally defined as mixed liquid suspended solids (MLSS)) oxidize the substrate (pollutants), present in the wastewater, due to the supply of oxygen. In the aeration tank, organic matter and nutrients (ammonium and phosphate) are removed from the wastewater. The efficiency of the activated sludge process is strongly related to the ability of the sludge aggregates to settle. The activated sludge process can be performed in a number of different ways, and the biochemical phase can have many different configurations, depending on the desired effluent quality. Biological nitrogen (nitrification and denitrification) and phosphorous removal stages can be achieved by expanding the classical activated sludge system with anoxic and anaerobic tanks. If the system is completely aerobic, the air supply allows the organic matter and ammonium oxidation, keeping the biomass in suspension. Otherwise, if anoxic zones are added, the denitrification takes place, and finally, anaerobiosis favors the presence of phosphorus accumulating bacteria.

The settling phase is a crucial step of the activated sludge process in biological wastewater treatment plants. This unit is usually associated to thickening, sludge recycle, effluent clarification, and sludge storage, producing an effluent with low suspended solids. Also, a fraction of the sludge could be removed from the system avoiding biomass overgrowth. Figure 2.1 depicts a scheme of a conventional activated sludge system.

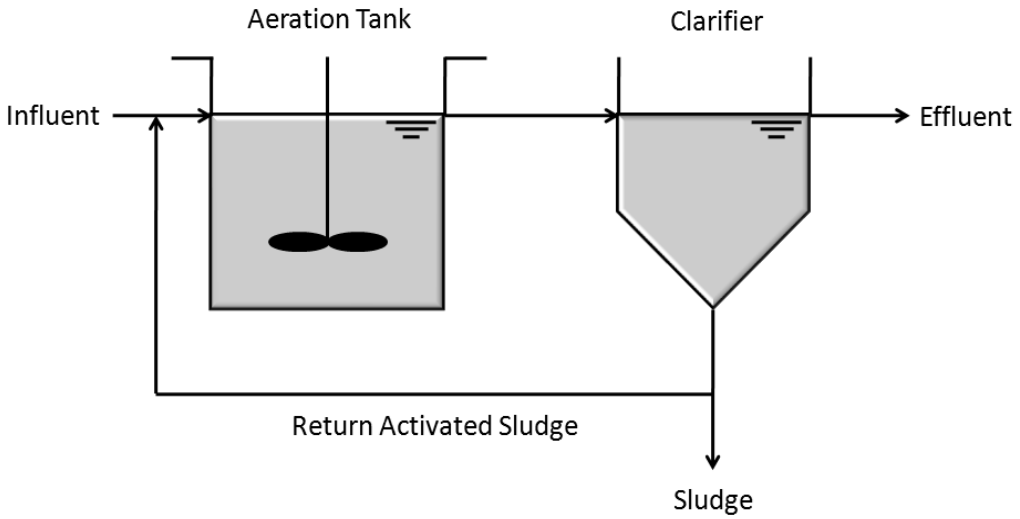

Figure 2.1. Conventional activated sludge process (adapted from Metcalf and Eddy, 2003).

2.3. ACTIVATED SLUDGE

2.3.1 ACTIVATED SLUDGE STRUCTURE

Usually, the biomass developed in activated sludge systems is comprised of approximately 95% bacteria and 5% other organisms (Richard, 1989). Depending on the external conditions, the microbial community composition responsible for the biological process in wastewater treatment plants could be diverse. The microbial community contains many species divided into bacteria, protozoa, metazoa, fungi, algae and viruses (Jenkins *et al.*, 2003). Although the bacteria are generally prevalent in the aeration tank, high concentrations of protozoa in the tank normally indicates a good performance. The species distribution and abundance have been pointed out as indicators of the quality of the effluent exiting an activated sludge plant, providing a useful mechanism to evaluate its performance (Ginoris *et al.*, 2007a). The population of microorganisms present depends on factors such as

wastewater composition, microbial growth rate, pH or temperature. The settling, flocculation and floc formation characteristics also influence the microbial diversity.

2.3.2 FLOCS COMPOSITION

Activated sludge microorganisms live and grow held together by a slime matrix comprised of extracellular polymeric substances (EPS) forming a three-dimensional aggregated microbial structure called floc, and by chemical biunding forces (Eikelboom, 2000). The EPS are typically composed of a variety of organic substances such as: proteins, humic substances, carbohydrates, nucleic acids and lipids (Frølund *et al.*, 1996; Liu and Fang, 2002). Free-living bacteria, protozoa and occasionally higher organisms are present around the flocs and in the liquor between them. Based on these characteristics, activated sludge aggregates are considered heterogeneous structures. Seviour and Blackall (1999) also revealed that the microbial floc composition and activity changes with wastewater composition and treatment plant operation.

EPS functions include the formation of a protective layer for the cells against the harmful external environment, such as biocides and sudden changes of pH, sorbing exogenous nutrients and organic molecules, and aggregating bacterial cells in flocs. Consequently, it is believed that EPS play a large role in the flocculation, settling and dewatering of activated sludge (Novák *et al.*, 1993; Urbain *et al.*, 1993; Liu and Fang, 2003).

2.3.3 FLOCS STRUCTURE ON SETTLING ABILITY

In activated sludge systems operations, sludge thickeningis crucial since the overall effectiveness of the process depends, in large scale, on the efficiency of the solid

separation step. To clarify the role of microbial aggregates in wastewater treatment, different studies were previously developed explaining their characteristics which exert direct or indirect influence on the sludge settling ability.

Settling ability is mainly dependent on the structure and nature of the activated sludge flocs, and these properties can change drastically depending on the operational conditions in WWTPs (Wilén and Balmer, 1999). Filamentous bacteria are also present in activated sludge flocs providing a backbone for the overall floc structure (Sezgin *et al.*, 1978). Figure 2.2 shows images from activated sludge systems using bright field microscopy. The success of a given activated sludge system depends on the ecosystem balance among floc-forming bacteria, such as *Pseudomonas* spp., *Zoogleal* spp., *Alcaligens* spp., and *Achromobacter* spp., and filamentous bacteria, such as *Nocardia* spp., *Rhodococcus* spp., Type 1863 and *Microthrix* spp. (Martins *et al.*, 2004; Rossetti *et al.*, 2005). However, the overgrowth of filamentous bacteria is also related to settling ability problems. A description of activated sludge aggregated and filamentous bacteria properties is further presented.

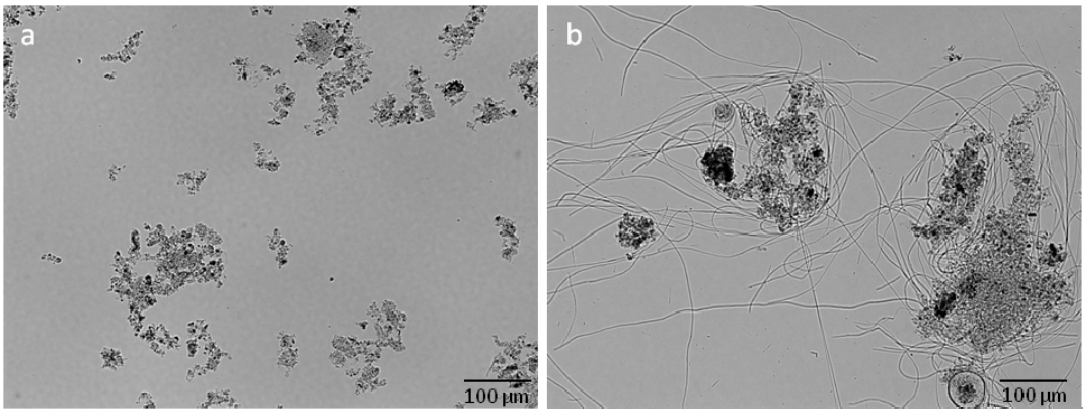

Figure 2.2. Bright field microscopy of activated sludge: (a) well-settling sludge; (b) sludge with filamentous bacteria excess (images obtained during this work).

To guarantee the success of a given activated sludge system, a global examination of the sludge must be performed. This characterization is based on morphological characteristics, which have a considerable variation in wastewater treatment. Recent developments on floc structure analysis associated to microscopy techniques give access to and in deep analysis of structural information (Jarvis *et al.*, 2005). The characteristics of bioaggregates, including their internal structure, chemical composition and microbial ecology, determine the transport properties and chemical reaction rates, and affect the overall performance of treatment processes involving aggregates (Chu and Lee, 2004). A variety of other factors are also known to affect settling rates due to flocculation and deflocculation processes (Wilén *et al.*, 2000).

The shape, structure, and strength of activated sludge flocs are three important characteristics for morphological characterization. The shape varies in its regularity, roundness and compactness, among others. If flocs present an irregular shape, consequently, the sludge settling velocity is reduced. However, the air supply in the aerated tank and the presence of protruding filamentous bacteria leads to flocs irregularly shaped. If compact flocs are present, bacteria are stacked close to one another, otherwise, open flocs give a reduced settling velocity. The presence of weak flocs can be easily detected by a large presence of cells at the edges of the flocs, promoting their deterioration (Eikelboom, 2000). Typical floc characteristics include irregular shape, broad distribution of particle sizes, fragile and easily compressible, highly porous (more than 99% porosity, based on the mass balance) and permeable to fluids, large specific surface area, inhomogeneous distribution of internal mass, networked structure, and poor dewaterability (Li and Ganczarczyk, 1989, 1990).

Other morphological, physical, and chemical factors, including particle surface properties, size, extracellular polymeric substances (EPS), and flocculating ability

were also referred as affecting factors in solid-liquid separation processes in activated sludge samples (Li and Ganczarczyk, 1987, 1991; Andreadakis, 1993; Barbusinski and Koscielniak, 1995; Grijspeerdt and Verstraete, 1997; Jin *et al.*, 2003; Li and Yang, 2007; Liao *et al.*, 2006; Urbain *et al.*, 1993; Wilén *et al.*, 2003; Jin *et al.*, 2004). Activated sludge flocs stability was also studied with particular attention to several process parameters, such as: solids retention time (SRT), organic loading rate (OLR), turbulence, and dissolved oxygen concentration (DO), as reported by Wilén and Balmer (1999). More recently, Wilén *et al.* (2008) studied a full-scale activated sludge system operated at transient conditions focusing in the chemical make-up of the flocs and relating these parameters with settling and compaction properties of the sludge. Recent research has demonstrated that variable biomass density (defined as the mass per biomass volume, not including pore spaces) can significantly affect activated biomass settling (Schuler and Jang, 2007a,b). Recently, Jones and Schuler (2010) revealed that variable biomass density is a significant contributor to seasonal variations in settleability in some full-scale activated sludge systems, independent to changes in filamentous bacteria contents.

2.3.4 SETTLING PROPERTIES

As discussed before, the basis of the activated sludge process is the growth of floc-forming bacteria upon a filamentous bacteria backbone which settle under gravity in the final clarifier, leaving a clarified supernatant and a thickened return sludge. Besides the overproduction of extracellular polymers by floc forming bacteria, a number of filamentous bacteria may also overgrow or undergrow, which can lead to operational problems. These problems occur mostly of the time when the operating conditions are not perfect mainly in terms of organic load, nutrients, and oxygen supply. The most common problems were reported by Jenkins *et al.* (2003): pinpoint flocs bulking; filamentous bulking; dispersed growth, and zoogleal (or

viscous) bulking (Table 2.1). Although the occurrence of any of the phenomena here described can negatively affect the clarifier performance, sludge bulking is one of the most significant problems. Bulking can cause severe operational problems, increasing the treatment costs and lowering the final effluent quality. The most critical problems involve poorer settleability, risk of sludge washout with the final effluent and deteriorated dewatering and thickening properties of the sludge (Eikelboom, 2000).

The effectiveness of bioflocculation and settling of activated sludge is often characterized by sludge volume index (SVI), which is defined as the volume (in mL) occupied by 1 g of sludge after 30 min of settling in a 1 L cylinder. Low SVI values indicate that the sludge is dense and thus the sludge has better settling abilities. According to Jenkins *et al.* (2003), an activated sludge with SVI lower than 120 mL/g is considered satisfactory and over 150 mL/g is considered bulking.

The bulking problem is frequently related to *Zoogleal* organisms (non-filamentous bulking) and/or filamentous microorganisms' overgrowth. The non-filamentous bulking was previously reported by Novak *et al.* (1993) where the SVI increase was not connected to an excessive growth of filamentous microorganisms. On the contrary, the number of filamentous microorganisms dropped during the experiment and a great amount of extracellular slime was found causing viscous bulking. Urbain *et al.* (1993) showed that the presence of filamentous microorganisms was observed in all samples analyzed, but was not always associated to poor settling ability, corroborating the theory that these microorganisms have an important role for the flocs structure stability.

Table 2.1. Problems associated with biomass separation in activated sludge systems (adapted from Jenkins *et al.*, 2003).

Settling problem	Cause	Effect
Filamentous bulking	Dificulties to maintain the required biomass concentration in the aeration basin Excessive large amounts of filamentous bacteria linking the flocs, interfering with compaction, settling and thickening	High SVI; sludge blanket overflows the secondary clarifier; solids handling processes become hydraulically overloaded
Pinpoint floc	Low organic loads; Extended aeration systems; Small, compact, weak, roughly spherical flocs are formed; large flocs settle rapidly; smaller flocs settle slowly	Low SVI and turbid effluent with high solids contents
Dispersed growth	Very high organic loading (high food to microorganism ratio, F/M); Presence of toxicity; Microorganisms form only small clusters or single cells	Turbid effluent and solids do not settle
Zoogleal bulking	Floc forming bacteria are present in large amounts producing extracellular material; in severe cases extracellular material imparts a slimy or jelly-like consistency to activated sludge	Reduced settling and compaction rates; overflow of sludge blanket from secondary clarifier; viscous foam may appear and poor sludge dewatering
Blanket rising	Denitrification in the secondary clarifier releases poorly soluble nitrogen gas which attaches to activated sludge flocs and floats them to surface	A scum of activated sludge forms on surface of the secondary clarifier and on aeration basin anoxic zones
Foam/scum formation	Caused by undegraded surfactants and by nocardioforms, *Microthrix parvicella*, or type 1863	Foam can float large amount of biomass to surface of treatment units

2.3.5 FILAMENTOUS BULKING

Plant surveys carried out in different countries have shown the divergence of relative importance of filamentous bacteria (Seviour and Blackall, 1999). The most common sludge bulking problem is caused by an excessive growth of different types of filamentous microorganisms. Figure 2.3 shows two images obtained during a filamentous bulking disturbance. Several factors emerge in determining the community composition of filamentous bacteria including a range of environmental factors and the nature of the influent being treated. Table 2.2 shows the main filamentous bacteria type that grows under each operational problem responsible for bulking conditions. These correlations support the idea that some filamentous bacteria appear to be selectively favored or unfavored under certain plant operational conditions.

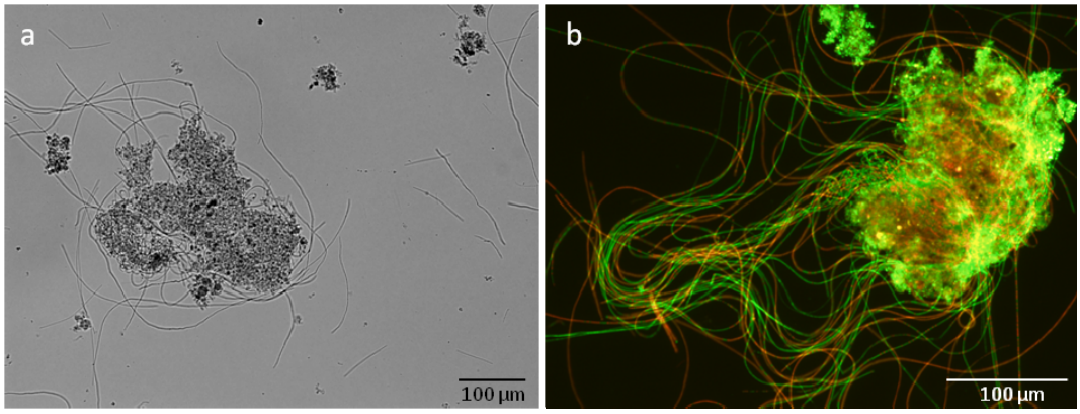

Figure 2.3. Filamentous bulking detection in (a) bright field and (b) epifluorescence microscopy (images obtained during this work).

Usually, filaments extend from flocs into the bulk solution and interfere with compaction, settling, thickening, and activated sludge concentration. In order to overcome filamentous bulking problems several authors (Chang *et al.*, 2004; and Agridiotis *et al.*, 2007) recommend the addition of chemicals such as chlorine and hydrogen peroxide, or even metal ions to the aerated tank. However, these additions only offer a short-term solution. Process design parameters are the key

factors that contribute to the growth of filamentous bacteria. Several works reported fundamental factors promoting filamentous microorganisms' overgrowth such as: oxygen and nutrient deficiencies, substrate composition (polymeric substrate), aeration tank configuration and long solids retention time (Martins *et al.*, 2003, 2004, 2010; Gaval and Pernelle, 2003; Tsai *et al.*, 2003; Liu and Liu, 2006).

Table 2.2. Dominant main filamentous bacteria type that grows under each operational problem responsible for each causative condition of bulking conditions (adapted from Madoni, 2005).

Operational problem	Dominant filamentous bacteria types
Low dissolved oxygen	Type 1701, *Sphaerotilus natans*, *Haliscomenobacter hydrossis*, Type 1863
Low organic loading rate (low F/M)	*Microthrix parvicella*, *Haliscomenobacter hydrossis*, Type 0041
Septic wastes/sulfides	*Thiothrix spp.*, Type 021N, *Beggioatoa*
Nutrient deficient (N and/or P)	*Thiothrix spp.*, Type 021N, *Sphaerotilus natans*, Type 0041

2.4. MICROSCOPY

Understanding filamentous bulking or other activated sludge failures, requires activated sludge characterization with respect to the aggregated and filamentous biomass contents, and filamentous microorganisms identification. Simple optical devices such as microscopes allow the visualization of activated sludge samples and, therefore, microscopic examination of activated sludge is quite useful for determining the physical nature of the aggregated biomass and the types and abundance of filamentous microorganisms (Jenkins *et al.*, 2003). In this manner, the

determination of the activated sludge biomass characteristics (related to settling and compaction) allows microscopy techniques to yield information related to activated sludge systems behavior in solid-liquid separation processes. Most microscopes are now fitted with a range of options allowing systems for bright field, phase contrast or fluorescence microscopy, which will be further explained in this section.

2.4.1 BRIGHT FIELD AND PHASE CONTRAST

Using bright field microscopy can give useful information about the state of the sludge. However, some transparent objects such as the majority of microbial cells cannot always be seen clearly since they lack contrast. To overcome this problem, increasing contrast (using phase contrast) could be a solution (Lindea *et al.*, 1999). Nevertheless, using bright field microscopy is also possible to improve contrast by using staining procedures.

Two staining procedures are used routinely in filamentous bacteria identification - the Gram stain and the Neisser stain (Figure 2.4). These staining methods, and the differentiation between positive and/or negative status, depend on the existence of a permeability barrier in bacteria based on the chemical and physical properties of their cell walls. The reaction to the Gram staining could be negative, strongly positive or weakly positive. Most filamentous microorganisms are Gram negative and commonly observed in activated sludge systems contributing to poor settlement of activated sludge flocs in secondary sedimentation tanks (Wagner et al., 1994). In the case of the Neisser staining, most of filamentous microorganisms are negative. However, the positive microorganisms possess intracellular granules and could be easier to identify.

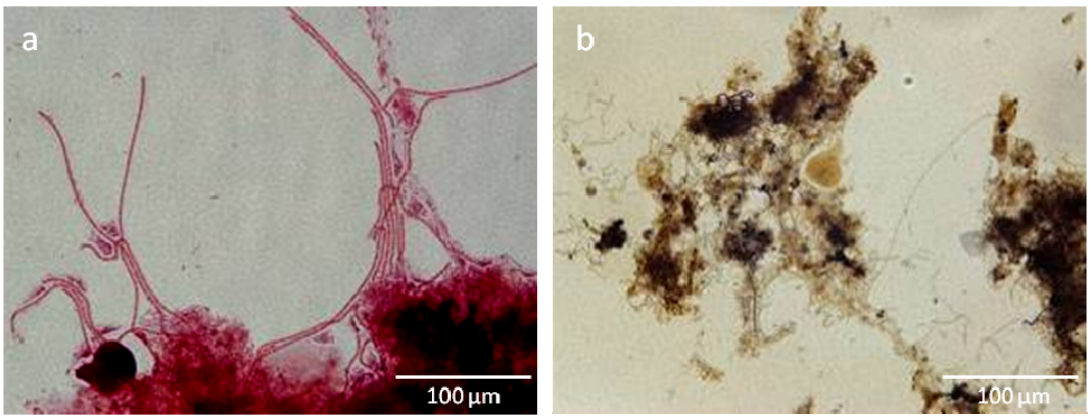

Figure 2.4. Gram (a) and Neisser (b) staining from an activated sludge sludge system (images obtained during this work).

Traditional Gram and Neisser staining techniques may be problematic since they require practice, accuracy and precision in the decoloration step, and may limit the assessment of large and dense flocs which do not decolorize correctly.

2.4.2 FLUORESCENCE MICROSCOPY

Bright field or phase contrast microscopy may not be able to answer some questions about the microbial communities based on species composition, structure and bacterial distribution. Fluorescence microscopy could be an expensive method, nevertheless, provides the user with a powerful ability for certain staining procedures for detecting different types of populations.

Using fluorescent probes is an excellent way to overcome some problems of studying microbial populations. On one hand, fluorescent probes are short sequences of DNA (16-20 nucleotides) labeled with fluorescent dyes, that recognize 16S rRNA sequences in fixed cells and hybridize with them *in situ* (DNA-RNA matching). On the other hand, other fluorescent stains can be used such is the case of DAPI (4',6-diamidino-2-phenylindole) or Acridine orange. DAPI is a blue nucleic

acid fluorescent stain forming fluorescent complexes with natural double-stranded DNA. It is used extensively in fluorescence microscopy (Invitrogen Molecular Probes, 2006). Acridine orange is a fluorescent dye and it can be useful to study viable/non-viable microorganisms by epifluorescence microscopy (Simpson *et al.*, 2006). Microorganisms can be identified and quantified in almost ecosystem with hybridization (Amann *et al.*, 1990). The combination of fluorescent microscopy with fluorescence *in situ* hybridization (FISH) probes is considered a powerful tool for *in situ* identification of microorganisms in activated sludge from WWTPs (Figure 2.5). Outstanding studies on the characterization of filamentous organisms were previously performed through the use of specific probes: *Sphaerotilus natans* and *Haliscomenobacter* spp. (Wagner *et al.*, 1994), *Eikelboom type 021N* group I, II, II (Kanagawa *et al.*, 2000), *Candidatus microthrix* (Erhart *et al.*, 1997), *Gordona amarae* (de los Reyes *et al.*, 1997). Moreover, more studies have focused on *Microthrix parvicella*, the dominant and the most studied filamentous bacteria present in WWTPs causing the common events of bulking and foaming phenomena in activated sludge treatment plants throughout the world (Blackall *et al.*, 1996; Rossetti *et al.*, 1997, 2005). Also, *Nostocoida limicola* is largely studied due to their incidence in foaming and bulking phenomena (Blackall *et al.*, 2000; Liu *et al.*, 2001; Seviour *et al.*, 2006). Recently, FISH analysis was also applied to samples obtained from a full-scale plant and related to floc stability (Wilén *et al.*, 2008). A review of molecular biology techniques can be found in Sanz and Köchling (2007) where emphasis is given to FISH analysis stating out advantages and disadvantages. Concluding, by using FISH, it is possible to observe the bacteria morphology and to quantify their number or equivalent biovolume (Nielsen *et al.*, 2009).

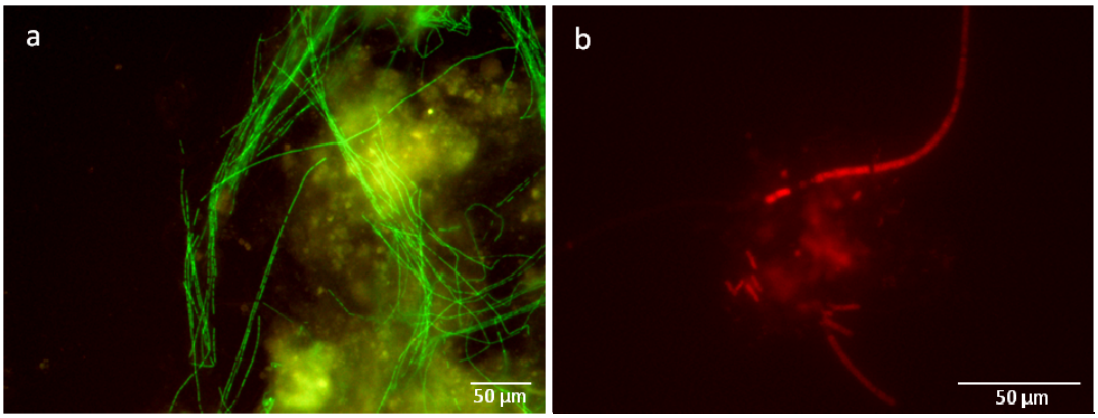

Figure 2.5. (a) *Haliscomenobacter hydrossis* and (b) *Thiothrix I* identification using FISH (images obtained during this work).

In addition to FISH, rapid fluorescent staining methods were developed to estimate both viable and damaged bacteria and the Gram status. Hexidium iodide (HI) is a novel fluorescent nucleic acid binding dye that allows the assessment of Gram status by differential absorption through bacterial cell walls, selectively staining gram-positive organisms without fixative methods (Invitrogen Molecular Probes, 2001). The use of such fluorescent dye may provide a robust, objective, and rapid alternative to traditional Gram staining in wastewater systems (Foster *et al.*, 2002). The Live/Dead® BacLight™ Bacterial Viability kit is a widely used method to measure viability and differentiates between living and dead cells by detecting if their membrane system is intact, even in a mixed population containing a broad range of bacterial types (Invitrogen Molecular Probes, 2004). This kit is comprised by two nucleic acid probes, green-fluorescent SYTO® 9 and red-fluorescent Propidium Iodide (PI). These stains differ both in their spectral characteristics and in their ability to penetrate healthy bacterial cells. This kit was used with success in drinking water and activated sludge samples (Boulos *et al.*, 1999; Lopez *et al.*, 2005; Louvet *et al.*, 2010). Images acquired using both staining kits are presented in Figure 2.6. For both images two long pass filters were used, the first in the green wavelength range with an excitation bandpass of 470-490 nm and emission at 516 nm, and the

second filter in the red wavelength range with an excitation bandpass of 530-550 nm and emission at 591 nm.

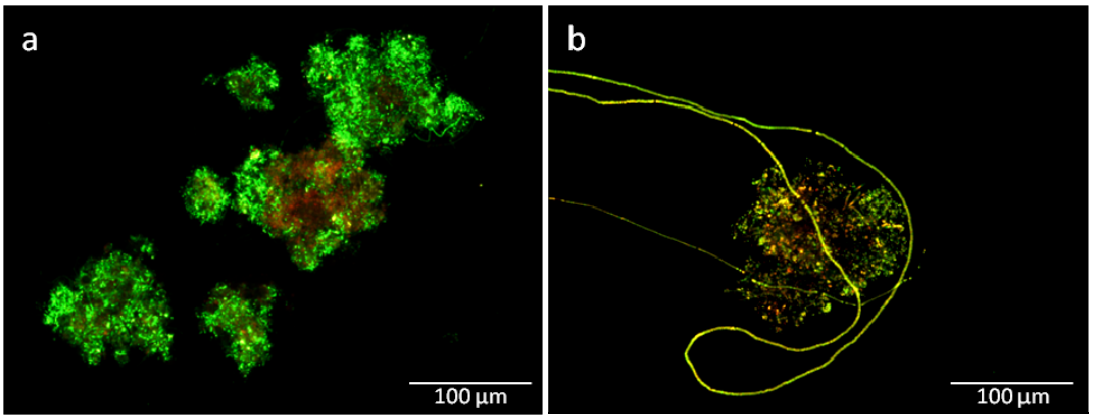

Figure 2.6. Images from epifluorescence microscopy (images obtained during this work): (a) Live/Dead® *Bac*Light™ (Green areas are live bacteria and red areas are damaged bacteria); (b) Live *Bac*Light™ Bacterial Gram (Green areas are negative bacteria and red areas are positive bacteria) (images obtained during this work).

2.4.3 CLSM MICROSCOPY

The confocal laser scanning microscopy (CLSM) is used for studying the microbial community structure. It is a technique where the light microscope has been transformed from an instrument able to view samples in two dimensions to one which can explore structures in three dimensions. Nowadays, CLSM is widely employed in studies with activated sludge especially to visualize and quantify FISH probed populations (Schmid *et al.*, 2003; Schuppler *et al.*, 1998; Larsen *et al.*, 2008; Davenport *et al.*, 2000). Other characteristics, such as the floc interior structure can also be studied using this technology (Chung and Lee, 2003; Chu *et al.*, 2005). Recently, CLSM was combined with fluorescent viability indicator (*Bac*Light® Bacterial Viability Kit, Molecular Probes) to study erythromycin toxicity in activated sludge (Louvet *et al.*, 2010).

2.5. IMAGE ANALYSIS PROCESS

Computers are key equipments for the analysis of large amounts of data, for tasks requiring complex computation, and for the extraction of quantitative information, opposite to the qualitative evaluation of human analysis. Today, the automatic analysis of numerical images captured by digital cameras enables to extract quickly quantitative information (Pons and Vivier, 1998). Thus, as a basic concept, image analysis is the extraction of significant information from images, by means of digital image processing techniques.

The diversity of digital image analysis applications are continuously growing through all areas of science including: medicine, such as dermoscopy (Blum *et al.*, 2008), radiography (Marijnissen *et al.*, 2008; Bould *et al.*, 1999) and endoscopy (Das *et al.*, 2008); chemistry, such as spectrophotometry (Vaughn *et al.*, 2009); biology (Berner and Paxson, 2003); and other fields of applications. In microscopy, Amaral (2003) reported several studies performed with digital image analysis which reside on the study of the morphology of fungal hyphae (Patankar *et al.*, 1993), yeasts characterization (Pons *et al.*, 1993; Kawasse *et al.*, 2003; Coelho *et al.*, 2004), enumeration and sizing of aquatic bacteria (David and Paul, 1989), colony texture analysis (Pons *et al.*, 1995), biomass and filamentous species determination (Pons and Vivier, 1998, 1999), fungus colonies biochemical and mycelium differentiation (Morrin and Ward, 1989), and mammalian cells (Pons *et al.*, 1992).

In image analysis, it is essential to take into account the previous phases of image capture/acquisition and image processing which are described below. These three steps are fundamental to improve image quality and extracting the most important information from images.

2.5.1 IMAGE ANALYSIS STEPS

2.5.1.1 IMAGE ACQUISITION

Image acquisition is performed through the use of digital cameras. The signal produced from the digital camera leads to the formation of a matrix of picture elements (called pixels) directly proportional to the light intensity received by each sensor (Amaral, 2003). The number of bits allocated to the pixels of any given image determines the number of colors in the image. For grayscale images it is common to represent each pixel in 8 bits corresponding to 256 grey levels. However, most real-world images are not monochrome, but full color. Light microscopes produce color images, and many preparation techniques make use of color, for example, to identify structure or localize chemical activity in novel ways. The use of color scales as a substitute for brightness values allows determining small changes locally, and to identify the same brightness values globally in an image (Russ, 2002). In fluorescence images, the color is an important factor and images are typically digitized as 24-bit RGB (red, green and blue channels), meaning that 8 bits or 256 (linear) levels of brightness for red, green, and blue are stored (Russ, 2002). Upon acquisition, images must be stored, where dozens of different storage formats are available for images. Different formats are able to store images in compact mode such as the CompuServe Bitmap format (GIF) or the Joint Photographers Expert Group format (JPEG) for single images and the Moving Pictures Expert Group format (MPEG and MPEG2) for video sequences. In image processing and analysis however this, although apparently small, loss of information is of critical importance and, therefore, not acceptable, leading to the use of the larger Tagged Image File format (TIFF) or Windows Bitmap format (BMP) (Amaral, 2003).

2.5.1.2 IMAGE PROCESSING

The purpose of the image processing step resides on obtaining a final image, usually grayscale or binary, holding significant information for a given application. First, images are pre-processed, the background is determined and removed or the

background light differences are minimized. Subsequently, segmentation is performed to separate the objects from the background. In order for segmentation to take place, a threshold value or values must be defined to allow the differentiation between the objects and the background. Obtaining a binary or mask image (pixel value of 1 for object and 0 for background) is a critical step, which should be automated as much as possible to avoid "subjectivity" (differences of appreciation between operators) (Pons and Vivier, 1999). A post-processing of the binary or mask images may be, at times, needed in order to solve some issues like the removal of border objects (cut-off by the borders of the image), removal of debris or the separation of touching objects, among several others (Amaral, 2003).

2.5.1.3 IMAGE ANALYSIS

This step is performed once the final images from the processing step are obtained. Usually, when working with binary images the most common parameters are related to the objects morphology such as equivalent diameter, area, number, perimeter, length, width, eccentricity, roundness, extent, convexity, compactness, and solidity. Sometimes, other parameters related to fractal dimensions are also taken into account when objects have highly irregular geometric shapes. However, the nature of the obtained parameters can differ significantly from application to application.

2.5.2 IMAGE ANALYSIS APPLICATION TO WASTEWATER

With a potential application in a number of different areas, unsurprisingly, image analysis has been increasingly used in wastewater treatment monitoring as a

complement tool for sludge characterization in aerobic and anaerobic processes, helping clarifying operational behavior and/or problems.

2.5.2.1 IMAGE ANALYSIS IN ANAEROBIC PROCESSES

The past two decades witnessed an extensive advance of anaerobic processes by the development of different types of reactors, mainly the upflow anaerobic sludge bed reactor (UASB) and the expanded granular sludge bed reactor (EGSB). A well-defined structured granular sludge allows for the enhancement of the system performance. Due to the granulation significance in anaerobic processes, research had been conducted to study different characteristics of granules, such as density, settling characteristics, strength and size, based on direct or indirect methods. However most of studies are mainly focused on size determinations based on image analysis (Bellouti *et al.*, 1997). Using image analysis procedures, Alves *et al.* (2000), studied anaerobic digesters characterizing aggregates and dispersed sludge, and quantifying the number and the length of free filaments. Image analysis was also applied to study the effect of increasing oleic acid concentration in anaerobic granulation and deterioration processes (Araya-Kroff *et al.*, 2004, Amaral *et al.*, 2004). Recently, image analysis procedures were applied to retrieve useful information on biomass activity recovery, organic loading disturbances, detergent shock loads, and organic solvents shock loads (Abreu *et al.*, 2007; Costa *et al.*, 2007, 2009a,b).

2.5.2.2 IMAGE ANALYSIS IN AEROBIC PROCESSES

An essential aspect of plant operation is the maintenance of activated sludge settling properties within an appropriate range. Sludge settling problems are largely related to morphological and surface characteristics of the flocs and to filamentous

bacteria proliferation. These problems are identified generally using the sludge volume index (SVI), which is the most used parameter to characterize the sludge settling ability. In fact, Sezgin *et al.* (1978, 1980) and Sezgin (1982) were the first authors demonstrating a strong relationship between SVI and filamentous bacteria contents.

A few settling problems leading to poor clarification and turbid effluents are due not only to filamentous bacteria proliferation, but due to bacteria inability to flocculate, named dispersed growth, or due to the break up of larger flocs and formation of quite small compact flocs which, however, do not settle properly (pin-point flocs). Results obtained for the floc structure and density revealed that the later is a function of their dispersion and those small flocs are denser than larger flocs (Li and Ganczarczyk, 1987). Later, Ganczarczyk (1994) established a correlation between the settling velocity with flocs size values, relating non-filamentous sludge settleability, in terms of SVI, to floc size and density. Moreover, a strong correlation between SVI and floc size and density was also observed by Andreadakis (1993).

The study of aerobic processes through the use of image analysis is performed mainly in activated sludge systems, which are the most frequently used in wastewater treatment. The relationship between the SVI and the sludge composition (aggregated biomass structure and contents and filamentous biomass contents) is considered of critical importance, making image analysis a powerful tool for biomass inspection and avoiding operator subjectivity.

Using image analysis techniques, Grijspeerdt and Verstraete (1997) found a good correlation between the form factor morphological parameter and the diluted sludge volume index (dSVI), and the possibility to estimate activated sludge concentration when the sludge is not too concentrated. Furthermore, Govoreanu *et al.* (2002) studied the most important parameters regarding the aggregated biomass characterization, aggregates size and shape. It is also known that

aggregates generated in activated sludge processes have some fractal features and several morphological properties of these aggregates can be characterized by fractal dimensions. The concept of fractal dimension was found applicable in the characterization of the aggregates' geometry, in the process of aggregation (Snidaro *et al.*, 1997; Li and Ganczarczyk, 1989, Thill *et al.*, 1998). Fractal dimension parameters were studied to find a threshold value in order to distinguish between the "weak" and "firm" flocs by Arelli *et al.* (2009), and a correlation between fractal dimensions and SVI was established. Further, a linear correlation between the relative abundance of filamentous species and SVI was also found by Arelli *et al.* (2009).

Cenens *et al.* (2002a,b), Banadda *et al.* (2005), and Jenné *et al.* (2002, 2003, 2004, 2005, 2007) also developed image analysis procedures for the activated sludge aggregated and filamentous biomass characterization. Banadda *et al.* (2005) established correlations between the SVI and the filamentous bacteria contents and aggregated biomass morphology. Following the works of Cenens *et al.* (2002a,b), which proposed an automatic thresholding algorithm for the classification of the aggregated and filamentous biomass, Jenné *et al.* (2002), studied the most important parameters to accurately classify the activated sludge biomass structure. Later Jenné *et al.* (2003, 2004, 2006), in lab-scale experiments, confirmed the relation between the filamentous bacteria abundance and the SVI values. Furthermore, and due to the recognition of significant changes in the floc shape due to bulking events, Jenné *et al.* (2005) characterized the debris fraction in the sludge, i.e. non-settleable material like very small flocs, short filaments, and small organic and inorganic waste material. This work further demonstrated the usefulness of image analysis information to assess the amount of suspended solids in the effluent.

The simultaneous characterization of the aggregated biomass structure and contents and filamentous bacteria contents was performed in full-scale WWTPs by Motta *et al.* (2001a,b, 2002, 2003) using image analysis procedures. The obtained results revealed the ability to detect filamentous bulking events based on the filamentous bacteria contents and allowed to establish a correlation between the settling ability properties and the filamentous bacteria contents and aggregates size. Following the work of these authors, full-scale activated sludge systems were also monitored in the last 5 years through the use of automated image analysis for aggregates and filaments recognition and characterization. Amaral and Ferreira (2005) and Mesquita *et al.* (2009a,b) found that automated image analysis allowed for the detection of bulking events, and established strong relationships between the sludge settling properties and image analysis parameters. Furthermore, a thorough analysis on the image acquisition methodology was performed, using samples from full-scale WWTPs, both in terms of the comparison between bright field and phase contrast microscopy (Mesquita *et al.,* 2010a), as well as on the effect of different dilutions and magnifications on the assessment of aggregated and filamentous bacterial contents and structure (Mesquita *et al.,* 2010b), in order to establish the best procedure.

Most of research described above was performed using monochrome images obtained from bright field or phase contrast microscopy. However, the use of color imaging is becoming more attractive because the use of color scales as a substitute for brightness values allows detecting small changes. This should be of great benefit, since these are among the goals of imaging methodologies (Russ, 2002). Currently image analysis procedures are being developed using color images for the detection of Gram staining characteristics of filamentous bacteria in activated sludge systems and storage of poly-β-hydroxybutyrate (PHB) in extended filamentous bacteria of activated sludge (Pandolfi and Pons, 2004; Pandolfi *et al.*, 2007). Also, fluorescent stains are used to quantify the relative abundance of

specific microorganisms, and to quantify the physiological state of these microorganisms. The study of Lopez *et al.* (2005) using epifluorescence microscopy, confocal laser scanning microscopy (CLSM), and two photon excitation laser scanning microscopy (TPE-LSM) revealed that for flocs with high cell density, the use of TPE-LSM is preferred, providing a clearer image of the internal structure of the aggregate. For typical activated sludge flocs epifluorescence and CLSM proved to be adequate.

2.6. CHEMOMETRICS IN WASTEWATER TREATMENT: PRINCIPAL COMPONENT ANALYSIS AND PARTIAL LEAST SQUARES

The term "chemometrics" is connected to mathematical, statistical, and chemical methods. Recent advances in computer and instrumentation techniques lead to the collection of large amounts of data from different processes that are best treated by the integration of these methods. Chemometrics tools can, therefore, be used for a wide variety of tasks, including the evaluation and interpretation of data, the optimization and the development of predictive models of processes and experiments, and finally the extraction of a maximum of information from the experimental data (Miller, 2005; Einax *et al.*, 1997). Nowadays, multivariate statistical analysis has become a very and important tool to extract useful information from datasets in order to improve process performance and product quality (Lee *et al.*, 2006).

In environmental analysis the processes and reactions are often non-stationary, irreversible, and take place in systems which are difficult to define, and impossible to describe using deterministic models (Einax *et al.*, 1997). The increasing use of chemometrics in environmental studies over the last two decades responds to the intensive research devoted to test and prove the power of data processing

techniques in this field. Three main areas of interest can be distinguished in chemometric environmental studies: quantitative chemical analysis, monitoring for environmental quality assessment and modeling and prediction of toxicological effects (Mas *et al.*, 2010). In WWTPs, and due to regulations for effluent quality, the detection of disturbances is very important for the enhancement of the process performance. Nowadays, the number of variables in WWTPs increased due to computerized measurement devices difficulting the interpretation of the dataset. Therefore, a more systematic way to handle and analyze data is needed to effectively extract relevant information for monitoring and supervision (Lee and Vanrolleghem, 2003). Multivariate statistical analysis has been already used to estimate influent chemical oxygen demand (COD) loads to a WWTP (Baeza *et al.*, 2002), allowing the differentiation between situations of low, normal and high influent loads, mainly based on oxygen uptake rate measurements in the activated sludge tanks.

Recently, the use of multivariate statistical analysis was found to be very important for WWTPs diagnosis in protozoa and metazoan identification. In fact, the recognition of these predators' organisms requires skilled operators and the identification task is labor and time consuming. For that purpose, Ginoris *et al.* (2007b,c) used image analysis associated to multivariate statistical analysis, such as discriminant analysis (DA), artificial neural networks (ANN) and decision trees (DT). In fact, ANN are considered to be very effective to capture non-linear relationships between variables in complex systems, and can be applied in situations where insufficient process knowledge is available to construct a model of the system (Gernaey *et al.*, 2004).

Some of the most applied multivariate statistical analysis in WWTPs operation, comprise principal component analysis (PCA) and partial least squares (PLS), given their ability to extract invaluable information regarding the state of the system

from large datasets. The work presented in this thesis addresses multidimensional data treatment using, primarily, these multivariate statistical methods, further described below.

2.6.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal component analysis (PCA) is among the most popular methods for extracting information from data, and has been applied in a wide range of fields. PCA is a multivariate statistical data analysis using projection into latent variables (LVs) to reduce high-dimensional and strongly correlated data to a much smaller dataset that can then be interpreted. PCA has been reported as a useful technique for determining correlations between large numbers of variables in a dataset. PCA acts by changing the high-dimensional datasets into low-dimensional datasets via the linear combination of variables constituting the original datasets. This approach is important for problems with a large number of input variables and features in chemical and biological processes. If some of the measured variables are linearly related or contaminated by noise, the first few components capture the relationship between the variables, and the remaining components are composed only of error. Thus, eliminating the less important components reduces the contribution of errors in the measured data and represents it in a compact manner (Bakshi, 1998).

Many researchers have already used PCA for wastewater treatment disturbances diagnosis (Rosen and Olsson, 1998; Tomita *et al.*, 2002; Yoo *et al.*, 2003; Choi and Lee, 2004; Lee *et al.*, 2004; Moon *et al.*, 2009; Aguado and Rosen, 2008) and for toxicity evaluation (Principi *et al.*, 2006). Furthermore, high-dimensional data from quantitative image analysis strengthen the need of multivariate statistical techniques for these purposes. Such was the case of Jenné *et al.* (2006) that applied PCA to find the parameters from image analysis datasets closer to the sludge

volume index (the most important parameters used for settling ability assessment). Recently, differentiation of detergent shock loads and/or organic loads using process operational parameters and image analysis datasets using PCA was studied by Costa *et al.* (2009c, 2010). PCA techniques have also been successfully applied for monitoring a wide range of wastewater treatment systems using spectroscopy (Dias *et al.*, 2008; Lourenço *et al.*, 2006, 2010).

2.6.2 PARTIAL LEAST SQUARES (PLS)

Partial least squares (PLS) is a linear multivariate statistical method particularly useful to predict a set of dependent (response) variables (**Y**) from a large set of independent (control) variables (**X**). In PLS regression, the model finds components (latent vectors) from **X** that are also relevant for **Y**, performing a simultaneous decomposition of **X** and **Y**. The regression method attempts that these components maximize the covariance between matrices **X** and **Y**. An optimal number of latent variables can be estimated by using cross-validation (CV) or separate test sets (Teppola *et al.*, 1998).

The study of Ferrer *et al.* (2008) revealed that with PLS approach it is possible to build predictive models for monitoring the performance of WWTPs, help in the diagnosis of a complex batch polymerization process, develop an automatic classifier based on image data, or assist in the empirical model building of a continuous polymerization process. A number of other attempts have been made to implement PLS modeling methodologies on WWTPs. Several applications are focused on predictions of quality parameters of the WWTP influent or effluent, and a considerable amount of the reported results are based on daily average values of the on-line measured variables combined with off-line measured variables

(Mujunen *et al.*, 1998; Lee *et al.*, 2006). PLS was found to predict deterioration of sludge settling ability properties (Mujunen *et al.*, 1998) and could easily isolate disturbances from the normal operating conditions (Teppola *et al.*, 1997). More recently, PLS approach was used to predict the sludge volume index in WWTPs during filamentous bulking events using only image analysis information (Amaral and Ferreira, 2005). PLS models were recently used for the estimation of several operational parameters using UV-Vis and near-infrared spectroscopy on a lab-scale activated sludge system and in wastewater treatment plants (Sarraguça *et al.*, 2009; Lourenço *et al.*, 2008, 2010).

The combination of different multivariate statistical methods is also commonly used. Pantsar-Kallio *et al.* (1999), combined PCA with PLS and this relationship proved to be an effective tool for analyzing and displaying the composition of sewage. The study showed that the pollutants from different sources could be characterized. The nature of sewage pollutants originating mainly from domestic waste was studied more closely and they were found to depend on many factors like the location of sampling areas, lifestyle of residents, day of week and sampling time. Singh *et al.* (2005) used chemometric techniques such as cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA) and partial least squares (PLS) to analyze the wastewater dataset and identify the factors which affect domestic sewage composition, spatial and temporal variations, and similarity and/or dissimilarity among wastewater characteristics.

2.7. REFERENCES

Abreu, A.A., Costa, J.C., Araya-Kroff, P., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for identifying structural changes during a revival process of anaerobic granular sludge. *Water Research* **41**, 1473-1480.

Agridiotis, V., Forster, C.F., and Carliell-Marquet, C. (2007) Addition of Al and Fe salts during treatment of paper mill effluents to improve activated sludge settlement characteristics. *Bioresource Technology* **98**, 2926-2934.

Aguado, D. and Rosen, C. (2008) Multivariate statistical monitoring of continuous wastewater treatment plants. *Engineering Applications of Artificial Intelligence* **21**, 1080-1091.

Alves, M.M., Cavaleiro, A.J., Ferreira, E.C., Amaral, A.L., Mota, M., da Motta, M., Vivier, H., and Pons, M.N. (2000) Characterization by image analysis of anaerobic sludge under shock conditions. *Water Science and Technology* **41** (12), 207-214.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990) Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations. *Applied and Environmental Microbiology* **56** (6), 1919-1925.

Amaral, A.L. (2003) *Image analysis in biotechnological processes: applications to wastewater treatment.* PhD Thesis, University of Minho, Braga, Portugal. (http://hdl.handle.net/1822/4506)

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

Amaral, A.L., Pereira, M.A., da Motta, M., Pons, M.-N., Mota, M., Ferreira, E.C., and Alves, M.M. (2004) Development of image analysis techniques as a tool to detect and quantify morphological changes in anaerobic sludge: II. Application to a granule deterioration process triggered by contact with oleic acid. *Biotechnology and Bioengineering* **87** (2), 194-199.

Andreadakis, A. (1993) Physical and chemical properties of activated sludge flocs. *Water Research* **12**, 1707-1714.

Araya-Kroff, P., Amaral, A.L., Neves, L., Ferreira, E.C., Pons, M.N., Mota, M., and Alves, M.M. (2004) Development of image analysis techniques as a tool to detect and quantify morphological changes in anaerobic sludge: I. Application to a granulation process. *Biotecnhnolgy and Bioengineering* **87** (2), 184-193.

Arelli, A., Luccarini, L., and Madoni, P. (2009) Application of image analysis in activated sludge to evaluate correlations between settleability and features of flocs and filamentous species. *Water Science and Technology* **59** (10), 2029-2036.

Baeza, J.A., Gabriel, D., and Lafuente, J. (2002) In-line fast OUR (oxygen uptake rate) measurements for monitoring and control of WWTP. *Water Science and Technology* **45** (4-5), 19-28.

Bakshi, B.R. (1998) Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal* **44** (7), 1596-1610.

Banadda, E.N., Smets, I.Y., Jenné, R., Van and Impe, J.F. (2005) Predicting the onset of filamentous bulking in biological wastewater treatment systems by exploiting image analysis information. *Bioprocess and Biosystems Engineering* **27**, 339-348.

Barbusinski, K. and Koscielniak, H. (1995) Influence of substrate loading intensity on floc size in activated sludge process. *Water Research* **29** (7), 1703-1710.

Bellouti, M., Alves, M.M., Novais, J.M., and Mota, M. (1997) Flocs vs granules: differentiation by fractal dimension. *Water Research* **31** (5), 1227-1231.

Berner, D.K. and Paxson, L.K. (2003) Use of digital images to differentiate reactions of collections of yellow starthistle (*Centaurea solstitialis*) to infection by *Puccinia jaceae. Biological Control* **28**, 171-179.

Blackall, L.L., Seviour, E.M., Bradford, D., Rossetti, S., Tandoi, V., and Seviour, R.J. (2000) '*Candidatus Nostocoida limicola*', a filamentous bacterium from activated sludge. *International Journal of Systematic Bacteriology* **50**, 703-709.

Blackall, L.L., Stratton, H., Bradford, D., Dot, T.D., Sjörup, C., Seviour, E.M., and Seviour, R.J. (1996) "*Candidatus Microthrix parvicella*," a Filamentous Bacterium from Activated Sludge Sewage Treatment Plants. *International Journal of Systematic Bacteriology* **46** (1), 344-346.

Blum, A., Zalaudek, I., and Argenziano, G. (2008) Digital Image Analysis for Diagnosis of Skin Tumors. *Seminars in Cutaneous Medicine and Surgery* **27**, 11-15.

Bould, M., Barnard, S., Learmonth, I.D., Cunningham, J.L., and Hardy, J.R.W., (1999) Digital image analysis: improving accuracy and reproducibility of radiographic measurement. *Clinical Biomechanics* **14**, 434-437.

Boulos, L., Prevost, M., Barbeau, B., Coallier, J., and Desjardins, R. (1999) LIVE/DEAD® *Bac*Light™: application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. *Journal of Microbiological Methods* **37**, 77-86.

Cenens, C., Van Beurden, K.P., Jenné, R., and Van Impe, J.F. (2002a) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **46** (1-2), 381-387.

Cenens, C., Jenné, R., and Van Impe, J.F. (2002b) Evaluation of different shape parameters to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **45** (4-5), 85-91.

Chang, W.C., Jou, S.J., Chien, C.C., and He, J.A. (2004) Effect of chlorination bulking control on water quality and phosphate release/uptake in an anaerobic-oxic activated sludge system. *Water Science and Technology* **50** (8), 177-183.

Choi, S.W. and Lee, I.B. (2004) Nonlinear dynamic processmonitoring based on dynamic kernel PCA. *Chemical Engineering Science* **59**, 5897-5908.

Chu, C.P. and Lee, D.J. (2004) Multiscale structures of biological flocs. *Chemical Engineering Science* **59**, 1875-1883.

Chu, C.P., Tsai, D.G., Lee, D.J., and Tay, J.H. (2005) Size-dependent anaerobic digestion rates of flocculated activated sludge: Role of intrafloc mass transfer resistance. *Journal of Environmental Management* **76**, 239-244.

Chung, H.Y. and Lee, D.J. (2003) Porosity and interior structure of flocculated activated sludge floc. *Journal of Colloid and Interface Science* **267**, 136-143.

Coelho, M.A.Z., Belo, I., Pinheiro, R., Amaral, A.L., Mota, M., Coutinho, J.A.P., and Ferreira, E.C. (2004) Effect of hyperbaric stress on yeast morphology: study by automated image analysis. *Applied Microbiology and Biotechnology* **66**, 318-324.

Costa, J.C., Abreu, A.A., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for monitoring structural changes of anaerobic granular sludge during detergent shock loads. *Biotechnology and Bioengineering* **98** (1), 60-68.

Costa, J.C., Moita, I., Abreu, A.A., Ferreira, E.C., and Alves, M.M. (2009a) Advanced monitoring of high-rate anaerobic reactors through quantitative image analysis of granular sludge and multivariate statistical analysis. *Biotechnology and Bioengineering* **102** (2), 445-456.

Costa, J.C., Moita, I., Ferreira, E.C., and Alves, M.M. (2009b) Morphology and physiology of anaerobic granular sludge exposed to an organic solvent. *Journal of Hazardous Materials* **167**, 393-398.

Costa, J.C., Alves, M.M., and Ferreira, E.C. (2009c) Principal component analysis and quantitative image analysis to predict effectsof toxics in anaerobic granular sludge. *Bioresource Technology* **100**, 1180-1185.

Costa, J.C. Alves, M.M., and Ferreira, E.C. (2010) A chemometric tool to monitor high-rate anaerobic granular sludge reactors during load and toxic disturbances. *Biochemical Engineering Journal* **53** (1), 38-43.

da Motta, M., Pons, M.N., and Roche, N. (2002) Study of filamentous bacteria by image analysis and relation with settleability. *Water Science and Technology* **46** (1-2), 363-369.

da Motta, M., Pons, M.N., Roche, N., and Vivier, H. (2001a) Characterization of activated sludge by automated image analysis. *Biochemical Engineering Journal* **9**, 165-173.

da Motta, M., Pons, M.N., and Roche, N. (2001b) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* **43** (7), 91-96.

da Motta, M., Pons, M.N., and Roche, N. (2003) Monitoring filamentous bulking in activated sludge systems fed by synthetic or municipal wastewater. *Bioprocess and Biosystem Engineering* **25**, 387-393.

Das, A., Nguyen, C.C., Li, F., and Li, B. (2008) Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. *Clinical Endoscopy* **67** (6), 861-867.

Davenport, R.J., Curtis, T.P., Goodfellow, M., Stainsby, F.M., and Bingley, M. (2000) Quantitative use of fluorescent *in situ* hybridization to examine relationships between mycolic acid-containing actinomycetes and foaming in activated sludge plants. *Applied and Environmental Microbiology* **66** (3), 1158-1166.

David, A.W. and Paul, J.H. (1989) Enumeration and sizing of aquatic bacteria by use of a silicon-intensified target camera linked image analysis system. *Journal of Microbiological Methods* **9**, 257-266.

Dias, A.M.A., Moita, I., Páscoa, R., Alves, M.M., Lopes, J.A., and Ferreira, E.C. (2008) Activated sludge process monitoring through in-situ NIR spectral analysis. *Water Science and Technology* **57** (10), 1643-1650.

de los Reyes, F.L., Ritter, W., and Raskin, L. (1997) Group-specific small-subunit rRNA hybridization probes to characterise filamentous foaming in activated sludge systems. *Applied and Environmental Microbiology* **63**, 1107-1117.

Eikelboom, D.H. (2000) *Process control of activated sludge plants by microscopic investigation.* IWA Publishing, London, England.

Einax, J.W., Zwanziger, H.W., and Geiss, S. (1997) *Chemometrics in Environmental Analysis*. VCH, Weinheim, Germany.

Erhart, R., Bradford, D., Seviour, R.J., Amann, R., and Blackall, L.L. (1997) Development and use of fluorescent *in situ* hybridization probes for the detection and identification of *''Microthrix parvicella''* in activated sludge. *Systematic and Applied Microbiology* **20**, 310-318.

Ferrer, A., Aguado, D., Vidal-Puig, S., Prats, M., and Zarzo, M. (2008) PLS: A versatile tool for industrial process improvement and optimization. *Applied stochastic models in business and industry* **24**, 551-567.

Foster, S., Snape, J.R., Lappin-Scott, H.M., and Porter, J. (2002) Simultaneous Fluorescent Gram Staining and Activity Assessment of Activated Sludge Bacteria. *Applied and Environmental Microbiology* **68** (10) 4772-4779.

Frølund, B., Griebe, T., and Nielsen, P.H. (1995) Enzymatic activity in the activated-sludge floc matrix. *Applied Microbiology and Biotechnology* **43**, 755-761.

Ganczarczyk, J. (1994) Microbial aggregates in wastewater treatment. *Water Science and Technology* **30** (8), 87-95.

Gaval, G. and Pernelle, J.J. (2003) Impact of the repetition of oxygen deficiencies on the filamentous bacteria proliferation in activated sludge. *Water Research* **37**, 1991-2000.

Gernaey, K.V., van Loosdrecht, M.C.M., Henze, M., Lind, M., and Jørgensen, S.B. (2004) Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software* **19**, 763–783.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007a) Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system. *Water Research* **41**, 2581-2589.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007b) Raw data pre-processing in the protozoa and metazoa identification by image analysis and multivariate statistical techniques. *Journal of Chemometrics* **21**, 156-164.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007c) Recognition of protozoa and metazoa using image analysis tools, discriminant analysis, neural networks and decision trees. *Analytica Chimica Acta* **595**, 160-169.

Govoreanu, R., Vandegehuchte, K., Saveyn, H., Nopens, I., De Clercq, B., van der Meeren, P., and Vanrolleghem, P. (2002) An automated image analysis system for on-line structural characterization of the activated sludge flocs. *Med. Fac. Landbouww. Univ. Gent* **67** (4), 75-178.

Grijspeerdt, K. and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31** (5), 1126-1134.

Invitrogen Molecular Probes (2001) *LIVE BacLight™ Bacterial Gram Kits*. Manuals and Product Inserts. http://probes.invitrogen.com/media/pis/mp07008.pdf.

Invitrogen Molecular Probes (2004) *LIVE/DEAD® BacLight™ Bacterial Viability Kits*. Manuals and Product Inserts. http://probes.invitrogen.com/media/pis/mp07007.pdf.

Invitrogen Molecular Probes (2006) *DAPI Nucleic Acid Stain*. Manuals and Product Inserts.

Jarvis, P., Jefferson, B., Gregory, J., and Parsons, S.A. (2005) A review of floc strength and breakage. *Water Research* **39**, 3121-3137.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Cenens, C., Geeraerd, A.H., and Van Impe, J.F. (2002) Towards on-line quantification of flocs and filaments by image analysis. *Biotechnology Letters* **24**, 931-935.

Jenné, R., Banadda, E.N., Philips, N., and Van Impe, J.F. (2003) Image Analysis as a Monitoring Tool for Activated Sludge Properties in Lab-Scale Installations. *Journal of Environmental Science and Health Part A—Toxic/Hazardous Substances & Environmental Engineering* **38** (10), 2009-2018.

Jenné, R., Banadda E.N., Smets, I.Y., and Van Impe, J.F. (2004) Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* **50** (7), 281-285.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jenné, R., Banadda, E.N., Smets, I.Y., Bamelis, A., Verdickt, L., and Van Impe, J.F. (2005) Activated sludge image analysis system: monitoring settleability and effluent clarity. *Water Science and Technology* **52** (10-11), 193-199.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Jin, B., Wilén, B.M., and Lant, P. (2003) A comprehensive insight into floc characteristics and their impact on compressibility and settleability of activated sludge. *Chemical Engineering Journal* **95**, 221-234.

Jin, B., Wilén, B.M., and Lant, P. (2004) Impacts of morphological, physical and chemical properties of sludge flocs on dewaterability of activated sludge. *Chemical Engineering Journal* **98**, 115-126.

Jones, P.A. and Schuler, A.J., (2010) Seasonal variability of biomass density and activated sludge settleability in full scale wastewater treatment systems. *Chemical Engineering Journal* **164**, 16-22.

Kanagawa, T., Kamagata, Y., Aruga, S., Kohno, T., Horn, M., and Wagner, M. (2000) Phylogenetic analysis of and oligonucleotide probe development for eikelboom type 021N filamentous bacteria isolated from bulking activated sludge. *Applied and Environmental Microbiology* **66**, 5043-5052.

Kawasse, F.M., Amaral, P.F., Rocha-Leão, M.H.M., Amaral, A.L., Ferreira, E.C., and Coelho, M.A.Z. (2003) Morphological analysis of Yarrowia lipolytica under stress conditions through image processing. *Bioprocess and Biosystems Engineering* **25**, 371-375.

Larsen, P., Nielsen, J.L., Otsen, D., and Nielsen, P.H. (2008) Amyloid-like adhesions produced by floc-forming and filamentous bacteria in activated sludge. Applied and *Environmental Microbiology* **74** (5), 1517-1526.

Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y.J., and Park, J.M. (2006) Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process Biochemistry* **41**, 2050-2057.

Lee, D.S. and Vanrolleghem, P.A., (2003) Monitoring of a Sequencing Batch Reactor Using Adaptive Multiblock Principal Component Analysis. *Biotechnology and Bioengineering* **82** (4), 489-497.

Lee, J.M., Yoo, C., and Lee, I.B. (2004) Statistical process monitoring with independent component analysis. *Journal of Process Control* **14**, 467-485.

Li, D. and Ganczarczyk, J. (1987) Stroboscopic determination of settling velocity, size and porosity of activated sludge flocs. *Water Research* **21**, 257-262.

Li, D. and Ganczarczyk J. (1989) Fractal geometry of particle aggregates generated in water and wastewater treatment processes. *Environmental Science Technology* **23**, 1385-1389.

Li, D. and Ganczarczyk, J. (1991) Size distribution of activated sludge flocs. *Journal of Water Pollution Control Federation* **63**, 806-814.

Li, D. and Ganczarczyk, J.J. (1990) Structure of activated sludge flocs. *Biotechnology and Bioengineering* **35**, 57-65.

Li, X.Y. and Yang, S.F. (2007) Influence of loosely bound extracellular polymeric substances (EPS) on the flocculation, sedimentation and dewaterability of activated sludge. *Water Research* **41** (5),1022-1030.

Liao, B.Q., Droppo, I.G., Leppard, G.G., and Liss, S.N. (2006) Effect of solids retention time on structure and characteristics of sludge flocs in sequencing batch reactors. *Water Research* **40** (13), 2583–2591.

Lindea, K.C., Seviour, E.M., Seviour, R.J., Blackall, L.L., and Soddell, J.A. (1999) *Practical methods for the examination and characterization of activated sludge*. In Seviour R.J., Blackall L.L. (Ed.) *The microbiology of activated sludge*, Kluwer Academic Publishers, London.

Liu, H. and Fang, H.H.P. (2002) Extraction of extracellular polymeric substances (EPS) of sludges. *Journal of Biotechnology* **95**, 249-256.

Liu, J.R., McKenzie, C.A., Seviour, E.M., Webb, R.I., Blackall, L.L., Saint, C.P., and Seviour, R.J. (2001) Phylogeny of the filamentous bacterium *'Nostocoida limicola'* III from activated sludge. *International Journal of Systematic Bacteriology* **51**, 195-202.

Liu, Y. and Liu, Q.S. (2006) Causes and control of filamentous growth in aerobic granular sludge sequencing batch reactors. *Biotechnology Advances* **24**, 115-127.

Liu, Y. and Fang, HP. (2003) Influence of extracellular polymeric substances on flocculation, settling and dewatering of activated sludge. *Critical Reviews in Environmental Science and Technology* **33** (3), 237-273.

Lopez, C., Pons, M.N., and Morgenroth, E. (2005) Evaluation of microscopic techniques (epifluorescence microscopy, CLSM, TPE-LSM) as a basis for the quantitative image analysis of activated sludge. *Water Research* **39**, 456-468.

Lourenço, N.D., Chaves, C.L., Novais, J.M., Menezes, J.C., Pinheiro, H.M., and Diniz, D. (2006) UV spectra analysis for water quality monitoring in a fuel park wastewater treatment plant. *Chemosphere* **65** (5), 786-791.

Lourenço, N.D., Menezes, J.C., Pinheiro, H.M., and Diniz, D. (2008) Development of PLS calibration models from UV-vis spectra for TOC estimation at the outlet of a fuel park wastewater treatment plant. *Environmental Technology* **29** (8), 891-898.

Lourenço, N.D., Paixão, F., Pinheiro, H.M., and Sousa, A. (2010) 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. *Applied Spectroscopy* **64** (9), 1061-1067.

Louvet, J.N., Heluin, Y., Attik, G., Dumas, D., Potier, O., and Pons, M.N. (2010) Assessment of erythromycin toxicity on activated sludge via batch experiments and microscopic techniques (epifluorescence and CLSM). *Process Biochemistry* **45** (11), 1787-1794.

Madoni, P. (2005) *Depurazione biologica nei fanghi attivi*. Enìa - Università degli Studi di Parma.

Marijnissen, A.C.A., Vincken, K.L., Vos, P.A.J.M., Saris, D.B.F., Viergever, M.A., Bijlsma, J.W.J., Bartels, L.W., and Lafeber, F.P.J.G. (2008) Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. *Osteoarthritis and Cartilage* **16**, 234-243.

Martins, A.M.P., Karahan, O., and van Loosdrecht, M.C.M. (2010) Effect of polymeric substrate on sludge settleability. *Water Research* **45** (1), 263-273.

Martins, A.M.P., Pagilla, K., Heijnen, J.J., and van Loosdrecht, M.C.M. (2004) Filamentous bulking sludge - a critical review. *Water Research* **38**, 793-817.

Martins, A.M.P., van Loosdrecht, M.C.M., and Heijnen, J.J. (2003) Effect of dissolved oxygen concentration on the sludge settleability. *Applied Microbiology Biotechnology* **62**, 586-593.

Mas, S., de Juan, A., Tauler, R., Olivieri, A., and Escandar, G. (2010) Application of chemometric methods to environmental analysis of organic pollutants: A review. *Talanta* **80**, 1052-1067.

Mesquita, D.P., Dias, O., Dias, A.M.A., Amaral, A.L., and Ferreira, E.C. (2009a) Correlation between sludge settleability and image analysis information using partial least squares. *Analytica Chimica Acta* **642** (1-2), 94-101.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2009b) Monitoring of activated sludge settling ability through image analysis: validation on full-scale wastewater treatment plants. *Bioprocess and Biosystems Engineering* **32** (3), 361-367.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2010a) A comparison between bright field and phase contrast image analysis techniques in activated sludge morphological characterization. *Microscopy and Microanalysis* **16** (2), 166-174.

Mesquita, D.P., Dias, O., Elias, R.A.V., Amaral, A.L., and Ferreira, E.C. (2010b) Dilution and magnification effects on image analysis applications in activated sludge characterization. *Microscopy and Microanalysis* **16** (5) 561-568.

Metcalf and Eddy (2003) *Wastewater engineering - treatment and reuse*. Fouth edition, McGraw-Hill international editions, New York.

Miller, C.E. (2005) *Chemometrics in Process Analytical Chemistry*. In Bakeev, A. *Process analytical technology*. Blackwell Publishing, Oxford, UK.

Moon, T.S., Kim, Y.J., Kim, J.R., Cha, J.H., Kim, D.H., and Kim, C.W. (2009) Identification of process operating state with operational map in municipal wastewater treatment plant. *Journal of Environmental Management* **90** (2), 772-778.

Morrin, M. and Ward, O.P. (1989) Studies on interaction of carbopol-934 with the hyphae of *Rhizopus arrhizus*. *Mycological Research* **92**, 265-272.

Mujunen, S.P., Minkkinen, P., Teppola, P., and Wirkkala, R.S. (1998) Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. *Chemometrics and Intelligent Laboratory Systems* **41** (1), 83-94.

Nielsen, P.H., Daims, H., and Lemmer, H. (2009) *FISH Handbook for Biological Wastewater Treatment: Identification and quantification of microorganisms in activated sludge and biofilms by FISH.* IWA Publishing, London, UK.

Novak, L., Larrea, L., Wanner, J., and Garcia-Heras, J.L. (1993) Non-filamentous activated sludge bulking in a laboratory scale system. *Water Research* **27** (8), 1339-1346.

Palm, J.C., Jenkins, D., and Parker, D.S. (1981) The relationship between organic loading, dissolved oxygen concentration and sludge settleability in the completely mixed activated sludge process. *Journal of Water Pollution Control Federation* **52**, 2484-2506.

Pandolfi, D. and Pons, M.N. (2004) Gram-staining characterization of activated sludge filamentous bacteria by automated colour analysis. *Biotechnology Letters* **26**, 1841-1846.

Pandolfi, D., Pons, M.N., and da Motta, M. (2007) Characterization of PHB storage in activated sludge extended filamentous bacteria by automated colour image analysis. *Biotechnology Letters* **29**, 1263-1269.

Pantsar-Kallio, M., Mujunen, S.P., Hatzimihalis, G., Koutoufides, P., Minkkinen, P., Wilkie, P.J., and Connor, M.A. (1999) Multivariate data analysis of key pollutants in sewage samples: a case study. *Analytica Chimica Acta* **393**, 181-191.

Patankar, D.B., Liu, T.C., and Oolman, T. (1993) A fractal model for the characterization of mycelial morphology. *Biotechnology and Bioengineering* **42**, 571-578.

Pons, M.N., Mona, H., Drouin, J.F., and Vivier, H. (1995) Texture characterization of colonies on solid substrate. *Proceedings of the 6th international conference on computer applications in biotechnology*, 189-194.

Pons, M.N., and Vivier, H. (1998) Beyond filamentous species. *Advances in Biochemical Engineering/Biotechnology* **60**, 61-95.

Pons, M.N., and Vivier, H. (1999) Biomass quantification by image analysis. *Advances in Biochemical Engineering/Biotechnology* **66**, 133-184.

Pons, M.N., Vivier, H., Rémy, J.F., and Dodds, J.A. (1993) Morphological characterization of yeast by image analysis. *Biotechnology and Bioengineering* **42**, 1352-1359.

Pons, M.N., Wagner, A., Vivier, H., and Marc, A. (1992) Application of quantitative image analysis to a mammalian cell line grown on microcarriers. *Biotechnology and Bioengineering* **40** (1) 187-193.

Principi, P., Villa, F., Bernasconi, M., and Zanardini, E. (2006) Metal toxicity in municipal wastewater activated sludge investigated by multivariate analysis and *in situ* hybridization. *Water Research* **40** (1), 99-106.

Richard, M. (1989) *Activated sludge microbiology*. The Water Pollution Control Federation, Alexandria, Virginia.

Rosen, C. and Olsson, G. (1998) Disturbance detection in wastewater treatment plants. *Water Science and Technology* **37** (12), 197-205.

Rosseti, S., Tomei, M.C., Nielson, P.H., and Tandoi, V. (2005) "*Microthrix parvicella*", a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. *FEMS Microbial Reviews* **29** (1), 49-64.

Rossetti, S., Christensson, C., Blackall, L.L., and Tandoi, V. (1997) Phentypic and phylogenetic description of an Italian isolate of "*Microthrix parvicella*". *Journal of Applied Microbiology* **82**, 405-410.

Russ J.C. (2002) *The Image Processing Handbook*. Fourth edition, CRC Press, Boca Raton.

Salton, M.R.J. (1964) *The Bacterial Cell Wall*. Elsevier, Amsterdam, The Netherlands.

Sanz, J.L. and Köchling, T. (2007) Molecular biology techniques used in wastewater treatment: An overview. *Process Biochemistry* **42**, 119-133.

Sarraguça, M.C., Paulo, A., Alves, M.M., Dias, A.M.A., Lopes, J.A., and Ferreira, E.C. (2009) Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy. *Analytical and Bioanalytical Chemistry* **395**, 1159-1166.

Schmid, M., Thill, A., Purkhold, U., Walcher, M., Bottero, J.Y., Ginestet, P., Nielsen, P.H., Wuertz, S., and Wagner, M. (2003) Characterization of activated sludge flocs by confocal laser scanning microscopy and image analysis. *Water Research* **37** (9), 2043-2052.

Schuler, A.J. and Jang, H. (2007a) Causes of variable biomass density and its effects on settling in full scale biological wastewater treatment systems. *Environmental Science and Technology* **41** (5), 1675-1681.

Schuler, A.J. and Jang, H. (2007b) Density effects on activated sludge zone settling velocities. *Water Research* **41** (8), 1814-1822.

Schuppler, M., Wagner, M., Schön, G., and Göbel, U.B. (1998) *In situ* identification of nocardioform actinomycetes in activated sludge using fluorescent rRNA-targeted oligonucleotide probes. *Microbiology* **144**, 249-259.

Seviour, E.M., Eales, K., Izzard, L., Beer, M., Carr, E.L., and Seviour, R.J. (2006) The *in situ* physiology of "*Nostocoida limicola*" II, a filamentous bacterial morphotype in bulking activated sludge, using fluorescence *in situ* hybridization and microautoradiography. *Water Science and Technology* **54** (1), 47-53.

Seviour, R. and Nielsen, P.H. (2010) *Microbial ecology of activated sludge*. IWA Publishing, London, UK.

Seviour, R.J. and Blackall, L.L. (1999) *The microbiology of activated sludge*. Kluwer academic publishers.

Sezgin, M. (1982) Variation of sludge volume index with activated sludge characteristics. *Water Research* **16**, 83-88.

Sezgin, M., Jenkins, D., and Palm, J.C. (1980) Floc size, filament length and settling properties of prototype activated sludge plants. *Progress in Water Technology* 12, 171-182.

Sezgin, M., Jenkins, D., and Parker, D.S. (1978) A unified theory of filamentous activated sludge bulking. *Journal of Water Pollution Control Federation* **50**, 362-381.

Simpson, J.M., Stroot, P.G., Gelman, S., Beydilli, I., Dudley, S., Oerther, D.B. (2006) 16S ribosomal RNA tools identify an unexpected predominance of Paenibacillus-like bacteria in an industrial activated sludge system suffering from poor biosolids separation. *Water Environmental Research* **78** (8), 864-871.

Singh, K.P., Malik, A., Mohan, D., Sinha, S., and Singh, V.K. (2005) Chemometric data analysis of pollutants in wastewater-a case study. *Analytica Chimica Acta* **532**, 15-25.

Snidaro, D., Zartarian, F., Jorand, F., Bottero, J.Y., Block, J.C., and Manem, J. (1997) Characterization of activated sludge flocs structure. *Water Science and Technology* **36** (4) 313-320.

Teppola, P., Mujunen, S.P., and Minkkinen, P. (1997) Partial least squares modeling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems* **38**, 197-208.

Teppola, P., Mujunen, S.P., and Minkkinen, P. (1998) A combined approach of partial least squares and fuzzy c-means clustering for the monitoring of an activated-sludge waste-water treatment plant. *Chemometrics and Intelligent Laboratory Systems* **41**, 95-103.

Thill, A., Veerapaneni, S., Simon, B., Wiesner, M., Bottero, J.Y., and Snidaro, D. (1998) Determination of structure of aggregates by confocal scanning laser microscopy. *Journal of Colloid and Interface Science* **204**, 357-362.

Tomita, R.K., Park, S.W., and Stomayor, O.A.Z. (2002) Analysis of activated sludge process using multivariate statistical tools - a PCA approach. *Chemical Engineering Journal* **90**, 283-290.

Tsai, M.W., Wentzel, M.C., and Ekama, G.A. (2003) The effect of residual ammonia concentration under aerobic conditions on the growth of *Microthrix parvicella* in biological nutrient removal plants. *Water Research* **37**, 3009-3015.

Urbain, V., Block, J.C., and Manem, D.J. (1993) Biofloculation in activated sludge: an analytical approach. *Water Research* **27** (5), 829-838.

Vaughn, M.R., van Oorschot, R.A.H., and Baindur-Hudson, S. (2009) A comparison of hair colour measurement by digital image analysis with reflective spectrophotometry. *Forensic Science International* **183**, 97-101.

Wagner, M., Amann, R., Kämpfer, P., Assmus, B., Hartmann, A., Hutzler, P., Springer, N., and Schleifer, K.H. (1994) Identification and *in situ* detection of gram-negative filamentous bacteria in activated sludge. *Systematic and Applied Microbiology* **17**, 405-417.

Wilén, B.M. and Balmer, P. (1999) The effect of dissolved oxygen concentration on the structure, size and size distribution of activated sludge flocs. *Water Research* **33** (2), 391-400.

Wilén, B.M., Jin, B., and Lant, P. (2003) The influence of key chemical constituents in activated sludge on surface and flocculating properties. *Water Research* **37**, 2127-2139.

Wilén, B.M., Lumley, D., Mattsson, A., and Mino, T. (2008) Relationship between floc composition and flocculation and settling properties studied at a full scale activated sludge plant. *Water Research* **42**, 4404-4418.

Wilén, B.M., Nielsen, J.L., Keiding, K., and Nielsen, P.H. (2000) Influence of microbial activity on the stability of activated sludge flocs. *Colloids and Surfaces B: Biointerfaces* **18**, 145-156.

Wilén, B.M., Onuki, M., Hermansson, M., Lumley, D., and Mino, T. (2008) Microbial community structure in activated sludge floc analysed by fluorescent *in situ* hybridization and its relation to floc stability. *Water Research* **42**, 2300-2308.

Yoo, C.K., Vanrolleghem, P.A., and Lee, I.B. (2003) Nonlinear modeling and adaptive monitoring with fuzzy and multivariate statistical methods in biological wastewatertreatment plants. *Journal of Biotechnology* **105**, 135-163.

Abstract

Different approaches using microscopy image analysis procedures were employed for activated sludge characterization. The approaches varied mainly on the type of visualization and acquisition methods. In this context, this study focused on the comparison of the two most common acquisition methods: bright field and phase contrast microscopy. Images were acquired from seven different wastewater treatment plants for a combined period of two years. Advantages and disadvantages of each acquisition technique are discussed in this chapter.

3. CHAPTER 3 – COMPARISON BETWEEN BRIGHT FIELD AND PHASE CONTRAST IMAGE ANALYSIS TECHNIQUES IN ACTIVATED SLUDGE MORPHOLOGICAL CHARACTERIZATION

3.1. INTRODUCTION

An activated sludge system includes a complex ecosystem composed of different types of microorganisms. A good balance between the different microorganisms is essential to guarantee good settling properties and a clear supernatant (Jenkins *et al.*, 2003). Generally, the evaluation of aerated tanks may be performed by visual inspection under an optical microscope coupled to automated image analysis methods. Activated sludge processes have been increasingly monitored through microscopic observations for aggregate contents and morphology and determination of protruding filamentous bacteria contents (da Motta *et al.*, 2001a; Jenné *et al.*, 2006). Subsequently, the gathered image analysis information was correlated with the sludge settling abilities (Ganczarczyk, 1994; Grijspeerdt and Verstraete, 1997; Banadda *et al.*, 2005), to assess biomass morphology changes (Jenné *et al.*, 2003) and to monitor bulking events in pilot plants (Jenné *et al.*, 2004, 2007; Amaral and Ferreira, 2005; da Motta *et al.*, 2001a,b).

The use of automated image analysis applications in activated sludge has increased in recent years, with two image acquisition methods standing out: phase contrast microscopy as proposed in the works of Cenens *et al.* (2002) and Jenné *et al.* (2006, 2007) among others; and bright field microscopy as in the works of da Motta *et al.* (2001a,b), Amaral and Ferreira (2005) and Mesquita *et al.* (2009a,b). In comparison, bright field microscopy is the cheapest and simplest method to examine activated sludge, whereas phase contrast microscopy requires more expensive equipment and a more skilled operator. Furthermore, the nature of phase contrast microscopy causes the aggregate borders to become ill-defined as the object's halo hinders the assessment of their boundaries. However, this method presents, at least in theory,

the advantage of a more precise determination of the protruding filamentous bacteria contents. Higher transparency of the filamentous bacteria may pose a contrast problem in bright field microscopy acquisition, regarding the clearer filament/dark background distinction in phase contrast microscopy. As a result of that previous studies have been performed using bright field acquisition methods to survey the aggregated biomass and phase contrast acquisition for the assessment of filamentous bacteria (Amaral, 2003; Abreu *et al.*, 2007; Costa *et al.*, 2007).

Based on the reported advantages and disadvantages of both methodologies, the present work aims to survey the aggregates and protruding filamentous bacteria contents of activated sludge using image analysis procedures for bright field and phase contrast microscopy. The best acquisition method for activated sludge characterization was determined by monitoring the activated sludge of seven different wastewater treatment plants for a period of 2 years.

3.2. MATERIALS AND METHODS

Activated sludge samples for each acquisition methodology study were collected from the aeration basins of seven wastewater treatment plants (WWTP), treating domestic effluents for a period of two years. Plants were located in the North of Portugal. A total of 128 samples were analyzed and for each, the biomass contents in terms of total suspended solids (TSS) were determined by weighing (APHA *et al.*, 1989). A total of 200 images for each sample was subsequently acquired and processed, in both phase contrast and bright field microscopy, in order to estimate the microbial aggregate and protruding filamentous bacteria contents by image analysis techniques.

3.2.1 IMAGE ACQUISITION

For each sample, a volume of 25 µL was placed on a slide and covered with a 20 mm × 20 mm cover slip for visualization and image acquisition in bright field and phase contrast microscopy. Sample deposition was performed by means of a calibrated micropipette with a tip that allowed passage of the largest aggregates. A total of 200 images (divided through 3 replicate slides) were acquired per sample to obtain significant data for both acquisition methods and to minimize sampling errors. Image acquisition of the aggregates and filaments on each slide was obtained by three horizontal passages at one quarter, half and three quarters of the slide. In all horizontal passages an image was acquired at each field length. In a total around 20 images were obtained for each slide. This method overcomes deposition of non-uniform aggregates and filaments on the slide.

3.2.2 BRIGHT FIELD MICROSCOPY

Images were acquired with a Leitz Laborlux S optic microscope (Leitz, Wetzlar), with 100x magnification, coupled to a Zeiss AxioCam (Zeiss, Oberkochen). Image acquisition was performed in 1300 × 1030 pixels and 8-bit format through the commercial software Axio Vision 3.1 (Zeiss, Oberkochen).

3.2.3 PHASE CONTRAST MICROSCOPY

Images were acquired with a Diaphot 300 microscope (Nikon Corporation, Tokyo) with 100x magnification, coupled to a Sony CCD AVC–D5CE (Sony, Tokyo) grey scale video camera. The images were acquired in 768 × 576 pixels and 8-bit format by a Data Translation DT 3155 (Data Translation, Marlboro) frame grabber using the commercial software package Image Pro Plus (Media Cybernetics, Silver Spring).

In order to compare both acquisition methods, calibration from pixels to the metric unit dimension was performed by means of a micrometer slide.

3.2.4 IMAGE PROCESSING AND ANALYSIS METHODOLOGY

The aggregated and filamentous bacteria contents and morphological descriptors were determined through the use of image processing and analysis programs adapted from previous routines first developed by Amaral (2003) and afterwards by Amaral and Ferreira (2005) in Matlab 7.3 (The Mathworks, Inc., Natick) language. Therefore, two different image processing routines were used, one dedicated to bright field images and the other to phase contrast images. The main stages of these programs comprise the image pre-treatment, segmentation, and debris elimination steps. Finally, the aggregated and protruding filamentous biomass binary images were saved for morphological characterization of the activated sludge in the image analysis step. A schematic description of the image processing programs is presented in Figure 3.1. A more detailed description of the image processing methods is presented below.

3.2.5 PRE-TREATMENT

The image pre-processing stage depends on the enhancement of the grayscale images by background determination (bright field and phase contrast images) and background removal (bright field images). In this stage, the original image is first divided by a background image to minimize background light differences. The aggregates and filaments are further enhanced by using local histogram equalization in order to improve the boundaries contrast.

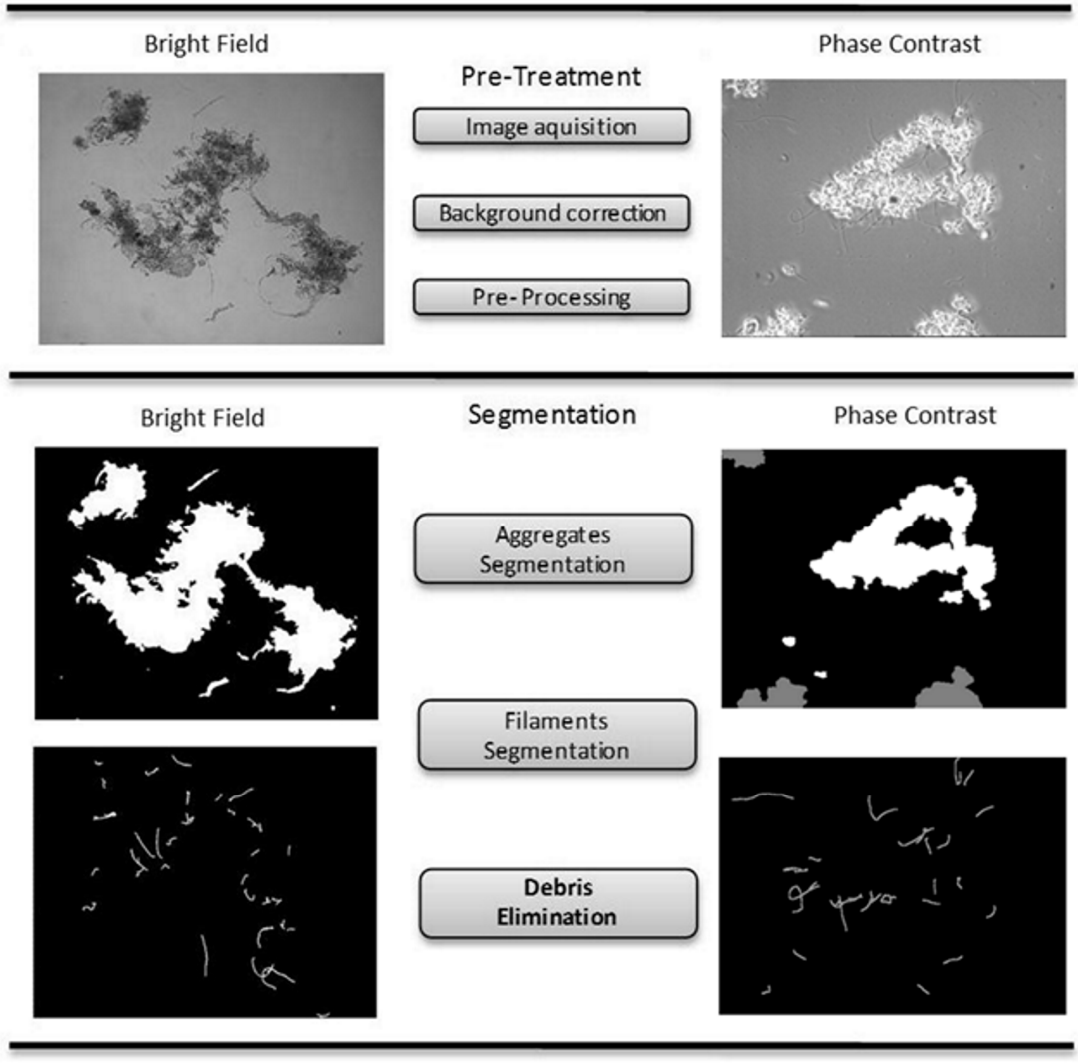

Figure 3.1. Schematic representation of the bright field and phase contrast image processing programs.

3.2.6 SEGMENTATION

This stage consists primarily in segmentation of the aggregates and filaments by the determination and simultaneous use of aggregate boundary and core images. The binary core image is obtained by segmenting the pre-treatment resulting image by a predefined 0.3 built-in threshold level. The boundary grayscale image is determined by the difference between dilated and eroded images of the pre-treatment resulting image. The binary boundary image is obtained by segmenting the resulting image by

a predefined 0.3 built-in threshold level. Finally, morphological opening (disk of radius 5) and reconstruction allow for the final aggregate binary image to be determined.

With respect to the filament segmentation, the aggregate binary image is used as a mask to eliminate the aggregates from the original grey scale image which is next segmented by a 92.5% percentile based threshold level. A series of erosion and dilation procedures on the aggregate binary image is then performed in order to enhance filament identification. Next, a predefined 0.2 built-in threshold level allows the determination of a filament marker image which is later used as a mask to reconstruct the filaments more accurately. Finally, a gyration radius based procedure is implemented to discard small filamentous-like debris by the use of a 1.2 cut-off value (Pons and Vivier, 1999).

3.2.7 DEBRIS ELIMINATION

The elimination of residual aggregates (smaller than 3.5 μm in diameter) and debris is performed by third order erosion and reconstruction operations and all the aggregates cut off by the image boundaries are removed.

3.2.8 MORPHOLOGICAL PARAMETERS

Following the image processing step, the recognized aggregated and filamentous bacteria from the collected images were analyzed in order to individually characterize them in terms of the most relevant morphological parameters described below. In order to select the best acquisition method, image analysis parameters related to aggregate and filament contents were determined, either directly from the

image analysis program or in association with the sludge physical properties for the whole set of 128 samples.

The individual aggregate area (Area) was determined as the pixel sum of each aggregate projected surface calibrated to metric units by a calibration factor F_{Cal} (µm/pixel) determined by the use of a micrometer slide. The filaments individual length (FL) was determined according to Walsby and Avery (1996), with N_{Thn} as the pixel sum of each thinned filament, N_{int} as the number of filament intersections and factor 1.1222 used to average the different measuring angles within the image. Once again the obtained values were calibrated to metric units by the use of the F_{Cal} (µm/pixel) calibration factor:

$$FL = \left(N_{Thn} + N_{int}\right) \times 1.1222 \times F_{cal} \quad (1)$$

Next, the total aggregate area per volume (TA/Vol) and total filament length per volume (TL/Vol) were determined for each replicate, respectively, as the sum of all aggregate areas per unit of volume and the sum of all filament lengths per unit of volume. Finally, for each sample, the average value of the 3 replicates was then determined. Furthermore, the total filament length per total aggregate area (TL/TA) was also determined alongside the total filament length per total suspended solids (TL/TSS). This characterized the aggregate and filament dynamics considered by Mesquita *et al.* (2009b) as the most relevant for sludge settling.

In order to compare both bright field and phase contrast acquisition methods the *F* test two-sample for variances was applied to the image analysis results using an α value of 0.05 (95% confidence). The *F* critical, *f* and *p*-values of the *F*-test were determined for the studied 128 different samples after normalization by the mean values of both methods. The *f* and *p* values were then compared with the *F* critical and α-values. For an $f < F$ critical and $p > \alpha$ the variances of both methods can be considered equal.

3.3. RESULTS AND DISCUSSION

3.3.1 DIRECT SAMPLE COMPARISON

Bright field and phase contrast methods were evaluated considering the parameters of total filament length and total aggregate area. These are suitable to represent protruding filamentous bacteria and aggregated biomass contents. Furthermore, the TL/TA and TL/TSS parameters, characterizing the aggregate and filament dynamics, were also compared due to their relevance for the assessment of sludge settling properties. For a clearer distinction between the values obtained for each acquisition method, the results for the above parameters were plotted by ascending bright field values considering the 128 samples (Figure 3.2).

Results revealed the relationship between both acquisition methods with respect to the four studied parameters, and helped to establish the main differences throughout the data set range (Figure 3.2). Aggregated biomass content values (Figure 3.2a), in terms of TA/Vol, obtained by phase contrast were systematically higher than for bright field methods. Over 90% of the data set was higher for the phase contrast method. The full data set revealed an average 38.8% higher TA/Vol values for phase contrast acquisition (average 1.83 $mm^2/\mu L$ for bright field and 2.54 $mm^2/\mu L$ for phase contrast). Given the poorer representation of the object boundaries in phase contrast microscopy, it seems reasonable to conclude that the phase contrast method overestimated the aggregated biomass contents.

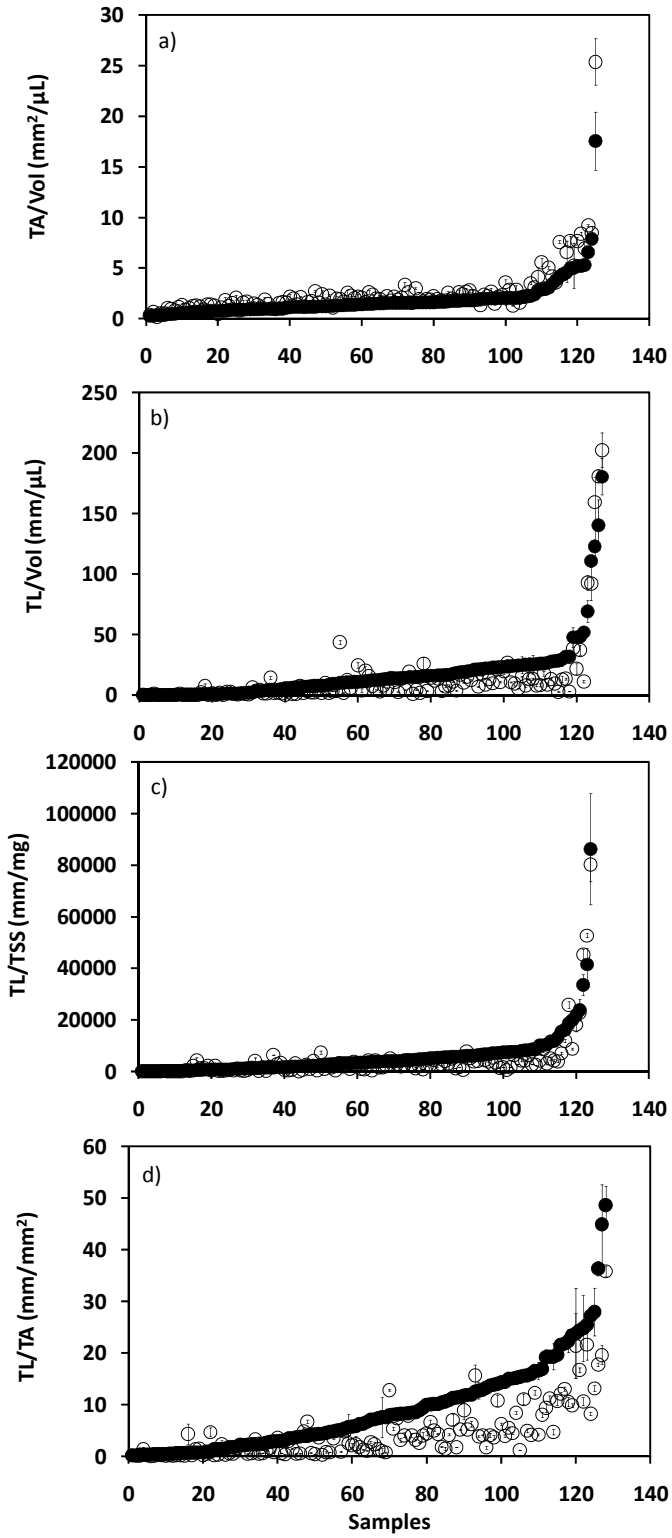

Figure 3.2. Bright field (●) and phase contrast (○) results for (a) TA/Vol (mm^2/µL), (b) TL/Vol (mm/µL), (c) TL/TSS (mm/mg) and (d) TL/TA (mm/mm^2) considering 128 samples plotted in order of ascending bright field values.

With respect to the protruding filamentous bacteria contents (Figure 3.2b), in terms of TL/Vol, for 75% of the data set there is an overestimation for bright field results when compared to phase contrast. This resulted in an average 23.9% lower TL/Vol value for phase contrast acquisition (average 17.32 mm/µL for bright field and 13.19 mm/µL for phase contrast). Given the better filament/background contrast obtained by phase contrast microscopy, these results were not expected. However, they can be explained by the fact that most of the activated sludge samples analyzed in this work presented short protruding filamentous bacteria. In this sense, the visible filament length in phase contrast would be shortened by the halo of the aggregates in phase contrast, which is a pressing matter in most of the studied activated sludge, presenting low (and short) protruding filamentous bacteria, as it can be seen in Figure 3.2b. However, when the analysis fell solely on samples presenting long protruding filaments (TL/Vol > 50 mm/µL, for samples 123 to 128 in Figure 3.2b), a clearer correspondence could be found between both methodologies. As a matter of fact, the TL/Vol determined by the bright field methodology on these samples were only 8.6% on average lower than the phase contrast methodology results (average 112.61 mm/µL for the bright field and 123.16 mm/µL for the phase contrast).

Schuler and Jassby (2007) showed that expressing filament contents per mass (TL/TSS) is probably the most useful way for comparing filament contents data from different studies and/or from samples with different biomass concentrations. Concentrations can vary greatly from one system to the other and this approach normalizes filament contents to biomass. Regarding this parameter behavior (Figure 3.2c) with biomass normalization, a similar trend between bright field and phase contrast methods arises when compared to the TL/Vol analysis. This results in an underestimation for phase contrast results when compared to bright field in 75% of the data set. Phase contrast methods for the TL/TSS ratio also underestimated TL/TSS ratio values an average 24.2% regarding bright field acquisition (average 5605.1 mm/mg for bright field and 4246.7 mm/mg for phase contrast). Given the

constraints of phase contrast on accurately determining the short protruding filamentous bacteria, an analysis was performed focusing on the long protruding filaments (TL/TSS > 20 000 mm/mg, for samples 123 to 128 in Figure 2c). As observed for the TL/Vol analysis, a much higher similarity was found between both methods and only a slight overestimation (0.6% on average) was found for the phase contrast method (average 37812 mm/mg for bright field and 38025 mm/mg for phase contrast).

The difference between bright field and phase contrast acquisition is also quite clear when observing filament length per aggregate area ratio (Figure 3.2d). Systematically (over 90% of the data set) higher TL/TA values were obtained for the bright field method. Phase contrast underestimates by an average 50% the TL/TA values obtained by bright field microscopy (average 9.08 mm/mm^2 for bright field and 4.55 mm/mm^2 for phase contrast) due to the total filament length underestimation and total aggregate area overestimation.

3.3.2 Regression analysis

The analysis of Figure 3.2 showed a clear correlation between the results for bright field and phase contrast microscopy acquisition. The correlation between bright field and phase contrast results was then studied to help establish the best acquisition method for both aggregated and protruding filamentous bacteria. Figure 3.3 represents the obtained correlations between bright field and phase contrast methods.

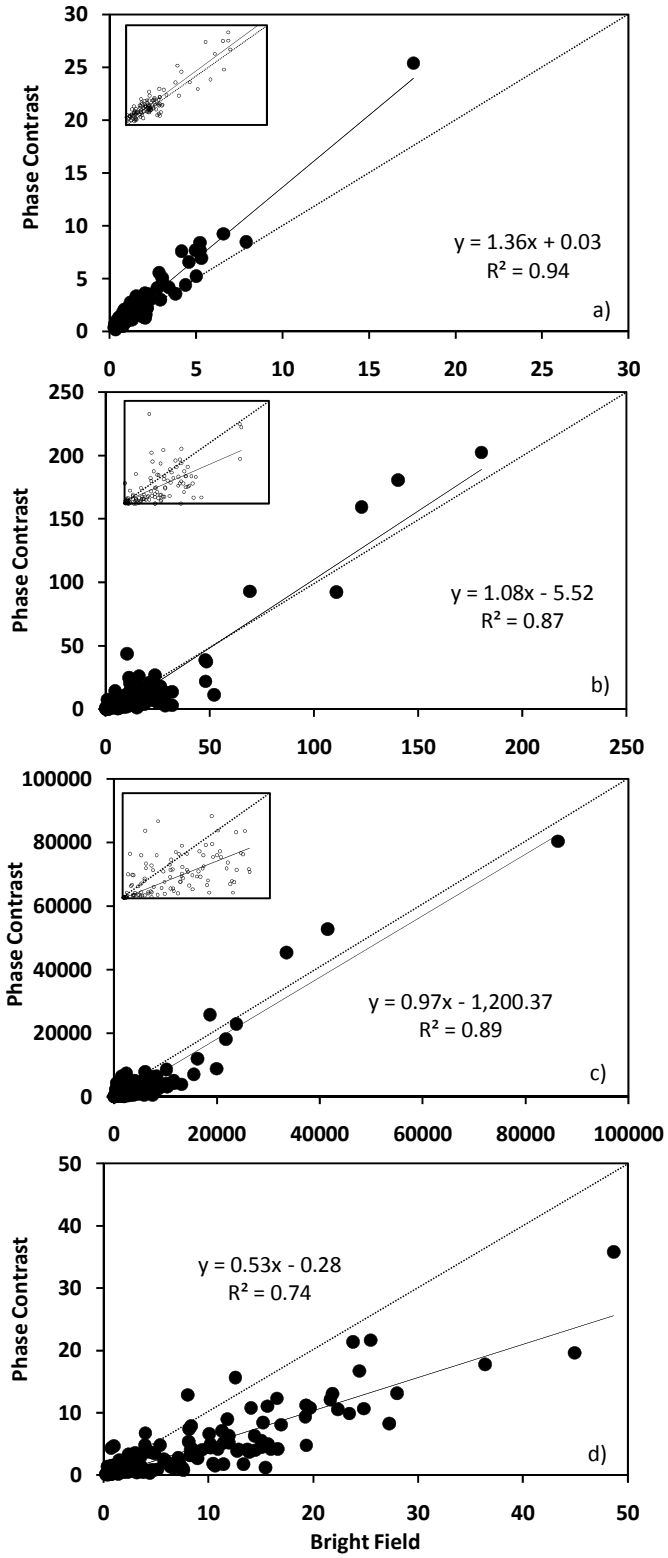

Figure 3.3. Bright field and phase contrast regressions for (a) TA/Vol ($mm^2/\mu L$), (b) TL/Vol ($mm/\mu L$), (c) TL/TSS (mm/mg) and (d) TL/TA (mm/mm^2). In (a), (b), and (c) a window is included using the lowest values.

Figure 3.3a shows a satisfactory correlation coefficient R^2 of 0.94 for the TA/Vol between bright field and phase contrast assessment. Once again, it was possible to conclude that the phase contrast method overestimated the aggregate detection with respect to the bright field acquisition method. Furthermore, the trend line for the TA/Vol estimation points towards a global overestimation around 36.3% (regression slope of 1.36), quite close to the 38.8% obtained in the direct comparison analysis.

With respect to recognition of the protruding filamentous bacteria (Figure 3.3b), the correlation of 0.93 (R^2 of 0.87) was not as satisfactory as for the TA/Vol analysis, but absolute values between the two methods were somewhat similar considering the global overestimation of 7.9% (regression slope of 1.08) for the phase contrast results. However, the regression equation seems to point towards an underestimation of 31.9% for the phase contrast method when compared to the bright field in the low filament content range. The higher dispersion (lower correlation coefficient) between the two methods can be explained, as mentioned above, because phase contrast likely underestimated short protruding filaments, and bright field slightly underestimated long protruding filaments, thus shifting low and high TL/Vol samples away from the global trend line. On the other hand, for high TL/Vol contents a reasonable correspondence was observed between the two methods as indicated by the low 8.6% and 7.9% overestimations with phase contrast in direct comparison and correlation analysis, respectively. The reason may be explained by the high importance of the larger, and further apart, TL/Vol samples with respect to the correlation determination by the least squares method. The lower values on larger filament contents of the bright field method balances, to some extent, the larger values obtained for smaller contents. It should be noted that, in samples with larger filament contents, the advantage of phase contrast recognition was less hindered by the ill defined aggregate borders leading to a more precise (and higher) filament determination. Therefore, the correspondence found between the

two methods indicated that bright field was indeed advantageous over phase contrast acquisition on well balanced activated sludge systems given that it does not lead to overall significant errors with respect to determination of the protruding filamentous bacteria. However, care should be taken in analyzing the results of bright field acquisition from activated sludge with very high filamentous contents.

Regarding the TL/TSS ratio results (Figure 3.3c), a correlation coefficient of 0.94 (R^2 of 0.89) was obtained. Absolute values for both samples were again somewhat similar considering the global underestimation of 3.1% (regression slope of 0.97) for the phase contrast results. However, the regression equation pointed to 21.4% underestimation for phase contrast regarding to the bright field method in the low filament contents per TSS range. Once more, the dispersion found between the two methods could be explained by the propensity of phase contrast to underestimate the short protruding filaments, and bright field slightly underestimated long protruding filaments, thus shifting the low and high TL/TSS samples away from the global trend line. Nonetheless, for high TL/TSS contents a good correspondence between the two methods could be found as shown by the slight 0.6% and 3.1% underestimations with phase contrast in direct comparison and correlation analysis, respectively. The reason for this is directly related to the protruding filamentous bacteria estimation differences found in the two methods and, the same caution must be applied in the TL/TSS ratio as in the TL/Vol assessment.

A considerable dispersion between bright field and phase contrast assessments was detected and a relatively poor 0.86 correlation coefficient (R^2 of 0.74) for the TL/TA ratio was obtained (Figure 3.3d). This result might be due to the cumulative sum of impreciseness both in terms of TL and TA resulting in a strong underestimation of the TL/TA ratio in phase contrast acquisition. This conclusion is further emphasized by the 0.53 slope obtained. This resulted in a global 47% underestimation (regression

slope of 0.53) of the TL/TA ratio by phase contrast which is in accordance with the 50% reduction found in the direct comparison analysis.

3.3.3 STATISTICAL ANALYSIS

To understand the relationship between bright field and phase contrast acquisition, the *F*-test two-sample for variances was performed for an α value of 0.05 (95% confidence). Table 3.1 presents the *f*-values and *p*-values of the statistical analysis comparing the normalized image analysis results for both bright field and phase contrast. A total of 128 samples for each method was analyzed resulting in 127 degrees of freedom and an *F* critical value of 1.34.

Table 3.1. *f* and *p*-values obtained by the *F-test* statistical analysis (α = 0.05) regarding the normalized aggregates and filaments contents as determined by bright field and phase contrast methodologies.

Variables	TA/Vol	TL/Vol	TL/TSS	TL/TA
f - value	1.02	2.30	1.83	1.53
p - value	0.45	1.84×10^{-6}	4.20×10^{-4}	0.01

Taking into consideration the *f*-value of 1.02 (lower than the *F* critical value of 1.34) and the *p*-value of 0.45 (larger than the α-value of 0.05) regarding the TA/Vol, it seems clear that both methods presented similar variances and are, therefore, statistically similar. Furthermore, these results were in accordance with the satisfactory R^2 value of 0.94 obtained in the regression analysis.

On the contrary the TL/Vol, TL/TSS and TL/TA statistical analysis had *f*-values higher than the *F* critical value of 1.34. This indicated that the sample variances for these parameters were statistically different, and therefore, the methods can be

considered as unrelated. The same conclusion can also be attained regarding the obtained *p*-values which were smaller than the statistical significance value (0.05). These results seem to corroborate the previous findings of a lower correlation between the bright field and phase contrast methods for the TL/Vol, TL/TSS and TL/TA.

A second *F*-test two-sample was also performed for an α-value of 0.05 (95 % confidence) regarding the samples with high (TL/Vol > 50 mm/µL, TL/TSS > 20000 mm/mg) and low (TL/Vol < 50 mm/µL, TL/TSS < 20000 mm/mg) filamentous contents (Table 3.2). The 6 samples in total for the TL/Vol > 50 mm/µL and TL/TSS > 20000 mm/mg analyses resulted in 5 degrees of freedom and consequently of an *F* critical value of 6.39, whereas the 121 samples total for the TL/Vol < 50 mm/µL and TL/TSS < 20000 mm/mg analyses resulted in 120 degrees of freedom and consequently an *F* critical value of 1.35.

Table 3.2. *f* and *p*-values obtained by the *F*-test statistical analysis (α = 0.05) regarding the two normalized TL/Vol and TL/TSS studied ranges, as determined by bright field and phase contrast methodologies.

Variables	**TL/Vol (mm/µL)**		**TL/TSS (mm/mg)**	
	< 50	> 50	< 20000	> 20000
f - value	1.57	1.14	1.70	1.24
F critical	1.35	6.388	1.35	6.39
p - value	0.01	0.45	0.00	0.42

For the TL/Vol parameter of high filamentous content samples (TL/Vol > 50 mm/µL), both methods had similar variances as shown by the *f*-value of 1.14 (lower than the *F* critical value of 6.39) and *p*-value of 0.45 (larger than the α value of 0.05). Therefore, it could be established that for the high TL/Vol range both methods are statistically

similar, as it was hypothesized in the previous direct sample comparison analysis. With respect to the low filamentous content samples (TL/Vol < 50 mm/μL), the *f*-value of 1.57 (higher than the *F* critical value of 1.35) and *p*-value of 0.01 (smaller than the α value of 0.05) led to the conclusion that the two methods were statistically different in the low TL/Vol range. Once again, these results corroborated the conclusions of the direct sample comparison analysis.

A similar result was also observed for the TL/TSS parameter. In high filamentous content samples (TL/TSS > 20000 mm/mg), the *f*-value of 1.24 (lower than the *F critical* value of 6.39) and *p*-value of 0.42 (larger than the α value of 0.05) point to statistical similarity of the two methods. In low filamentous content samples (TL/TSS < 20000 mm/mg), the *f*-value of 1.70 (higher than the *F* critical value of 1.35) and *p* value of 0 (smaller than the α value of 0.05) point to the opposite conclusion regarding the low TL/TSS range. As above, these results also corroborated the conclusions of the direct sample comparison analysis.

3.3.4 ACCURACY ANALYSIS

In order to determine the accuracy of both methods the average of the TA/Vol, TL/Vol, TL/TA and TL/TSS for all samples, the average of the standard deviations for all samples, and the average error percentages for all samples were determined (Table 3.3). The parameter and their standard deviations were determined as the average of the 128 samples for that parameter. The accuracy error percentage was determined as the ratio between the average parameter standard deviation and the average parameter value, multiplied by 100.

Table 3.3. Average values, average standard deviation and accuracy error percentage values regarding the aggregates and filaments contents as determined by bright field (BF) and phase contrast (PC) methodologies.

Variables	TA/Vol ($mm^2/\mu L$)		TL/Vol (mm/µL)		TL/TSS (mm/mg)		TL/TA (mm/mm^2)	
	BF	**PC**	**BF**	**PC**	**BF**	**PC**	**BF**	**PC**
Average values	1.83	2.54	17.32	13.19	5605.1	4246.7	9.08	4.55
Average standard deviation	0.19	0.23	1.91	1.28	602.2	304.9	0.76	0.42
Average error %	10.18	9.02	11.06	9.70	10.74	7.18	8.36	9.13

Accuracy error percentages for all parameters for both methods showed no significant variations, although a slight advantage for the phase contrast method was apparent, with exception of the TL/TA parameter. These results indicate that both methodologies are almost equivalent. For both phase contrast and bright field methods the accuracy error percentage remained around 10 % of the average value and within a window of 2%. Only the TL/TSS difference was more significant (10.74% against 7.18%). The TL/Vol accuracy error percentage for the two methods was much closer (11.06% against 9.70%) and this difference may be partly explained by a higher error associated with the TSS measurement regarding the bright field samples.

3.3.5 ROBUSTNESS ANALYSIS

This study also addressed the robustness evaluation of both methods in terms of aggregated and protruding filamentous bacteria detection by the developed image processing programs. Default values for the detection thresholds of aggregates and filaments, loaded in the beginning of both programs were analyzed throughout monitoring of the seven wastewater treatment plants. After each processing step, and to avoid inaccurate assessment, the final labeled images were scrutinized by visual inspection. Thus, during the monitoring period, the final binary aggregate and

filament images that were not in accordance with the original images were reprocessed, with minor adjustments on the detection thresholds. Table 3.4 displays the recognition results for both bright field and phase contrast methods regarding the detection of aggregates and protruding filaments, with default and changed values.

Table 3.4. Number and percentage of images treated with default and changed values during the processing step for bright field and phase contrast programs for both aggregates and protruding filamentous bacteria detection.

		# Default	# Changed	% Default	% Changed
Bright field	Aggregates	33168	0	100%	0%
	Filaments	32000	1168	96%	4%
Phase contrast	Aggregates	33230	694	98%	2%
	Filaments	29309	4615	86%	14%

Data in Table 3.4 points out the high robustness for detection of both aggregate and filamentous bacteria with the default values provided from the bright field and phase contrast programs. Filaments detection proved to be most prone to errors with 4% and 14% adjustments needed for bright field and phase contrast methods, respectively. This result can be explained by the small width of the filaments and higher microscopy focusing dependence. Comparing the two acquisition methods, phase contrast seems to be less robust than bright field leading to a higher percentage of changes in default values for recognition of both aggregates and filamentous bacteria. The robustness of the bright field method was 100% for aggregate recognition and 96% for filament recognition. Only 98% and 86% robustness values were attained for the same parameters regarding the phase contrast method.

3.4. CONCLUSIONS

This study demonstrated that, with respect to the protruding filamentous bacteria, bright field acquisition results are somewhat similar to the phase contrast results, mainly for high filaments contents. However, a clear discrepancy between methods in the low filament samples was found. This may be due to the existence, in phase contrast acquisition, of a halo surrounding the aggregates, which hinders their correct determination by this method. If the halo of an aggregate is too bright then the part of the filament that falls within the halo area may not be visible as it is quite thin. Furthermore, it may not be simple to discern the border between the halo and the actual aggregate making it difficult to estimate the length of the filament encompassed by the aggregate's halo. In the assessment of the aggregated biomass contents both methods are statistically similar, although with a clear overestimation of the aggregates area for the phase contrast acquisition. An accuracy analysis also demonstrated no significant variations for both methods. Although both methods provided very high robustness, bright field microscopy surpassed phase contrast results in this matter. Considering the advantages and disadvantages of each acquisition method bright field microscopy proved to be a more simple and inexpensive method that provided the best overall results.

3.5. REFERENCES

Abreu, A.A., Costa, J.C., Araya-Kroff, P., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for identifying structural changes during a revival process of anaerobic granular sludge. *Water Research* **41**, 1473-1480.

Amaral, A.L. (2003) *Image analysis in biotechnological processes: applications to wastewater treatment.* PhD Thesis, University of Minho, Braga, Portugal. (http://hdl.handle.net/1822/4506).

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

APHA, AWWA, and WPCF. (1989) *Standard Methods for the Examination of Water and Wastewater*. 17th Edition, American Public Health Association, Washington D.C.

Banadda, E.N., Smets, I.Y., Jenné, R., and Van Impe, J.F. (2005) Predicting the onset of filamentous bulking in biological wastewater treatment systems by exploiting image analysis information. *Bioprocess Biosystems Engineering* **27**, 339-348.

Cenens, C., Van Beurden, K.P., Jenné R., and Van Impe, J.F. (2002) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technololy* **46** (1-2), 381-387.

Costa, J.C., Abreu, A.A., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for monitoring structural changes of anaerobic granular sludge during detergent shock loads. *Biotechnology and Bioengineering* **98** (1), 60-68.

da Motta, M., Pons, M.N., and Roche, N. (2001a) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technololy* **43** (7), 91–96.

da Motta, M., Amaral, A.L., Casellas, M., Pons, M.N., Dagot, C., Roche, N., Ferreira, E.C., and Vivier, H. (2001b) Characterization of activated sludge by automated image analysis: validation on full-scale plants. *IFAC Computer Applications in Biotechnology*, Québec City, Canada, 427-431.

Ganczarczyk, J.J. (1994) Microbial Aggregates in Wastewater Treatment. *Water Science and Technology* **30**, 87-95.

Grijspeerdt, K. and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31**, 1126-1134.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Banadda, E.N., Philips, N., and Van Impe, J.F. (2003) Image Analysis as a Monitoring Tool for Activated Sludge Properties in Lab-Scale Installations. *Journal of Environmental Science and Health, Part A - Toxic/Hazardous Substances and Environmental* **38** (10), 2009-2018.

Jenné, R., Banadda, E.N., Smets, I.Y., and Van Impe, J.F. (2004) Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* **50** (7), 281-285.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2009a) Monitoring of activated sludge settling ability through image analysis: validation on full-scale wastewater treatment plants. *Bioprocess Biosystems Engineering* **32** (3), 361-367.

Mesquita, D.P., Dias O., Dias, A.M.A., and Amaral, A.L., Ferreira, E.C. (2009b) Correlation between sludge settleability and image analysis information using partial least squares. *Analytica Chim Acta* **642** (1-2), 94-101.

Pons, M.N. and Vivier, H. (1999) Biomass quantification by image analysis. *Advances in Biochemical Engineering/Biotechnology* **66**, 133-184.

Schuler, A.J. and Jassby, D. (2007) Filament content threshold for activated sludge bulking: Artifact or reality? *Water Research* **41**, 4349-4356.

Walsby, A.E. and Avery, A. (1996) Measurement of filamentous cyanobacteria by image analysis. *Journal of Microbiological Methods* **26**, 11-20.

Abstract

This chapter reports activated sludge systems operating with high biomass contents. For an accurate sludge characterization based on image analysis, sometimes samples must be diluted. Furthermore, the selection of the best image acquisition magnification is directly related to the amount of biomass screened. It was found that assessments of biomass contents and structure were affected by dilutions. Therefore, the correct operating dilution requires careful consideration. Moreover, the acquisition methodology comprising a 100x magnification allowed data on aggregated and filamentous biomass to be determined and smaller aggregates to be identified and characterized, with a good accuracy regarding biomass representativeness.

4. CHAPTER 4 – DILUTION AND MAGNIFICATION EFFECTS ON IMAGE ANALYSIS APPLICATIONS IN ACTIVATED SLUDGE CHARACTERIZATION

4.1. INTRODUCTION

Wastewater treatment plants (WWTP) are frequently operated with activated sludge systems, consisting of an aerated tank for aerobic sludge treatment and a clarifier for the settling stage to allow aggregated biomass to separate. An adequate balance between the different types of bacteria is essential for the formation of aggregates with acceptable properties of structure, stability, size and density. This results in effective settling of the sludge and low suspended solid effluent levels (Jenkins *et al.*, 2003; Jin *et al.*, 2003; Wilén *et al.*, 2008). Recently, owing to advances in imaging hardware, automated image analysis methodologies have been increasingly adopted to monitor activated sludge properties. Activated sludge processes are currently monitored via microscopic observation for (a) determination of aggregated and filamentous bacterial contents (da Motta *et al.*, 2001; Cenens *et al.*, 2002; Amaral, 2003; Jenné *et al.*, 2006, 2007; Banadda *et al.*, 2005), (b) prediction of settling ability based on image analysis and multivariate statistical techniques (Amaral and Ferreira, 2005; Mesquita *et al.*, 2009a,b), and (c) protozoan and metazoan identification (Amaral *et al.*, 2008; Ginoris *et al.*, 2007a,b).

The fact that a number of activated sludge systems operate with relatively high biomass contents suggests that during image acquisition samples should be diluted for adequate sludge characterization. When samples are not diluted, aggregates and filaments may appear super-imposed on images, which can be misleading (da Motta *et al.*, 2002). Therefore, dilutions aid microscopic inspection of sludge, and when using image analysis methodologies they allow the system to be evaluated accurately. However, dilution techniques can be problematic as the amount of

screened biomass decreases and they could lead to morphological changes of the aggregated biomass.

A high percentage of published studies regarding activated sludge characterization rely on image acquisition procedures with microscopic visualization at 100x magnification (Abreu *et al.*, 2007; Costa *et al.*, 2007; Mesquita *et al.*, 2009a, 2010). The advantage of using such a magnification is that it allows aggregated and filamentous bacterial contents and morphological data to be determined from the same image. However, the relationship between the image analysis data acquired at 100x magnifications and at a lower and more representative magnification, regarding aggregated biomass contents, should be investigated.

The present study describes the use of an image analysis procedure, developed for characterizing activated sludge systems using bright field microscopy, to investigate the effect of five different dilutions and two magnifications on image analysis data.

4.2. MATERIALS AND METHODS

For the dilution study, three samples were collected from an aeration tank in the municipal WWTP located in Braga (Portugal) treating domestic effluents. Total suspended solids (TSS) were determined by weight reflecting samples covering the range of normal operating TSS values (4708 mg/L, 3000 mg/L and 2310 mg/L). For each sample, an undiluted sample and four dilutions were prepared as follows: one part sludge to one (1:2), four (1:5), seven (1:8) and nine parts water (1:10). To investigate the magnification effect, two magnifications were applied (20x and 100x), in a total of 13 samples taken from two aerated basins of a WWTP treating domestic effluents, located in Bragança (Portugal). The contents and morphology of bacterial aggregates and protruding filamentous bacteria were studied by bright field image acquisition and analysis.

4.2.1 IMAGE ACQUISITION

In all cases a volume of 25 µL was placed on a slide and covered with a 20 mm x 20 mm cover slip for visualization and image acquisition using bright field microscopy. 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 samples on the slides.

For the dilution analysis, 200 images per sample (divided between three slides) were acquired in order to obtain significant data for all dilutions studied. Images were visualized using a Leitz Laborlux S optical microscope (Leitz, Wetzlar) with 100x total magnification, and acquired using a Zeiss AxioCam (Zeiss, Oberkochen) digital camera. Images were acquired using 1300 × 1030 pixels in 8-bit format through the commercial software Axio Vision 3.1 (Zeiss, Oberkochen).

Regarding the magnification analysis, approximately 25 images per sample were acquired for each magnification. The images were acquired with a NIKON SMZ 800 - DIA stereomicroscope (Nikon, Tokyo) for the 20x total magnification (10x magnification apochromatic objective and 2x zoom factor), and with a NIKON NIK 50i CC/CF optic microscope (Nikon, Tokyo) for the 100x total magnification (10x magnification apochromatic objective). In both cases images were acquired using a NIKON DS-5M-U1 (Nikon, Tokyo) digital camera with 1280 × 960 pixels in 8-bit format through the commercial software NIS-ELEMENTS D (Nikon, Tokyo). Images of each dilution studied are presented in Figure 4.1, acquired at 100x magnification, and an image acquired at 20x magnification is presented in Figure 4.2d.

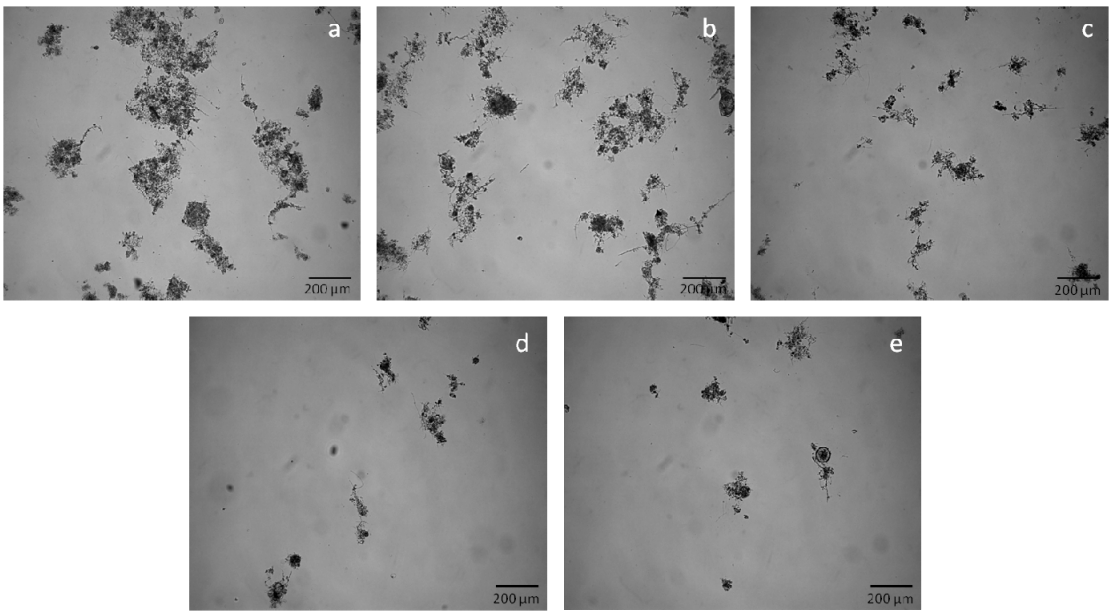

Figure 4.1. (a) Undiluted sample, (b) dilution 1:2, (c) dilution 1:5, (d) dilution 1:8, e) dilution 1:10.

4.2.2 IMAGE PROCESSING AND ANALYSIS

Aggregated and filamentous bacteria contents and morphology were determined using image processing and analysis programs, adapted from Amaral and Ferreira (2005) in Matlab 7.3 language (The Mathworks, Inc., Natick). An image processing routine was used to obtain and save binary images of aggregated and protruding filamentous biomass, and an image analysis program determined the aggregate and filament contents and morphological parameters. The image processing program comprises three stages: image pre-treatment; segmentation; debris elimination. Figure 4.2 presents an example of original and binary images concerning the recognition of aggregated biomass (100x and 20x magnification) and filamentous bacteria (100x magnification).

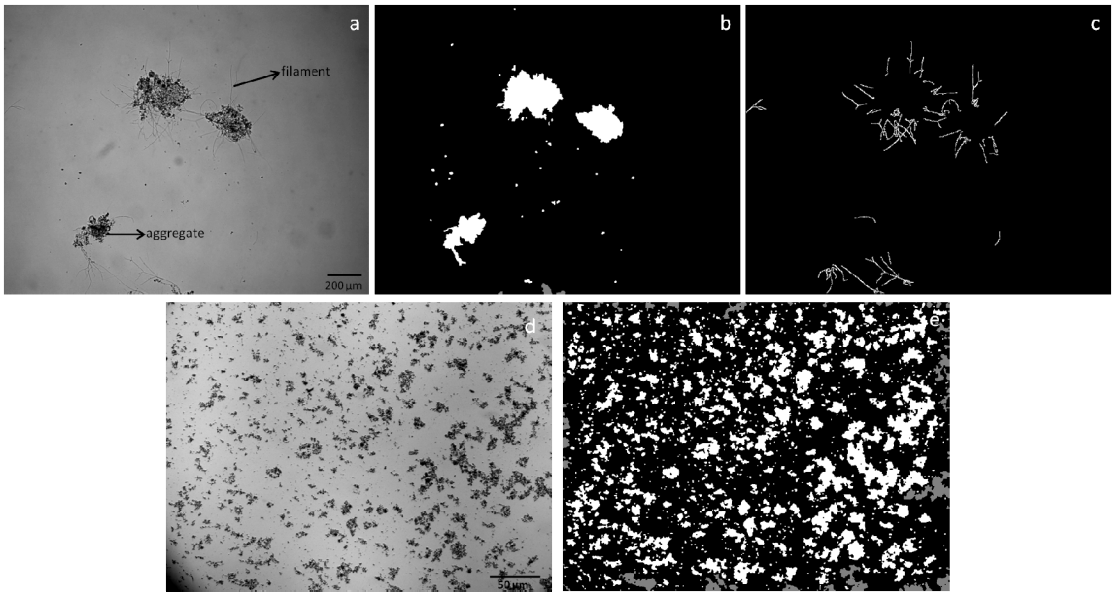

Figure 4.2. (a) Original image at 100x magnification, (b) binary image of aggregates, (c) binary image of filaments, (d) original image at 20x magnification, (e) binary image of aggregates.

The same image processing program was used for both magnifications, but no filamentous bacteria could be identified in the 20x magnification images, due to their small size.

4.2.3 MORPHOLOGICAL PARAMETERS

Characterization of activated sludge systems in terms of filamentous and aggregated biomass contents and morphology can be obtained using image analysis methodologies (Amaral and Ferreira, 2005). For the dilution study, the area (Area) of each recognized aggregate (aggregate not bisected by the image boundaries) was calculated as the projected surface of the aggregate, whereas the total area of recognized aggregates per unit volume (TRA/Vol) was calculated as the sum of recognized aggregate areas per unit volume. Total filament contents per unit volume (TL/Vol) was calculated as the total length of the filaments per unit volume, and the number of recognized aggregates per unit volume (Number/Vol) was calculated as

the total number of recognized aggregates per unit volume (Amaral, 2003). Total filament content per total area of recognized aggregates (TL/TRA ratio) was obtained by dividing the TL/Vol by the TRA/Vol parameter. Equivalent diameter (Deq) was calculated on the basis of the Area parameter using the following equation:

$$D_{eq} = 2 F_{Cal} \sqrt{\frac{Area}{\pi}} \quad (1)$$

where F_{Cal} is the calibration factor (µm per pixel).

Furthermore, aggregates were classified according to size in order to understand the dilution effect on biomass structure: small aggregates (D_{eq} < 25 µm); intermediate aggregates (25 µm < D_{eq} < 250 µm); large aggregates (D_{eq} > 250 µm). This classification has been used previously (Mesquita *et al.* 2009a). For each class, recognized aggregates area percentage (Area %) was calculated as the ratio between the sum of each recognized aggregate area belonging to that class and the sum of the overall recognized aggregate areas. The aggregates recognition percentage parameter (Recognition %) was calculated as the ratio between the total recognized aggregates area and the total aggregates area (including aggregates bisected by the image boundaries).

Several parameters were used to investigate the relationship between the magnifications used in this study: total recognized aggregates area per volume (TRA/Vol); number of recognized aggregates per image (Number/Image); number percentage of recognized aggregates (Number %), calculated as the ratio between the total number of recognized aggregates in a class and the total number of recognized aggregates; area percentage of recognized aggregates (Area %).

4.2.4 STATISTICAL ANALYSIS

For dilution analysis, a *runs test* was applied (Bradley, 1968) to image analysis data to calculate the number of different runs, the length of the runs and the *z*-value. An α-value of 0.05 (95% confidence) was used and the *z*-value was compared with the *Critical z*-value for that α-value (1.96). For a *z*-value < 1.96 data exhibited random behavior, and for a *z*-value > 1.96 exhibited non-random behavior.

To compare the 20x and 100x acquisition methodologies, the two-sample *t test* for averages was applied to the image analysis data to calculate the Critical *t*-value, *t*-, and *p*-value. The number of observations was equal to the number of samples studied (13), and an α-value of 0.05 (95% confidence) was used. *t*- and *p*-values were compared with the Critical *t*-value and α-values. For $t <$ Critical *t*-value and $p > \alpha$, the averages of both methods were equal.

4.3. RESULTS AND DISCUSSION

4.3.1 DILUTION STUDY

When a dilution is performed, the amount of screened biomass decreases, which can lead to an incorrect characterization of the biomass within the biological system (da Motta *et al.* 2002). Furthermore, the modification of the osmotic pressure when a dilution is performed can trigger biomass to deflocculate, causing subsequent release of floc-forming and filamentous bacteria to the mixed liquor, which can change the aggregate size and morphology. Therefore, precautions must be taken when diluting samples. However, dilutions could be advantageous in high biomass content systems, as smaller aggregates and protruding filaments can be hidden by larger aggregates in non-diluted samples. As presented in Figure 4.1, increasing dilutions caused aggregates to be more spaced out, allowing smaller aggregates and filamentous

bacteria hidden by larger aggregates in undiluted samples to be identified. However, increasing dilutions lead to a decrease of the screened biomass in each field of view, and may hinder the representativeness of each sample. The TL/Vol, TRA/Vol and TL/TRA parameters, presented in Figure 4.3, were used to assess the dilution effect on activated sludge within normal operating TSS values (ranging from 2310 to 4708 mg/L).

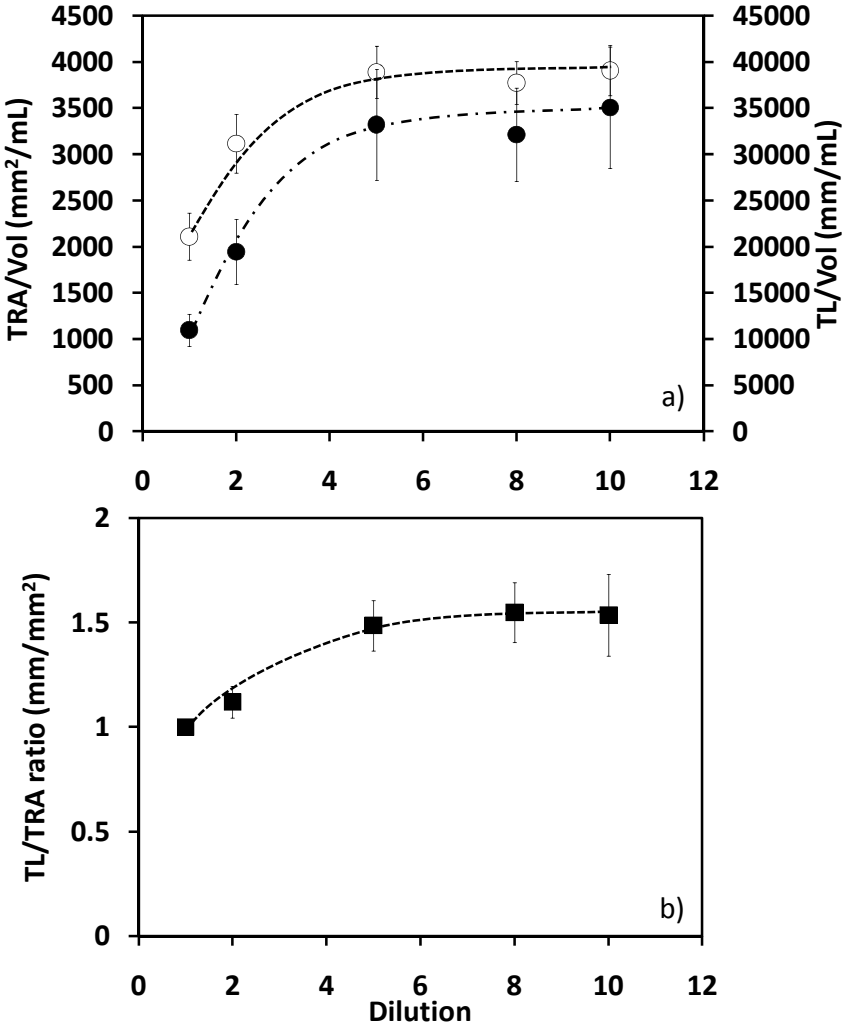

Figure 4.3. (a) Average values of the three samples for the five dilutions regarding TL/Vol (●), TRA/Vol (○), and (b) TL/TRA ratio (■).

The results depicted in Figure 4.3a demonstrate that TRA/Vol and TL/Vol converged to an asymptotic value after the fivefold (1:5) dilution. Both parameters were strongly influenced by dilution, increasing from the non-diluted sample to the fivefold dilution (run of three consecutive ascending values in both cases), probably due to the release of enclosed filamentous bacteria and aggregates, and biomass deflocculation phenomena. The TRA/Vol almost doubled (1.84 times) its non-diluted value when the fivefold dilution was performed, while TL/Vol increased 3.03 times for the same dilution. Therefore, it could be concluded that, despite both TRA and TL parameters converged to an asymptotic value after the fivefold dilution, TL was more affected by dilution than TRA. Considering the trend lines for TL and TRA, it was found a correlation between the two parameters when increasing dilutions are performed (TL/TRA ratio dependence with dilution; see Figure 4.3b). A well-defined trend line was obtained modeling the dependence of TL/TRA ratio with dilution, and as expected from the previous results, a threshold value (around 1.5 mm/mm^2) was obtained for the fivefold dilution. These results suggest that for the present study an optimal dilution (fivefold) was determined. This fivefold dilution unveiled all enclosed filamentous bacteria and aggregates, as verified by the absence of further increase in TRA and TL parameters. In the above cases, no *z*-values from the *runs test* were obtained to verify data randomness because there were too few data points (fewer than five) in each situation (below and above the fivefold dilution).

As discussed by da Motta *et al.* (2002), dilutions are related to assessment of the size and number of aggregates by image analysis techniques. This dependence on dilutions was evaluated by assessing the behavior of each aggregates class (smaller, intermediate and larger) after dilution. Number/Vol, TRA/Vol, Area % and equivalent diameter for each class of samples are depicted in Figure 4.4a, Figure 4.4b, Figure 4.4c and Figure 4.4d, respectively.

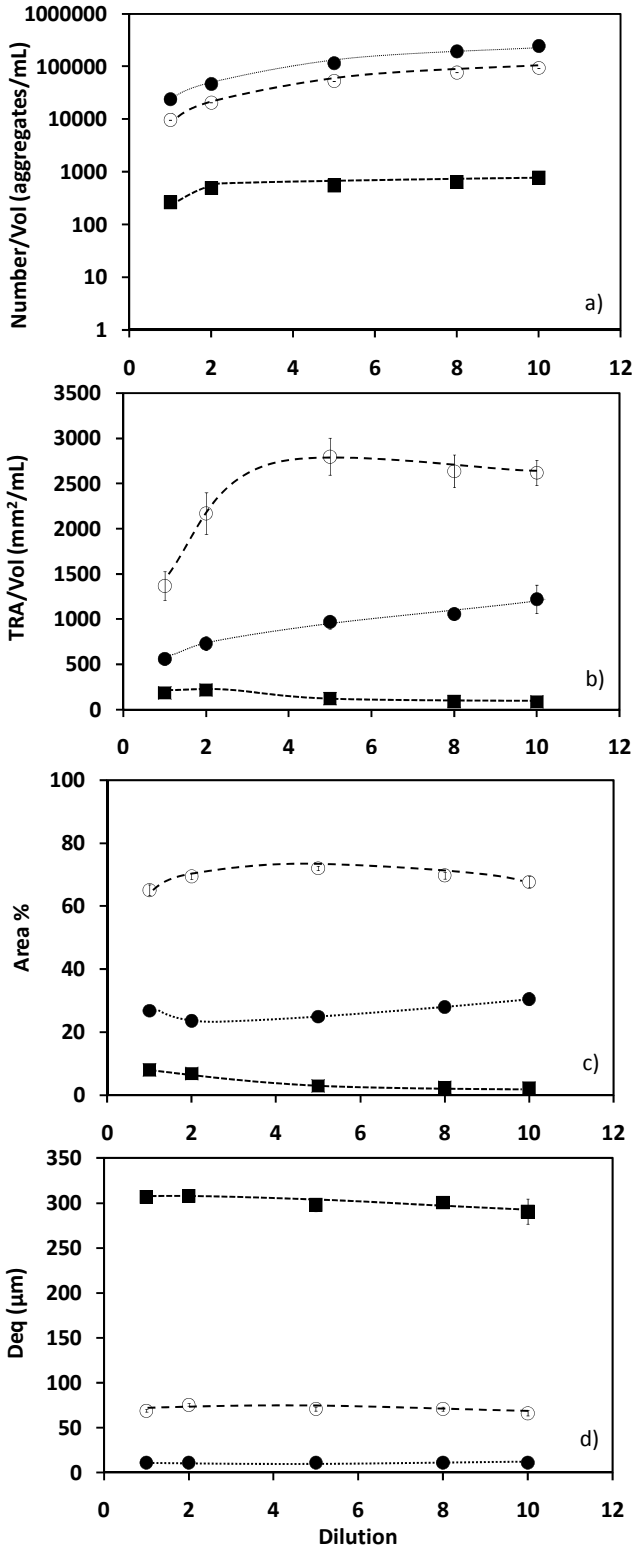

Figure 4.4. (a) Average number of aggregates per volume; (b) TRA/Vol, (c) average area percentage, (d) aggregates equivalent diameter for small (●), intermediate (○) and large (■) aggregates.

The results depicted in Figure 4.4a revealed a stabilization of the number of larger aggregates after a twofold dilution, following an initial increase (from 264 to 496 aggregates/mL). However, the analysis of the TRA/Vol behavior for larger aggregates (Figure 4.4b) demonstrates that from the twofold dilution onwards there was an apparent decrease in the contents of those aggregates (run of four consecutive descending values), which could represent a deflocculating mechanism. With respect to the number of intermediate and small aggregates, both increased with dilution (*z*-values of 1.993 in both cases), pointing to either a deflocculating process or a disclosure of aggregates. Furthermore, until the tenfold dilution, no asymptotic value was found, and their numbers reached a value 10.19 times larger than the non-diluted sample for small aggregates, and 9.73 times larger for intermediate aggregates. Comparing these results with the TRA/Vol behavior (Figure 4.4b), it could be seen that the twofold dilution caused the disclosure of enclosed aggregates, particularly of intermediate aggregates (marked increase from 1366 mm^2/mL to 2169 mm^2/mL in the TRA/Vol of intermediate aggregates). From the twofold to the fivefold dilution, the increase on the contents of intermediate (20540 to 53352 aggregates/mL and 2169 to 2798 mm^2/mL) and smaller (46457 to 114832 aggregates/mL and 729 to 969 mm^2/mL) aggregates behaved in the opposite manner to the larger aggregates (218 to 122 mm^2/mL). This suggests that there was some deflocculation of the larger aggregates (decrease in their contents), coupled with further disclosure of the intermediate and smaller aggregates (as shown by the overall increase of biomass contents, from the twofold to the fivefold dilution; see Figure 4.3a). From the fivefold dilution onwards there seems to be a change in the driving force behind the aggregated biomass behavior. Larger aggregates continued to deflocculate, and to some extent the intermediate aggregates may have also deflocculated, as pointed by the decrease (run of three consecutive descending values) of the TRA/Vol values of the larger and intermediate aggregates until the tenfold dilution. The increase (run of three consecutive ascending values) of the

contents of smaller aggregates (Figure 4.4b) confirms this hypothesis. Except for the behavior of the number of intermediate and small aggregates throughout the dilution range, no *z*-values from the *runs test* were obtained to verify data randomness because there were too few data points (fewer than five) in the remaining cases.

The aggregates area percentage parameter (Figure 4.4c) established the contribution of each aggregate class to the overall aggregated biomass. The increase of the intermediate area percentage for the twofold dilution (from 65.09% to 69.51%) was in opposition to the behavior of the smaller (26.84% to 23.61%) and larger aggregates (8.08% to 6.89%), and supports a prevalent disclosure of intermediate aggregates. With respect to the fivefold dilution, the significant decrease (6.89% to 3.00%) in the larger aggregates area percentage, and the increase in the intermediate (69.51% to 72.02%) and smaller (23.61% to 24.98%) aggregates, is consistent with an intermediate and smaller aggregates disclosure and a possible deflocculation of the larger aggregates. From the fivefold dilution onwards the small aggregates area percentage increased (run of three consecutive ascending values) while it decreased for the intermediate and larger aggregates (runs of three consecutive descending values), consistent with the deflocculation of the intermediate and larger aggregates.

The behavior of the aggregates equivalent diameter (Figure 4.4d) demonstrated that the small and intermediate aggregates remained somewhat stable with dilution (*z*-values of 0.66 for both cases), thus not fully confirming the possible deflocculation of the intermediate aggregates from the fivefold dilution onwards. However, the equivalent diameter of the larger aggregates decreased from the twofold dilution onwards (run of four consecutive descending values), sustaining the deflocculation of these aggregates with the dilutions increase.

In order to allow the establishment of a procedure to select the best dilution factor, the percentage aggregate recognition was determined, for each of the images acquired within each dilution, and is presented in Figure 4.5.

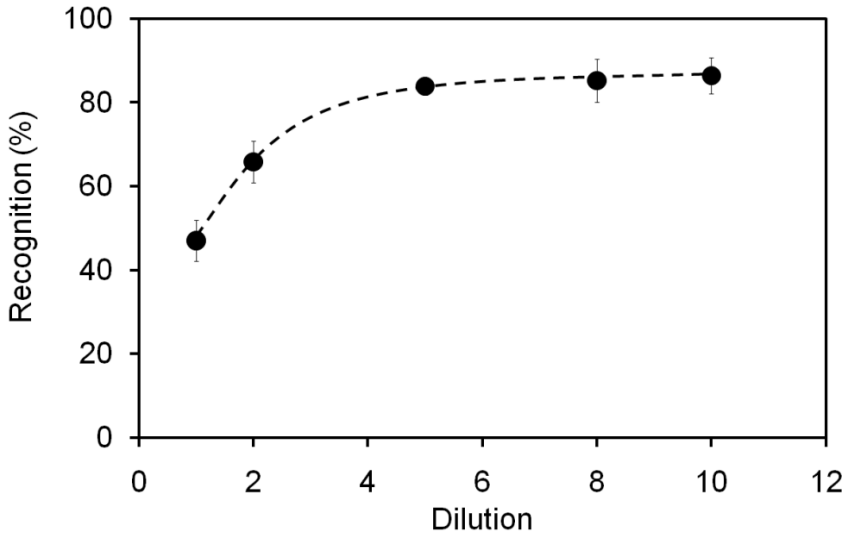

Figure 4.5. Aggregates recognition behavior for each dilution factor.

The aggregates recognition percentage increased until the fivefold dilution (run of three consecutive ascending values), remaining fairly stable from this dilution onwards (Figure 4.5), in agreement with the TL/Vol, TRA/Vol and TL/TRA parameters behavior. The recognition percentage attained a maximum of approximately 85%, demonstrating an improved ability to separate enclosed and overlapping aggregates up to the fivefold dilution (increase of the aggregates recognition percentage). Furthermore, the 85% recognition percentage value may point to a maximum threshold for this parameter, considering the likelihood of aggregates being bisected by the image boundaries (implying recognition percentages below 100%) even at low magnification. Therefore, it seems that the recognition percentage parameter could be considered a valuable method for identifying the optimal dilution factor to assess aggregated biomass contents correctly.

4.3.2 MAGNIFICATION STUDY

The effect of magnification on the determination of aggregates and protruding filamentous bacteria in the samples was determined. Given the poor pixel representation of any given aggregate under lower magnification, no morphological comparison, except for size, was performed between the two magnifications studied. As shown in Figure 4.2, the magnification change from 100x to 20x does not allow the aggregates contour to be recorded accurately and no filamentous bacteria can be recognized.

As shown in Figure 4.6, a total of 13 samples, comprising both magnifications (approximately 25 images for each sample), were used to determine the number of aggregates per image and the aggregates number percentage, for sizes ranging from 0 to 250 μm (25 μm intervals).

Figure 4.6a demonstrates that more aggregates per image are obtained when images are acquired using 20x magnification. Therefore it would be advantageous when working with higher magnifications such as 100x to increase the number of acquired images in order to increase the representativeness of the acquired data. In terms of the aggregates number percentage for each magnification, Figure 4.6b and Table 4.1 demonstrate an increase in the small aggregates for the 100x magnification, which was expected given the size constraints of the 20x magnification for smaller aggregates. In opposition the higher percentage of large aggregates in the 20x magnification is due to the size constraint of the 100x magnification for larger aggregates (Table 4.1).

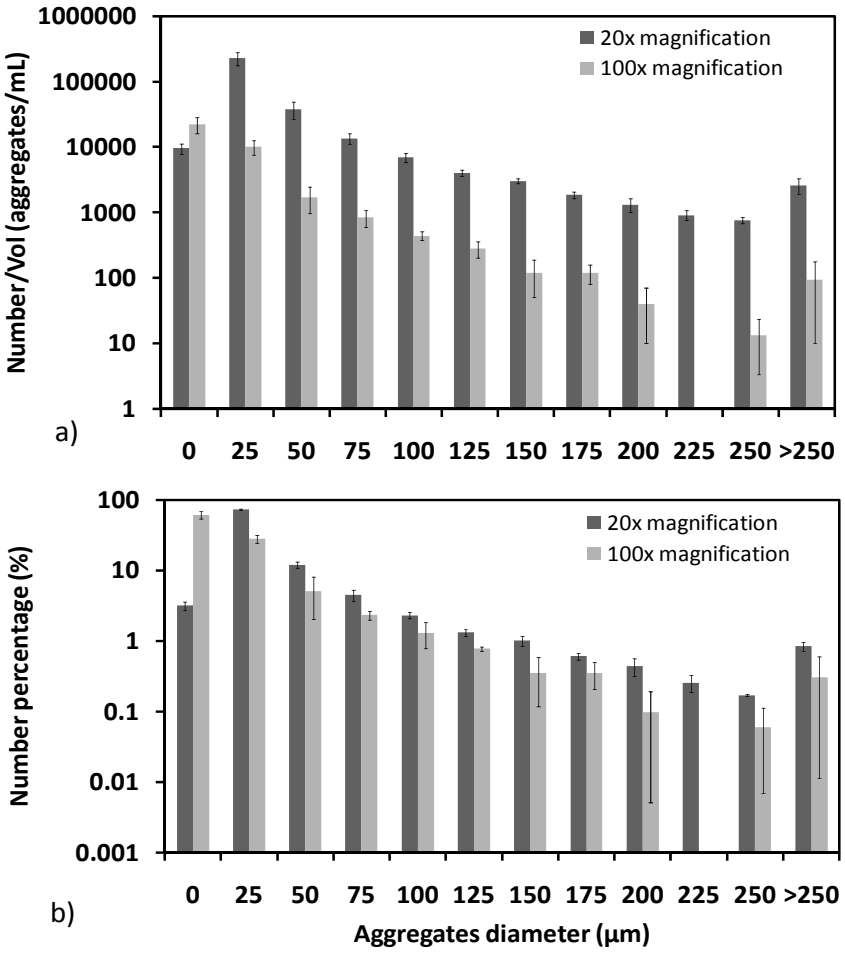

Figure 4.6. Histogram showing the behavior of aggregates based on size for 20x and 100x magnifications in total number of aggregates per image (a) and aggregates number percentage (b).

Table 4.1. Aggregates percentage for 20x and 100x magnifications for the small, intermediate and large aggregates. Standard deviations are presented in brackets.

	Magnification	Small aggregates (<25 µm)	Intermediate aggregates (25-250 µm)	Large aggregates (>250 µm)
Number %	20x	50.18 (4.47)	48.89 (4.40)	0.95 (0.30)
	100x	85.88 (3.26)	14.02 (3.25)	0.38 (0.17)
Area %	20x	24.11 (4.28)	67.70 (4.55)	8.36 (2.53)
	100x	55.87 (8.03)	42.62 (7.36)	4.30 (2.11)

Table 4.2 presents the results of the *t-test* two-sample for averages (performed for an α-value of 0.05) in terms of the number and area percentages for the small, intermediate and large aggregates, as determined by the 20x and 100x magnifications methods. The number of observations was 13 in all cases, resulting in a Critical *t*-value of 1.78. As Table 4.2 demonstrates, the two methods led to different results as, in all cases, the *t*-value obtained surpassed the Critical *t*-value, and the *p*-value was always below 0.05. The conjugation of the results presented in Table 4.1 and Table 4.2 demonstrates that 20x magnification does not allow smaller aggregates to be identified accurately, whereas 100x magnification can lead to underestimation of larger aggregate contents. Therefore, although adequate for most situations, in a scenario where large aggregates are present in representative numbers, a unique 100x magnification acquisition technique could present some limitations.

Table 4.2. *t*- and *p*-values obtained by the *t-test* two-sample for averages (α=0.05) regarding the number and area aggregates percentage for small, intermediate and large aggregates, as determined by the 20x and 100x magnifications methodologies.

	Variables	***t*-value**	***p*-value**
Number %	Small aggregates	26.59	2.45×10^{-12}
	Intermediate aggregates	26.12	3.03×10^{-12}
	Large aggregates	4.83	2.05×10^{-4}
Area %	Small aggregates	12.74	1.23×10^{-8}
	Intermediate aggregates	8.44	1.08×10^{-6}
	Large aggregates	5.52	6.63×10^{-5}

In order to study the effect of magnification on the representativeness of overall biomass contents, the correlation of TRA/Vol between 20x and 100x magnification acquisitions was calculated and is presented in Figure 4.7, whereas the average TRA/Vol and standard errors are presented in Table 4.3.

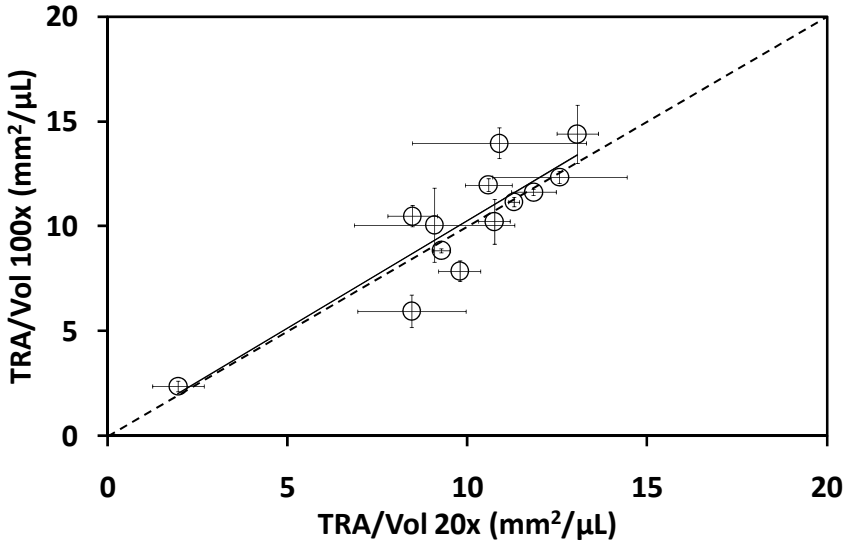

Figure 4.7. Correlation analysis of TRA/Vol between the 20x and 100x magnification acquisitions.

Table 4.3. Average TRA/Vol for the 20x and 100x magnification acquisitions and corresponding standard deviations.

	Magnification		
	20x	**100x**	**Error %**
TRA/Vol ($mm^2/\mu L$)	9.85	10.07	2.55
Standard Deviation ($mm^2/\mu L$)	1.69	1.07	-

The results presented in Figure 4.7 reveal a good correlation (slope of 1.03) between the total aggregated biomass determined by the 100x and the 20x acquisition methods, although some dispersion was evident (R^2 of 0.78). Furthermore, the average TRA/Vol values determined for the whole datasets (Table 4.3) demonstrated a high correlation between the two magnifications, resulting in an overestimation of 2.55% for the 100x magnification. In addition, the standard error obtained for the 100x magnification results was 36.7% lower than that for the 20x magnification, indicating that the 100x magnification acquisition method was more stable than the determination of aggregates contents. A two-sample *t-test* for averages (α-value of

0.05) was performed for the TRA/Vol parameter, as determined by the 20x and 100x magnifications methods. The number of observations was 13, resulting in a Critical *t*-value of 1.78. Analyzing the *t*- and *p*-values obtained (0.59 and 0.28 respectively), demonstrates that the two methods gave similar results, as the *t*-value was lower than the Critical *t*-value and the *p*-value larger than 0.05. Therefore, it may be concluded that the 100x magnification and acquisition methodology allows both aggregates and filament data to be determined from the same slide, and smaller aggregates to be identified and characterized with no significant loss of accuracy regarding the representativeness of total biomass contents. However, when large aggregates are present in representative numbers, a complementary smaller magnification analysis would be advantageous.

4.4. CONCLUSIONS

This study demonstrated the vulnerability of image analysis to dilutions. It was shown that it is rather difficult to predict and quantify the dilution effect for the aggregated and filamentous biomass as a whole, on the basis of a single correction factor. Furthermore, when the assessment of biomass contents is no longer affected by further dilutions, this may not be true for the biomass structure. Therefore, the optimal operating dilution must be carefully established, and the determination of the aggregates recognition percentage could be a valuable methodology for identifying this dilution factor. Regarding the image acquisition methodology, it seems that the 100x magnification causes no loss in accuracy compared with the 20x magnification in terms of biomass representativeness, for most cases, with the advantages of allowing aggregated and filamentous bacteria to be determined from a single image, and smaller aggregates to be identified and characterized.

4.5. REFERENCES

Abreu, A.A., Costa, J.C., Araya-Kroff, P., Ferreira, E.C., and Alves, M.M. (2007) Quantitative image analysis as a diagnostic tool for identifying structural changes during a revival process of anaerobic granular sludge. *Water Research* **41**, 1473-1480.

Amaral, A.L. (2003) *Image Analysis in Biotechnological Processes: Applications to Wastewater Treatment*. PhD Thesis, University of Minho, Braga, Portugal (http://hdl.handle.net/1822/4506).

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

Amaral, A.L., Ginoris, Y.P., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2008) Stalked protozoa identification by image analysis and multivariable statistical techniques. *Analytical and Bioanalytical Chemistry* **391**, 1321-1325.

Banadda, E.N., Smets, I.Y., Jenné, R., and Van Impe, J.F. (2005) Predicting the onset filamentous bulking in biological wastewater treatment systems exploiting image analysis information. *Bioprocess and Biosystems Engineering* **27**, 339-348.

Bradley, J.V. (1968) *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs (N.J.)

Cenens, C., Van Beurden, K.P., Jenné, R., and Van Impe, J.F. (2002) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **46** (1-2), 381-387.

Costa, J.C., Abreu, A.A., Ferreira E.C., and Alves, M.M. (2007) Quantitative Image Analysis as a Diagnostic Tool for Monitoring Structural Changes of Anaerobic Granular Sludge During Detergent Shock Loads. *Biotechnolology and Bioengineering* **98** (1), 60-68.

da Motta, M., Amaral, A.L., Neves, L., Araya-Koff, P., Ferreira, E.C., Alves, M.M., Mota, M., Roche, N., Vivier, H., and Pons, M.N. (2002) Dilution effects on biomass characterization by image analysis. *In Proceedings of the 14th Brazilian Congress on Chemical Engineering,* CD-ROM (p. 9), Natal (Brazil).

da Motta, M., Pons, M.N., and Roche, N. (2001) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* **43** (7), 91-96.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007) Recognition of protozoa and metazoa using image analysis tools, discriminant analysis, neural networks and decision trees. *Analytica Chimica Acta* **595**, 160-169.

Ginoris, Y.P., Amaral, A.L., Nicolau, A., Coelho, M.A.Z., and Ferreira, E.C. (2007) Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system. *Water Research* **41**, 2581-2589.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technololy* **54** (1), 167-174.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Jin, B., Wilén, B.M., and Lant, P. (2003) A comprehensive insight into floc characteristics and their impact on compressibility and settleability of activated sludge. *Chemical Engineering Journal* **95**, 221-234.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2009a) Monitoring of activated sludge settling ability through image analysis: validation on full-scale wastewater treatment plants. *Bioprocess and Biosystems Engineering* **32** (3), 361-367.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2009b) Correlation between sludge settling ability and image analysis information using Partial Least Squares. *Analytica Chimica Acta* **642** (1-2), 94-101.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2010) A Comparison between bright field and phase contrast image analysis techniques in activated sludge morphological characterization. *Microscopy and Microanalysis* **16** (2), 166-174.

Wilén, B.M., Lumley, D., Mattsson, A., and Mino, T. (2008) Relationship between floc composition and flocculation and settling properties studied at a full scale activated sludge plant. *Water Research* **42**, 4404-4418.

Abstract

In this chapter, eight different wastewater treatment plants were studied for a combined period of two years regarding the aggregated and filamentous bacteria morphology and contents. An image analysis program was used to characterize the biomass and a multivariatre partial least squares method allowed the prediction of the sludge volume index. The obtained results allowed the detection of filamentous bulking events and subsequent estimation of the sludge volume index. The developed image analysis methodology proved to be a feasible method for the continuous monitoring of activated sludge systems and identification of disturbances.

5. CHAPTER 5 – MONITORING OF ACTIVATED SLUDGE SETTLING ABILITY THROUGH IMAGE ANALYSIS: VALIDATION ON FULL-SCALE WASTEWATER TREATMENT PLANTS

5.1. INTRODUCTION

Biological processes are very sensitive to external and internal variations, affecting their optimum working conditions. In activated sludge systems, an adequate balance between the different types of bacteria is essential to ensure an efficient pollution removal, good sludge settling ability and low suspended solids effluent levels. After the oxidation of the organic matter in the aerated tank, the flocculated biomass is separated from the treated effluent by means of their settling ability in a settling tank. The settling phase is considered a critical stage of the process in which filamentous bulking and deflocculation processes are the most common problems, causing the decrease of the sludge settling ability and effluent quality deterioration (Sezgin *et al.*, 1978; Sezgin, 1982; Ganczarczyk, 1994; Li and Yuan, 2002; Jenkins *et al.*, 2003; Schuler and Jang, 2007a,b).

The solid-liquid separation in wastewater treatment plants (WWTP) is commonly assessed through the sludge volume index (SVI) evaluation, characterizing the sludge settling ability. However, as the SVI only provides a macroscopic evaluation of the activated sludge, the microscopic characteristics of the sludge should be also determined by examination under an optical microscope. Therefore, it comes as no surprise that microscopic image analysis procedures have been used, in recent years, to complement the wastewater treatment processes mainly on the morphological characterization and evaluation of aggregated and filamentous bacteria contents. As the activated sludge process efficiency is dependent on the aggregates physical properties, some basic parameters have been used to study the settling ability of the microbial aggregates such as the filamentous bacteria contents, aggregates contents, size, shape, density, porosity and settling velocity, as well as consistency and specific

surface area (Cenens *et al.*, 2002a,b; Jenné *et al.*, 2006). In fact, image analysis procedures are, nowadays, considered to be a feasible method to characterize quantitatively aggregates and protruding filamentous bacteria, and subsequently used to relate to the settling abilities (Grijspeerdt and Verstraete, 1997; Banadda *et al.*, 2005; Jenné *et al.*, 2003, 2007; da Motta *et al.*, 2001a), assess biomass morphology changes and monitor filamentous bulking events in pilot and full-scale plants (da Motta *et al.*, 2001a,b; Amaral and Ferreira, 2005).

Following past studies (da Motta *et al.*, 2001a; Amaral and Ferreira, 2005; Amaral, 2003), the present work aims to survey the filamentous bacteria and aggregates contents and morphology of the biomass collected from 8 full-scale wastewater treatment plants during a combined period of two years. For that purpose, an image processing and analysis program was developed for bright field microscopy, providing the necessary data for monitoring the biological system. The results based on image analysis procedures will be related on one hand to the operating conditions, and on the other, to the reported events in the WWTPs.

5.2. MATERIALS AND METHODS

The biomass used in this study was collected from the aeration basins of eight WWTP (WWTP1, WWTP2, WWTP3, WWTP4, WWTP5, WWTP6, WWTP7 and WWTP8), treating domestic effluents, located in the North of Portugal. Samples were taken to perform physical measurements, on one hand, and microscopic observations, on the other, in order to estimate the contents and morphology of the microbial aggregates and protruding filamentous bacteria by image acquisition and analysis. The biomass settling ability was measured through the determination of the sludge volume index in a 10 L cylindrical column (15 cm diameter), with the sludge height variation

monitored for 30 min. Total suspended solids (TSS) were determined by weight, and used to determine the SVI parameter (APHA, 1989).

5.2.1 IMAGE ACQUISITION

For each sample, a volume of 25 µL was placed on a slide and covered with a 20mm × 20mm cover slip for visualization and image acquisition for all the surveyed wastewater treatment plants. The volume deposition was performed by means of a recalibrated micropipette with a sectioned tip at the end to a diameter allowing the passage of even the larger aggregates. For each monitored day three individual sample slides were obtained and screened in the upper, median and lower third each, with images grabbed for each four fields of view.

Except for WWTP8, around 200 images per sample were acquired in bright field microscopy to obtain representative information of the sludge. All the images were acquired in a Leitz Laborlux S optic microscope (Leitz, Wetzlar), with 100x magnification, coupled to a Zeiss AxioCam HRc (Zeiss, Oberkochen) camera. The image acquisition was performed in 1300x1030 pixels and 8-bit format through the AxioVision 3.1 (Zeiss, Oberkochen) software.

The image acquisition of the WWTP8, had already been performed for an earlier study and, relied on 25 images per sample acquired in bright field microscopy in a SZ 4045TR-CTV Olympus stereo microscope (Olympus, Tokyo), with 40x magnification, coupled to a Sony CCD AVC–D5CE (Sony, Tokyo) grey scale video camera. Filament image acquisition was accomplished through phase contrast microscopy on a Zeiss Axioscop microscope (Zeiss, Oberkochen) with a 100x magnification, for an average 30 images per sample. The image acquisition was performed in 768 × 576 pixels and 8-bit format by a Data Translation DT 3155 (Data Translation, Marlboro) frame grabber using the commercial software package Image Pro Plus (Media Cybernetics, Silver Spring).

5.2.2 IMAGE PROCESSING AND ANALYSIS METHODOLOGY

The image processing and analysis program for aggregated and filamentous bacteria, was developed in Matlab 7.3 (The Mathworks, Inc., Natick) language, adapting a previous version developed by Amaral and Ferreira (2005). Primarily, the image processing step determined the binary images from the aggregated biomass and the protruding filamentous bacteria and thereafter, the morphological parameters were determined. Figure 5.1 shows an example of the original, binary and labeled images resulting of the main steps of the program, comprising the image pre-treatment, segmentation, and debris elimination whereas the image analysis program was oriented to the aggregates and filaments contents determination.

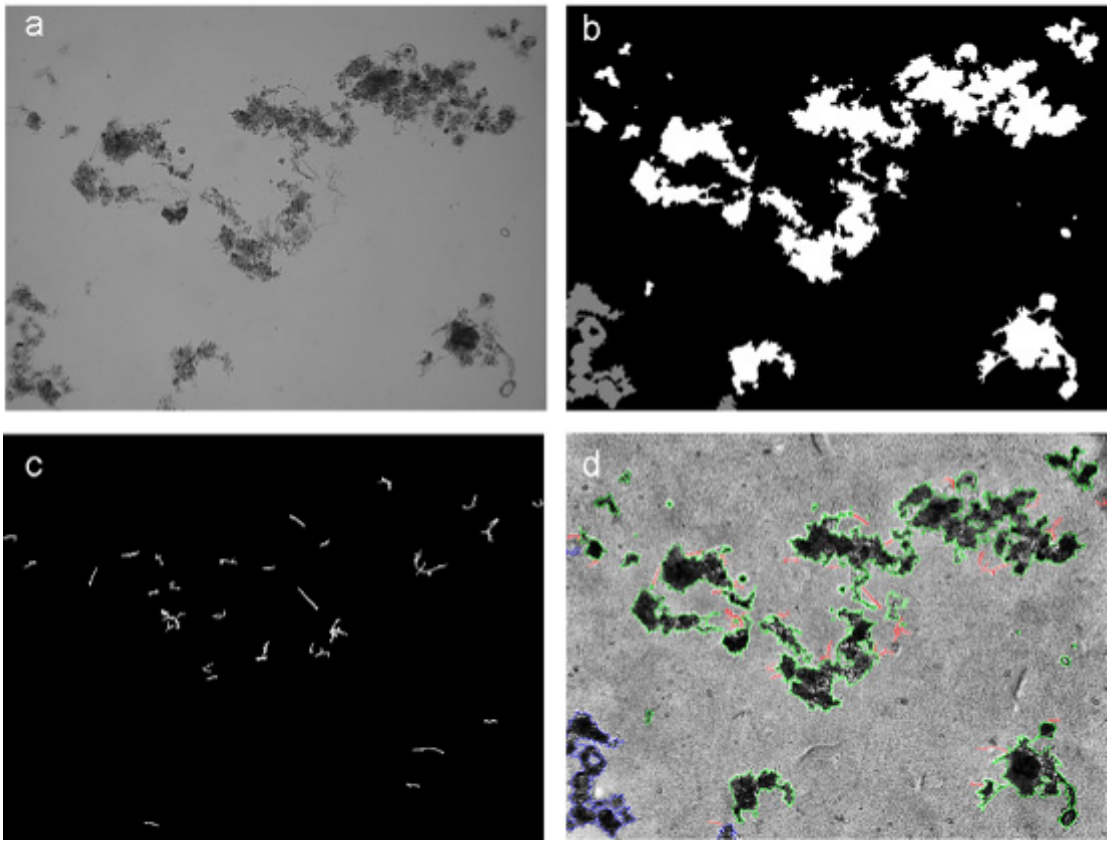

Figure 5.1. Original image from the activated sludge system with 100x magnification (a), binary aggregates image (b), binary filaments image (c) and final labeled image (d).

5.2.3 IMAGE ANALYSIS PARAMETERS

Supported on the previous study of Amaral and Ferreira (2005), 36 parameters were determined either directly from the image analysis program either in association with the sludge physical properties for a total of around 400 000 aggregates. Total aggregates number per sludge volume, total aggregates area per sludge volume, aggregates area, total filaments length per sludge volume, filaments length, aggregates length, aggregates perimeter and aggregates equivalent diameter were determined for all the samples collected. The morphological descriptors convexity, solidity, roundness, and eccentricity were also determined by the image analysis methodology. Furthermore, the aggregates characterization was subsequently divided in three classes in order to differentiate between small low settling structures (aggregates below 25 μm in equivalent diameter), normal sized good settling structures (aggregates between 25 and 250 μm in equivalent diameter), and large connected low settling structures (aggregates above 250 μm in equivalent diameter). The total filaments length per sludge volume, filaments average length per aggregates average area ratio, filaments average length per aggregates average equivalent diameter ratio and the total filaments length per total aggregated area ratio were determined alongside the total filaments length per volatile suspended solids ratio characterizing the filaments dynamics. A more detailed description of each parameter can be found in Amaral (2003).

5.2.4 PARTIAL LEAST SQUARES

The collected data set (5184 data points from 144 samples x 36 parameters) was correlated by Partial Least Squares regression (PLS), an iterative algorithm that extracts linear combinations of the essential features of the original **X** data while modeling the **Y** data dependence on the work set, being therefore well suited for multivariate calibration. PLS have been shown to be an efficient approach in

monitoring complex processes since the high dimensional strongly cross-correlated data can be reduced to a much smaller and interpretable set of latent variables (Teppola *et al.*, 1997). To perform the PLS analysis from the data set, *SIMCA 8.0* (*Umetrics*, *Umëa*) software package was used. *SIMCA* computes the variable importance in the projection (VIP) as the sum over all model dimensions (PLS components) of the variable influence. The parameters that are found to be the most important are the ones presenting a VIP value larger than 1. A more detailed description about this method can be found in Umetri (1998).

5.3. RESULTS AND DISCUSSION

The PLS analysis allowed to predict the SVI values from the image analysis data with a reasonable correspondence as it can be seen in Figure 5.2. Analyzing the results it was found that the multiple correlation coefficient (goodness of fit) attained for the SVI prediction was 0.90 (R^2), allowing to predict, at some extent, the sludge volume index values from the obtained image analysis data.

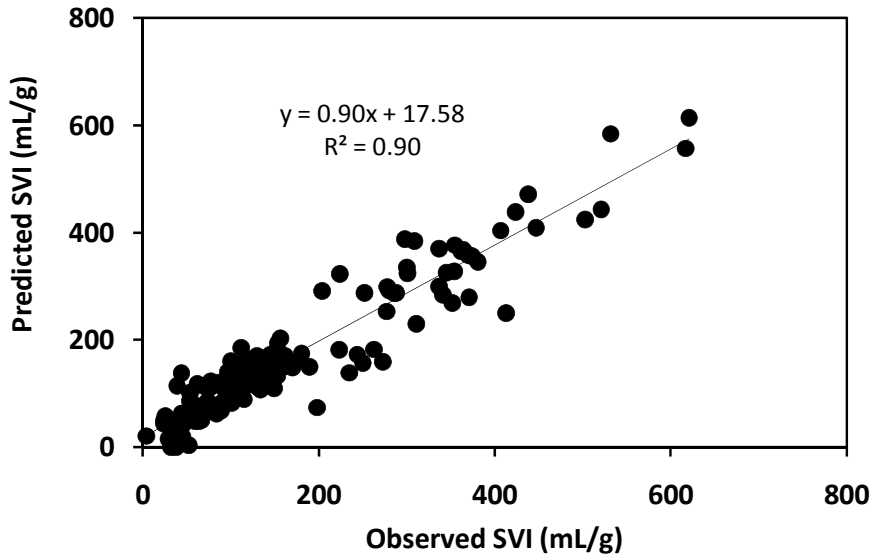

Figure 5.2. Global correlation between the observed and predicted SVI data points.

It must be noticed that, above the 200 mL/g SVI threshold, there seems to be a higher dispersion regarding the observed versus the predicted results and, hence, the accuracy of the prediction will not be as high as for lower SVI values. However, this behavior may be explained by the fact the majority of the data points above 200 mL/g were acquired in a previous study with an acquisition of an average of 25 to 30 images per sample (acquisition for WWTP8) instead of the average 200 images for the other wastewater treatment plants. In accordance to this explanation, no significant correlations were found between the errors of the predicted SVI and the SVI values within the WWTP8 data set (R^2 of 0.02) and between the errors of the predicted SVI and the SVI values within the remaining WWTP data set (R^2 of 0).

The observed SVI values monitored throughout the eight different wastewater treatment plants are presented in Figure 5.3, alongside the predicted SVI values.

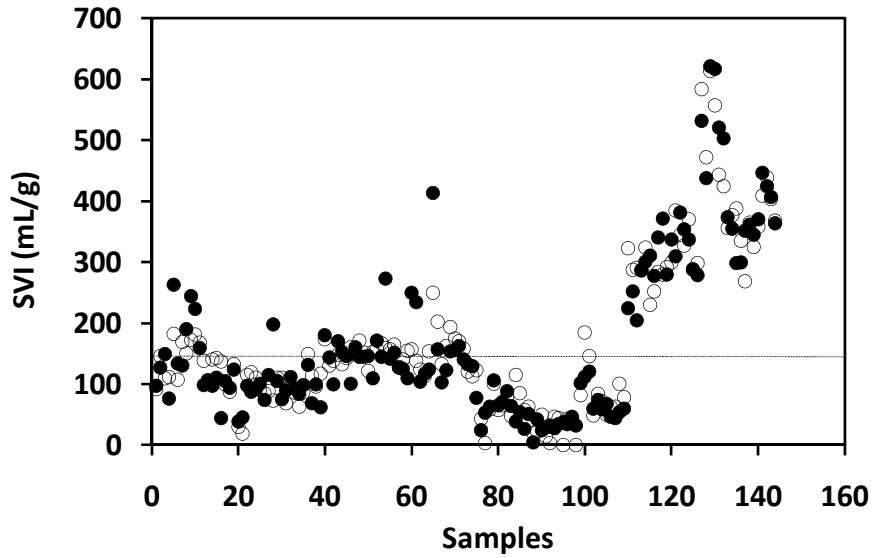

Figure 5.3. Global observed (•) and predicted (o) SVI data points from the WWTP aerated tanks.

From the analysis of Figure 5.3 it can be perceived that the predicted SVI values follow, at some extent, the observed SVI trends corroborating the findings of the PLS analysis results. Furthermore, it was found that the observed and predicted SVI

values differ between each other around 31 mL/g in average. It was also found that the difference between the values is smaller for the lower SVI values, around 24 mL/g for the set of SVI values below 200 mL/g, and larger for higher SVI values, around 49 mL/g for the set of SVI values above 200 mL/g. Once again this behavior may be explained by the fact the majority of the data points above 200 mL/g rely on an average of 25 to 30 images per sample instead of the average 200 images for the data points below 200 mL/g.

In most cases for the SVI below the bulking limit (150 mL/g according to Jenkins *et al.*, 2003) there is a correspondence to the predicted SVI (presenting also values below 150 mL/g). An analogous trend is also found for the values indicating filamentous bulking phenomena with predicted SVI values above the 150 mL/g limit. In fact for the 89 data points that presented SVI values below 150 mL/g, this methodology correctly predicted 77 (86.5%) of these as non bulking situations, whereas for the 55 data points presented SVI values above 150 mL/g, this methodology correctly predicted 49 (89.1%) of these as bulking situations. Given the 24 mL/g uncertainty for the SVI values below 200 mL/g and the fact that the bulking limit is set on 150 mL/g, a closer analysis of the values outside this uncertainty area (150±24 mL/g) reveals that only 6 (6.7%) values below the lower limit and 3 (5.5%) values above the upper limit are incorrectly classified.

For an in-depth analysis of each of the WWTP, their individual SVI values (observed and predicted) are represented in Figure 5.4. From this figure, it can be withdrawn that the eight aerated tanks have different operating conditions, and some with strong fluctuations in terms of the sludge settling abilities.

For WWTP1, the SVI results, here depicted on the Figure 5.4a, presented values mainly below the 150 mL/g threshold limit for bulking events. Only five days (days 28, 49, 56, 63, and 70) presented SVI values higher than 150 mL/g (bulking phenomena), ranging from 160 to 265 mL/g. The predicted values for this WWTP

were, for the most part, in accordance with the observed SVI values with larger deviations occurring on days 28, 56 and 105 (all above 50 mL/g). With respect to the bulking prediction ability, only 2 days (out of 23) were incorrectly classified, with one of those days falling into the uncertainty area (150±24 mL/g), thus meaning that only one day was effectively misclassified.

With respect to the WWTP2 (Figure 5.4b), low SVI values predominated throughout the survey, as the SVI remained below the bulking threshold (150 mL/g) during all the monitoring period with exception of days 14 and 149 with 198 mL/g and 181 mL/g, respectively. Such was also the case of the WWTP5 (between 20 and 100 mL/g), WWTP6 (below 50 mL/g) and WWTP7 (between 40 and 120 mL/g), depicted on Figure 5.4e, Figure 5.4f and Figure 5.4g, respectively. Once again, the predicted values for these WWTPs were in accordance with the observed SVI values with a larger deviation (above 50 mL/g) occurring solely on days 14 and 128 of the WWTP2 and day 9 of the WWTP7. With respect to the bulking prediction ability, only two days (out of 51) were misclassified, falling outside the uncertainty area (150±24 mL/g).

Regarding the WWTP3 (Figure 5.4c) and WWTP4 (Figure 5.4d), the SVI values were always on the threshold of configuring a bulking phenomenon and in some days presented mild bulking phenomena. Furthermore, in the last monitored day of WWTP3 and on days 35 to 42 and day 105 of WWTP4 presented considerable bulking problems. Again, the predicted values for these two WWTPs were, for the most part, in accordance with the observed SVI values with larger deviations occurring on days 56 and 112 for WWTP3 and on days 35, 42 and 105 for WWTP4 (above 50 mL/g). With respect to the bulking prediction ability, 14 days (out of 35) were incorrectly classified, but 8 of those days fell into the uncertainty area (150±24 mL/g), thus meaning that only 6 days were effectively misclassified.

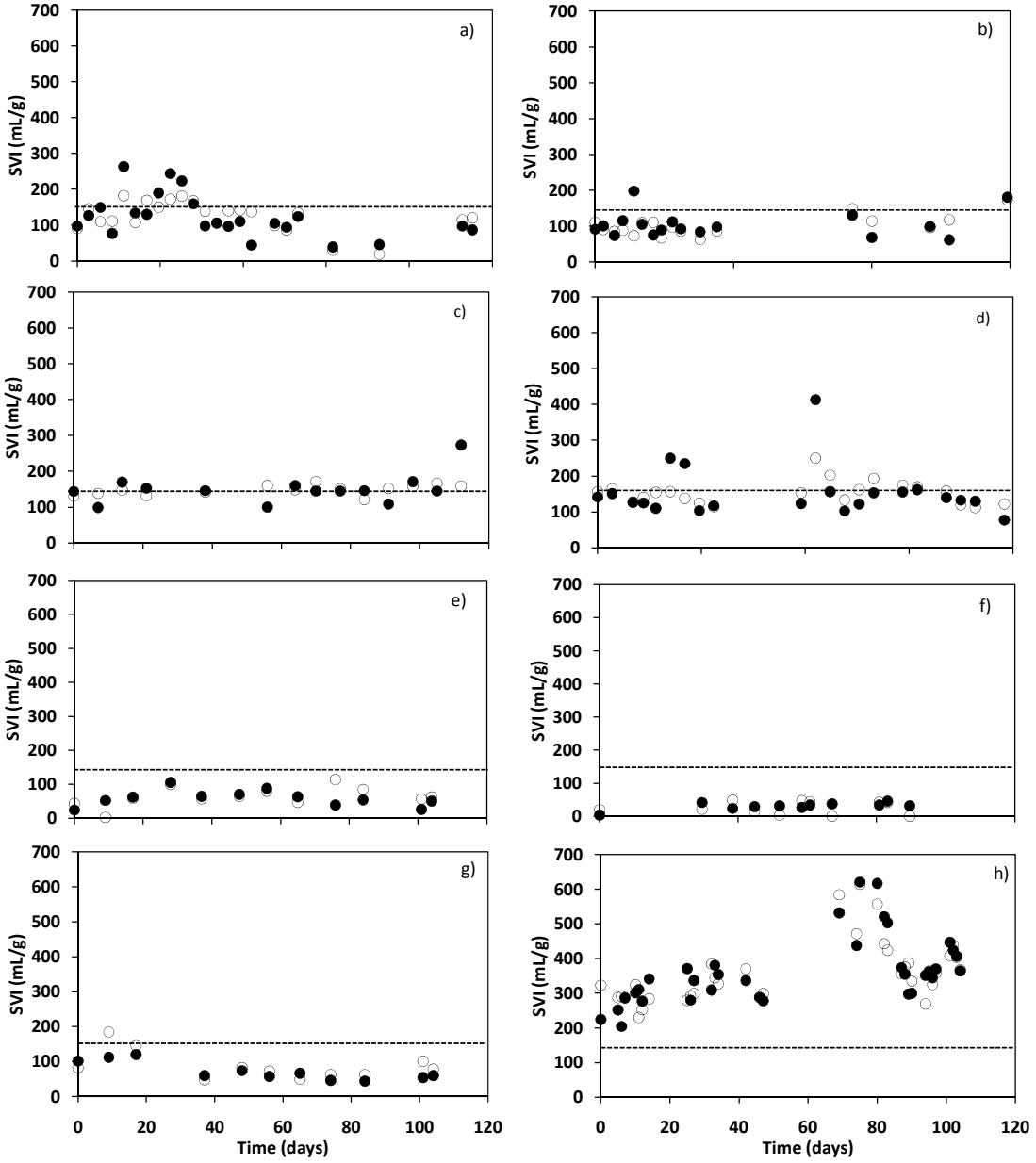

Figure 5.4. Observed (•) and predicted (o) SVI data points from the eight WWTP aerated tanks. a) WWTP1 b) WWTP2 c) WWTP3 d) WWTP4 e) WWTP5 f) WWTP6 g) WWTP7 h) WWTP8.

The SVI behavior in WWTP8, depicted on Figure 5.4h, revealed a severe bulking problem, throughout all the monitoring period, with values ranging from 204 mL/g up to 617 mL/g. It was found a reasonable correspondence between the predicted and observed SVI values for this WWTP with larger deviations (above 50 mL/g)

occurring on 12 of the 35 days (days 1, 6, 11, 14, 25, 32, 69, 80, 82, 83, 89 and 94). Again it should be emphasized the fact that this WWTP data points rely on an average of 25 to 30 images per sample instead of the average 200 images for the remaining WWTP, hence, explaining the poorer correspondence between the predicted and observed SVI values. However, and with respect to the bulking prediction ability, not a single day (out of 35) was incorrectly classified. Furthermore, it should also be stressed that, the predicted SVI followed the observed SVI behavior throughout all the monitoring period.

In a global analysis, it should also be noticed that, the two WWTP that were chronically on the threshold of bulking events (WWTP3 and WWTP4) accounted for 14 of the 18 days incorrectly classified, and taking into account the uncertainty area (150±24 mL/g) for 6 of 9 days effectively misclassified.

5.4. CONCLUSIONS

The PLS analysis of the image analysis data allowed to predict the SVI values with a reasonable correspondence as it can be inferred by the 0.947 correlation coefficient obtained between the observed and predicted SVI values. It should be kept in mind, however, that although the obtained correlation holds true for the SVI range presented in the surveyed WWTPs (up to 700 mL/g) no direct extrapolations can be performed for higher SVI values regarding more severe bulking phenomena.

It was also found that, for the current methodology, the observed and predicted SVI values distanced from each other an average 24 mL/g for the set of SVI values below 200 mL/g, which comprises the bulking threshold limit. This fact allowed the establishment of an uncertainty area (150±24 mL/g) regarding the effectiveness of classifying bulking/non-bulking situations. That being the case, the results revealed

that only 9 out of 144 (6.3%) samples outside the uncertainty area were, in fact, incorrectly classified according to this methodology.

In conclusion, the obtained results showed that the developed image analysis and PLS methodology proved to be adequate for continuous examination of sludge settling properties, in terms of the sludge volume index. Furthermore, this methodology was able to predict, for the most part, the filamentous disturbances within the aerated tank of eight different wastewater treatment plants.

5.5. REFERENCES

Amaral, A.L. (2003) *Image analysis in biotechnological processes: applications to wastewater treatment*. PhD Thesis, University of Minho, Braga, Portugal (http://hdl.handle.net/1822/4506).

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

APHA, AWWA, and WPCF (1989) *Standard Methods for the Examination of Water and Wastewater*. 17th Edition, American Public Health Association, Washington D.C.

Banadda, E.N., Smets, I.Y., Jenné, R., and Van Impe, J.F. (2005) Predicting the onset filamentous bulking in biological wastewater treatment systems exploiting image analysis information. *Bioprocess Biosystem Engineering* **27**, 339-348.

Cenens, C., Jenné, R., and Van Impe, J.F. (2002a) Evaluation of different shape parameters to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **45** (4-5), 85-91.

Cenens, C., Van Beurden, K.P., Jenné, R., and Van Impe, J.F. (2002b) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **46** (1-2), 381-387.

da Motta, M., Amaral, A.L., Casellas, M., Pons, M.N., Dagot, C., Roche, N., Ferreira, E.C., and Vivier, H. (2001a) Characterization of activated sludge by automated image analysis: validation on full-scale plants. *Proceedings of the IFAC Computer Applications in Biotechnology*, Québec City, Canada, 427-431.

da Motta, M., Pons, M.N., and Roche, N. (2001b) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* **43** (7), 91-96.

Ganczarczyk, J.J. (1994) Microbial Aggregates in Wastewater Treatment. *Water Science and Technology* **30**, 87-95.

Grijspeerdt, K. and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31**, 1126-1134.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jenné, R., Banadda, E.N., Philips, N., and Van Impe, J.F. (2003) Image Analysis as a Monitoring Tool for Activated Sludge Properties in Lab-Scale Installations. *Journal of Environmental Science and Health, Part A - Toxic/Hazardous Substances and Environmental* **38** (10), 2009-2018.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Jenné, R., Banadda, E.N., Smets, I.Y., and Van Impe, J.F. (2004) Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* **50** (7), 281-285.

Li, X. and Yuan, Y. (2002) Settling velocities and permeabilities of microbial aggregates. *Water Research* **36**, 3110-3120.

Schuler, A.J. and Jang, H. (2007a) Causes of variable biomass density and its effects on settling in full scale biological wastewater treatment systems. *Environmental Science and Technology* **41** (5), 1675-1681.

Schuler, A.J. and Jang, H. (2007b) Density effects on activated sludge zone settling velocities. *Water Research* **41** (8), 1814-1822.

Sezgin, M. (1982) Variation of Sludge Volume Index with Activated Sludge Characteristics. *Water Research* **16**, 83-88.

Sezgin, M., Jenkins, D., and Parker, D. (1978) A unified theory of filamentous activated sludge bulking. *Journal of Water Pollution Control Federation* **50**, 362-381.

Teppola, P., Mujunen, S.P., and Minkkinen, P. (1997) Partial least squares modeling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems* **38**, 197-208.

Umetri, A.B. (1998) *User's Guide to SIMCA*. CD-ROM.

Abstract

The large amounts of activated sludge data collected with the image analysis implementation can be treated by multivariate statistical procedures such as PLS. In this chapter the implementation of image analysis and PLS techniques has shown to provide important information for better understanding the behavior of activated sludge processes, and to predict, at some extent, the sludge volume index. The obtained results allowed explaining the strong relationships between the sludge settling properties and the free filamentous bacteria contents, aggregates size and aggregates morphology, establishing relevant relationships between macroscopic and microscopic properties of the biological system.

6. CHAPTER 6 – CORRELATION BETWEEN SLUDGE SETTLING ABILITY AND IMAGE ANALYSIS INFORMATION USING PARTIAL LEAST SQUARES

6.1. INTRODUCTION

In activated sludge systems, an adequate balance between the different types of bacteria is necessary to ensure an efficient pollution removal, good sludge settling abilities and low suspended solids levels in the final effluent. Poor settling biomass is normally attained by an improper aggregate's formation and filamentous bacteria proliferation, resulting in lower clarifier efficiency. Usually, some malfunctions may occur within the activated sludge system such as pinpoint flocs formation, filamentous bulking, dispersed growth and zoogleal bulking (Bitton, 1994).

Image analysis procedures, based on microscopic observations, is nowadays considered to be a feasible method to characterize quantitatively aggregates and filamentous bacteria, and subsequently used to monitor bulking events in pilot and full plants (da Motta *et al.*, 2001a; Amaral and Ferreira, 2005), exceeding the initial manual quantification of filamentous bacteria proposed by Jenkins *et al.* (1986). Bulking can be caused by both filamentous and non-filamentous factors, affecting in different ways the sludge settling ability which can be detected by image analysis methodologies (Amaral and Ferreira, 2005; Jenné *et al.*, 2004, 2006, 2007; da Motta *et al.* 2001b,c; Palm *et al.*, 1980). The combination of settling properties and the parameters obtained from image analysis may offer powerful information enabling immediate interventions on the biological system. In fact, the study developed by Sezgin (1982) established that the Sludge Volume Index (SVI) is strongly influenced by floc size and filamentous bacteria contents. Other authors (Matsui and Yamamoto, 1984; Ganczarczyk, 1994; Grijspeerdt and Verstraete, 1997) used automated image analysis to relate the microorganism's morphology in biological systems with the

sludge settling properties. The settling ability can be subsequently related with the microscopic parameters using multivariable statistical techniques, such as partial least squares regression (PLS) and principal component analysis (PCA) (Amaral and Ferreira, 2005; Jenné *et al.*, 2006). A close correlation between the filamentous bacteria per suspended solids ratio and the SVI was indeed achieved by Amaral and Ferreira (2005) during filamentous bulking events.

Encouraged by the success of image analysis procedures over the last years in a broad range of different areas, the present work uses an automated image analysis method to characterize the activated sludge structure, focusing on the prediction ability of sludge settling properties, in both good settling and filamentous bulking periods. In this sense, the collected images were treated in order to characterize the aggregated and filamentous bacteria, thus originating the different morphological parameters, further used as independent **X** variables in the PLS regression for the sludge volume index (SVI) model determination.

6.2. MATERIALS AND METHODS

The activated sludge samples analyzed in this work were collected during several months from the aeration basins of eight municipal wastewater treatment plants (WWTP) treating domestic effluents located in the North of Portugal (WWTP1, WWTP2, WWTP3, WWTP4, WWTP5, WWTP6, WWTP7, WWTP8). A total of 142 samples (representing different days in different WWTP) were collected (17 from WWTP1, 33 from WWTP2, 14 from WWTP3, 23 from WWTP4, 21 from WWTP5, 11 from WWTP6, 12 from WWTP7 and 11 from WWTP8) for a period between 4 to 7 months and time spans varying from 3 until 7 days between each sample. From each sample the total suspend solids (TSS) was measured by weight (APHA *et al.*, 1989)

and further used to calculate the sludge volume index (SVI) in a 10 L cylindrical column (15 cm diameter) for 30 min, as follows:

$$\mathrm{SVI} = \frac{\mathrm{H}_{30}}{\mathrm{H}_0 \times \mathrm{TSS}} \quad (1)$$

where H_0 and H_{30} are the height in the time 0 and 30 min, respectively. The SVI is highly used for sludge settling ability evaluation, varying inversely with the sludge ability to settle, making it an interesting parameter within the system efficiency assessment.

Microscopic observations were performed for all samples in order to determine the morphological parameters of microbial aggregates and filaments contents using digital image analysis. The maximum period of time between sample collection and image acquisition did not exceed 2 h with no longer than 30 min without aeration.

6.2.1 IMAGE ACQUISITION

For image acquisition, a volume of 25 µL from each sample was distributed on a slide and covered with a 20 mm × 20 mm cover slip for visualization and image acquisition. The volume deposition was performed by means of a recalibrated micropipette with a sectioned tip at the end to a diameter allowing the passage of even the larger aggregates. During the survey, three different slides (replicates) were screened in order to minimize sampling errors, and images were acquired for 22 different positions in the upper, middle and lower parts of the slide. Therefore, a total around 200 images per sample (22 x 3 positions x 3 replicates) were acquired in bright field microscopy to obtain representative information of the sludge. All the images were acquired in a Leitz Laborlux S optic microscope (Leitz, Wetzlar), with 100x magnification, coupled to a Zeiss AxioCam HRc (Zeiss, Oberkochen) camera. The

image acquisition was performed in 1300 × 1030 pixels and 8-bit format through the AxioVision 3.1 (Zeiss, Oberkochen) software. This acquisition methodology was performed for WWTP1, WWTP2, WWTP3, WWTP4, WWTP5, WWTP6 and WWTP7.

A previous survey studied by Amaral and Ferreira (2005) was also included in this work representing an earlier period of filamentous bulking conditions in one of the studied WWTP (described as WWTP8 samples). The image acquisition of the WWTP8 survey relied on 25 images per sample (three slides were also used with the purpose described above) acquired in bright field microscopy in a SZ 4045TR-CTV Olympus stereo microscope (Olympus, Tokyo), with 40x magnification, coupled to a Sony CCD AVC–D5CE (Sony, Tokyo) grey scale video camera. Filament image acquisition was accomplished through phase contrast microscopy on a Zeiss Axioscop microscope (Zeiss, Oberkochen) with a 100 times magnification, for an average 30 images per sample. The image acquisition was performed in 768 × 576 pixels and 8-bit format by a Data Translation DT 3155 (Data Translation, Marlboro) frame grabber using the commercial software package Image Pro Plus (Media Cybernetics, Silver Spring). In order to compare both acquisition methodologies, calibration from pixels to the metric unit dimensions was performed by means of a micrometer slide.

6.2.2 IMAGE PROCESSING

The image processing and analysis program for aggregated and filamentous bacteria characterization was developed in Matlab 7.3 (The Mathworks, Inc., Natick) language, adapting a previous version developed by Amaral (2003). Primarily, the image processing step established the binary images from the aggregated biomass and the protruding and freely dispersed filamentous bacteria and thereafter, morphological parameters were determined. In total, and during the current work, a

universe around 400 000 aggregates, from the overall 142 samples dataset of the 8 WWTPs, was acquired and individually characterized.

Figure 6.1 shows a schematic representation of the main steps of the program, comprising the image pre-treatment, segmentation, and debris elimination whereas the image analysis program was oriented to the aggregated and filamentous bacteria characterization and contents determination.

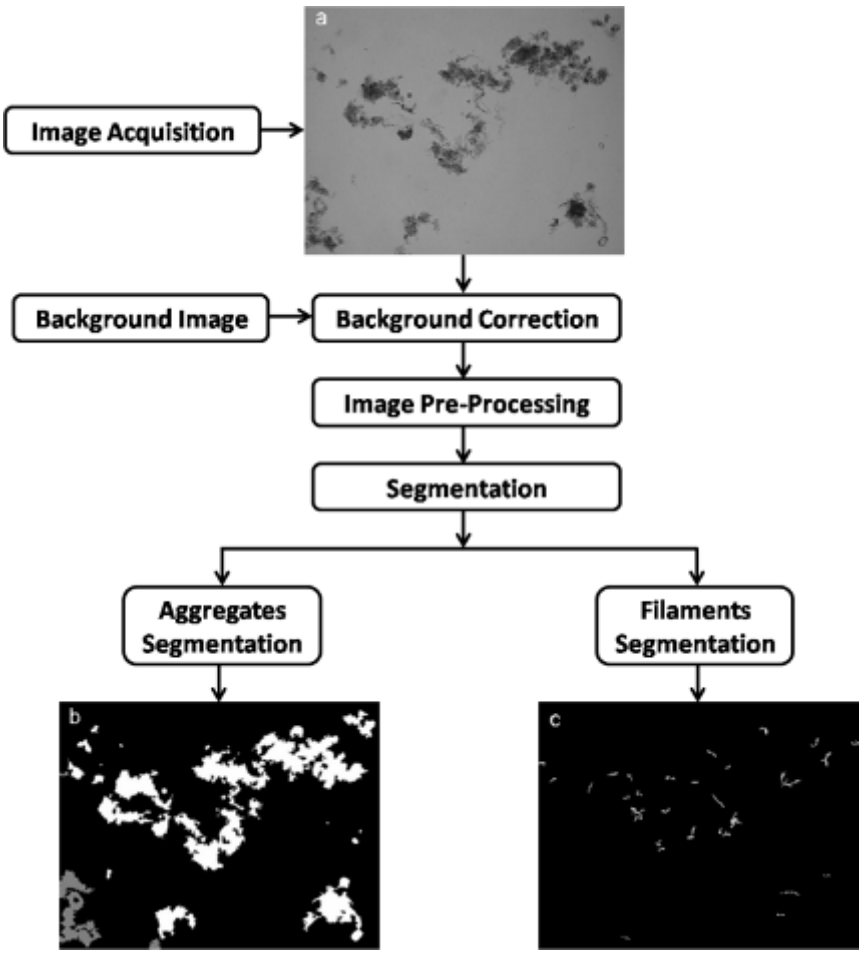

Figure 6.1. Schematic representation of the image processing procedures.

6.2.3 IMAGE ANALYSIS PARAMETERS

Following the image processing step, the recognized aggregated and filamentous bacteria from the collected images, were treated in order to individually characterize each aggregate and filamentous bacteria, in terms of the most relevant morphological parameters. Therefore, and based on the previous study of Amaral and Ferreira (2005) 36 morphological parameters were determined (in addition to the 2 sludge physical properties, TSS and SVI), either directly from the image analysis program, either in association with the sludge physical properties. The overall dataset of 38 parameters is presented in Table 6.1.

Total aggregates number per sludge volume (Nb/Vol), total aggregates area per sludge volume (TA/Vol), aggregates individual area (Area) and total filaments length per sludge volume (TL/Vol), described in Amaral (2003), were determined for all the samples collected alongside the filaments individual length (FL), aggregates individual length (L), aggregates individual perimeter (P) and aggregates individual equivalent diameter (Deq) described next.

The filaments individual length (FL) was determined according to (Walsby and Avery, 1996), with N_{Thn} as the pixel sum of each thinned filament, N_{int} as the number of filaments intersections and factor 1.1222 used to average the different measuring angles within the image:

$$FL = \left(N_{Thn} + N_{int}\right) \times 1.1222 \times F_{cal} \quad (2)$$

The aggregates individual equivalent diameter (Deq) was calculated based on the area determination with F_{Cal} as the calibration factor ($\mu m.pixel^{-1}$) (Glasbey and Horgan, 1996; Russ, 1995):

$$D_{eq} = 2F_{Cal}\sqrt{\frac{Area}{\pi}} \quad (3)$$

The aggregates individual perimeter was calculated by Walsby and Avery (1996) with N_{Per} as the pixel sum of the objects and factor 1.1222 used to average the different measuring angles within the image:

$$P = N_{Per} \times 1.1222 \times F_{Cal} \quad (4)$$

The aggregates individual length (L) was given by the maximum Feret Diameter, which is the maximum distance between two parallel tangents touching opposite borders of the object, converted to metric units (Glasbey and Horgan, 1996; Russ, 1995).

The morphological descriptors convexity (Conv), solidity (Solid), roundness (Round), and eccentricity (Ecc) were also determined for each individual aggregate, by the image analysis methodology, as follows.

Eccentricity was calculated using the area and the second order moments (M_2) of each individual aggregate as (Glasbey and Horgan, 1996):

$$Ecc = \frac{(4\pi)^2 (M_{2X} - M_{2Y})^2 + 4M_{2XY}^2}{Area^2} \quad (5)$$

Convexity was determined as the ratio between the convex perimeter (P_{Conv}) and the perimeter (P) of each individual aggregate by (Glasbey and Horgan, 1996):

$$Conv = \frac{P_{Conv}}{P} \quad (6)$$

Roundness was calculated using the area and the convex perimeter of each individual aggregate by (Glasbey and Horgan, 1996):

$$Round = \frac{4\ \pi\ Area}{P_{Conv}^2} \quad (7)$$

Solidity was determined as the ratio between each individual aggregate area and convex envelope area ($Area_{Conv}$) (Russ, 1995).

$$\text{Solid} = \frac{\text{Area}}{\text{Area}_{\text{conv}}} \quad (8)$$

Furthermore, the aggregates morphological characterization was subsequently divided in three classes in order to differentiate between small low settling structures (aggregates below 25 µm in equivalent diameter), intermediate good settling structures (aggregates between 25 µm and 250 µm in equivalent diameter), and large connected low settling structures (aggregates above 250 µm in equivalent diameter). For each of these aggregates classes (small, intermediate and large) the number of aggregates per volume was determined as the total number of the aggregates belonging to that class present in a given activated sludge volume, as well as the aggregates area percentage as the ratio between the total area of the aggregates belonging to that class and the total area of all aggregates, Those parameters were determined in order to account for the contents significance of each aggregates class, both in terms of their number and area.

Finally, total filaments length per sludge volume (TL/Vol), filaments average length per aggregates average area ratio ($FL_{avg}/Area_{avg}$), filaments average length per aggregates average equivalent diameter ratio (FL_{avg}/Deq_{avg}) and total filaments length per total aggregated area ratio (TL/TA) were determined alongside the total filaments length per total suspended solids (TL/TSS) ratio characterizing the filaments dynamics.

For clarity purposes, the dataset was classified, according to the physical and morphological meaning of each parameter, in five main descriptor groups: free filamentous bacteria contents; free filamentous bacteria characterization; aggregates contents; aggregates size and aggregates morphology. Furthermore, for the

aggregates contents, aggregates size, and aggregates morphology, a more detailed analysis was performed, including data of all aggregates, labeled as overall (ovr), intermediate (int) and large aggregates (lrg). For the determination of these values, the average (mean value) of, respectively, all (ovr), intermediate (int) and large (lrg) individual aggregates parameters, was determined. The poor pixel representation of the small aggregates within the collected images, due to their small size, means that the determination of the morphological properties is quite prone to large errors, and therefore, was not included in this study. In order to simplify the identification, Table 6.1 shows all the parameters description taken from the program including those associated with physical properties.

Table 6.1. Parameters collected from the eight wastewater treatment plants.

Parameter	Symbol	Description
Predictor (**X**)	***Free filamentous bacteria contents***	
	TL/Vol	Total filaments length per volume
	TL/TSS	Total filaments length per total suspended solids
	TL/TA	Total filaments length per total aggregates area
	Free filamentous bacteria characterization	
	$FL_{avg}/Area_{avg}$	Filaments average length per aggregates average area
	FL_{avg}	Filaments average length
	FL_{avg}/Deq_{avg}	Filaments average length per aggregates average equivalent diameter
	Aggregates contents	
	TA/Vol	Total aggregates area per volume
	Nb_{tot}/Vol	Total number of aggregates per volume
	Nb_{sml}/Vol	Number of small aggregates per volume
	Nb_{int}/Vol	Number of intermediate aggregates per volume
	Nb_{lrg}/Vol	Number of larger aggregates per volume
	TSS	Total suspended solids
	Aggregates size	
	Overall	
	$Area_{ovr}$	Aggregates average area
	% $Area_{sml}$	Area percentage of small aggregates
	% $Area_{int}$	Area percentage of intermediate aggregates

	% $Area_{lrg}$	Area percentage of larger aggregates
	Deq_{ovr}	Aggregates average equivalent diameter
	P_{ovr}	Aggregates average perimeter
	L_{ovr}	Aggregates average length
	Intermediate aggregates	
	Deq_{int}	Equivalent diameter of intermediate aggregates (mean value)
	P_{int}	Perimeter of intermediate aggregates (mean value)
	L_{int}	Length of intermediate aggregates (mean value)
	Large aggregates	
	Deq_{lrg}	Equivalent diameter of larger aggregates (mean value)
	P_{lrg}	Perimeter of larger aggregates (mean value)
	L_{lrg}	Length of larger aggregates (mean value)
	Aggregates morphology	
	Overall	
	$Conv_{ovr}$	Aggregates average Convexity
	$Round_{ovr}$	Aggregates average Roundness
	$Solid_{ovr}$	Aggregates average Solidity
	Ecc_{ovr}	Aggregates average Eccentricity
	Intermediate aggregates	
	$Conv_{int}$	Convexity of intermediate aggregates (mean value)
	$Round_{int}$	Roundness of intermediate aggregates (mean value)
	$Solid_{int}$	Solidity of intermediate aggregates (mean value)
	Ecc_{int}	Eccentricity of intermediate aggregates (mean value)
	Large aggregates	
	$Conv_{lrg}$	Convexity of large aggregates (mean value)
	$Round_{lrg}$	Roundness of large aggregates (mean value)
	$Solid_{lrg}$	Solidity of large aggregates (mean value)
	Ecc_{lrg}	Eccentricity of large aggregates (mean value)
Response (**Y**)	SVI	Sludge volume index

6.2.4 DATA REDUCTION AND PARTIAL LEAST SQUARES

A cross-correlation analysis between the collected data was then performed in order to reduce the dataset, leading to the exclusion of one variable (physical or morphological parameter) for each pair presenting a correlation factor above 90%.

The reduced dataset was further treated by Partial Least Squares regression (PLS), an iterative algorithm that extracts linear combinations of the essential features of the original **X** data while modeling the **Y** data dependence on the work set, in order to perform a multivariate calibration of the SVI variable (**Y**) from the remaining dataset (**X** variables) (Teppola *et al.*, 1997; Wold *et al.*, 2001). PLS has been shown to be an efficient approach in monitoring complex processes since the high dimensional strongly cross-correlated data can be reduced to a much smaller and interpretable set of latent variables (Teppola *et al.*, 1997; Aarnio and Minkkenen, 1986). PLS reduces the dimension of the predictor variables by extracting factors or latent variables that are correlated with **Y** while capturing a large amount of the variations in **X**. This means that PLS maximizes the covariance between matrices **X** and **Y**. In PLS, the scaled matrices **X** and **Y** are decomposed into score (**t** and **u**) and loading (**p** and **q**) vectors, and residual error matrices (**E** and **F**):

$$\mathbf{X} = \sum_{i=1}^{a} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} + \mathbf{E} \tag{9}$$

$$\mathbf{Y} = \sum_{i=1}^{a} \mathbf{u}_i \mathbf{q}_i^{\mathrm{T}} + \mathbf{F} \tag{10}$$

where *a* is the number of latent variables. In an inner relation the score vector **t** is linearly regressed against the score vector **u**.

$$\mathbf{u}_i = \mathrm{b}_i \mathbf{t}_i + \mathrm{h}_i \tag{11}$$

where *b* is regression coefficient which is determined by minimizing the residual *h*.

It is crucial to determine the optimal number of latent variables (components) and cross-validation (CV) is a practical and reliable way to test the predictive significance of each PLS regression. In order to do so, part of the training dataset is kept out of the model development, predicted by the model and finally compared with the

actual values (cross-validation). The prediction error sum of squares (PRESS) is the sum of the squared differences between the observed and predicted values for the data kept out of the model fitting. This procedure is repeated until every sample has been kept out once and only once. Therefore, the final PRESS has contributions from all data (Teppola *et al.*, 1997; Wold *et al.*, 2001).

To perform the PLS analysis from the data set, *SIMCA 8.0 (Umetrics, Umeå)* software package was used. This software first mean centers the data and scales it by the use of the standard deviation. A cross-validation method is used in which 7 subgroups are sequentially removed from the model development in each new PLS latent variable (component) determination. Furthermore, for every dimension, *SIMCA* computes the overall PRESS/SS, where SS is the residual sum of squares of the previous dimension. A component is then considered significant if PRESS/SS is statistically smaller than 1.0. *SIMCA* also displays the Q^2 value which is the fraction of the total variation of the **X**'s that can be predicted by a component, as estimated by cross-validation, and is computed as (1.0-PRESS/SS). Furthermore, *SIMCA* also determines the variable importance in the projection (VIP), which represents the influence of each variable (*k*) of the data matrix (**X**) on the results matrix (**Y**), so that the variables with a VIP larger than 1 have an influence above average on the result and are, therefore, the most relevant for explaining **Y**. *SIMCA* computes the VIP as the sum over all model dimensions (PLS components) of the variable influence. A more detailed description about this method can be found in (Umetri, 1998).

In the present study, the predictor matrix **X** consisted of 37 variables and the response matrix **Y** of the key variable SVI (Table 1) in a total of 142 input samples in total. The sludge volume index was selected as **Y** variable since it is a key indicator of the sludge settling ability. For the current PLS application, a random 95 samples were used for training (2/3 of the dataset), and the remaining 47 samples (1/3 of the dataset) for the validation of the developed models.

6.3. RESULTS AND DISCUSSION

The cross-correlation analysis on the full dataset (5396 data points from 142 samples x 38 variables, including 1 **Y** variable) allowed to exclude 13 **X** variables (TL/TA, $FL_{avg}/Area_{avg}$, FL_{avg}/Deq_{avg}, Nb_{tot}/Vol, $Area_{ovr}$, Deq_{int}, L_{int}, P_{lrg}, L_{lrg}, $Round_{ovr}$, $Solid_{ovr}$, $Solid_{lrg}$ and Ecc_{lrg}) due to the existence of a correlation factor above 0.9 with other variables. The exclusion of one variable, for each pair presenting a correlation factor above 0.9, was determined after a full 38 variables dataset PLS regression was performed, allowing to establish the most important variable (higher VIP), of the pair, for the SVI prediction. The results of the PLS regression variables importance (VIP) for 3 components (best fit according to Q^2 values), and the correlation factors, for each studied pair of variables is presented in Table 6.2. The remaining variables were automatically used for the further approach leading to a reduced dataset of 25 variables (24 **X** variables and 1 **Y** variable) used in the subsequent PLS analysis.

Table 6.2. Variables presenting correlation factors above 0.9 and respective VIP values for the PLS regression with 3 components.

Variable kept	(VIP)	Variable excluded	(VIP)	Correlation
TL/TSS	2.11	TL/TA	2.00	0.93
FL_{avg}/Deq_{avg}	0.84	$FL_{avg}/Area_{avg}$	0.73	0.90
% $Area_{sml}$	1.11	FL_{avg}/Deq_{avg}	0.84	0.91
Nb_{sml}/Vol	0.86	Nb_{tot}/Vol	0.57	0.92
Deq_{ovr}	1.27	$Area_{ovr}$	0.92	0.94
L_{ovr}	1.28	L_{lrg}	1.27	0.93
P_{ovr}	1.08	P_{lrg}	1.06	0.98
L_{int}	0.17	Deq_{int}	0.16	0.92
P_{int}	0.24	L_{int}	0.17	0.93
Ecc_{ovr}	1.01	Ecc_{lrg}	0.84	0.92
$Round_{ovr}$	0.99	$Solid_{ovr}$	0.85	0.97
$Round_{lrg}$	1.08	$Round_{ovr}$	0.99	0.95
$Round_{lrg}$	1.08	$Solid_{ovr}$	0.85	0.90
$Solid_{ovr}$	0.85	$Solid_{lrg}$	0.64	0.94

Subsequently a second PLS regression was applied using the reduced 25 variables dataset (24 **X** variables), to describe the SVI, and seemingly evaluated for a number of components (latent variables) ranging up to 22 (until a negligible correlation increase between **X** and **Y** was found). According to the results, presented in Table 6.3, a total of 2 components (latent variables) allowed to obtain the best SVI model, given the fact that only the two first components presented Q^2 values higher than the 0.097 limit (cross-validation threshold for the current 95 samples training dataset PLS analysis).

Table 6.3. Values of the cumulative R^2X, R^2Y and Q^2 values for the model and slope and regression factors for the training, validation and training + validation datasets, for the 25 variables dataset PLS regression. (Slope values are equal to the R^2 values in the model built on the training set).

Component	Model (training set)				Validation set		Training + validation sets	
	Slope and R^2	R^2X	R^2Y	Q^2	Slope	R^2	Slope	R^2
1	0.65	0.28	0.65	0.62	0.78	0.79	0.69	0.69
2	0.73	0.15	0.08	0.13	0.84	0.84	0.76	0.76
3	0.77	0.14	0.05	0.09	0.86	0.85	0.80	0.79
4	0.81	0.09	0.03	-0.04	0.86	0.90	0.82	0.84
5	0.82	0.16	0.01	0.03	0.86	0.92	0.83	0.85
6	0.84	0.03	0.02	-0.16	0.88	0.91	0.85	0.86
7	0.85	0.05	0.01	-0.03	0.87	0.92	0.86	0.87
8	0.86	0.02	0.01	-0.09	0.87	0.91	0.87	0.88
9	0.87	0.02	0.00	-0.20	0.85	0.89	0.86	0.87
10	0.87	0.01	0.00	-0.25	0.88	0.90	0.87	0.88
11	0.87	0.01	0.00	-0.24	0.88	0.89	0.87	0.88
12	0.87	0.01	0.00	-0.19	0.90	0.88	0.88	0.88
13	0.88	0.01	0.00	-0.13	0.90	0.88	0.88	0.88
14	0.88	0.01	0.00	-0.12	0.90	0.88	0.88	0.88
15	0.88	0.00	0.00	-0.11	0.90	0.88	0.88	0.88
16	0.88	0.00	0.00	-0.11	0.90	0.87	0.88	0.88
17	0.88	0.00	0.00	-0.10	0.90	0.88	0.88	0.88
18	0.88	0.00	8.6e-05	-0.06	0.91	0.88	0.89	0.88
19	0.88	0.00	3.7e-05	-0.06	0.91	0.88	0.89	0.88
20	0.88	0.00	7.8e-06	-0.06	0.91	0.88	0.89	0.88
21	0.88	0.00	2.6e-06	-0.03	0.91	0.88	0.89	0.88
22	0.88	0.00	5.1e-08	-0.02	0.91	0.88	0.89	0.88

However, it was found that the 2 components PLS regression presented quite low cumulative fractions on the variation of the **Y**'s variables (0.73 for cumulative R^2Y), for a cumulative Q^2 of 0.67. Furthermore, the results obtained for the regression analysis between the predicted and observed SVI values (Table 6.3), were shown to present a mediocre prediction ability with a correlation coefficient R^2 of 0.84 for the validation dataset, and even lower for the training dataset with a correlation coefficient R^2 of 0.73. Also, when a regression analysis between the predicted and observed SVI values from both the validation and the training datasets was performed (training + validation results in Table 6.3) a mediocre prediction ability was found with a correlation coefficient R^2 of 0.76. The inclusion of further components on the model was studied, taking into account overfitting problems that may arise by the inclusion of too many components. Thus, the number of components was chosen as the best compromise of high correlation coefficient and slope close to one, and always taking the validation set as reference. Analyzing Table 6.3, it was found that the 6 components PLS regression presented the best set of conditions regarding the combination of high correlation coefficient (0.91) and slope value close to one (0.88). Figure 6.2 presents the regression analysis between the predicted and observed SVI values for the 25 variables dataset (24 **X** variables) PLS regression with 6 components, where the attained aggregated dataset correlation coefficient R^2 was 0.86 for a regression slope of 0.85.

Subsequently, a selection was carried out, regarding the variables presenting variable importance (VIP) values higher than 1, in order to perform a third PLS study, resulting in a total of 10 variables for both 2 and 6 components PLS regression, as presented in Table 6.4.

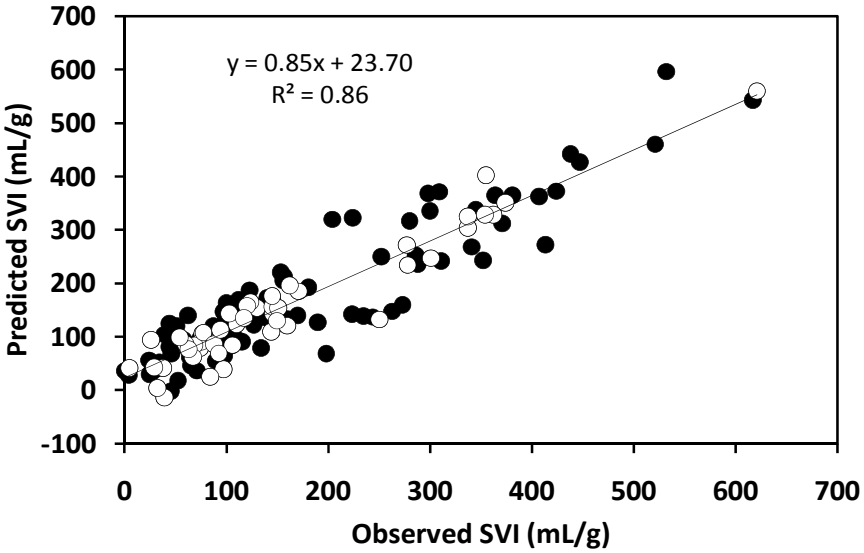

Figure 6.2. Relationship between the predicted and observed SVI for the 25 variables dataset PLS regression with 6 components. Training dataset (•) and validation dataset (o).

Table 6.4. Variables presenting a VIP larger than 1, for the 25 variables dataset PLS regression with 2 and 6 components.

Descriptor group	Variable	VIP 2 components	VIP 6 components
Free filamentous bacteria contents	TL/TSS	2.13	2.01
	TL/Vol	1.67	1.64
Aggregates size	L_{ovr}	1.29	1.23
	Deq_{ovr}	1.27	1.23
	% $Area_{sml}$	1.11	1.08
	P_{ovr}	1.07	1.02
Aggregates morphology	$Solid_{int}$	1.40	1.41
	$Conv_{int}$	1.39	1.31
	$Round_{lrg}$	1.08	1.04
	Ecc_{ovr}	1.02	1.08

The results showed that the TL/TSS was the most important variable (larger VIP), corroborating the findings of Amaral and Ferreira (2005), followed by the TL/Vol, both representing the free filamentous bacteria contents, thus clearly stating the

importance of this descriptor group. It was also found that 8 other **X** variables presented a VIP higher than 1 including the intermediate aggregates convexity and solidity, the overall eccentricity and the large aggregates roundness for the aggregates morphology descriptor group, the overall equivalent diameter, perimeter and length, and the small aggregates area % for the aggregates size descriptor group. Furthermore, from these results it could be established that both the free filamentous bacteria characterization and the aggregates contents seemed to play no significant role in predicting the SVI, as no variable from these descriptor groups had a VIP larger than 1 for a 2 and 6 components PLS regressions.

Subsequently a third PLS regression was applied using the 11 variables dataset (10 **X** variables and 1 **Y** variable), to describe the SVI, and seemingly evaluated for a number of components (latent variables) ranging up to 10 (number of **X** variables). According to the results, presented in Table 6.5, a total of 2 components (latent variables) allowed to obtain the best SVI model, given the fact that only the two first components presented Q^2 values higher than the 0.097 Q^2 limit (cross-validation threshold for the current 95 samples training dataset PLS analysis).

However, it was found that the 2 components PLS repression presented quite low cumulative fractions on the variation of the **Y**'s variables (0.77 for cumulative R^2Y), for a cumulative Q^2 of 0.74. Furthermore, the results obtained for the regression analysis between the predicted and observed SVI values (Table 6.5), were shown to present a mediocre prediction ability with a correlation coefficient R^2 of 0.83 for the validation dataset, and even lower for the training dataset with a correlation coefficient R^2 of 0.77. Also, when a regression analysis between the predicted and observed SVI values from both the validation and the training datasets was performed (training + validation results in Table 6.5) a mediocre prediction ability was found with a correlation coefficient R^2 of 0.80. Following the same strategy of the full model, the number of components was increased to 5 regarding the

combination of high correlation coefficient (0.89) and slope value close to one. Figure 6.3 presents the regression analysis between the predicted and observed SVI values for the 11 variables dataset (10 **X** variables) PLS regression with 5 components, where the attained aggregated dataset correlation coefficient R^2 was 0.84 for a regression slope of 0.83. With these results it is clear that the decrease in dimensionality from the full data set model to the models with 24 or even 10 variables does not significantly worsen the SVI prediction ability in terms of correlation coefficient or regression slope in the representation of $Y_{predicted}$ vs $Y_{observed}$.

Table 6.5. Values of the cumulative R^2X, R^2Y and Q^2 values for the model and slope and regression factors for the training, validation and training + validation dataset, for the 11 variables dataset PLS regression. (Slope values are equal to the R^2 values in the model built on the training set).

Component	Model (training set)				Validation set		Training + validation sets	
	Slope and R^2	R^2X	R^2Y	Q^2	Slope	R^2	Slope	R^2
1	0.66	0.54	0.66	0.64	0.75	0.81	0.69	0.70
2	0.77	0.11	0.11	0.27	0.83	0.88	0.79	0.80
3	0.80	0.13	0.04	0.08	0.83	0.88	0.81	0.83
4	0.81	0.08	0.01	-0.03	0.83	0.89	0.82	0.84
5	0.82	0.04	0.01	-0.05	0.85	0.89	0.83	0.84
6	0.83	0.04	0.00	-0.07	0.84	0.88	0.83	0.85
7	0.83	0.02	0.00	-0.14	0.85	0.89	0.83	0.85
8	0.83	0.02	0.0003	-0.12	0.85	0.88	0.84	0.85
9	0.83	0.01	4.0e-05	-0.08	0.85	0.89	0.84	0.85
10	0.83	0.01	7.3e-07	-0.05	0.85	0.89	0.84	0.85

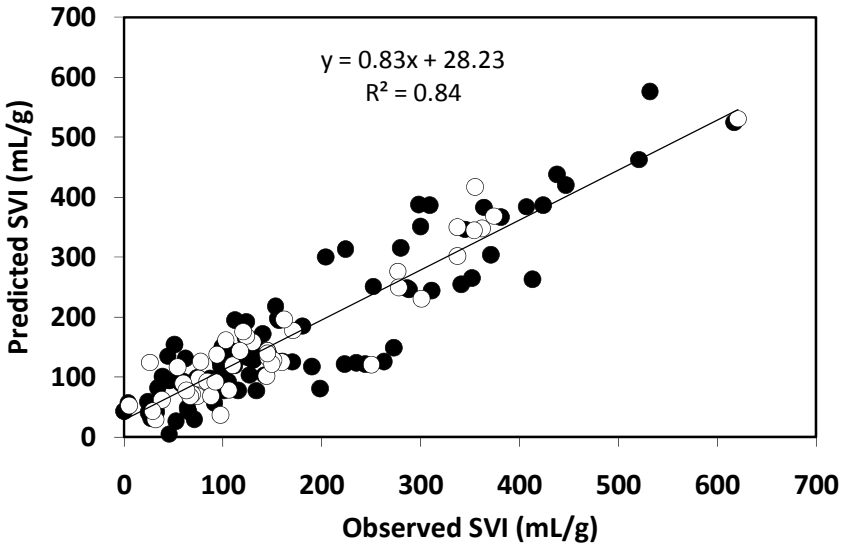

Figure 6.3. Relationship between the predicted and observed SVI for the 11 variables dataset PLS regression with 5 components. Training dataset (•) and validation dataset (o).

It should be stressed that the PLS prediction, apart from providing an SVI estimation as a function of the X dataset, also helps to identify the most relevant parameters linked to the increase or decrease of the SVI. Therefore, the most important variables regarding SVI prediction (higher VIP values) were sought, resulting in a total of 2 variables for the 2 components PLS regression, and 4 variables for the 5 components PLS regression, as presented in Table 6.6. The results showed that the TL/TSS was the most important variable (larger VIP), once again corroborating the findings of Amaral and Ferreira (2005), followed by the TL/Vol, both representing the free filamentous bacteria contents, thus clearly stating the importance of this descriptor group. It was also found that 2 other **X** variables presented a VIP higher than 1 in the 5 components PLS regression, namely the intermediate aggregates convexity and solidity both regarding the aggregates morphology descriptor group. In addition, the regression coefficients of 0.61, 0.16, -0.50 and 0.22, for the 5 components PLS regression, regarding respectively the TL/TSS, TL/Vol, intermediate aggregates solidity and intermediate aggregates convexity, allowed to establish the relationships

between these variables and the SVI. In that sense, it could be established that an increase on the filaments contents and filaments contents per suspended solids, as well as on the intermediate aggregates borders roughness (convexity) seems to lead to the increase of the SVI and, therefore of low settling abilities. Inversely, the increase on the intermediate aggregates solidity seems to lead to the decrease of the sludge volume index. In fact, the obtained relationships could be expected for filamentous bulking phenomena, characterized by high filaments to aggregates ratios.

Table 6.6. Variables importance values for the 11 variables dataset PLS regression for 2 and 5 components.

Descriptor group	Variable	VIP 2 components	VIP 5 components
Free filamentous bacteria contents	TL/TSS	1.57	1.52
	TL/Vol	1.16	1.17
Aggregates size	L_{ovr}	0.95	0.92
	Deq_{ovr}	0.90	0.91
	% $Area_{sml}$	0.76	0.78
	P_{ovr}	0.85	0.85
Aggregates morphology	$Solid_{int}$	0.98	1.05
	$Conv_{int}$	0.90	1.01
	$Round_{lrg}$	0.83	0.82
	Ecc_{ovr}	0.74	0.74

6.4. CONCLUSIONS

Relating to the work of Amaral and Ferreira (2005), in the course of this survey, it was possible to study a wider range of SVI data, comprising good and poor settling ability properties of the sludge, whereas in the earlier study the SVI values comprised solely values larger than 250 mL g^{-1}. In this way, and given the attained correlation

coefficients and established relationships, it seems reasonable to infer that the PLS regressions performed during this study showed some promising results.

Moreover, the obtained results allowed explaining the strong relationships between the sludge settling properties, as described by the SVI, and some filamentous bacteria and aggregated biomass morphological descriptor groups such as the free filamentous bacteria contents, aggregates size and aggregates morphology, establishing relevant relationships between macroscopic and microscopic properties of the biological system. Furthermore, the performed parameters reduction study allowed the establishment of a 10 independent variables procedure for SVI estimation, after the identification of the most important variables regarding SVI prediction, and the demonstration of the utmost importance of the free filamentous bacteria contents descriptor group on the SVI prediction ability.

6.5. REFERENCES

Aarnio, P., and Minkkinen, P. (1986) Application of partial least-squares modelling in the optimization of a waste-water treatment plant. *Analytica Chimica Acta* **191**, 457-460.

Amaral, A.L., and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544** (1-2), 246-253.

Amaral, A.L. (2003) *Image analysis in biotechnological processes: applications to wastewater treatment.* PhD Thesis, University of Minho, Braga, Portugal. (http://hdl.handle.net/1822/4506)

APHA, AWWA, and WPCF (1989) *Standard Methods for the Examination of Water and Wastewater*. 17th Edition, American Public Health Association, Washington D.C.

Bitton, G. (1994) *Wastewater Microbiology*, Wiley-Liss, New York.

da Motta, M., Pons, M.N., and Roche, N. (2001a) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* **43** (7), 91-96.

da Motta, M., Amaral, A.L., Casellas, M., Pons, M.N., Dagot, C., Roche, N., Ferreira, E.C., and Vivier, H. (2001b) Characterization of activated sludge by automated image analysis: validation on full-scale plants. *Proceedings of the 8th International Conference on Computer Applications in Biotechnology* Québec city, 427-431.

da Motta, M., Pons, M.N., Roche, N., and Vivier, H. (2001c) Characterization of activated sludge by automated image analysis. *Biochemical Engineering Journal* **9**, 165-173.

Dagot, C., Pons, M.N., Casellas, M., Guibaud, G., Dollet, P., and Baudu, M. (2001) Use of image analysis and rheological studies for the control of settleability of filamentous bacteria: application in SBR reactor. *Water Science and Technology* **43** (3), 27-33.

Ganczarczyk, J.J. (1994) Microbial Aggregates in Waste-Water Treatment. *Water Science and Technology* **30** (8), 87-95.

Glasbey, C.A., and Horgan, G.W. (1995) *Image Analysis for the Biological Sciences*. John Wiley and Sons, Chichester.

Grijspeerdt, K., and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31** (5), 1126-1134.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jenné, R., Banadda, E.N., Smets, I., Deurinck, J., and Van Impe, J. (2007) Detection of filamentous bulking problems: Developing an image analysis system for sludge composition monitoring. *Microscopy and Microanalysis* **13** (1), 36-41.

Jenné, R., Banadda, E.N., Smets, I.Y., and Van Impe, J.F. (2004) Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* **50** (7), 281-285.

Matsui, S., and Yamamoto, R. (1984) The use of a color TV technique for measuring filament length and investigating sludge bulking causes. *Water Science and Technology* **16** (10-11), 69-81.

Palm, J.C., Jenkins, D., and Parker, D.S. (1980) Relationship between organic loading, dissolved-oxygen concentration and sludge settleability in the completely-mixed activated-sludge process. *Journal Water Pollution Control Federation* **52** (10), 2484-2506.

Russ, J.C. (2002) *The Image Processing Handbook*. Fourth edition, CRC Press, Boca Raton.

Sezgin, M. (1982) Variation of sludge volume index with activated sludge characteristics. *Water Research* **16**, 83-88.

Teppola, P., Mujunen, S.P., and Minkkinen, P. (1997) Partial least squares modeling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems* **38**, 197-208.

Umetri, AB (1998) *User's Guide to SIMCA*. CD-ROM.

Walsby, A.E., and Avery, A. (1996) Measurement of filamentous cyanobacteria by image analysis. *Journal of Microbiological Methods* **26** (1-2), 11-20.

Wold, S., Sjöström, M., and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58** (2), 109-130.

Abstract

This chapter proposes the detection of different types of disturbances in an activated sludge system. Most reported image analysis methodologies have been used in activated sludge processes with the aim of filamentous bulking detection, though, other disturbances could be expected in WWTPs. Such disturbances can lead to fluctuations in biomass contents, affecting the total suspended solids (TSS) and the sludge settling ability (sludge volume index - SVI). Four experiments were conducted simulating filamentous bulking, zoogleal or viscous bulking, pinpoint flocs formation, and normal operating conditions. Alongside the TSS and SVI determination, the aggregated and filamentous biomass contents and morphology was studied, as well as the biomass Gram and viability status, by means of image analysis.

7. CHAPTER 7 – DISTURBANCES DETECTION IN A LAB-SCALE ACTIVATED SLUDGE SYSTEM USING IMAGE ANALYSIS

7.1. INTRODUCTION

Typically, an activated sludge system is a continuous biological process extensively used in wastewater treatment plants (WWTPs), which includes primarily a metabolic conversion of the contaminants in a biochemical stage. Subsequently, the sludge is separated from the treated effluent in a clarifier. For an effective treatment, sludge flocculation and rapid settling in the clarifier is crucial to the successful operation of an activated sludge system. As most of biological processes, activated sludge systems are sensitive to sudden changes leading to large economical and environmental consequences. In an attempt to monitor and control the biological process, the past decades have witnessed a significant growth of activated sludge inspection by means of optical microscopy. In this context, research has raised regarding the determination of the complex nature of microbial communities and activated sludge structure, including morphological, physical and chemical parameters, referred as affecting factors in solid-liquid separation processes (Li and Ganczarczyk, 1987; Andreadakis, 1993; Barbusinski and Koscielniak 1995; Grijspeerdt and Verstraete, 1997; Jin *et al.*, 2003; Li and Yang, 2007; Liao *et al.*, 2006; Urbain *et al.*, 1993; Novak *et al.*, 1993; Wilén *et al.*, 2003; Jin *et al.*, 2004).

Depending on the state of the sludge, some malfunctions are common (Jenkins *et al.,* 2003): formation of pinpoint flocs due to the absence of filamentous bacteria, leading to small flocs with poor settling ability properties; filamentous bulking due to filamentous bacteria surplus, leading to large interconnected flocs with poor settling ability properties; dispersed growth due to floc-forming bacteria not flocculating and leading to a non-settling turbid effluent; and zoogleal or viscous bulking due to a

surplus of exopolyssacharide production, leading to large flocs with poor settling and compaction properties and viscous effluents presenting high organic contents.

In recent years, microscopy observations are becoming important methods to monitor and control activated sludge systems and, as a result of that, these techniques are becoming widespread for the characterization of activated sludge microbial aggregates (Andreadakis, 1993; Barbusinski and Koscielniak, 1995). Furthermore, the association of image processing and analysis methodologies with microscopy visualization allows an accurate evaluation of the activated sludge status (Li and Ganczarczyk, 1991; Grijspeerdt and Verstraete, 1997). Indeed, several authors (da Motta *et al.*, 2002; Cenens *et al.*, 2002; Amaral and Ferreira, 2005; Jenné *et al.*, 2006; and Arelli *et al.*, (2009) have already proposed a range of image analysis procedures to characterize and relate operating parameters, such as SVI, with the biomass structure of activated sludge systems, mainly in terms of aggregated and filamentous biomass. However, up to the present, most studies relating sludge volume index (SVI) and sludge morphological properties have been focused in filamentous bulking conditions solely (da Motta *et al.*, 2002; Amaral and Ferreira, 2005). Moreover, the emphasis of such studies has been focused on the biomass contents and morphology, and little attention has been given to the Gram type and physiological status (viable or damaged) of such biomass.

However, when acquiring images from dyed samples, traditional grayscale imaging is not appropriate, and color imaging must be used. In fact, color imaging can provide more information than gray images and the use of color scales allows seeing small changes locally (Cheng *et al.*, 2001; Russ, 2002).

The traditional Gram stain is the most widely used stain method in microbial classification and is a fundamental technique in the examination of activated sludge samples (Jenkins *et al.*, 2003). Gram-positive organisms differ from Gram-negative

organisms depending on the existence of a permeability barrier in bacteria after treatment by ethanol. With particular attention to activated sludge systems, Gram staining can be problematic regarding large and dense flocs which may not decolorize correctly. Furthermore, the reaction to the Gram staining could be negative, strongly positive or weakly Gram-positive, which could be problematic for Gram status identification through image analysis procedures. SYTO® 9 and hexidium iodide (HI) are novel fluorescent nucleic acid binding dyes that allow the assessment of Gram status by differential absorption through bacterial cell walls and selectively staining Gram-positive organisms without fixative methods. The use of such fluorescent dyes may provide a robust, objective, and rapid alternative to traditional Gram staining in wastewater systems (Foster *et al.*, 2002).

The Live/Dead BacLight™ Bacterial Viability kit is a widely used method to measure viability which differentiates live from dead cells by detecting if their membrane system is intact, even in a mixed population containing a broad range of bacterial types (Invitrogen Molecular Probes, 2004). This kit is comprised by two nucleic acid probes, green-fluorescent SYTO® 9 and red-fluorescent Propidium Iodide (PI). These stains differ both in their spectral characteristics and in their ability to penetrate healthy bacterial cells.

The present study aims the identification and quantification of aggregated and filamentous biomass from a lab-scale activated sludge system using image processing and analysis procedures. The biomass composition on Gram-positive and Gram-negative bacteria, as well as viable and damaged bacteria, is also evaluated by image analysis coupled to fluorescent staining. Finally, the obtained image analysis data is used to identify activated sludge malfunctions (pinpoint flocs formation, filamentous bulking and zoogleal bulking), and further correlated with settling ability (SVI) and biomass contents (TSS) parameters.

7.2. MATERIALS AND METHODS

7.2.1 ACTIVATED SLUDGE SYSTEM DESCRIPTION

Experimental data was obtained from a lab-scale activated sludge system based on a 17 L aerated tank with suspended biomass, followed by a 2.5 L cylindrical clarifier. The aerated tank was inoculated with activated sludge from a wastewater treatment plant located in Braga (Portugal). The system was fed with a synthetic medium prepared with the following composition (mg/L): $NaCH_3COO \cdot 3H_2O$, 2073; $(NH_4)_2SO_4$, 140; $MgSO_4 \cdot 7H_2O$, 25; KH_2PO_4, 44; $K_2HPO_4 \cdot 2H_2O$, 59; $CaCl_2 \cdot 2H_2O$, 30; $FeCl_3 \cdot 6H_2O$, 18.8; $NaHCO_3$, 105. A micronutrients solution was also added to the system to guarantee the biomass maintenance. The pH of the system was controlled with a pH meter and a control pump (Model BL 7916-BL 7917, Hanna Instruments, Woonsocket, RI, USA) with 0.01 M HCl solution. Complete mixing inside the reactor was guaranteed by supplying a continuous inflow of air bubbles through an air diffuser placed at the bottom of the reactor. An oxygen probe (TriOmatic 690, WTW, Weilheim, Germany) was used to measure the amount of dissolved oxygen. The concentration of dissolved oxygen was maintained around 7 mg/L. Sludge recirculation from the settler to the reactor was guaranteed by an air pump. Figure 7.1 displays the schematic layout of the plant used for this study.

7.2.2 OFF-LINE PROCESS MONITORING

The study of the four above-mentioned conditions (filamentous bulking, pinpoint flocs, zoogleal bulking, and normal conditions) was sequentially conducted. Between each experiment, the system was recharged with biomass to guarantee a rapid establishment of the new condition, by changing the organic loading rate (OLR). During each experiment, total suspended solids (TSS) measurements were conducted in accordance with the procedures described in Standard Methods (APHA *et al.*,

1989). The biomass settling ability was measured through the determination of the SVI in a 1 L Imhoff cone, with the sludge height variation monitored for 30 min.

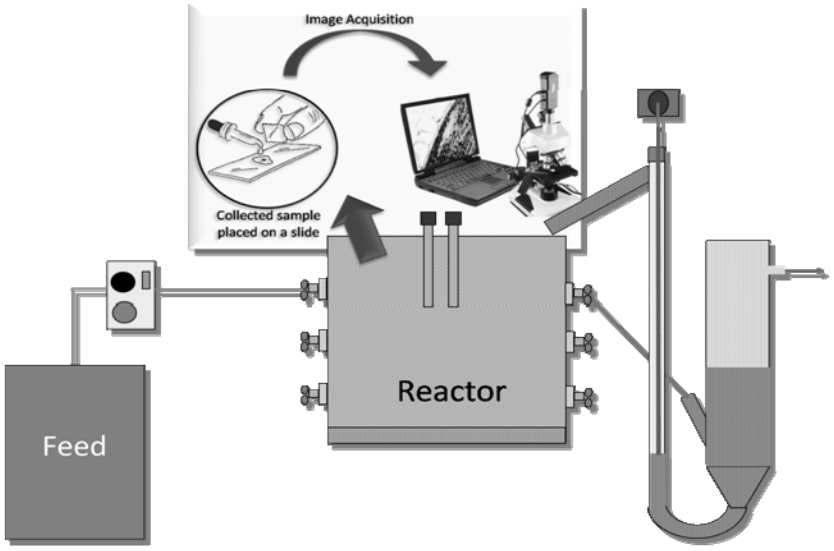

Figure 7.1. Schematic layout of the plant and image acquisition procedures.

7.2.3 SAMPLE PREPARATION FOR STAINING PROCEDURES

For both staining procedures, a sterile solution of 0.85% NaCl was prepared. 1.5 µL of each dye was put into 5 mL of the NaCl solution and the tube was wrapped with aluminum foil (staining solution). A volume of 100 µL of undiluted biomass suspension was mixed with 50 µL of staining solution and incubated in darkness for 15 min at room temperature. A preliminary experiment allowed concluding that this concentration was sufficient for staining the overall bacteria population. The bacteria population was then visualized through fluorescence microscopy.

7.2.4 LIVE/DEAD® STAINING

The Live/Dead® BacLight™ bacterial viability kit was used to differentiate viable and damaged bacteria (Invitrogen Molecular Probes, 2004). The kit utilizes a mixture of

SYTO® 9 green-fluorescent nucleic acid stain and a red-fluorescent nucleic acid stain, propidium iodide (PI). Viable bacteria are stained by SYTO® 9 and damaged bacteria are stained by PI.

7.2.5 LIVE BACLIGHT™ BACTERIAL GRAM STAINING

The Live Baclight™ bacterial Gram stain allows to easily classify bacteria as Gram-positive or Gram-negative without using fixatives (Invitrogen Molecular Probes, 2001). This kit utilizes a mixture of SYTO® 9 green-fluorescent nucleic acid stain and a red-fluorescent nucleic acid stain, hexidium iodide (HI). Gram-negative bacteria are stained by SYTO® 9 and Gram-positive bacteria are stained by HI.

7.2.6 BRIGHT FIELD IMAGE ACQUISITION

The microbial community was observed by means of an optical microscope Olympus BX51 (Olympus, Tokyo, Japan), at 100x magnification, coupled with a camera Olympus DP25 (Olympus, Tokyo, Japan). Images were acquired at 1360 × 1024 pixels and 8-bit format through the commercial software Cell^B (Olympus, Tokyo, Japan). Samples were taken from the aerated tank, and 3 slides per sample were used in a total of 150 images (3 × 50 images). 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 samples on the slides. For each slide, a volume of 10 µL was covered with a 20 mm × 20 mm cover slip, for visualization and image acquisition. Images were acquired in the upper, middle and bottom of the slide in order to improve the representativeness of the microbial community in the system.

7.2.7 FLUORESCENCE IMAGE ACQUISITION

Slides with stained sludge samples (10 µL on each slide) were observed in an epifluorescence microscope (Olympus BX51, Tokyo, Japan) at 200x magnification. Two filters were used, one in the green wavelength range with an excitation bandpass of 470-490 nm and emission at 516 nm (long pass filter), and the second filter in the red wavelength range with an excitation bandpass of 530-550 nm and emission at 591 nm (long pass filter). Images were acquired at 1360 × 1024 pixels, and 24-bit RGB format (8 bit red, 8 bit green and 8 bit blue channels) through the commercial software Cell^B (Olympus, Tokyo, Japan), and 2 slides per sample were used in a total of 100 images (2 × 50 images).

7.2.8 BRIGHT FIELD IMAGE PROCESSING

The image analysis program used for the recognition of the aggregated and filamentous biomass in the grayscale images was adapted from a previous version of Amaral (2003), developed in Matlab 7.3 (The Mathworks, Inc., Natick, USA). Figure 7.2 shows the original image, the binary image of the microbial aggregates and the binary image of the filamentous bacteria, after the processing step.

Figure 7.2. (a) Original Image, (b) Aggregates binary image, (c) Filaments binary image.

7.2.9 FLUORESCENCE IMAGE PROCESSING

A new program was developed in Matlab 7.3 (The Mathworks, Inc., Natick, USA) to recognize and characterize the aggregated and filamentous biomass in the fluorescent images, comprising the extraction of the green channel from the original RGB image collected at 516 nm and the red channel from the original RGB image collected at 591 nm, followed by background correction, image segmentation and aggregated and filamentous biomass recognition, for each of the above channels. The determination of the SYTO® 9-stained cells was performed through the use of the green channel of the RGB image collected at 516 nm), whereas the determination of the propidium iodide and the hexidium iodide-stained bacteria was performed through the use of the red channel of the RGB image collected at 591 nm. The developed program also allowed for the calculation of a fluorescence-based intensity image, for both green and red channels, directly correlated with the fluorescence of the aggregated and filamentous biomass. These images were subsequently used, regarding the filamentous biomass, to accurately identify the Gram and viability status of overlapping SYTO® 9 and propidium or hexidium iodide stained cells. Figure 7.3 presents the original image in the green channel, the intensity image and the aggregates and filaments recognition. A more detailed description of the image processing methods is presented below.

7.2.10 PRE-TREATMENT

The image pre-processing stage depends on the enhancement of each of the two color channels (green channel from the original RGB image collected at 516 nm and the red channel from the original RGB image collected at 591 nm) by background determination and removal. In this stage, a block process methodology was applied to determine the 200 x 200 pixels minimum, surrounding each pixel in each of the

two channel images, and used as background images to minimize background light differences.

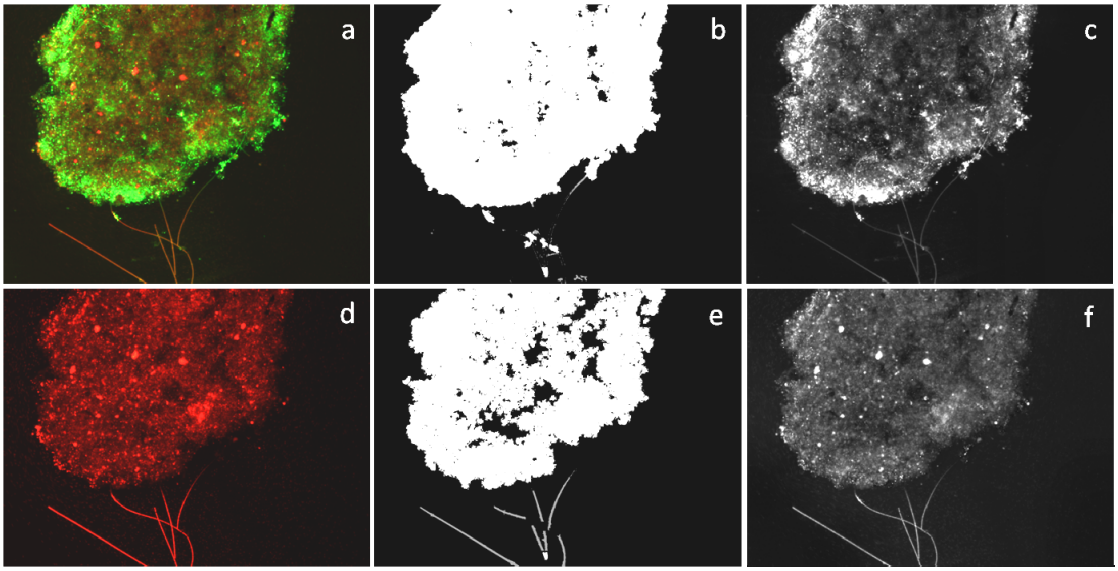

Figure 7.3. (a) Original image from the green filter, (b) area detection image, (c) intensity image, (d) original image from the red filter, (e) area detection image (f) intensity image.

7.2.11 SEGMENTATION AND DEBRIS ELIMINATION

This stage consisted primarily in segmentation of the aggregates and filaments in each of the two channel images. Different predefined threshold levels were used for both the filaments and aggregates, as well as for each channel image. In each channel image, first the aggregates were segmented by the respective aggregates threshold level, and then morphological opening (disk of radius 12) and reconstruction was applied for the elimination of small debris and attached filaments. Regarding the filaments identification, after an adaptative histogram equalization of the original two channel images, the determination of the filaments in each channel image was performed by the segmentation by the respective filaments threshold level. Finally, after a morphological opening (disk of radius 3) was

applied, a gyration radius based procedure was implemented to discard small filamentous-like debris by the use of a 1 cut-off value (Pons and Vivier, 1999).

7.2.12 SUPERIMPOSED FILAMENTS IDENTIFICATION

In order to prevent a single filament to be identified in both the red and green channel, and thus be characterized as both Gram negative and positive or viable and damaged at the same time, superimposed filaments in both images had to be recognized. First the superimposed filaments were recognized by a morphological procedure combining both filaments binary channel images. Next, the two channel images resulting from the background removal were compared and each channel intensity of the superimposed filaments was determined. The filament was then attributed to the channel in which it presented the highest intensity value and, thus, deleted from the other filament binary channel image.

7.2.13 DETERMINATION OF THE FLUORESCENCE-BASED INTENSITY IMAGES

The last step of the image analysis program consisted in determining the fluorescence-based intensity images of the aggregated and filamentous bacteria in both channels. For that purpose, the aggregates and filaments binary images, in each channel, were used as masks to respectively identify the aggregates and filaments intensity images.

7.2.14 MORPHOLOGICAL PARAMETERS DETERMINATION

The binarization of the bright field grayscale images allowed the determination of the aggregated and filamentous biomass morphology and contents. Aggregates were

classified according to size in (Eikelboom, 2000): small aggregates (D_{eq} < 25 µm); intermediate aggregates (25 µm < D_{eq} < 250 µm); large aggregates (D_{eq} > 250 µm). For each studied class, aggregates area percentage (Area %) was calculated, as defined by Amaral and Ferreira (2005). The aggregates total area per volume (TA/Vol), filaments total length per volume (TL/Vol), filaments total length per aggregates total area ratio (TL/TA), and filaments total length per total suspended solids ratio (TL/TSS), were also determined according to Mesquita *et al.* (2010). The ratio between Gram-negative and Gram-positive filamentous bacteria was determined as the filaments area ratio (G_AG/AR) from the colored images acquired in fluorescent microscopy of Live Baclight™ Gram stained samples. The ratio between viable and damaged filamentous bacteria was also determined based on the filaments area ratio (LD_AG/AR) from the colored images acquired in fluorescent microscopy of Live/Dead® BacLight™ bacterial viability stained samples.

7.3. RESULTS AND DISCUSSION

7.3.1 SLUDGE VOLUME INDEX (SVI) AND TOTAL SUSPENDED SOLIDS (TSS)

The filamentous bulking, pinpoint floc formation, zoogleal bulking and normal conditions experiments were performed sequentially, upon the introduction of fresh inoculum between each experiment. Two distinct filamentous conditions were studied, the first one identified as FB1, and the second as FB2. The first step of the data analysis aimed at identifying the period in which each different condition was established. For that purpose, alongside a microscopy inspection, an SVI analysis allowed the identification of bulking conditions, while a TSS inspection was used for the pinpoint floc phenomena recognition. The effectiveness of bioflocculation and settling of activated sludge is often characterized by SVI. Low SVI values indicate that the sludge is dense and thus the sludge has better settling ability. According to

Jenkins *et al.* (2003), an activated sludge with SVI lower than 120 mL/g is considered satisfactory and over 150 mL/g is considered bulking. The determination of bulking conditions was, therefore, based on this 150 mL/g threshold. Figure 3 depicts the SVI and TSS results for each experiment.

Viscous or zoogleal bulking is caused by an excessive amount of extracellular polysaccharides (EPS), and has been shown to have a negative effect on the biomass thickening and compaction due to the water-retentive nature of EPS, making the activated sludge flocs density closer to that of the surrounding water, thus increasing the sludge SVI (Jobbagy *et al.*, 2001; Jin *et al.*, 2003; Jenkins *et al.*, 2003; Novak *et al.*, 1993; Peng *et al.*, 2003). In extreme, severe instances of viscous bulking were shown to result in the absence of solids separation (Jenkins *et al.*, 2003).

Filamentous organisms are an essential part of the floc population in an activated sludge process, forming the backbone to which floc-forming bacteria adhere. However, filamentous bulking may take place when filamentous bacteria overgrow, leading to poor sludge settling ability and poor thickening characteristics of the sludge. Filamentous bulking is one of the most studied problems regarding activated sludge as proven by the literature (Eikelboom, 2000; Jenkins *et al.*, 2003; Martins *et al.*, 2004; Li *et al.*, 2008; Schuler and Jassby, 2007).

Regarding the establishment of bulking conditions, all samples from the first and second periods of filamentous bulking (FB1 and FB2), as well as zoogleal bulking (ZB), presented SVI values higher than 150 mL/g (with the exception of the last sample of the first filamentous bulking condition with 100 mL/g). It was also found that filamentous bulking phenomena originated SVI values higher (in average) than zoogleal bulking which is in accordance with the results obtained by Novak *et al.* (1993). Furthermore, a microscopy inspection allowed identifying a large amount of filamentous bacteria emerging from medium sized flocs in the filamentous bulking

experiments, in contrast with the lower filamentous bacteria contents emerging from larger flocs in the zoogleal bulking experiment (further corroborated by the image analysis results). It could also be seen, in the current work that the zoogleal bulking experiment presented a constant TSS increase related to the high EPS production. Thus, the comparison of the SVI and TSS data, alongside the microscopy inspection and image analysis results (referred below) allowed the distinction between filamentous and zoogleal bulking conditions.

Pinpoint flocs phenomena are characterized by the formation of small and mechanically fragile activated sludge flocs, presenting low settling properties, formed by floc-forming bacteria and lacking the filamentous bacteria backbone (Jenkins *et al.*, 2003). As the majority of these flocs do not settle, at all, the SVI parameter is not strongly affected by the formation of pinpoint flocs. However, in such instances, the clarifier's supernatant presents high solids concentration, thus leading to the biomass washout from the reactor, and a correspondent decrease on the reactor TSS. Analyzing the TSS and SVI results for the pinpoint experiment, it was clear that the SVI remained well below the bulking threshold value, and that the TSS contents steadily decreased throughout the monitoring period, as expected.

With respect to the normal conditions (NC) experiment, the obtained SVI (lower than 150 mL/g), and TSS values (between 3 to 4 g TSS/L) fell within the expected ranges. It should be noticed that, the sharp initial TSS decrease (from 6-7 to 3-4 g TSS/L) was due to the introduction of a highly concentrated inoculum parameter at the beginning of the experiment, and that afterwards the TSS contents remained approximately constant.

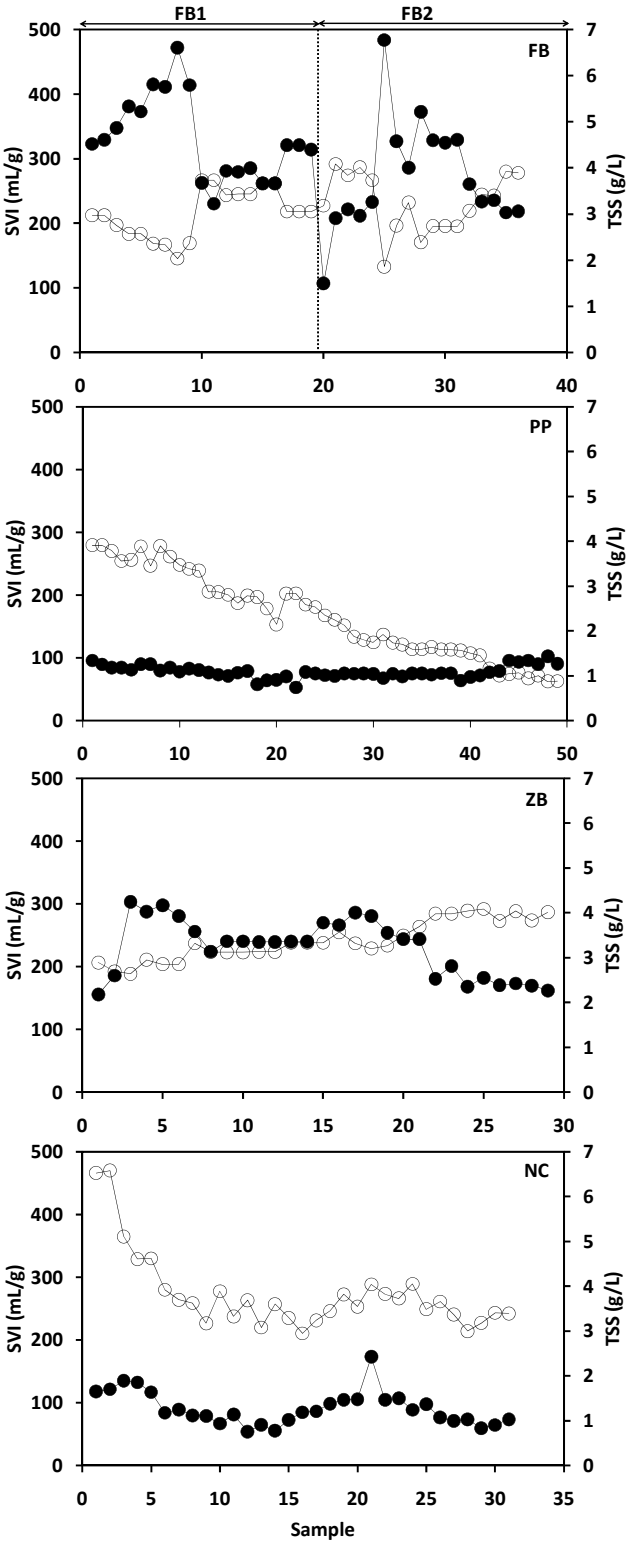

Figure 7.4. SVI (•) and TSS (o) behavior for all experiments. (FB1 – filamentous bulking 1; FB2 – filamentous bulking 2; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

7.3.2 MORPHOLOGICAL PARAMETERS

Image processing techniques are powerful tools in activated sludge systems for biomass inspection, characterization, and quantification in terms of aggregates and filaments. In the present work, the use of such techniques allowed the determination of the biomass structure in terms of the aggregated and filamentous bacteria composition. Furthermore, the size distribution of the aggregated biomass was also studied regarding the percentage of small (Deq < 25 µm), intermediate (25 µm < Deq < 250 µm), and large (Deq > 250 µm) aggregates, as depicted in Figure 7.5. During the filamentous bulking experiments (FB1 and FB2), small aggregates prevailed, either dispersed into the mixed liquor, either linked to filamentous bacteria. In contrast, the small aggregates originated by the pinpoint flocs (PP) experiment were almost exclusively dispersed in the mixed liquor. Furthermore, throughout this experiment the intermediate aggregates prevailed with respect to the small flocs, which may be explained by the washout of a considerable percentage of the smaller aggregated biomass. In fact, looking back at the TSS results (Figure 7.4), the strong washout phenomena that occurred during the pinpoint floc experiment, due to the low compaction and settling abilities of the sludge, could easily have led to the removal of the small aggregates from the reactor. During the zoogleal bulking (ZB) experiment, the percentage of large aggregates was higher, on average, than the residual percentages of the other experiments. This result was expected due to the overproduction of exopolymers, thus increasing the aggregates size. Finally, looking at the structure of the biomass during the normal conditions (NC) experiments, the results revealed a predominance of the intermediate aggregates regarding the small aggregates, as expected, and residual percentages of large aggregates.

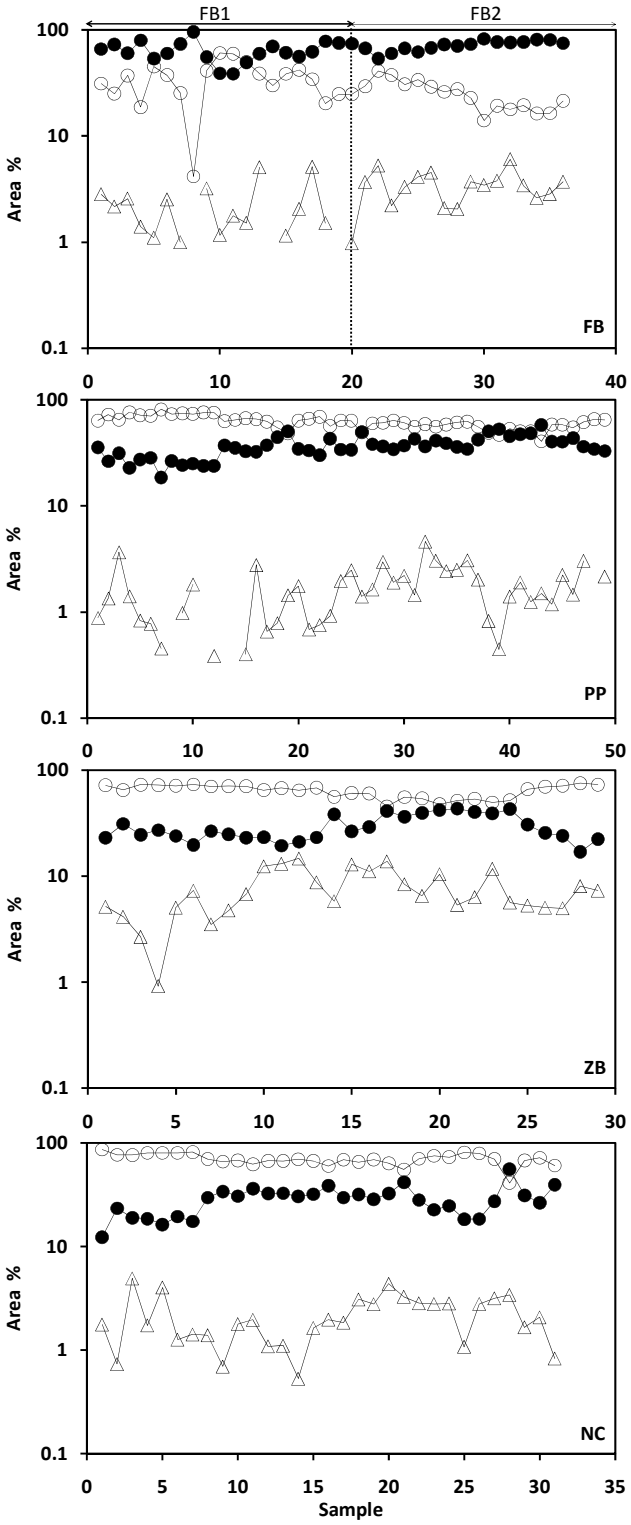

Figure 7.5. Area percentage behavior for each condition studied for small (•); intermediate (o); and large aggregates (Δ). (FB – filamentous bulking; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

The aggregated biomass (TA/Vol) and filamentous bacteria (TL/Vol) contents were also studied and are shown in Figure 7.6. The filamentous bulking experiments (FB1 and FB2) presented the highest values of the filamentous bacteria contents (TL/Vol), which could be expected given that, by definition, filamentous bulking phenomena occur upon the excessive growth of filamentous bacteria, both inside and extending from the flocs. It could also be seen that the two filamentous bulking events were different from each other, mainly in terms of the aggregated biomass contents, much higher in the second event (FB2). Regarding the zoogleal bulking (ZB) experiment the aggregated biomass (TA/Vol) contents increased throughout the experiment, in agreement with the TSS increase previously discussed. Furthermore, the filamentous bacteria contents (TL/Vol) presented moderate values, although decreasing during the experiment, which could be expected given that the conditions associated with zoogleal bulking favors the growth of aggregated rather than filamentous bacteria. During the pinpoint flocs formation (PP) period, the aggregated biomass contents (TA/Vol) decreased, as expected given the decrease of the TSS, configuring biomass washout from the reactor. Furthermore, the filamentous bacteria contents (TL/Vol) remained low, throughout this experiment. Analyzing the biomass structure during the normal condition experiment, low filaments contents (TL/Vol) were obtained throughout the experiment, whereas the aggregated biomass contents (TA/Vol) fluctuated between moderate values, in accordance with the TSS behavior.

The ratios between filamentous and aggregated bacteria (ln(TL/TA)) and between filamentous bacteria and suspended solids (TL/TSS) were also studied and are presented in Figure 7.7. Analyzing Figure 7.7, it seems clear the close relationship between ln(TL/TA) and TL/TSS, presenting a high correlation factor (R^2) of 0.9372 (Figure 7.8). The close relationship between these two parameters led to the analysis of the possibility of assessing the TL/TSS by the ln(TL/TA), with the advantage, for activated sludge systems inspection, that the ln(TL/TA) is independent of the suspended solids determination.

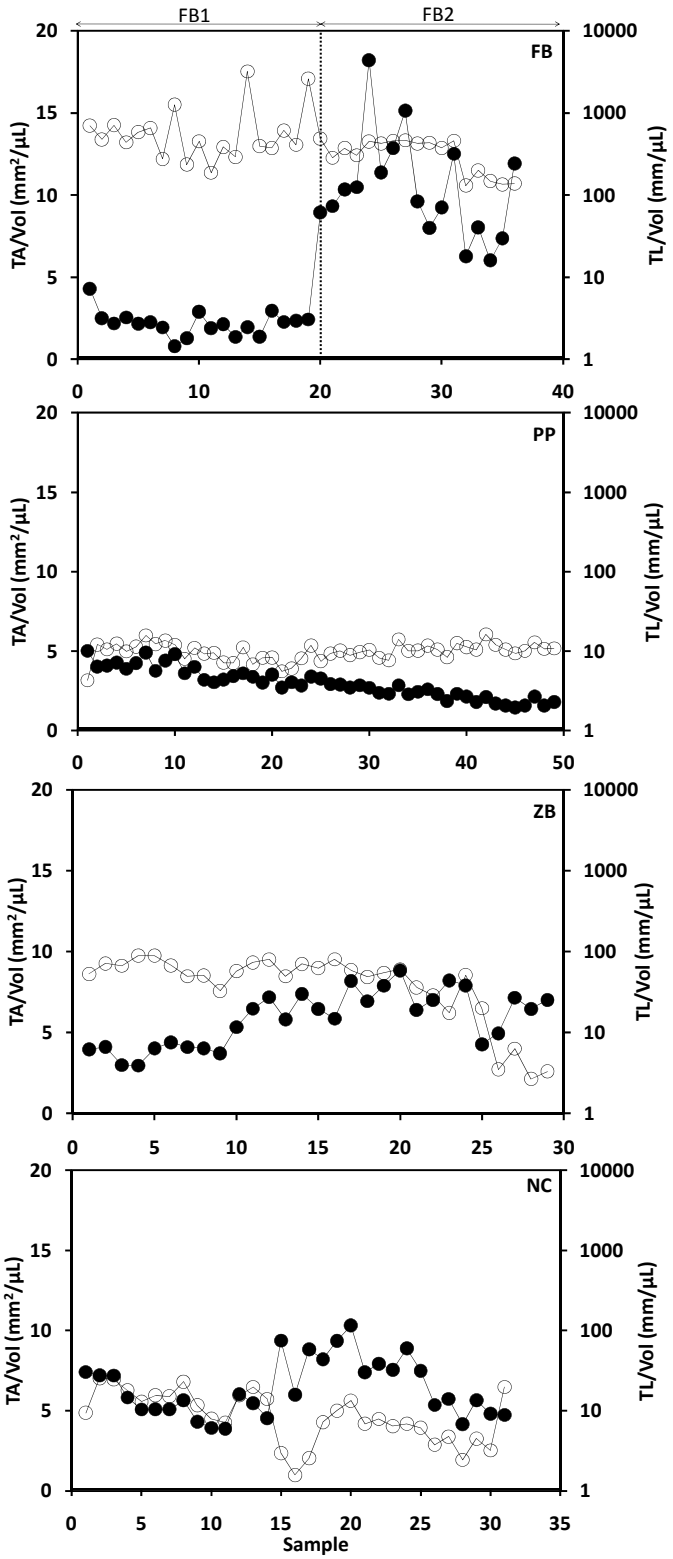

Figure 7.6. Experimental behavior of TA/Vol (•) and TL/Vol (o). (FB – filamentous bulking; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

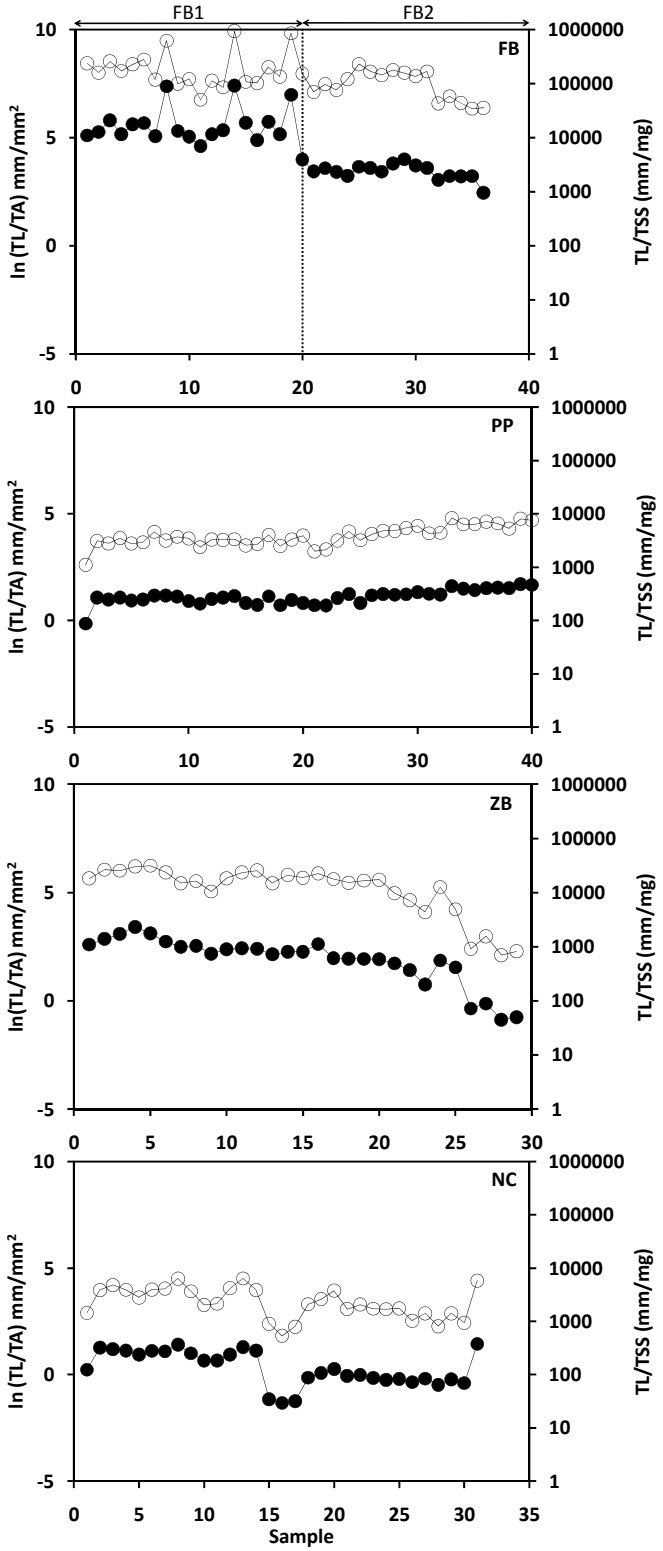

Figure 7.7. Experimental behavior of TL/TSS (o) and ln(TL/TA) (•). (FB – filamentous bulking; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

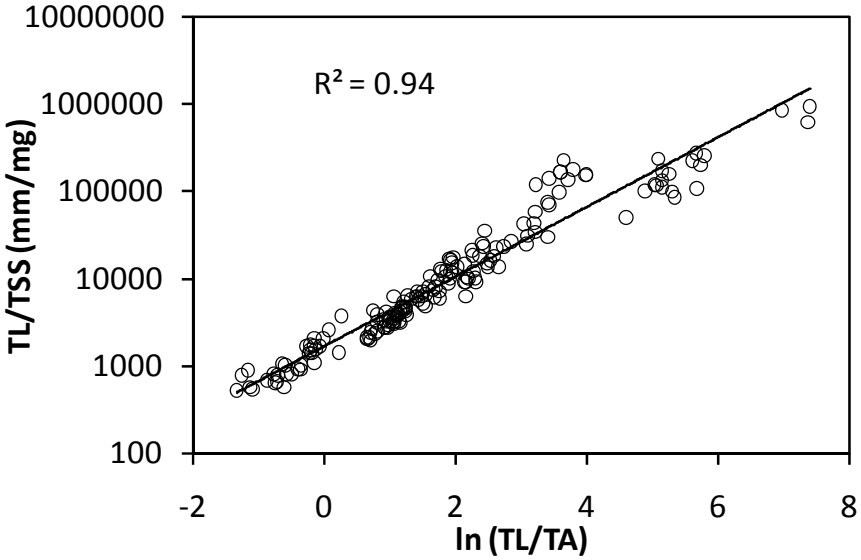

Figure 7.8. Relationship between TL/TSS and ln(TL/TA).

The filamentous bulking experiments (FB1 and FB2) presented the highest values for the ratios for the filamentous bacteria contents versus aggregated biomass (ln(TL/TA)) and total suspended solids (TL/TSS). This fact strengthens the identification of these two periods as filamentous bulking phenomena based on the high percentages of filamentous bacteria regarding the overall biomass. In distinction to the filamentous bulking experiments, the ln(TL/TA) and TL/TSS presented lower values for the zoogleal bulking experiment and even decreased slightly throughout this period. These results were expected, since the conditions associated to zoogleal bulking favors the growth of aggregated biomass contrasting to filamentous bacteria. Regarding the pinpoint flocs experiment, the biomass washout from the reactor led to a continuous growth of the ln(TL/TA) and TL/TSS parameters (although within the range of moderate values) due to the decrease of suspended solids. With respect to the normal conditions experiment, and in accordance to the observed low filaments contents, the filamentous bacteria ratios to the aggregated biomass and total suspended solids was kept low throughout the experiment as demonstrated by the ln(TL/TA) and TL/TSS values.

7.3.3 SVI ASSESSMENT

The image analysis dataset was further used to predict the sludge volumetric index (SVI). In order to do so, SVI values were plotted against the filamentous bacteria contents (TL/Vol) and ratios to the aggregated biomass (ln (TL/TA)) and total suspended solids (TL/TSS). The best results were obtained for the TL/Vol and TL/TSS correlations, and are presented in Figure 7.9. It was found a relationship between the SVI and the filamentous bacteria contents (correlation factor R^2 of 0.83), above a TL/Vol threshold value of 7 mm/µL. Below this value the amount of filamentous bacteria seems not to present significant relevance in SVI values. Furthermore, this correlation is influenced the most by the filamentous and zoogleal bulking conditions, given the fact that their values extend throughout the majority of the Y data range. The same analysis was performed for the correlation between SVI and TL/TSS, which was widely used in previous studies (Amaral *et al.*, 2003, Amaral and Ferreira, 2005, Lee *et al.* 1983; Matsui and Yamamoto, 1984). A direct dependence (correlation factor R^2 of 0.80) for TL/TSS above 2000 mm/mg was achieved, and below this threshold, TL/TSS values did not seemed to influence the SVI. Again, filamentous and zoogleal bulking conditions, extending throughout the majority of the Y data range, influenced the most the SVI values. Although, past studies have proven the validity of the TL/TSS parameter to assess SVI values in filamentous bulking, the present results, encompassing a wider range of conditions, seem to point to the TL/Vol parameter as the most adequate overall. Regarding the adequateness of the ln(TL/TA) parameter to capably substitute the TL/TSS parameter, this could not be fully confirmed given the obtained lower correlation value R^2 of 0.74.

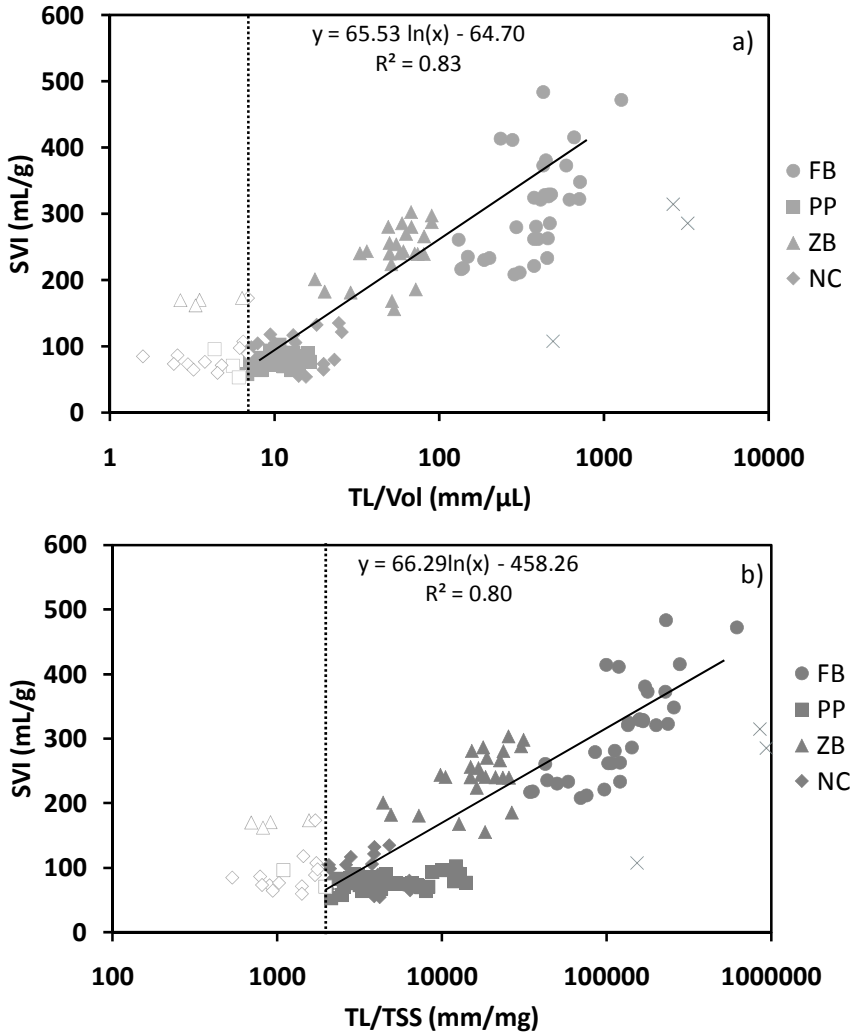

Figure 7.9. Correlation between SVI and (a) TL/Vol, (b) TL/TSS. Discarded outliers are represented by cross. (FB – filamentous bulking; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

7.3.4 TSS ASSESSMENT

In the course of this work it was also attempted to elucidate the dependence of total suspended solids (TSS) with the TA/Vol parameter, as shown in Figure 7.10. Given the high percentage of filamentous bacteria contributing to TSS and not accounted for the TA/Vol (representing solely the aggregates area), all filamentous bulking samples

were discarded from this analysis. Therefore, for the pinpoint flocs, zoogleal bulking and normal conditions TSS showed a directly dependence (correlation factor R^2 of 0.83) with the aggregated biomass content, but only up to a threshold value of 5 $mm^2/\mu L$. Thus, in filamentous bulking conditions or in the case of high aggregated biomass contents it seems not possible to predict TSS contents from the TA/Vol parameter. Therefore, only for pinpoint flocs, normal conditions and zoogleal bulking phenomena with low TA/Vol contents, this parameter may be used for the TSS assessment.

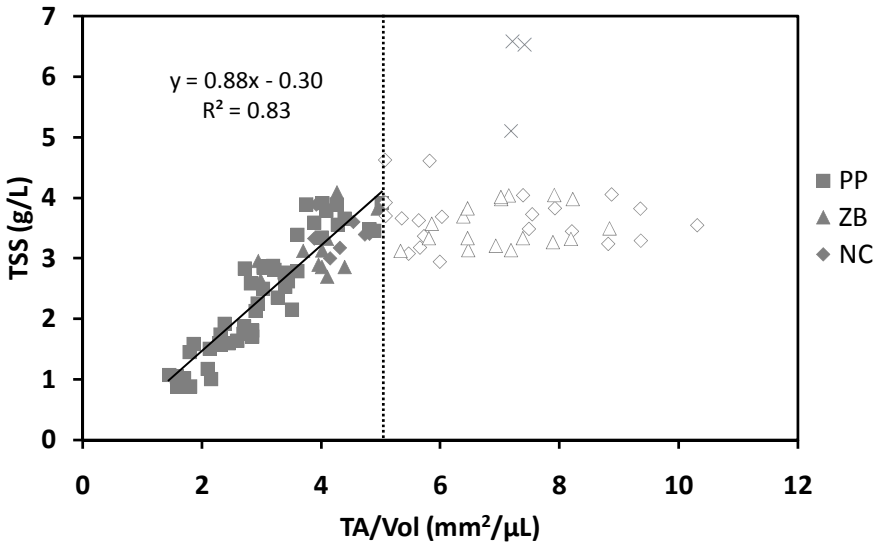

Figure 7.10. Correlation between TSS and TA/Vol. Discarded outliers are represented by cross. (PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

7.3.5 PHYSIOLOGICAL STATUS

Results from fluorescence microscopy are depicted in Figure 7.11, presenting the ratios between filamentous Gram-negative and positive bacteria (G_AG/AR) and between filamentous viable and damaged bacteria (LD_AG/AR). A slight difference between the filamentous bulking experiment 1 and 2 was detected. Although in both cases Gram-negative bacteria predominated, during FB1 their number varied from 10

to 100 fold the Gram-positive bacteria, whereas in FB2 their number was smaller, varying from 1 to 10 fold. Furthermore, regarding the filamentous bacteria viability, and although the vast majority of cells were viable, FB1 and FB2 also differed considerably. The two periods presented opposite tendencies with increasing filamentous bacteria viability throughout FB1 and decreasing filamentous bacteria viability throughout FB2. Regarding the zoogleal bulking experiment, a similar trend to the second filamentous bulking condition was observed with the predominance of Gram-negative bacteria, varying from 1 to 100 fold and decreasing filamentous bacteria viability throughout the experiment, although maintaining a large majority of viable cells. In the pinpoint experiment, a shift on the filamentous bacteria community was observed from initially predominant Gram-negative bacteria towards a majority of Gram-positive bacteria at the end of the experiment. Regarding the filamentous bacteria viability no clear trend was observed throughout the experiment, maintaining a large majority of viable cells. The normal condition experiment started with the predominance of filamentous Gram-negative bacteria that progressively shifted towards the equilibrium with the Gram-positive bacteria at the end of the experiment. Regarding the filamentous bacteria viability a decreasing trend in the filamentous bacteria viability was noticed throughout the experiment, although maintaining a large majority of viable cells.

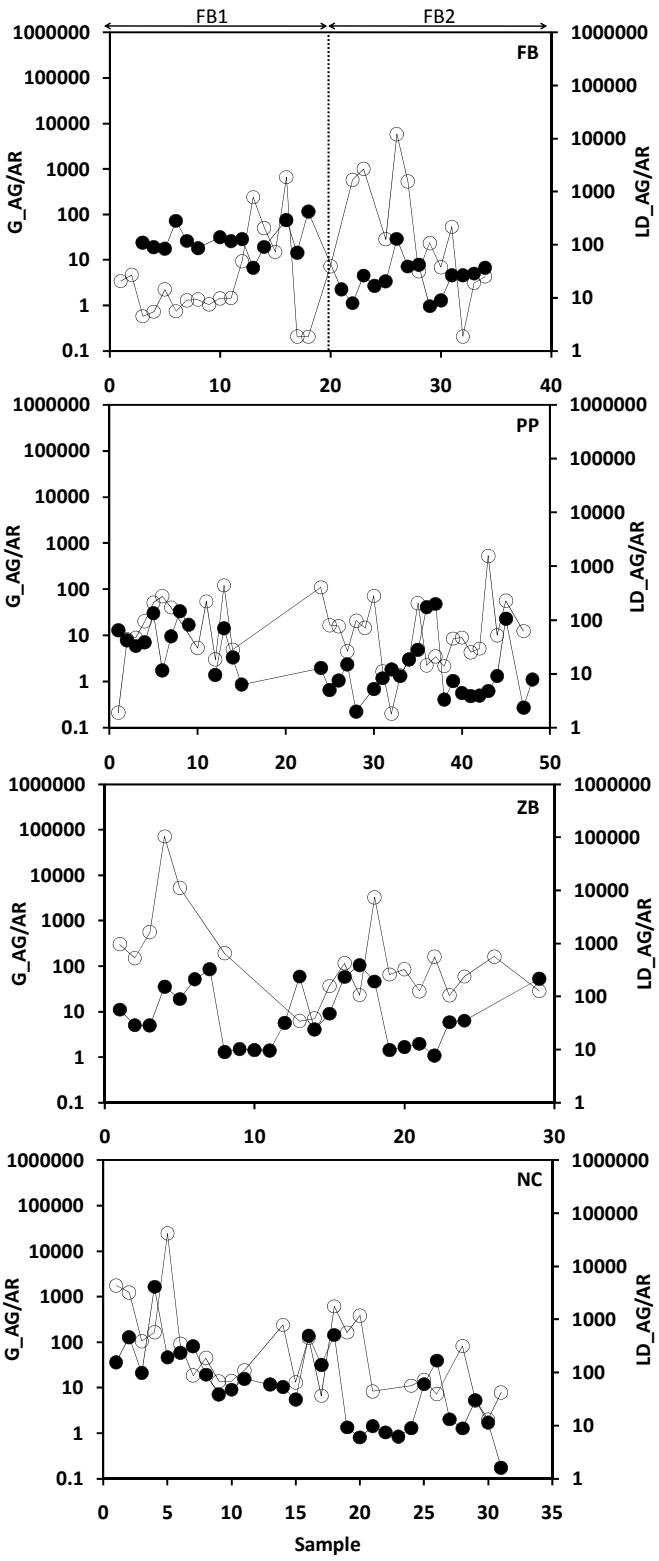

Figure 7.11. Experimental behavior of G_AG/AR (•) and LD_AG/AR (o). (FB – filamentous bulking; PP – pinpoint floc; ZB – zoogleal bulking; NC – normal conditions).

7.4. CONCLUSIONS

Image analysis parameters were used to elucidate several disturbances that may occur in an activated sludge system, and the possibility of measuring two of the most important operational parameters (SVI and TSS) for biomass characterization. During both filamentous bulking experiments the majority of the aggregated biomass formed small aggregates, either dispersed in the mixed liquor, either attached to filamentous bacteria. Furthermore, filamentous bulking presented the largest filamentous bacteria contents, as well as their ratio versus aggregated bacteria and suspended solids. Differences between the filamentous bacteria composition and viability from filamentous bulking experiment 1 and 2 were also detected. Although in both cases viable Gram-negative bacteria predominated, FB1 presented a larger Gram-negative fraction with increasing viability throughout the experiment, opposite to the decreasing trend in FB2. The zoogleal bulking experiment presented the largest contents of large aggregates alongside the increase of both the aggregated biomass and TSS. The filamentous bacteria contents as well as their ratio versus aggregated bacteria and suspended solids presented a decreasing trend throughout this period, since the conditions associated to zoogleal bulking favors the growth of aggregated biomass contrasting to filamentous bacteria. Furthermore it was detected a predominance of Gram-negative bacteria, and decreasing filamentous bacteria viability throughout this experiment, although maintaining a large majority of viable cells. Regarding the pinpoint flocs experiment, it was possible to observe a biomass washout phenomenon leading to the decrease of biomass contents mirrored by the TSS and aggregated biomass behaviors, which, in turn led to a continuous growth of the ln(TL/TA) and TL/TSS parameters (although within the range of moderate values). During this experiment, a large majority of viable cells was observed although the filamentous bacteria community shifted from initially predominant Gram-negative bacteria towards a majority of Gram-positive bacteria at

the end of the experiment. The biomass structure under normal operating conditions showed a predominance of intermediate flocs, low filamentous bacteria contents and respective ratios versus aggregated bacteria and suspended solids. Furthermore, it was noticed a shift from an initial clear predominance of filamentous Gram-negative bacteria towards the equilibrium with the Gram-positive bacteria at the end of the experiment. Although slightly decreasing throughout the experiment, a large majority of viable cells was maintained.

Regarding the SVI assessment, a relationship (correlation factor R^2 of 0.83) between the SVI and the filamentous bacteria contents was found, for TL/Vol values above 7 mm/µL. Although, past studies have proven the validity of the TL/TSS parameter to assess SVI values in filamentous bulking, the present results, encompassing a wider range of conditions, seem to point to the TL/Vol parameter as the most adequate overall. In the current work, the assessment of TSS by image analysis was limited to pinpoint flocs, normal conditions and zoogleal bulking phenomena with low TA/Vol contents (5 $mm^2/\mu L$), thus excluding filamentous bulking and high aggregated biomass contents systems. In conclusion it may be inferred that the image analysis methodology clearly revealed the nature of each studied condition during this work.

7.5. REFERENCES

Amaral, A.L. (2003) *Image Analysis in Biotechnological Processes: Applications to Wastewater Treatment.* PhD Thesis, University of Minho, Braga, Portugal (http://hdl.handle.net/1822/4506)

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

Andreadakis, A. (1993) Physical and chemical properties of activated sludge flocs. *Water Research* **12**, 1707-1714.

APHA, AWWA, and WPCF (1989) *Standard methods for the examination of water and wastewater*, 17th Edition, American Public Helth Association, Washington D.C.

Arelli, A., Luccarini, L., and Madoni, P. (2009) Application of image analysis in activated sludge to evaluate correlations between settleability and features of flocs and filamentous species. *Water Science and Technology* **59** (10), 2029-2036.

Banadda, E.N., Smets, I.Y., Jenné, R., and Van Impe, J.F. (2005) Predicting the onset of filamentous bulking in biological wastewater treatment systems by exploiting image analysis information. *Bioprocess and Biosystems Engineering* **27**, 339-348.

Barbusinski, K., and Koscielniak, H. (1995) Influence of substrate loading intensity on floc size in activated sludge process. *Water Research* **29** (7), 1703-1710.

Cenens, C., Van Beurden, K.P., Jenné, R., and Van Impe, J.F. (2002) On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology* **46** (1-2), 381-387.

Cheng, H.D., Jiang, X.H., Sun, Y., and Wang, J. (2001) Color image segmentation: advances and prospects. *Pattern Recognition* **34**, 2259-2281.

da Motta, M., Pons, M.N., and Roche, N. (2002) Study of filamentous bacteria by image analysis and relation with settleability. *Water Science and Technology* **46** (1-2), 363-369.

Eikelboom, D.H. (2000) *Process control of activated sludge plants by microscopic investigation*. First edition, IWA Publishing, London.

Foster, S., Snape, J.R., Lappin-Scott, H.M., and Porter, J. (2002) Simultaneous Fluorescent Gram Staining and Activity Assessment of Activated Sludge Bacteria. *Applied and Enviromental Microbiology* **68** (10), 4772-4779.

Grijspeerdt, K. and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31** (5), 1126-1134.

Invitrogen Molecular Probes (2001) *LIVE BacLight™ Bacterial Gram Kits*. Manuals and Product Inserts. http://probes.invitrogen.com/media/pis/mp07008.pdf.

Invitrogen Molecular Probes (2004) *LIVE/DEAD BacLight™ Bacterial Viability Kits*. Manuals and Product Inserts. http://probes.invitrogen.com/media/pis/mp07007.pdf.

Jenkins, D., Richard, M.G., and Daigger, G. (2003) *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. Lewis publishing, Boca Raton, FL.

Jenné, R., Banadda, E.N., Gins, G., Deurinck, J., Smets, I.Y., Geeraerd, A.H., and Van Impe, J.F. (2006) Use of image analysis for sludge characterization: studying the relation between floc shape and sludge settleability. *Water Science and Technology* **54** (1), 167-174.

Jin, B., Wilén, B.M., and Lant, P. (2003) A comprehensive insight into floc characteristics and their impact on compressibility and settleability of activated sludge. *Chemical Engineering Journal* **95**, 221-234.

Jin, B., Wilén, B.M., and Lant, P. (2004) Impacts of morphological, physical and chemical properties of sludge flocs on dewaterability of activated sludge. *Chemical Engineering Journal* **98**, 115-126.

Jobbagy, A., Literathy, B., and Tardy, G. (2002) Implementation of glycogen accumulating bacteria in treating nutrient-deficient wastewater. *Water Science and Technology* **46** (1-2), 185-190.

Lee, S.E., Koopman, B., Bode, H., and Jenkins, D. (1983) Evaluation of alternative sludge settleability indexes. *Water Research* **17** (10), 1421-1426.

Li, D. and Ganczarczyk, J. (1987) Stroboscopic determination of settling velocity, size and porosity of activated sludge flocs. *Water Research* **21**, 257-262.

Li, D. And Ganczarczyk, J. (1991) Size distribution of activated sludge flocs. *Journal of Water Pollution Control Federation* **63**, 806-814.

Li, X.Y., and Yang, S.F. (2007) Influence of loosely bound extracellular polymeric substances (EPS) on the flocculation, sedimentation and dewaterability of activated sludge. *Water Research* **41** (5), 1022-1030.

Liao, B.Q., Droppo, I.G., Leppard, G.G., and Liss, S.N. (2006) Effect of solids retention time on structure and characteristics of sludge flocs in sequencing batch reactors. *Water Research* **40** (13), 2583-2591.

Martins, A.M.P., Pagilla, K., Heijnen, J.J., and van Loosdrecht, M.C.M. (2004) Filamentous bulking sludge - a critical review. *Water Research* **38** (4), 793-817.

Matsui, S. and Yamamoto, R. (1984) The use of color TV technique for measuring filament length and investigating sludge bulking causes. *Water Science Technology* **16** (10-1), 69-81.

Mesquita, D.P., Dias, O., Dias, A.M.A., Amaral, A.L., and Ferreira, E.C. (2009a) Correlation between sludge settleability and image analysis information using partial least squares. Analytica Chimica Acta **642** (1-2), 94-101.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2009b) Monitoring of activated sludge settling ability through image analysis: validation on full-scale wastewater treatment plants. *Bioprocess and Biosystems Engineering* **32** (3), 361-367.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2010) A comparison between bright field and phase contrast image analysis techniques in activated sludge morphological characterization. *Microscopy and Microanalysis* **16** (2), 166-174.

Novak, L., Larrea, L., Wanner, J., and Garcia-Heras, J.L. (1993) Non-filamentous activated sludge bulking in a laboratory scale system. *Water Research* **27** (8), 1339-1346.

Peng, Y., Gao, C., Wang, S., Ozaki, M., and Takigawa, A. (2003) Non-filamentous sludge bulking caused by a deficiency of nitrogen in industrial wastewater treatment. *Water Science and Technology* **47** (11), 289-295.

Pons, M.N., and Vivier, H. (1999) Biomass quantification by image analysis. *Advances in Biochemical Engineering/Biotechnology* **66**, 133-184.

Russ, J.C. (2002) *The Image Processing Handbook*. Fourth edition, CRC Press, Boca Raton.

Schuler, A.J., and Jassby, D. (2007) Filament content threshold for activated sludge bulking: artifact or reality? *Water Research* **41** (19), 4349-4356.

Urbain, V., Block, J.C., and Manem, D.J. (1993) Biofloculation in activated sludge: an analytical approach. *Water Research* **27** (5), 829-838.

Wilén, B.M., Jin, B., and Lant, P. (2003) The influence of key chemical constituents in activated sludge on surface and flocculating properties. *Water Research* **37**, 2127-2139.

Abstract

This chapter focuses on the use of chemometric techniques for identifying activated sludge abnormalities and parameters prediction. Combining chemometric methods with image analysis can improve monitoring activated sludge systems and minimize the need for analytical measurements. For that purpose data was colected from the aggregated and filamentous biomass, biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria, and operational parameters. Principal component analysis was subsequently applied to identify activated sludge abnormalities and partial least squares models were developed for key parameters prediction, namely sludge volume index, chemical oxygen demand, nitrate concentration, and total suspended solids.

8. CHAPTER 8 – APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND PARTIAL LEAST SQUARES

8.1. INTRODUCTION

Activated sludge systems are very sensitive to unexpected changes in operating conditions, mainly in terms of organic load, nutrients and oxygen contents. The modification of the operating conditions mostly affects the solid-liquid separation, a stage that is often characterized by the sludge volume index (SVI). Problems related to the sludge settling ability were previously reported encompassing pinpoint flocs formation, filamentous bulking, dispersed growth, and zoogleal or viscous bulking as the most common phenomena in activated sludge systems. Based on microscopy inspection the SVI, regulating the solid-liquid separation, has been shown to be related to the floc structure in activated sludge systems (Palm *et al.*, 1980; Dagot *et al.*, 2001), especially on filamentous bulking surveys (Lee *et al.*, 1983; Jenné *et al.*, 2004, 2007; Amaral and Ferreira, 2005; da Motta *et al.*, 2001a,b).

Image processing and analysis methodologies have been increasingly used for activated sludge characterization. Sezgin (1982) established that the SVI is strongly influenced by floc size and filamentous bacteria contents. Other studies such as Matsui and Yamamoto (1984), Ganczarczyk (1994) and Grijspeerdt and Verstraete (1997) used automated image analysis to relate the microorganism's morphology in biological systems with the sludge settling properties. Recently, image analysis has been the basis of the assessment of biomass morphology changes using different types of microscopy techniques (da Motta *et al.*, 2001a; Amaral and Ferreira, 2005; Jenné *et al.*, 2007; Mesquita *et al.*, 2010a). The combination between settling properties and image analysis data, such as the aggregated and filamentous bacteria contents and morphology, Gram-positive/Gram-negative and viable/damaged bacteria ratios, may offer powerful information, as well as a mechanism to decide if a given intervention should be operated in the system.

Monitoring an activated sludge system, may represent the determination of huge amounts of data, regarding the system biotic and abiotic characterization. To make sense of such complexity chemometric techniques are, quite often, indispensable. In fact, due to the high amount of information provided by image analysis techniques, chemometric techniques are quite valuable to organize such information in order to better characterize the activated sludge system. As matter of fact, principal component analysis (PCA) and partial least squares (PLS) have proven to be important tools for the identification of operating conditions and the prediction of certain parameters (Singh *et al.*, 2005, 2010; Lee *et al.*, 2006; Sousa *et al.*, 2007; Woo *et al.*, 2009).

Principal components analysis (PCA) is a tool used for data compression and information extraction. PCA finds combinations of variables describing major trends in the data (Wise *et al.,* 2004). The aim of PCA is to project the high dimensional space into a more visual low dimensional one, and by doing this find the key variables of the system. Thus, there will be a number of new principal components, which describe most of the variance in the process in a space of fewer dimensions than the original (Rosen and Olsson, 1998). Recently, PCA techniques have been successfully applied for monitoring a wide range of wastewater treatment systems (Dias *et al.*, 2008; Lourenço *et al.*, 2006, 2010; Costa *et al.*, 2009).

Partial least squares (PLS) can be seen as a supervised model to assess and monitor any given predicted variable, thus detecting deviations of a certain interest (Rosen, 2001). PLS models were recently used for the estimation of chemical oxygen demand (COD), total organic carbon (TOC), nitrate concentration ($N\text{-}NO_3^-$), total suspended solids (TSS) using UV-Vis and near-infrared spectroscopy on a wastewater treatment system (Sarraguça *et al.*, 2009; Lourenço *et al.*, 2008, 2010).

Among the operational parameters most commonly used for assessing the quality of wastewaters, the chemical oxygen demand (COD), as an indirect measure of the

organic matter concentration, is of the utmost importance. Ammonium, nitrite and nitrate concentrations are also widely assessed, thus reflecting the extent of nitrification/denitrification processes. Total suspended solids (TSS) are a measure of the non-diluted organic and inorganic matter present in a wastewater. Another parameter widely used in activated sludge systems is the sludge volume index (SVI) reflecting the sludge settling ability. However, the traditional methodologies of most of the above parameters are costly, labor and time-consuming, and a few may even present environmental risks due to formed residues. Therefore, a methodology that could predict these parameters, in a safer, time and labor effective manner is of great interest.

In the present work, a large amount of data, reflecting the operation of an activated sludge system in four different experimental conditions (pinpoint flocs formation, filamentous bulking, zoogleal bulking and normal conditions), was treated using chemometric methods. PCA and PLS multivariate statistical analysis were applied integrating image analysis parameters from aggregated and filamentous biomass characterization, biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria, alongside operational parameters. The potential of PCA was examined for the identification of each of the lab-scale activated sludge condition, and PLS was then used to predict sludge settling ability (SVI) and other parameters related to the biological process (TSS, COD and nitrate concentrations) for each condition.

8.2. MATERIALS AND METHODS

8.2.1 ACTIVATED SLUDGE SYSTEM DESCRIPTION

The same lab-scale activated sludge system described in Chapter 7, section 7.2.1, was used in this work.

8.2.2 OFF-LINE PROCESS MONITORING

The system was regularly monitored for parameters as total suspended solids (TSS), sludge volume index (SVI), chemical oxygen demand (COD), ammonium ($N\text{-}NH_4^+$), nitrite ($N\text{-}NO_2^-$) and nitrate ($N\text{-}NO_3^-$) concentrations. The TSS measurements were conducted in accordance with the procedures described in Standard Methods (APHA *et al.*, 1989). The SVI was determined in a 1 L Imhoff cone, with the sludge height variation monitored for 30 min. Samples for COD, ammonium, nitrite and nitrates quantification were collected and filtered, from the feed and the reactor. COD was measured with Hach Lange COD cell tests (LCK 414 and LCK 514) on a spectrophotometer (Hach Lange DR 5000). Nitrite was determined with Griess-Hosvay method, similar to the colorimetric method from Standard Methods (APHA *et al.*, 1989). Ammonium was determined according to Nessler's method (APHA *et al.*, 1989). Nitrate was determined by HPLC (Jasco, Tokyo, Japan) with automatic injection, UV detector (210 nm), and a Varian (Palo Alto, CA, USA) Metacarb 67H column operating at room temperature of 60°C. The eluent was a solution of sulfuric acid (0.005 mol/L) with a flow rate of 0.7 mL/min and a pressure of between 70-80 kg/cm^2.

8.2.3 IMAGE ANALYSIS INFORMATION

The aggregated and filamentous biomass contents and morphology were assessed by images acquired through bright field microscopy, and biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria by fluorescence microscopy. The methodology for image analysis parameters acquisition is presented in Chapter 7, from section 7.2.3 to 7.2.9. Morphological parameters determination can be found in Chapter 6, section 6.2.3. However, more parameters were used in this study, further presented.

The Width (W) of an object is given as the minimum Feret Diameter (F_{min}) converted to metric units, where the Feret Diameter of an object is the maximum distance between two parallel tangents touching opposite borders of the object (Glasbey and Horgan, 1995).

The Feret Factor (FF) is defined as the ratio between the maximum Feret Diameter F_{Max} and the Feret Diameter at 90° F_{Max90} of F_{Max} (Noesis, 1998).

Compactness (Comp) is determined by the following equation (Russ, 2002):

$$\mathrm{Comp} = \frac{\sqrt{\frac{4}{\pi}\mathrm{A}}}{\mathrm{F_{max}}} \tag{1}$$

The Extent (Ext) is defined as the ratio between the object area and its bounding box area.

$$\mathrm{Ext} = \frac{\mathrm{A}}{\mathrm{W_{BB}} \times \mathrm{L_{BB}}} \tag{2}$$

where W_{BB} is the Bounding Box Width and L_{BB} the Bounding Box Length respectively.

The Robustness (Rob) is given by the following equation (Pons *et al.*, 1997):

$$\mathrm{Rob} = \frac{2\mathrm{er_{obj}}}{\sqrt{\mathrm{A}}} \tag{3}$$

where er_{obj} is the number of erosions needed to delete an object.

The Concavity index (LrgC) is given by the following equation (Pons *et al.*, 1997):

$$\mathrm{LrgC} = \frac{2\mathrm{er_{comp}}}{\sqrt{\mathrm{A}}} \tag{4}$$

where er_{comp} is the number of erosions needed to delete the complement of an object (in relation to its Convex Envelope).

The Area ratio (RelArea) is given as the ratio of the area of the holes in an aggregate and the area of the aggregate.

The information for the physiological parameters was obtained from the green channel (of the RGB image collected at 516 nm) and the red channel (of the RGB image collected at 591 nm). The green channel is associated to gram-negative and viable bacteria, and the red channel is associated to gram-positive and damaged bacteria. Gram negative/positive and viable/damaged aggregates and filaments were characterized both in terms of projected area and in mean intensity value. The ratios between gram positive/negative and viable/damaged bacteria, both in terms of projected area and mean intensity value, were next determined.

Table 8.1 presents the overall dataset of parameters used in this work.

Table 8.1. Operational parameters, morphological parameters, and fluorescent parameters included in the dataset.

Variable	Name
Operational parameters	
COD [a]	Chemical Oxygen Demand
$N\text{-}NO_3^-$ [a]	Nitrate
$N\text{-}NH_4^+$ [a]	Ammonium
$N\text{-}NO_2^-$ [a]	Nitrite
OLR	Organic Loading Rate
TSS	Total Suspended Solids
SVI	Sludge Volume Index
Morphological parameters (from bright field microscopy) [b]	
Deq	Equivalent Diameter
P	Perimeter
L	Length
W	Width
FF	Form Factor
Conv	Convexity

Comp	Compactness
Round	Roundness
Solid	Solidity
Ext	Extent
Ecc	Eccentricity
Rob	Robustness
LrgC	Largest concavity
RelArea	Ratio between hole and object area
% Nb	Aggregates number percentage
% Area	Aggregates area percentage
Nb/Vol	Aggregates number per volume
TA/Vol	total aggregate area per volume
TL/Vol	total filament length per volume
TL/TA	total filament length per total aggregate area
TL/TSS	total filament length per total suspended solids
Physiological parameters (from epifluorescence microscopy)	
G_AA_R; G_AA_G	Gram-positive aggregated bacteria area; Gram-negative aggregated bacteria area
G_INA_R; G_INA_G	Gram-positive aggregated bacteria intensity; Gram-negative aggregated bacteria intensity
G_AF_R; G_AF_G	Gram-positive filaments area; Gram-negative filaments area
G_INF_R; G_INF_G	Gram-positive filaments intensity; Gram-negative filaments intensity
LD_AA_R; LD_AA_G	Damaged aggregated bacteria area; Viable aggregated bacteria area
LD _INA_R, LD_INA_G	Damaged aggregated bacteria intensity; Viable aggregated bacteria intensity
LD _AF_R; GR_AF_G	Damaged filaments area; Viable filaments area
LD _INF_R; LD _INF_G	Damaged filaments intensity; Viable filaments intensity
G_AA_G/AA_R	Ratio between gram-negative and gram-positive aggregates area
G_INA_G/INA_R	Ratio between gram-negative and gram-positive aggregates intensity
G_AF_G/AF_R	Ratio between gram-negative and gram-positive filaments area
G_INF_G/INF_R	Ratio between gram-negative and gram-positive filaments intensity
LD_AA_G/AA_R	Ratio between viable and damaged aggregates area
LD _INA_G/INA_R	Ratio between viable and damaged aggregates intensity
LD _AF_G/AF_R	Ratio between viable and damaged filaments area
LD _INF_G/INF_R	Ratio between viable and damaged filaments intensity

[a] determined inside the reactor
[b] for small (Deq<25 µm), intermediate (25 µm<Deq<250 µm) and large (Deq>250 µm)

8.2.4 PRINCIPAL COMPONENT ANALYSIS (PCA) AND PARTIAL LEAST SQUARES (PLS)

Before PCA decomposition, variables were mean centered (adjusted to zero mean by subtracting off the original mean of each column). Further, variables were autoscaled

(adjusted to unit variance by dividing each column by its standard deviation) to give them equal influence on the model since variables are expressed in different units and display substantially different absolute value ranges (Rosen, 2001).

PCA decomposes the data matrix **X** as the sum of the outer product of vectors $\boldsymbol{t}_i$ (containing the scores) and $\boldsymbol{p}_i$ (containing the loadings) plus a residual matrix **E**:

$$\mathbf{X=TP'+E} \quad (1)$$

Each of the principal components (PCs) captures as much as possible the variation which has not been explained by the former PCs, i.e., the first PC maximizes the covariance in the original data and the subsequent PCs maximize the covariance in the residual matrices which are left after extracting the former PCs. The dimensional reduction is based on the fact that the principal components are orthogonal and hence uncorrelated. It should be kept in mind that the principal components are linear combinations of the original variables. Therefore, they are abstract variables which are used to visualize latent structures and latent phenomena in the data. In this way, the original data is projected into a new coordinate system in which the objects are described by the scores and the variables by the loadings (Teppola *et al.*, 1998).

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 of the *X*-space are rotated so that the predictive power of the *X*-space with regard to the *Y*-space is enhanced (Rosen, 2001). In PLS it is crucial to determine the optimal number of latent variables (LVs) and cross-validation (CV) is a practical and 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 (cross-validation). A thorough description of PLS methodology can be visualized on Materials and Methods section in Chapter 6

In this study, Matlab™ 7.3 (The Mathworks, Natick, MA) was used to perform both principal component analysis (PCA) and partial least squares (PLS) analyses.

8.3. RESULTS AND DISCUSSION

8.3.1 PCA RESULTS

A Principal Components Analysis (PCA) was applied to the normalized data of the operation of the lab-scale activated sludge system. Experiments were conducted within the activated sludge system in order to observe four different operational conditions (filamentous bulking, zoogleal bulking, pinpoint flocs formation and normal conditions). The PCA study was performed using solely morphological parameters obtained from bright field microscopy (1st dataset), solely physiological parameters obtained from fluorescence microscopy (2nd dataset), combining morphological and physiological parameters (3rd dataset) and combining all the information obtained from image analysis alongside operational parameters (4th dataset). Table 8.2 presents the percentage of variance captured by each Principal Component (PC) as a function of the number of PCs. The number of components used in a PCA model represents a measure of the data complexity and can be regarded as the number of independent underlying phenomena. In the present study the number of components was determined by the eigenvalue-one criterion (analysis of the eigenvalues of the covariance matrix), according to which only the PC's with eigenvalues greater than one are considered relevant (Lourenço *et al.*, 2006).

Table 8.2. Percentage variance captured for each PC in the PCA model of mean centered data.

1st - Morphological parameters			
PC	**Eigenvalues**	**Variance (%)**	**Cumulative (%)**
1	15.72	28.58	28.58
2	14.54	26.43	55.01
3	7.17	13.03	68.04
4	**4.63**	**8.42**	**76.46**
2nd - Physiological parameters			
PC	**Eigenvalues**	**Variance (%)**	**Cumulative (%)**
1	4.71	19.63	19.63
2	3.34	13.93	33.57
3	2.99	12.48	46.05
4	2.26	9.40	55.45
5	1.67	6.96	62.41
6	1.49	6.22	68.64
7	1.17	4.89	73.52
8	**1.05**	**4.39**	**77.91**
3rd - Morphological + Physiological parameters			
PC	**Eigenvalues**	**Variance (%)**	**Cumulative (%)**
1	18.95	23.69	23.69
2	17.81	22.56	45.94
3	8.59	10.74	56.68
4	6.09	7.61	64.29
5	4.07	5.09	69.39
6	2.74	3.42	72.81
7	2.14	2.67	75.49
8	**1.87**	**2.34**	**77.83**
4th - Morphological + Physiological + operational parameters			
PC	**Eigenvalues**	**Variance (%)**	**Cumulative (%)**
1	20.89	24.01	24.01
2	18.23	20.95	44.96
3	9.23	10.76	55.72
4	9.36	7.77	68.27
5	6.76	4.77	68.27
6	4.15	3.33	71.59
7	2.89	2.89	74.49
8	**2.52**	**2.27**	**76.77**

The number of PCs presenting eigenvalues larger than 1 ascended to four (explaining 76.5% of the total variance) with respect to the morphological parameters study, and to eight in the physiological, morphological plus physiological and image analysis plus operational parameters, corresponding to total variances of respectively 77.9%, 77.8% and 76.8%.

A more explicit way to analyze and detect changes or disturbances in the measured data is through a score plot analysis. A score plot is any pair of score vectors plotted against each other, in which each sample appears as a data point and closely interrelated samples appear clustered together. For clarity purposes, only the first three PCs scatter plot will be presented next for each studied condition.

Figure 8.1 shows the first three PCs capturing 68% of the total variance using solely morphological parameters. PC1 and PC2 score plot (Figure 8.1a) evidenced two filamentous bulking clusters, clearly distinct with respect to the remaining studied conditions. All the remaining conditions (pinpoint flocs formation, zoogleal bulking and normal conditions) were able to be separated in different clusters, although only barely, by the use of the PC1 and PC3 scatter plots (Figure 8.1b).

The PCA study using solely the physiological results is depicted in Figure 8.2, representing the first three PCs and capturing only 46% of the total variance. PC1 and PC3 score plot (Figure 8.2a) allowed distinguishing the two filamentous bulking clusters with respect to each other and to the remaining studied conditions. PC2 and PC3 score plot (Figure 8.2b) further allowed distinguishing between the zoogleal bulking and pinpoint flocs formation conditions. The normal operation conditions could only be scarcely separated with respect to the zoogleal bulking phenomena by PC2 and PC3 score plot, as seen in Figures 8.2c. No clear distinction between the pinpoint flocs formation and normal operational conditions was obtained by the use of the first three principal components.

The combination of the morphological and physiological parameters, regarding the first three PCs, provided 56.7% of the total variance, as shown in Figure 8.3. Although the total variance, explained by these PCs was smaller than the obtained by the use of solely the morphological parameters, this can be explained by the larger original dataset. However, only a very slight improvement regarding the individualization of the two filamentous bulking clusters by PC1 and PC2 score plot (Figure 8.3a), as well as between the remaining conditions by PC2 and PC3 (Figure 8.3b), could be attained.

In Figure 8.4, although the first three PCs (PC1, PC2 and PC3) only explained 55% of the total variance, larger original dataset, they allowed to clearly distinguish all four studied conditions from each other. In fact, the presented PC1, PC2 and PC3 score plots reflected the four main clusters found in the data set, corresponding to the filamentous bulking, zoogleal bulking, pinpoint flocs formation and normal conditions. All the remaining conditions (pinpoint flocs formation, zoogleal bulking and normal conditions) were able to be separated in different clusters by the use of the PC2 and PC3 scatter plot (Figure 8.4b). However, and despite the larger original dataset, it could also be seen a slightly clearer distinction of these conditions regarding the use of the morphological and physiological dataset or even solely the morphological dataset.

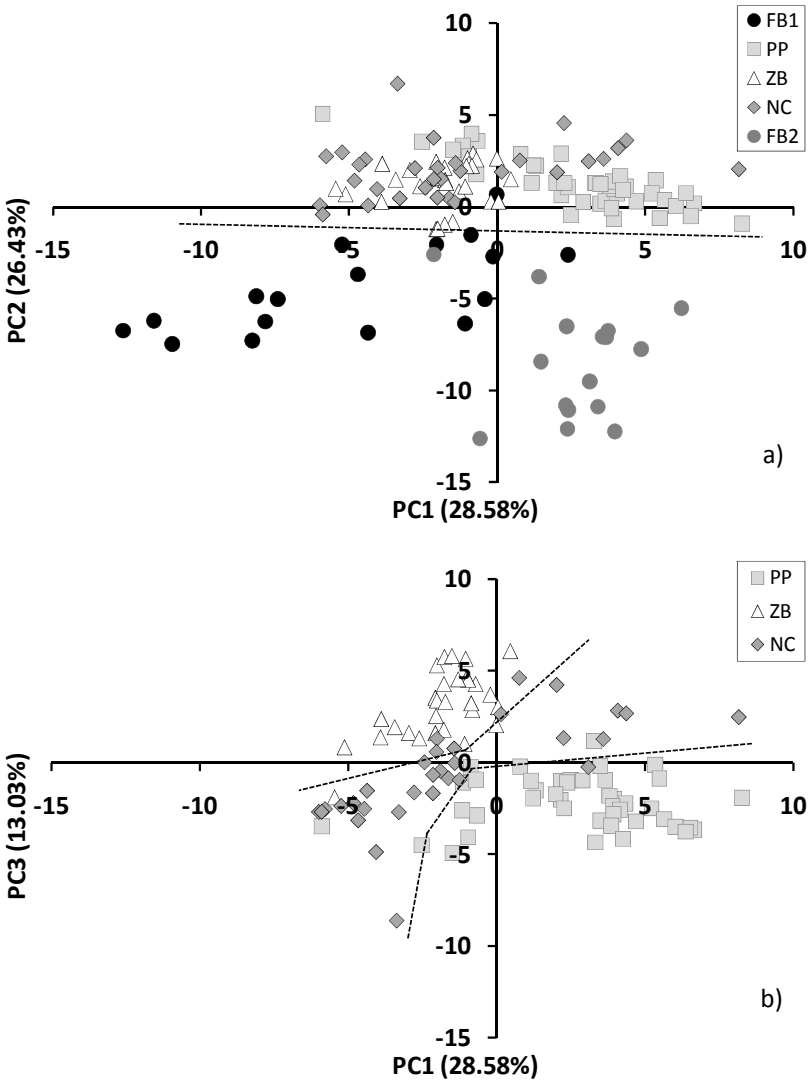

Figure 8.1. PCA score plot of the PC1 *vs.* PC2 (a), PC1 *vs.* PC3 (b) for the recognition of all phenomena for the 1st dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zoogleal bulking; NC – normal conditions.

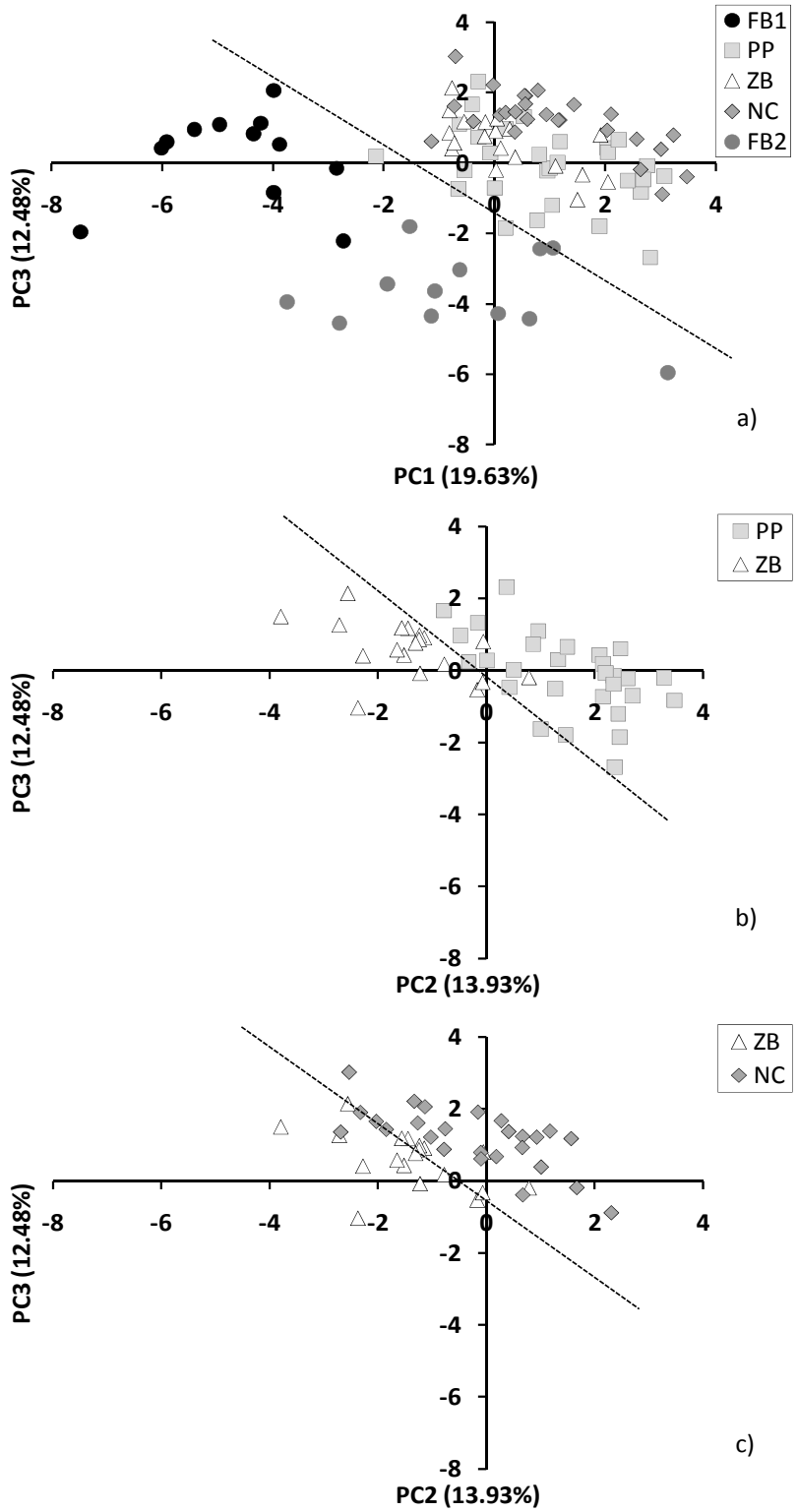

Figure 8.2. PCA score plot of the PC1 *vs.* PC3 (a), PC2 *vs.* PC3 (b), PC2 *vs.* PC3 (c) for the recognition of all phenomena using for the 2nd dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zoogleal bulking; NC – normal conditions.

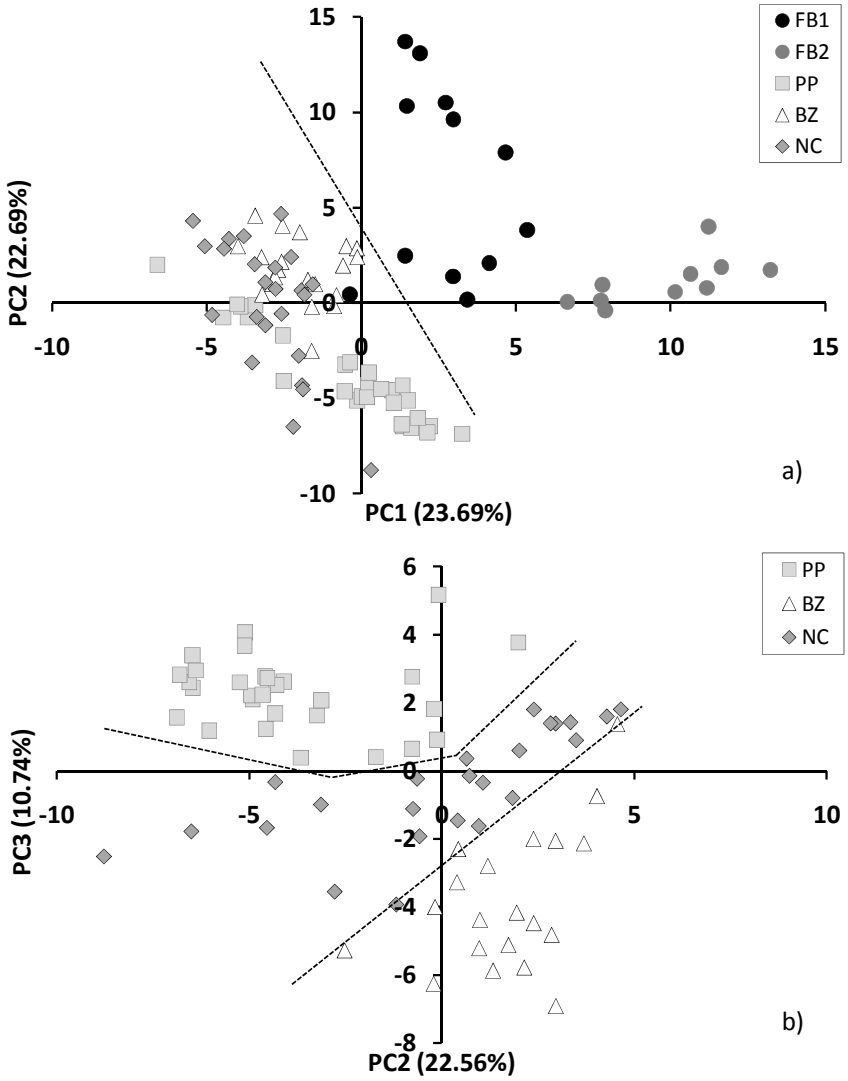

Figure 8.3. PCA score plot of the PC1 *vs.* PC2 (a), PC2 *vs.* PC3 (b) for the recognition of all phenomena for the 3rd dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zoogleal bulking; NC – normal conditions.

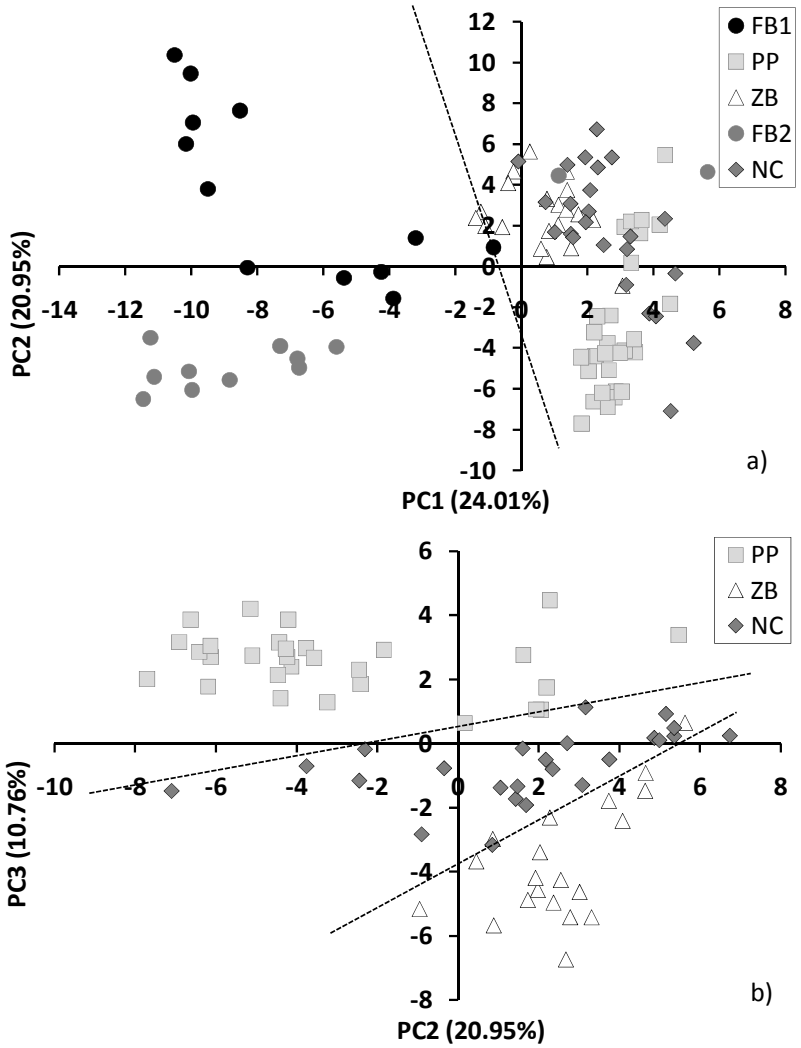

Figure 8.4. PCA score plot of the PC1 *vs.* PC2 (a), PC2 *vs.* PC3 (b) for the recognition of all phenomena for the 4th dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zoogleal bulking; NC – normal conditions.

Table 8.3 and Table 8.4 shows the variables presenting the larger principal component coefficients of each PC derived by PCA, regarding morphological parameters and the combination of morphological and physiological parameters, respectively.

Table 8.3. Identification of the variables presenting the larger principal component coefficients in the morphological parameters PCA analysis.

Variables	PC1	Variables	PC2	Variables	PC3
Sol_int	**0.238**	% Area_int	**0.223**	Deq_int	**0.301**
Round_int	**0.236**	% Area_s	**0.223**	L_int	**0.298**
LrgC_int	**0.233**	W_s	**0.208**	W_int	**0.291**
Ext_int	0.229	% Nb_s	0.204	% Area_lg	0.268
Comp_int	0.227	Nb/Vol_s	0.203	P_int	0.260
Rob_int	0.222	% Nb_int	0.202	L_s	0.250
FF_int	0.220	FF_s	0.191	P_s	0.233
FF_lg	0.200	Deq_s	0.191	% Nb_lg	0.216
RelArea_lg	0.196	Comp_s	0.190	Nb/Vol_lg	0.205
Ecc_int	0.194	Round_s	0.185	Deq_s	0.193
Per_lg	0.191	Rob_lg	0.184	Conv_s	0.183

s – small aggregates (Deq<25 µm); int – intermediate aggregates (25 µm<Deq<250 µm); lg – large aggregates (Deq>250 µm)

Table 8.4. Identification of the variables presenting the larger principal component coefficients in the morphological and physiological parameters PCA analysis.

Variables	PC1	Variables	PC2	Variables	PC3
Comp_s	**0.200**	Ext_int	**0.225**	Deq_int	**0.276**
% Area_s	**0.198**	Rob_int	**0.223**	W_int	**0.258**
% Area_int	**0.198**	FF_int	**0.215**	L_int	**0.255**
Round_s	0.198	Sol_int	0.215	% Area_lg	0.252
TL/VOL	0.197	LrgC_int	0.215	P_int	0.211
TL/TSS	0.196	FF_lg	0.204	% Nb_lg	0.207
FF_s	0.195	Conv_lg	0.203	L_s	0.207
W_s	0.186	P_lg	0.203	LD_AA_R	0.199
Ecc_s	0.181	Round_int	0.197	TA/VOL	0.195
LD_AF_G	0.176	RelArea_lg	0.197	Nb/Vol_lg	0.194
Ext_s	0.176	Comp_int	0.186	Per_s	0.190

s – small aggregates (Deq<25 µm); int – intermediate aggregates (25 µm<Deq<250 µm); lg – large aggregates (Deq>250 µm)

The use of solely morphological parameters (Table 8.3) revealed that PC2, the main component responsible for the separation of the filamentous bulking clusters, was strongly influenced by intermediate and small aggregates size characterization. PC1 and PC3, responsible for the separation of the remaining clusters, were strongly

affected the of the intermediate aggregates morphology and size characterization, respectively.

Analyzing Table 8.4, it was found that PC1, the main component responsible for the separation of the filamentous bulking clusters, was strongly affected by the morphology of the small aggregates, by the aggregated biomass size distribution, and by the filamentous bacteria contents. PC2 and PC3, responsible for the separation of the remaining clusters, were strongly influenced by the morphology of intermediate aggregates, and by their size characterization and aggregated biomass size distribution, respectively.

Considering the combination of morphological, physiological and operational parameters, Table 8.5 shows the variables presenting the larger principal component coefficients of each PC derived by PCA.

Table 8.5. Identification of the variables presenting the larger principal component coefficients in the morphological, physiological and operational parameters PCA analysis.

Variables	PC1	Variables	PC2	Variables	PC3
TL/Vol	**0.200**	Rob_int	**0.210**	Deq_int	**0.247**
TL/TSS	**0.199**	Ext_int	**0.210**	W_int	**0.228**
SVI	**0.188**	Conv_lg	**0.208**	LD_AF_R	**0.228**
Comp_s	0.185	P_lg	0.203	L_int	0.228
Round_s	0.184	FF_lg	0.202	% Area_lg	0.226
LD_AF_G	0.177	FF_int	0.199	TA/VOL	0.226
FF_s	0.176	Sol_lg	0.192	$N\text{-}NO_3^-\_r$	0.203
% Area_int	0.175	Conv_int	0.192	% Nb_lg	0.189
%_Area_s	0.174	Rob_lg	0.190	G_INA_G	0.184
Ecc_s	0.173	LrgC_int	0.190	P_int	0.184
ln(TL/TA)	0.173	RelArea_lg	0.188	Nb/Vol_lg	0.181

s – small aggregates (Deq<25 μm); int – intermediate aggregates (25 μm<Deq<250 μm); lg – large aggregates (Deq>250 μm); r – measurement inside the reactor

From Table 8.5, it was found that PC1 was strongly affected by the filamentous bacteria contents and sludge settling ability (SVI), contributing for the clear distinction of the filamentous bulking phenomena. PC2 was strongly influenced by the morphology of large and intermediate aggregates, and PC3 by the intermediate aggregates size characterization. PC2, combined with PC3, was responsible for the distinction of the zoogleal bulking conditions from the normal operating conditions, thus reflecting the importance of the intermediate and large aggregates morphology and size characterization. PC3 was responsible mainly for the distinction of the pin point flocs conditions from the non-filamentous bulking conditions, emphasizing the importance of the intermediate aggregates size characterization for this purpose.

Looking back at the obtained results, the performed PCA analysis allowed identifying the studied conditions, namely filamentous bulking, zoogleal bulking, pinpoint flocs formation and normal conditions. This fact assumes higher relevance to evaluate the activated sludge condition, and act accordingly, and to allow for a better prediction of operational parameters such as SVI, TSS, COD and other by PLS, as reported in the following section.

8.3.2 PLS RESULTS

Following the results of the PCA model presented above, the possibility of using image analysis data to predict SVI, COD, $N\text{-}NO_3^-$, and TSS through PLS analysis was sought. The collected data from the aggregated and filamentous biomass characterization (morphological parameters), biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria (physiological parameters), and operational parameters were used for that purpose. The total number of samples used for training and validating the PLS model as well as the concentration ranges for each of the studied parameters are presented in Table 8.6.

Table 8.6. Characterization of the dataset used to develop the PLS models for SVI, COD, N-NO_3^-, and TSS.

Parameter	PLS model		Range of values	
	Training	Validation	Min	Max
SVI (mL/g)	71	35	55.5	483
COD (mg/L)			2.4	180
N-NO_3^- (mg/L)			0.02	313
TSS (g/L)			0.94	6.5

The choice of the number of latent variables (LV) is crucial for the monitoring and prediction outcome and typically cross-validation (leave-one-out) is used to select the appropriate number. In this method, the number of model components (latent variables) is usually chosen by the lowest root mean square error of cross-validation, RMSECV. However, it was found during the present work that the number of latent values depicted by this method did not allow obtaining a satisfactory training model. Furthermore, the use of such methodology may under or overfit the validation data results. In order to overcome such problems, in this work, the number of latent variables selection was performed based on the regression coefficient of the overall (training and validation) data. Furthermore, at least 1000 different training and validation datasets from the original dataset were tested, obtaining random training and validation datasets.

The partial least square analyses results for COD, N-NO_3^-, TSS, and SVI prediction, using the morphological dataset (1st dataset), the physiological dataset (2nd dataset), the morphological plus physiological dataset (3rd dataset), and the overall (image analysis alongside operational parameters – 4th dataset) parameters dataset, are depicted in Figure 8.5, 8.6, 8.7, and 8.8, respectively. Furthermore, the presented regression analyses and coefficients were obtained for the predicted and measured values from the overall (training and validation) sample dataset.

Regarding only the morphological parameters from the aggregated and filamentous biomass characterization (Figure 8.5), the regression analysis between predicted and measured COD and TSS presented low prediction abilities with correlation coefficients R^2 of 0.61 and 0.74, respectively, indicating that only morphological parameters from image analysis are not able to predict COD and TSS, when no prior identification on the sample operating condition is performed. However, the regression between predicted and measured for $N\text{-}NO_3^-$ and SVI was quite satisfactory with correlation coefficients R^2 of 0.91 and 0.91, respectively. It should be noticed, though, that for the larger $N\text{-}NO_3^-$ validation samples a large dispersion is observed thus hindering the found prediction ability.

Figure 8.6 shows the regression analysis using the physiological parameters obtained from epifluorescence microscopy. Lowest correlation coefficients R^2 of 0.82, 0.70, 0.87, and 0.41, were achieved for SVI, COD, $N\text{-}NO_3^-$, and TSS, respectively, demonstrating the low prediction ability of physiological parameters solely to estimate these operational parameters.

Combining the morphological and physiological parameters (Figure 8.7) resulted in an improvement on the prediction ability, regarding the two previous studies. However, regarding the COD and TSS prediction abilities, the obtained correlation coefficients R^2 of 0.71 and 0.80, respectively, were still rather low. The correlation coefficients R^2 for $N\text{-}NO_3^-$ and SVI of 0.94 and 0.93, respectively, and the dispersion reduction of the larger $N\text{-}NO_3^-$ validation samples, may allow predicting, at some extent, these parameters. Nevertheless, care should be taken regarding the $N\text{-}NO_3^-$ prediction ability, due to the fact that the large majority of samples clustered in a narrow range of values (below 100 mg/L and representing one third of the total range span), and thus being strongly and positively influenced by the few large $N\text{-}NO_3^-$ values.

Regarding the overall (image analysis alongside operational parameters) parameters dataset, only a slight improvement was obtained (Figure 8.8). Regarding the SVI prediction ability, a good correlation coefficient R^2 of 0.94 and 0.99 regressions slope were obtained, demonstrating the ability of the overall parameters dataset to predict the SVI parameter, even when no prior identification on the sample operating condition is performed. Furthermore, the regression analysis between the predicted and measured COD values demonstrated lower prediction abilities with a correlation coefficient R^2 of 0.69 and 0.82 regression slope. During the operation of the lab-scale activated sludge system, good COD removal rates were obtained leading to a narrow range of low COD values, explaining the poor prediction ability found for the COD parameter. Good prediction ability was obtained between the predicted and measured $N\text{-}NO_3^-$ values, presenting a correlation coefficient R^2 of 0.96 and a 1 regression slope. However, and given that the $N\text{-}NO_3^-$ values clustered in a narrow range (below 100 mg/L), care must be taken analyzing these results. In fact, solely 11 samples spanned across two thirds of the $N\text{-}NO_3^-$ values, whereas the remaining 95 samples were clustered in the other third. This fact may have artificially increased the prediction correlation coefficient, thus hindering the confidence of this analysis. Finally, the TSS analysis resulted in a satisfactory correlation between the predicted and measured TSS values, presenting a correlation coefficient R^2 of 0.88 and 0.98 regression slope, inferring that the TSS parameter may be, at some extent, predicted by the use of the overall parameters dataset even when no prior identification on the sample operating condition is performed.

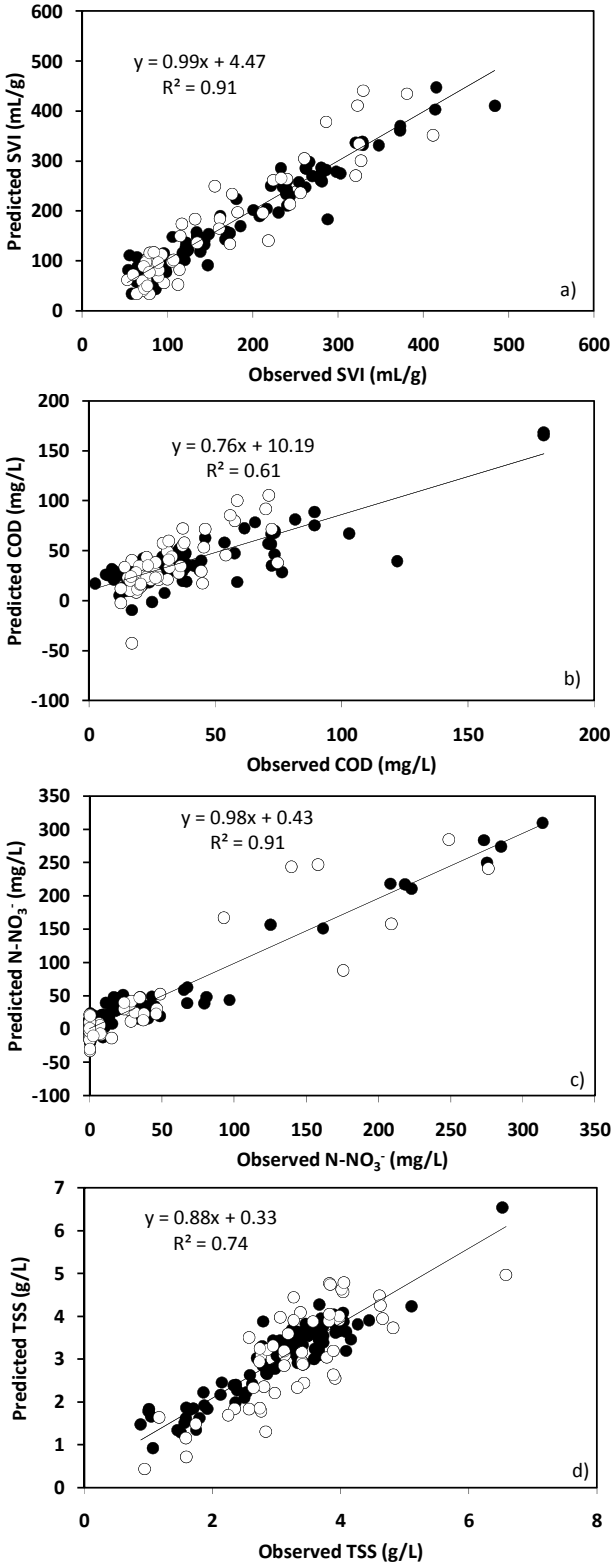

Figure 8.5. PLS regression for (a) SVI, (b) COD, (c) N-NO_3^-, and (d) TSS using the 1st dataset. Training values (•) Validation values (o).

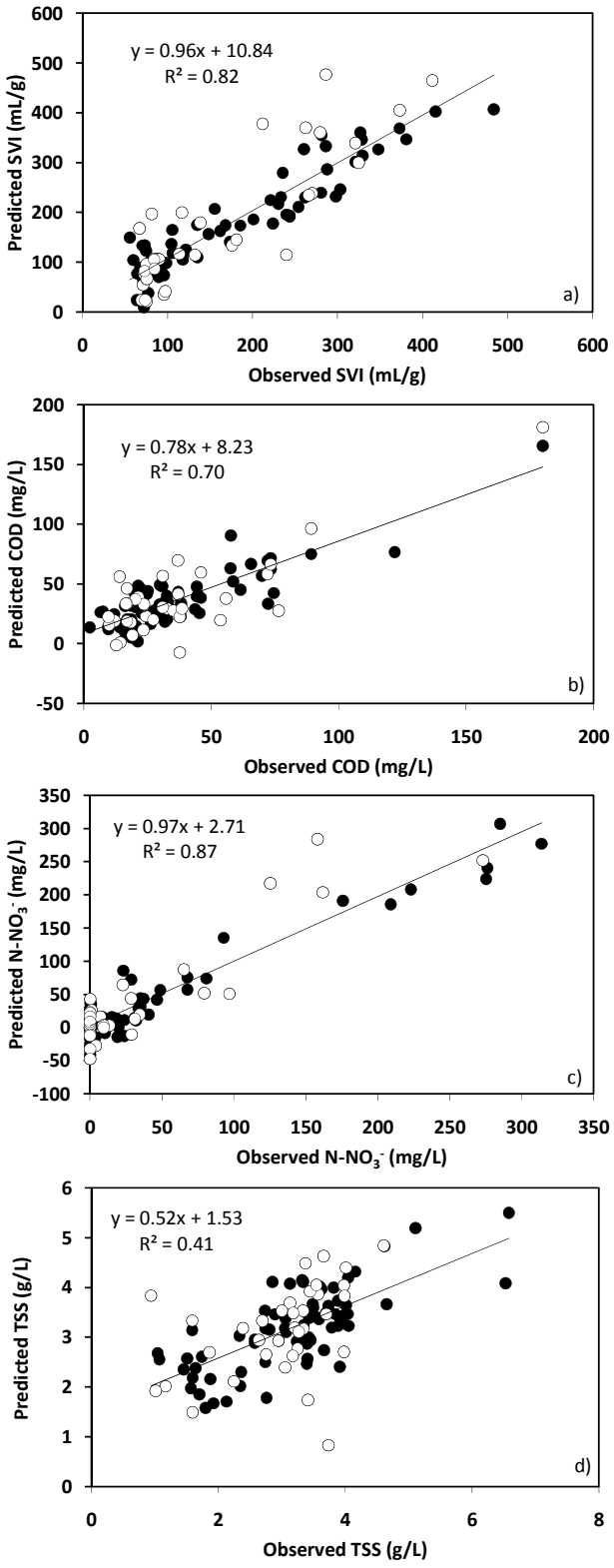

Figure 8.6. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using the 2nd dataset. Training values (●) Validation values (o).

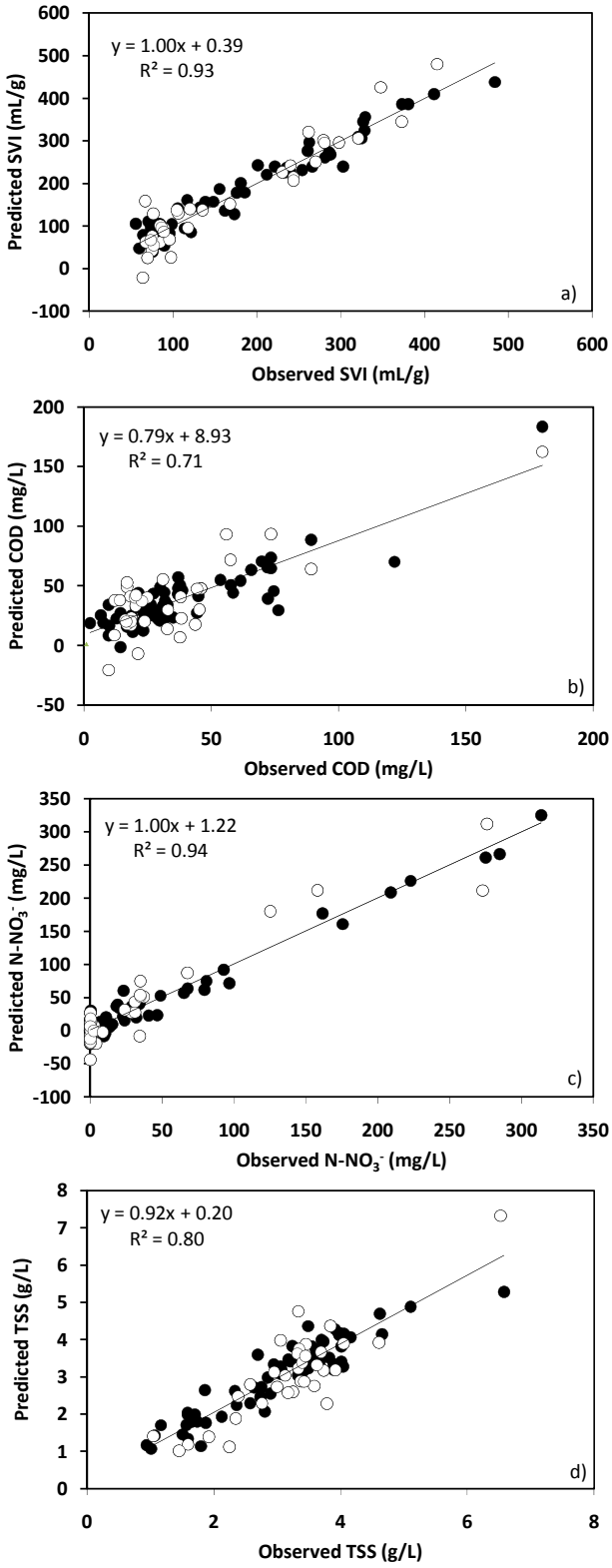

Figure 8.7. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using the 3rd dataset. Training values (•) Validation values (o).

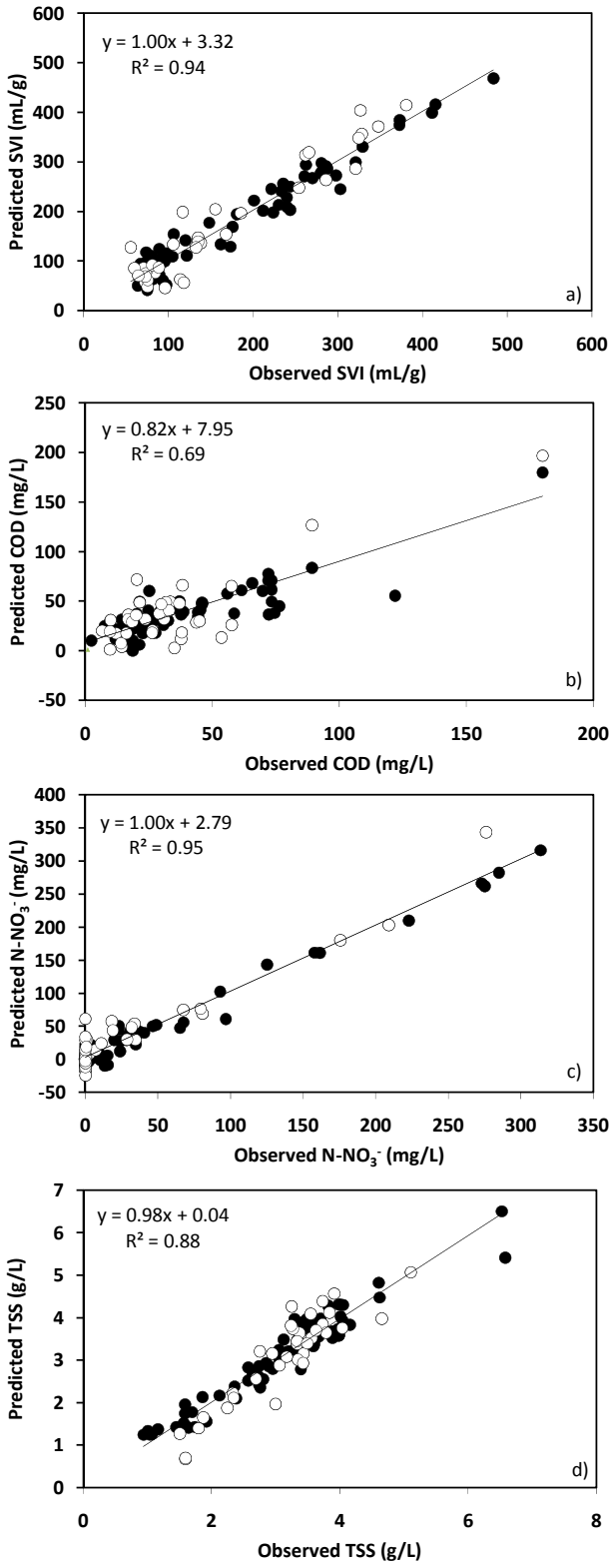

Figure 8.8. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using the 4th dataset. Training values (•) Validation values (o).

In order to improve the prediction ability for SVI, COD, N-NO_3^-, and TSS, each of the four individual conditions (filamentous bulking, pinpoint flocs formation, zoogleal bulking, and normal conditions), as described in Chapter 7, was modeled by a PLS analysis. Keeping in mind the 2/3 training to 1/3 validation dataset ratios, Table 8.7 presents the number of samples used for the calibration and validation datasets. The range of values obtained for each condition and for each parameter studied is presented in Table 8.8. The number of original variables was reduced upon the determination of its variable importance in the projection (VIP), in order to keep the number of variables below the number of training samples, as required for the PLS analysis. The VIP represents the influence of each variable of the data matrix (**X**) on the results matrix (**Y**), so that the variables with a VIP larger than 1 have an influence above average on the result and are, therefore, the most relevant for explaining **Y**. A more detailed description about this method can be found in Umetri (1998). Furthermore, alongside the VIP determination, a correlation analysis was carried out between the original variables to eliminate all those who showed a correlation factor higher than 0.9 with a larger VIP variable.

Table 8.7. Number of variables and samples used for the PLS model in each condition. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions).

Condition	Variables	Samples	
		Training	Validation
FB	8	16	7
PP	10	20	9
ZB	7	13	6
NC	9	18	8

Table 8.8. Range of values obtained for each parameter studied. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions).

Parameters	Condition	Range of Values	
		Min	Max
SVI (mL/g)	FB	220	483
	PP	63	96
	ZB	155	303
	NC	55	113
COD (mg/L)	FB	18	180
	PP	6.6	122
	ZB	9.7	72.3
	NC	9.8	74.6
$N\text{-}NO_3^-$ (mg/L)	FB	0.04	313
	PP	18	97
	ZB	0.04	15.5
	NC	0.02	15.2
TSS (g/L)	FB	1.8	3.8
	PP	0.9	3.9
	ZB	2.6	4.0
	NC	2.9	6.5

The results obtained with the PLS model for each individual condition are compiled in Table 8.9, Table 8.10, Table 8.11, and Table 8.12, where the number of PLS latent variables was chosen based on the regression coefficient of the overall (training and validation) data. Table 8.9 displays the PLS modeling results using only morphological parameters obtained from bright field microscopy. Table 8.10 presents PLS modeling results using only physiological parameters from epifluorescence microscopy. The PLS results obtained using the combination of morphological and physiological results is shown in Table 8.11. Finally, Table 8.12 shows the PLS modeling results using morphological, physiological and operational parameters.

In Table 8.9, the results obtained for the morphological parameters, PLS modeling show low SVI correlation factors (for the overall samples dataset), between 0.65 and 0.75, with the higher values corresponding to the filamentous bulking and zoogleal bulking samples due to their larger range. Although these results do not seem, at a

glance, very promising, it should be stressed that each of the studied condition represents only a small part of the overall analysis, and thus, the absolute regression values should not be compared with the ones found in the previous analysis. Such comparison may only be performed upon the representation of the overall conditions dataset, as individually determined by this methodology, which is performed further on in this chapter. Regarding the COD concentration even lower correlation factors were obtained, thus not allowing for the COD prediction in each individual condition, at this stage. Observing the $N\text{-}NO_3^-$ results, satisfactory to good correlation factors were obtained for zoogleal bulking, filamentous bulking and normal conditions, thus allowing for good prediction abilities. Contrary to the pinpoint flocs condition, the prediction ability for zoogleal bulking was quite satisfactory, in spite of the narrow range of values. Finally, it was also found that the morphological parameters were able to predict TSS during pinpoint flocs, zoogleal bulking and normal conditions, but were not as satisfactory for the filamentous bulking conditions.

PLS modeling results presented in Table 8.10 regarding SVI, $N\text{-}NO_3^-$, TSS and COD prediction abilities show, in half of the cases, an increase, and a decrease in the other half,. Comparing these results with the obtained for the morphological parameters, it is shown an increase of the SVI prediction ability for three conditions (FB, ZB, and NC), with the higher values corresponding to the filamentous bulking (R^2 of 0.81) and zoogleal bulking (R^2 of 0.90) samples due to their larger range. Regarding the COD concentration, the use of physiological parameters increased the correlation factors obtained for filamentous bulking, pinpoint flocs and normal conditions, but only for filamentous bulking was quite satisfactory. Correlation factors higher than 0.8 were obtained regarding $N\text{-}NO_3^-$ and TSS prediction ability on most cases. However, only for pinpoint flocs in the case of $N\text{-}NO_3^-$ and filamentous bulking in the case of TSS the prediction ability improved for the physiological parameters, regarding the previous morphological parameters analysis.

Combining the morphological with the physiological results, as presented in Table 8.11, it was found a clear improvement of the coefficient regressions regarding the SVI values. However, and apart from the zoogleal bulking conditions, no satisfactory prediction abilities could be retrieved from the use of these parameters, at this point, and a further analysis was performed upon the representation of the overall conditions dataset, as individually determined by this methodology. Such was also the case for the COD assessment, where, although a clear improvement was also observed, no satisfactory prediction abilities were found by combining the morphological with the physiological results. Regarding the $N\text{-}NO_3^-$ analysis, and with the exception of the pinpoint flocs condition, no advantages were found concerning the remaining conditions by the introduction of the physiological parameters. A similar conclusion could be obtained for the TSS assessment, where, with the exception of the filamentous bulking condition, no advantages were found regarding the remaining conditions by the introduction of the physiological parameters.

It is clear from Table 8.12 that the PLS models developed for SVI, COD, $N\text{-}NO_3^-$, and TSS generally improved when the reactor operational parameters were added to the analysis. In the case of SVI results, the introduction of the operational parameters allowed for good prediction abilities regarding the filamentous and bulking conditions. Although no satisfactory prediction abilities could be retrieved for the pinpoint flocs and normal conditions, at this point, a further analysis was performed upon the representation of the overall conditions dataset, as individually determined by this methodology. Correlations obtained for COD are, in three of the four cases, lower than 0.9, thus not allowing good prediction abilities with the exception of the filamentous bulking conditions. The narrow range of COD values (ZB between 9.7 and 72.3 mg/L; NC between 9.8 and 74.6 mg/L) is considered one of the major contributors to this poor correlation. With respect to the $N\text{-}NO_3^-$ and TSS concentrations good correlation coefficients (higher than 0.9) were obtained for most cases, indicating that the introduction of the operational parameters allowed

for good prediction abilities regarding these two parameters. Furthermore, a clear improvement regarding the morphological parameters study was obtained for the N-NO_3^- assessment in pinpoint flocs condition and for the TSS assessment in filamentous bulking conditions. A single exception was noticed for the TSS assessment in the normal conditions that presented lower prediction ability.

Table 8.9. PLS modeling results for SVI, COD, N-NO_3^-, and TSS and for each condition studied using the 1st dataset. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions)

	Condition	PLS LVs	R^2 Training set	R^2 Validation set	R^2 overall (training + validation) set
SVI (mL/g)	FB	5	0.76	0.72	0.75
	PP	7	0.70	0.73	0.65
	ZB	4	0.76	0.84	0.73
	NC	5	0.66	0.76	0.64
COD (mg/L)	FB	5	0.62	0.71	0.66
	PP	7	0.42	0.38	0.35
	ZB	5	0.77	0.63	0.63
	NC	5	0.79	0.36	0.48
N-NO_3^- (mg/L)	FB	5	0.94	0.92	0.92
	PP	5	0.41	0.14	0.23
	ZB	5	0.91	0.92	0.90
	NC	5	0.97	0.88	0.92
TSS (g/L)	FB	5	0.72	0.71	0.68
	PP	6	0.98	0.93	0.97
	ZB	5	0.91	0.92	0.90
	NC	5	0.97	0.88	0.94

Table 8.10. PLS modeling results for SVI, COD, N-NO_3^-, and TSS and for each condition studied using the 2nd dataset. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions).

	Condition	PLS LVs	R^2 Training set	R^2 Validation set	R^2 overall (training + validation) set
SVI (mL/g)	FB	4	0.84	0.88	0.81
	PP	5	0.74	0.64	0.59
	ZB	3	0.92	0.88	0.89
	NC	4	0.74	0.69	0.68
COD (mg/L)	FB	4	0.94	0.84	0.87
	PP	5	0.67	0.80	0.53
	ZB	3	0.87	0.63	0.56
	NC	4	0.71	0.67	0.63
N-NO_3^- (mg/L)	FB	4	0.89	0.93	0.88
	PP	4	0.91	0.84	0.84
	ZB	3	0.80	0.70	0.75
	NC	4	0.90	0.85	0.88
TSS (g/L)	FB	4	0.86	0.89	0.84
	PP	4	0.95	0.86	0.91
	ZB	3	0.88	0.77	0.85
	NC	4	0.91	0.85	0.80

Table 8.11. PLS modeling results for SVI, COD, N-NO_3^-, and TSS and for each condition studied using the 3rd dataset. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions).

	Condition	PLS LVs	R^2 Training set	R^2 Validation set	R^2 overall (training + validation) set
SVI (mL/g)	FB	4	0.89	0.89	0.86
	PP	5	0.80	0.87	0.76
	ZB	2	0.92	0.93	0.91
	NC	4	0.87	0.73	0.81
COD (mg/L)	FB	4	0.92	0.90	0.88
	PP	5	0.66	0.43	0.54
	ZB	3	0.69	0.61	0.66
	NC	4	0.79	0.72	0.64
N-NO_3^- (mg/L)	FB	4	0.71	0.63	0.61
	PP	5	0.77	0.71	0.69
	ZB	3	0.94	0.94	0.92
	NC	4	0.77	0.59	0.63
TSS (g/L)	FB	4	0.77	0.99	0.80
	PP	3	0.97	0.96	0.96
	ZB	3	0.84	0.90	0.82
	NC	3	0.90	0.93	0.87

Table 8.12. PLS modeling results for SVI, COD, N-NO_3^-, and TSS and for each condition studied using the 4th dataset. (FB – filamentous bulking, PP – pinpoint flocs, ZB – zoogleal bulking, NC – normal conditions).

	Condition	PLS LVs	R^2 Training set	R^2 Validation set	R^2 overall (training + validation) set
SVI (mL/g)	FB	3	0.97	0.96	0.97
	PP	5	0.86	0.77	0.79
	ZB	3	0.92	0.97	0.93
	NC	4	0.85	0.80	0.79
COD (mg/L)	FB	3	0.96	0.91	0.94
	PP	5	0.74	0.85	0.73
	ZB	3	0.76	0.77	0.73
	NC	4	0.61	0.69	0.53
N-NO_3^- (mg/L)	FB	2	0.99	0.86	0.95
	PP	4	0.99	0.98	0.99
	ZB	3	0.94	0.97	0.95
	NC	4	0.98	0.96	0.97
TSS (g/L)	FB	3	0.99	0.97	0.99
	PP	4	0.89	0.90	0.85
	ZB	3	0.94	0.83	0.90
	NC	4	0.96	0.86	0.93

Following the individual study of each experimental condition and parameter, the obtained results, for each studied parameter were collected into an overall analysis, and the regression values were compared to the ones represented in Figures 8.5 to 8.8. As above, the presented regression analysis and coefficients were obtained for the predicted and measured values from the overall (training and validation) sample dataset.

The use of the morphological parameters dataset depicted in Figure 8.9 shows high regression coefficients of 0.95, 0.93, and, 0.93 for SVI, N-NO_3^-, and TSS, respectively. However, such was not the case for the COD prediction ability presenting a low regression coefficient of 0.65. Comparing these results with the ones obtained without the individual condition analyses (Figure 8.5), an increased prediction ability for each studied parameter, could be observed.

Figure 8.10 depicts SVI, COD, N-NO_3^- and TSS regressions between observed and predicted values using only the physiological dataset. The regression coefficients R^2 of 0.97, and 0.77 obtained for SVI and COD, respectively, were higher than the ones achieved using the morphological dataset. Observing the SVI results obtained for the individual conditions (Table 8.10) where regression coefficients lower than 0.9 were always achieved, it is clear an improvement of the SVI prediction ability. However, in spite of a good prediction ability of N-NO_3^- and TSS (R^2 of 0.92, and 0.92, respectively), comparing with the obtained with the morphological dataset, lower regression coefficients were obtained indicating that using only physiological parameters does not improve the prediction ability of N-NO_3^- and TSS. Again, it is clear an enhancement of the prediction ability of these parameters when compared with the results obtained without the individual condition analyses (Figure 8.6).

The combination of morphological and physiological datasets is represented in Figure 8.11. Again, good regression coefficients R^2 of 0.98, 0.91, and 0.94 were obtained for SVI, N-NO_3^-, and TSS, respectively, as for the morphological dataset analysis, but a larger and satisfactory regression R^2 of 0.81 was now obtained for the COD prediction ability. Furthermore, apart from the N-NO_3^- prediction ability, the remaining parameters prediction ability also improved regarding the physiological dataset analysis. Once more, comparing these results with the ones obtained without the individual condition analyses (Figure 8.7) an increased prediction ability for each studied parameter apart for the N-NO_3, could be observed.

Regarding the overall (image analysis alongside operational parameters) parameters dataset, presented in Figure 8.12, N-NO_3^-, TSS, SVI and parameters presented very high regression coefficients R^2 of 0.97, 0.98 and 0,99, indicating quite good prediction abilities. Although not quite satisfactory, the overall COD prediction ability also improved, after the individual COD determination for each condition, to a regression coefficient R^2 of 0.84. However, care should be taken regarding the N-

NO_3^- and COD prediction abilities, due to the fact that the large majority of the samples were clustered in a narrow range of values (one third and half of the total range span, respectively), and thus being strongly and positively influenced by the few large N-NO_3^- (11 out of 106 samples in the two higher thirds) and COD (4 out of 106 samples in the higher half) values.

Overall, the comparison of the regression coefficients from Figures 8.5 to 8.8 with Figures 8.9 to 8.12, allowed concluding that a previous identification of the disturbances occurring in an activated sludge systems improved the prediction ability of the studied operational parameters. It could also be observed that stand alone analysis for both the morphological and the physiological parameters allowed for good prediction abilities regarding SVI, N-NO_3^-, and TSS, but not for COD. The further use of the operational parameters presented no significant advantages for the SVI and TSS estimation, but was invaluable for the COD and N-NO_3^- assessment. However, it should also be kept in mind that care should be taken regarding the COD and N-NO_3^- results, as previously discussed.

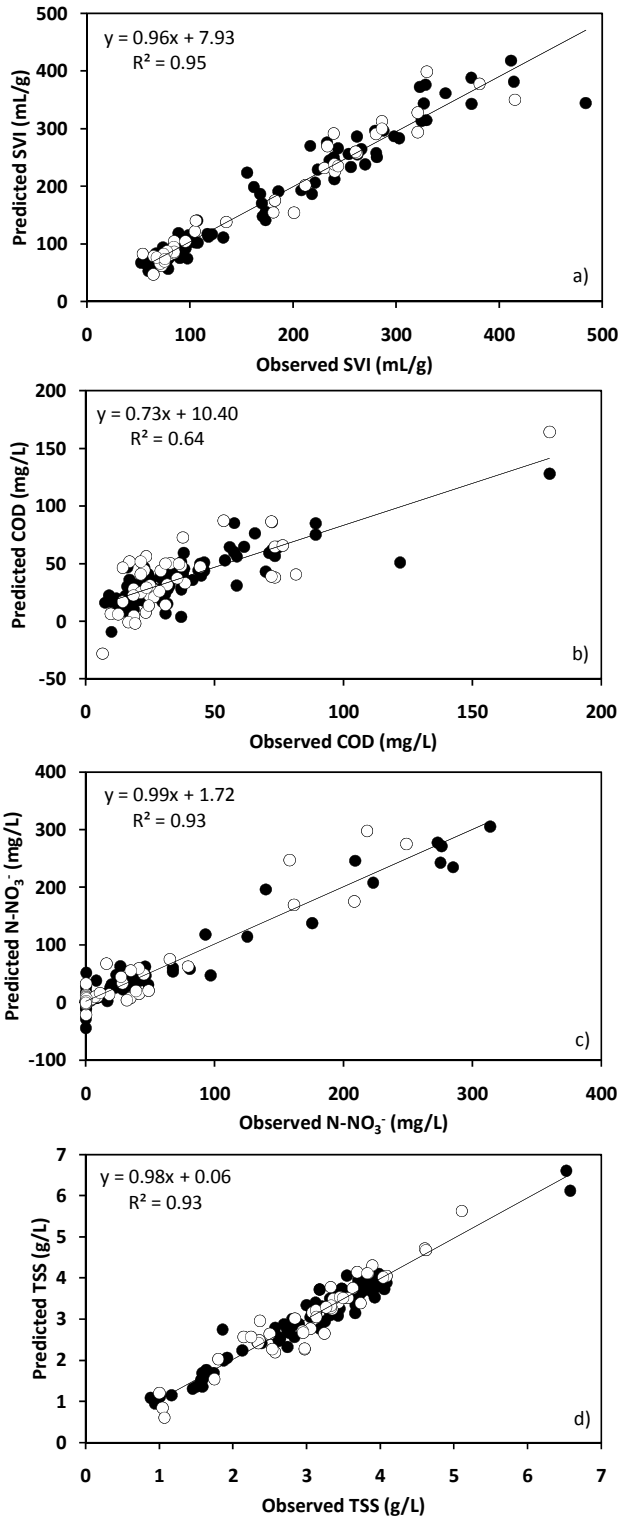

Figure 8.9. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using FB, PP, ZB and NC measured and predicted results for the 1st dataset. Training values (•) Validation values (o).

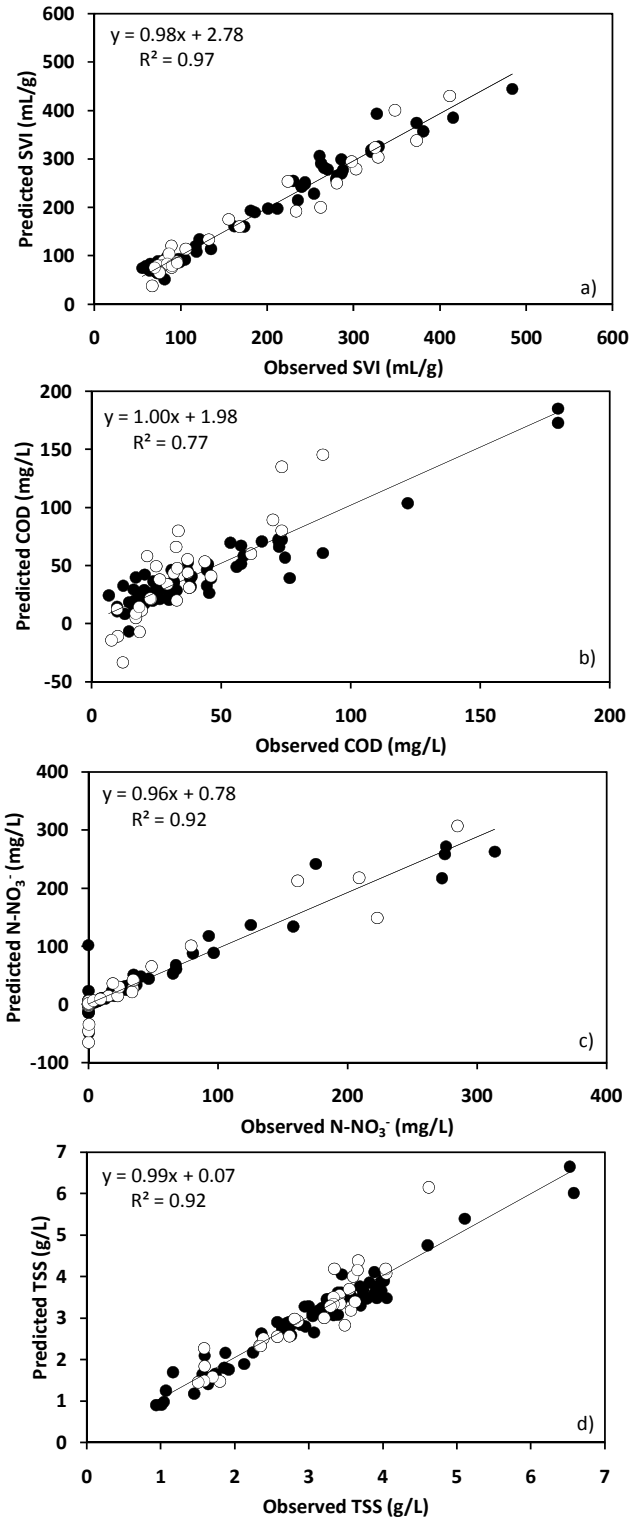

Figure 8.10. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using FB, PP, ZB and NC measured and predicted results for the 2nd dataset. Training values (•) Validation values (o).

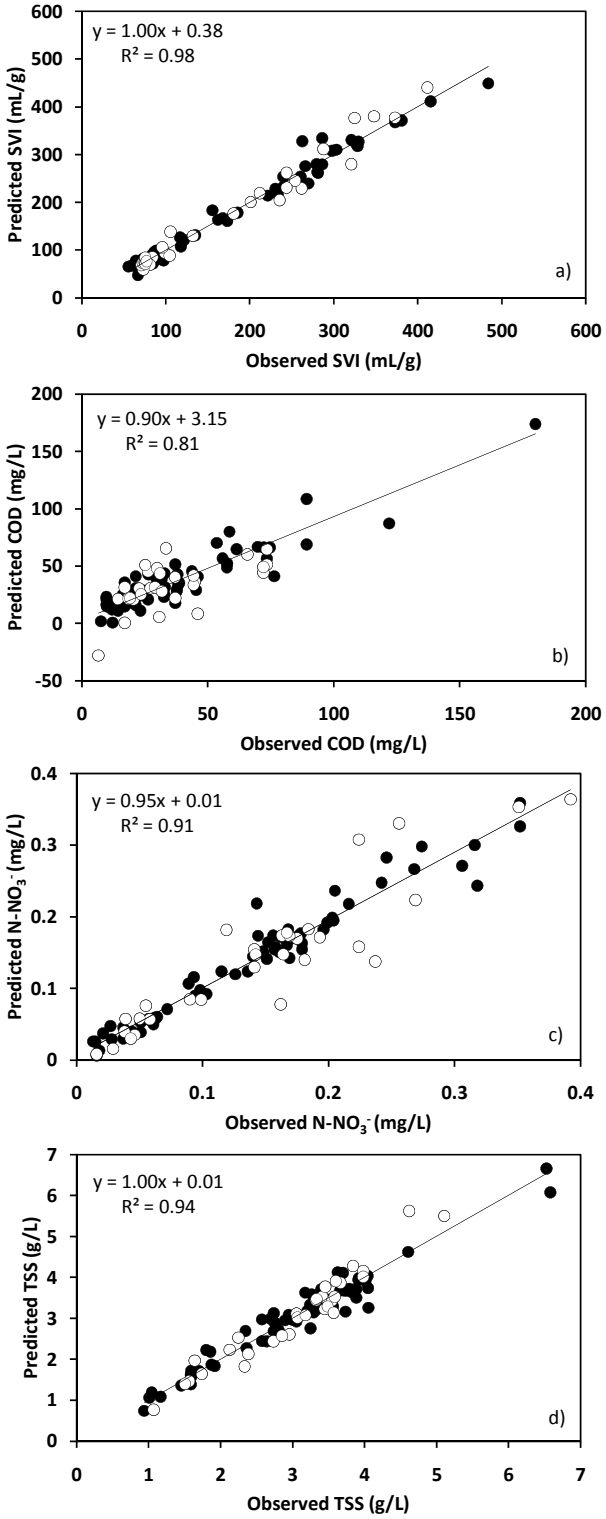

Figure 8.11. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using FB, PP, ZB and NC measured and predicted results for the 3rd dataset. Training values (•) Validation values (o).

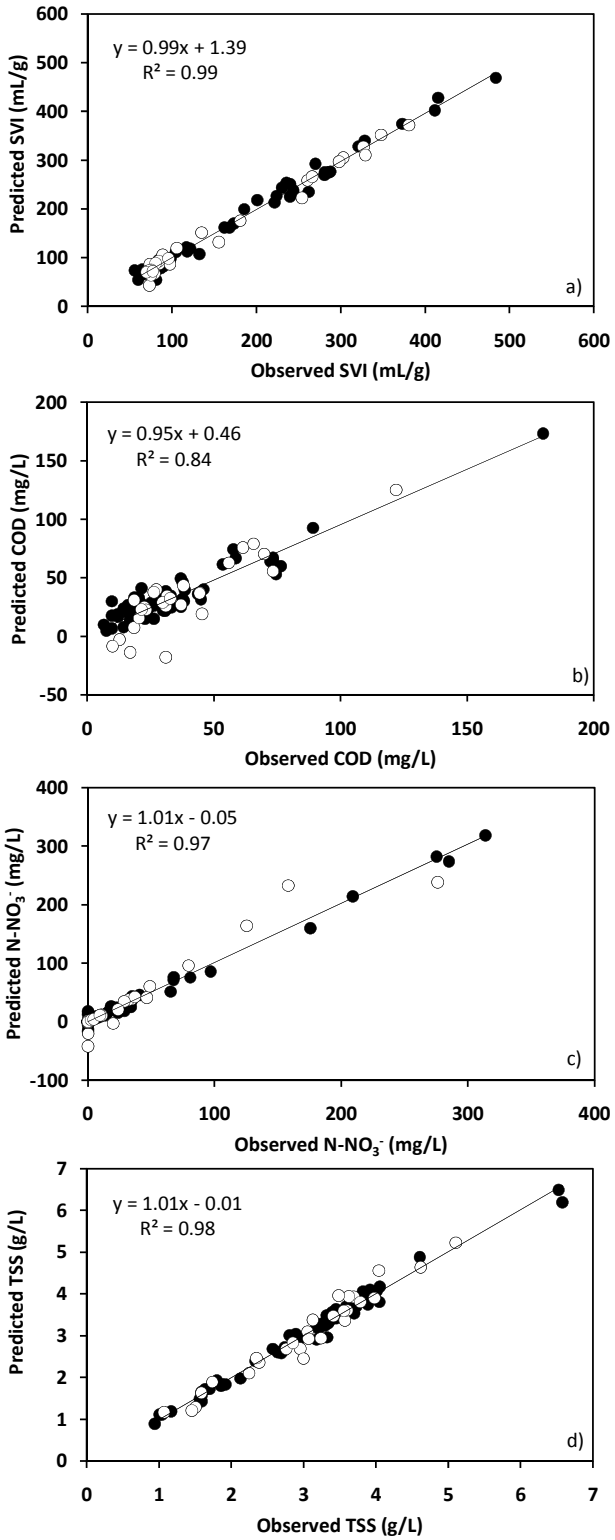

Figure 8.12. PLS regression for (a) SVI, (b) COD, (c) $N\text{-}NO_3^-$, and (d) TSS using FB, PP, ZB and NC measured and predicted results for the 4th dataset. Training values (•) Validation values (o).

8.4. CONCLUSIONS

This work evaluated the ability of image analysis information to monitor activated sludge systems. The performed PCA analysis allowed the identification of filamentous bulking, pinpoint flocs, zoogleal bulking and normal conditions. Using PLS some key parameters such as SVI, COD, $N\text{-}NO_3^-$, and TSS were estimated. It was found that a previous identification of the disturbances in the lab-scale activated sludge system, improved the prediction ability of these parameters. Regarding the four analyses performed, it was concluded that the use of morphological and physiological parameters allowed for good SVI, $N\text{-}NO_3^-$, and TSS prediction abilities. Due to the narrow range of values, the addition of operational parameters was essential for COD estimation. However, it could also be seen that care should be taken regarding the COD and $N\text{-}NO_3^-$ results, due to the cluster of the large majority of samples in a narrow range of values. These results support the use of image analysis procedures in quality assessment of activated sludge systems monitoring and control. From these studies it can thus be concluded that image analysis associated with chemometric techniques can be a valuable tool for wastewater quality monitoring.

8.5. REFERENCES

Amaral, A.L. and Ferreira, E.C. (2005) Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* **544**, 246-253.

APHA, AWWA, and WPCF (1989) *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington, DC.

Costa, J.C., Alves, M.M., and Ferreira, E.C. (2009) Principal component analysis and quantitative image analysis to predict effects of toxics in anaerobic granular sludge. *Bioresource Technology* **100**, 1180-1185.

da Motta, M., Amaral, A.L., Casellas, M., Pons, M.N., Dagot, C., Roche, N., Ferreira, E.C., and Vivier, H. (2001b) Characterization of activated sludge by automated image analysis: validation on full-scale plants. *IFAC Computer Applications in Biotechnology*, Québec City, Canada, 427-431.

da Motta, M., Amaral, A.L., Neves, L., Araya-Koff, P., Ferreira, E.C., Alves, M.M., Mota, M., Roche, N., Vivier, H., and Pons, M.N. (2002) Dilution effects on biomass characterization by image analysis. *In Proceedings of the 14th Brazilian Congress on Chemical Engineering,* CD-ROM (p. 9), Natal (Brazil).

da Motta, M., Pons, M.N., and Roche, N. (2001a) Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technololy* **43** (7), 91–96.

Dagot, C., Pons, M.N., Casellas, M., Guibaud, G., Dollet, P., and Baudu, M. (2001) Use of image analysis and rheological studies for the control of settleability of filamentous bacteria: application in SBR reactor. *Water Science and Technology* **43** (3), p. 27-33.

Dias, A.M.A., Moita, I., Páscoa, R., Alves, M.M., Lopes, J.A., and Ferreira, E.C. (2008) Activated sludge process monitoring through in-situ NIR spectral analysis. *Water Science and Technology* **57** (10), 1643-1650.

Ganczarczyk, J.J. (1994) Microbial Aggregates in Wastewater Treatment. *Water Science and Technology* **30**, 87-95.

Glasbey, C.A. and Horgan, G.W. (1995) *Image Analysis for the Biological Sciences*. John Wiley and Sons, Chichester.

Grijspeerdt, K. and Verstraete, W. (1997) Image analysis to estimate the settleability and concentration of activated sludge. *Water Research* **31**, 1126-1134.

Jenné, R., Banadda, E.N., Smets, I.Y., and Van Impe, J.F. (2004) Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* **50** (7), 281-285.

Jenné, R., Banadda, E.N., Smets, I.Y., Deurinck, J., and Van Impe, J.F. (2007) Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis* **13**, 36-41.

Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y.J., and Park, J.M. (2006) Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process Biochemistry* **41**, 2050-2057.

Lee, S.E., Koopman, B., Bode, H., and Jenkins, D. (1983) Evaluation of alternative sludge settleability indexes. *Water Research* **17** (10), 1421-1426.

Lourenço, N.D., Chaves, C.L., Novais, J.M., Menezes, J.C., Pinheiro, H.M., and Diniz, D. (2006) UV spectra analysis for water quality monitoring in a fuel park wastewater treatment plant. *Chemosphere* **65** (5), 786-791.

Lourenço, N.D., Menezes, J.C., Pinheiro, H.M., and Diniz, D. (2008) Development of PLS calibration models from UV-vis spectra for TOC estimation at the outlet of a fuel park wastewater treatment plant. *Environmental Technology* **29** (8), 891-898.

Lourenço, N.D., Paixão, F., Pinheiro, H.M., and Sousa, A. (2010) 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. *Applied Spectroscopy* **64** (9), 1061-1067.

Martins, A.M.P., Pagilla, K.R., Heijnen, J.J., and Van Loosdrecht, M.C.M. (2004) Bulking filamentous sludge - a critical review. *Water Research* **38** (4), 793–817.

Matsui, S. and Yamamoto, R. (1984) The use of a color TV technique for measuring filament length and investigating sludge bulking causes. *Water Science and Technology* **16**, 69-81.

Mesquita, D.P., Dias, O., Amaral, A.L., and Ferreira, E.C. (2010a) A comparison between bright field and phase contrast image analysis techniques in activated sludge morphological characterization. *Microscopy and Microanalysis* **16** (2), 166-174.

Mesquita, D.P., Dias, O., Elias, R.A.V., Amaral, A.L., and Ferreira, E.C. (2010b) Dilution and magnification effects on image analysis applications in activated sludge characterization. *Microscopy and Microanalysis* **16** (5) 561-568.

Noesis, (1998) *Visilog 5 Documentation Set*.

Palm, J.C., Jenkins, D., and Parker, D.S. (1980) Relationship between organic loading, dissolved-oxygen concentration and sludge settleability in the completely-mixed activated-sludge process. *Journal of Water Pollution Control Federation* **52** (10), 2484-2506.

Pons, M.N., Vivier, H., and Dodds, J. (1997) Particle shape characterization using morphological descriptors. *Particle System Characterization* **14** (6), 272-277.

Rosen, C. (2001) *A chemometric approach to process monitoring and control with applications to wastewater treatment operation.* PhD Thesis, Lund Institute of Technology, Sweden.

Rosen, C. and Olsson, G. (1998) Disturbance detection in wastewater treatment plants. *Water Science and Technology* **37** (12), 197-205.

Russ, J.C. (2002) *The Image Processing Handbook*. Fourth edition, CRC Press, Boca Raton.

Sarraguça, M.C., Paulo, A., Alves, M.M., Dias, A.M.A., Lopes, J.A., and Ferreira, E.C. (2009) Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy. *Analytical and Bioanalytical Chemistry* **395**, 1159-1166.

Schuler, A. and Jassly, D. (2007) Filament content threshold for activated sludge bulking: Artifact or Reality? *Water Research* **41**, 4349-4356.

Sezgin, M. (1982) Variation of Sludge Volume Index with Activated Sludge Characteristics. *Water Research* **16**, 83-88.

Singh, K.P., Basant, N., Malik, A., and Jain, G. (2010) Modeling the performance of "up-flow anaerobic sluge blanket" reactor based wastewater treatment plant using linear and nonlinear approaches – a case study. *Analytica Chimica Acta* **658**, 1-11.

Singh, K.P., Malik, A., Mohan, D., Sinha, S., and Singh, V.K. (2005) Chemometric data analysis of pollutants in wastewater—a case study. *Analytica Chimica Acta* **532**, 15-25.

Sousa, A.C., Lucio, M., Neto, O., Marcone, G., Pereira, A., Dantas, E., Fragoso, W.D., Araújo, M., and Galvão, R. (2007) A method for determination of COD in a domestic wastewater treatment plant by using near-infrared reflectance spectrometry of seston. *Analytica Chimica Acta* **588**, 231-236.

Teppola, P., Mujunen, S.P., Minkkinen, P., Puijola, T., and Pursiheimo, P. (1998) Principal component analysis, contribution plots and feature weights in the monitoring of sequential process data from a paper machine's wet end. *Chemometrics and Intelligent Laboratory Systems* **44**, 307-317.

Umetri, AB (1998) *User's Guide to SIMCA*. CD-ROM.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., and Windig, W. (2004) *PLS_Toolbox Version 3.5 for use with MATLAB™*. Eigenvector Research, Inc..

Woo, S.H., Jeon, C.O., Yun, Y.S., Choi, H., Lee, C.S., and Lee, D.S. (2009) On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant. *Journal of Hazardous Materials* **161**, 538-544.

Abstract

This chapter reports the main conclusions of the thesis and also presents suggestions for future work.

9. CHAPTER 9 – CONCLUSIONS AND SUGGESTIONS

9.1. CONCLUSIONS

Most of wastewater treatment plants are, nowadays, operated with activated sludge systems. However, this system is highly affected by changes on hydraulic and organic loads, nutrients and biomass contents affecting the plant performance, settling ability properties and biomass contents and morphology, among others. As a result of that the activated sludge biomass structure and contents are subject to sudden shifts that are directly dependent on the disturbance type. Recently, the use of image analysis is becoming a widespread procedure capable of monitoring the shifts of aggregated and filamentous bacteria structure and contents. Furthermore, and given the size of the datasets collected by image analysis procedures, the use of chemometric techniques in wastewater treatment monitoring and control has, in recent years, increasingly implemented. As a matter of fact, chemometric techniques allow comprehensive data to be organized in a visual friendly lesser dimensional space, as well as to model variables dependences. This thesis focused developing and optimizing image analysis procedures to determine the biomass contents, structure, viability and Gram status in order to timely detect disturbances, such as filamentous bulking, zoogleal bulking and pinpoint flocs formation, and predict crucial parameters (SVI, TSS, COD, $N\text{-}NO_3^-$) in an activated sludge system.

From the comparison between bright field and phase contrast methods (CHAPTER 3), it was found that, regarding filamentous bacteria detection, bright field acquisition results were found somewhat similar to the phase contrast results in medium to high filamentous contents, but much more accurate regarding low filamentous contents. Regarding the assessment of the aggregated biomass size it was found that the phase contrast methodology clearly overestimated this parameter. Furthermore both techniques revealed to be very robust in terms of the obtained results. In conclusion, bright field microscopy proved to be a more simple and inexpensive method while providing the best overall results.

Another point of interest in image analysis procedures is the dependence on sample dilutions and microscopy magnifications (CHAPTER 4). The study of different dilution factors revealed that the optimal operating dilution must be carefully established, and the determination of the aggregates recognition percentage could be a valuable methodology for identifying the best dilution factor. Furthermore, the biomass structure and contents presented different variations with the dilution, thus implying that each parameter must be studied individually. The use of a 100x magnification was found to cause no accuracy loss regarding biomass representativeness. Furthermore, this magnification allowed for the detection of both aggregated and filamentous bacteria from a single image, and to smaller aggregates to be identified and characterized.

In CHAPTER 5 and CHAPTER 6 the partial least square chemometric technique is used to predict the sludge volume index in eight full-scale activated sludge systems in both normal and filamentous bulking conditions. After the determination of the predicted SVI uncertainty range the classification of the collected samples was found satisfactory inside this range, and quite accurate outside. Furthermore, a model for the prediction of the SVI values was successfully obtained by PLS analysis, and, the obtained results allowed determining a strong relationship between the SVI and free filamentous bacteria contents.

The biomass contents, structure, viability and Gram status were determined by image analysis techniques and used to elucidate several disturbances that may occur in a WWTP, such as filamentous bulking, pinpoint flocs formation and zoogleal bulking, apart from normal conditions, simulated in a lab-scale activated sludge system (CHAPTER 7). Among several other traditional operation parameters, SVI and TSS were also measured and used to correctly classify the collected sample in each of the studied conditions. Furthermore strong correlations were found for the SVI and TSS with the filamentous and aggregated biomass contents, respectively.

The application of chemometric techniques to the biomass contents, structure, viability and Gram status data and reactor operation parameters was discussed in CHAPTER 8. Principal component analysis (PCA) allowed the identification of several disturbances, namely filamentous bulking, pinpoint flocs, zoogleal bulking as well as normal conditions by grouping the collected samples in corresponding clusters. Using partial least squares (PLS), some operational parameters such as SVI, COD, $N\text{-}NO_3^-$, and TSS were predicted. The previous disturbances identification seemed to be crucial for the parameters prediction ability improvement. Furthermore, the association between image analysis and chemometric techniques revealed significant results for activated sludge monitoring and control.

9.2. SUGGESTIONS FOR FUTURE WORK

In this section, some suggestions for future work are presented in order to improve WWTPs characterization based on image analysis and chemometric techniques.

Future work needs to be focused on the enhancement of image analysis procedures through the use of other staining procedures such as Neisser, India Ink, and PHB as well as in FISH techniques in order to identify the predominant bacterial groups. The information provided by such techniques should be further studied with respect to their contribution regarding the detection of operational disturbances, both stand alone as well as combined with morphological and physiological information.

Due to the role played by protozoa in wastewater treatment, new studies should be addressed focusing on improving the recognition of sessile protozoa, and thus contributing for the global characterization of an activated sludge system.

Furthermore, the use of other chemometric techniques, such as decision trees, discriminant analysis and artificial neural networks, should be studied in full-scale WWTPs, integrating morphological, physiological, and operational parameters to allow for an early-warning system for the detection of operating problems.

Finally, as different wastewater types and complexity require a variety of strategies to allow for clean and secure final effluents, the extent of the actual work to other biological wastewater treatment systems should be considered for future work.

SCIENTIFIC OUTPUT

Papers in journals with peer review:

- Mesquita D.P., Dias O., Elias R.A.V., Amaral A.L., Ferreira E.C. (2010) Dilution and magnification effects on image analysis applications in activated sludge characterization. *Microscopy and Microanalysis* 16 (5) 561-568.
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. (2010) A comparison between bright field and phase contrast image analysis techniques in activated sludge morphological characterization. *Microscopy and Microanalysis* 16 (2), 166-174.
- Mesquita D.P., Dias O., Dias A.M.A., Amaral A.L., Ferreira E.C. (2009) Correlation between sludge settleability and image analysis information using partial least squares. *Analytica Chimica Acta* 642 (1-2), 94-101.
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. (2009) Monitoring of activated sludge settling ability through image analysis: validation on full-scale wastewater treatment plants. *Bioprocess and Biosystems Engineering* 32 (3), 361-367.
- Mesquita D.P., Amaral A.L. Ferreira E.C., Coelho M.A.Z. (2009) Study of saline wastewater influence on activated sludge flocs through automated image analysis. *Journal of Chemical Technology & Biotechnology* 84 (4), 554-560.

Publications in conference proceedings:

- Amaral A.L., Mesquita D.P., Ferreira E.C. Distinction of activated sludge bulking phenomena by image analysis. *14º Encontro Nacional de Saneamento Básico, 16 pp, 26-29 October, Porto, Portugal, 2010.*
- Mesquita D.P., Dias O., Dias A.M.A., Amaral A.L., Ferreira E.C. Correlation between sludge settleability and image analysis information using Partial Least Squares. *Proceedings of the 11th International Conference on Chemometrics in Analytical Chemistry, p. 175-179, 30 June - 4 July, Vol. 4, Montpellier, France, 2008.*
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. Bright field versus phase contrast microscopy in activated sludge image acquisition methodologies. *Proceedings of the 10th International Chemical and Biological Engineering Conference, p. 1778-1783, 04-06 September, Braga, Portugal, 2008.*
- Dias O., Mesquita D.P., Amaral A.L. Ferreira E.C. Dilution effects on aggregates and filaments contents in automated image analysis methodologies. *Proceedings CD of the International Symposium on Sanitary and Environmental Engineering, 8 pp, 24-27 June, Florence, Italy, 2008.*
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. Evaluation of activated sludge systems by image analysis procedures. *Proceedings CD of the International Symposium on Sanitary and Environmental Engineering, 8 pp, 24-27 June, Florence, Italy, 2008.*

- Mesquita D.P., Ribeiro R.R., Teixeira D., Ferreira E.C., Coelho M.A.Z. SBR performance for synthetic and fishery wastewater treatment. *4th IWA Specialised Conference on Sequencing Batch Reactor, p. 21-24, 7-10 April, Rome,Italy, 2008.*
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. Relationship between sludge volume index and biomass structure within activated sludge systems. *XVII Brazilian Congress on Chemical Engineering, 7 pp, 14-17 September, Recife, Brazil, 2008.*

Abstracts and Posters in conferences:

- Amaral A.L., Mesquita D.P., Ferreira E.C. Predicting SVI from activated sludge systems in different operating conditions through quantitative image analysis. *2nd Meeting of the Institute for Biotechnology and Bioengineering, 23-24 October, Braga, Portugal, 2010.*
- Amaral A.L., Mesquita D.P., Ferreira E.C. Characterization of activated sludge abnormalities by image analysis and multivariate statistics. *12th International Conference on Chemometrics in Analytical Chemistry, 18-21 October, Atwerpen, Belgium, 2010.*
- Mesquita D.P., Amaral A.L., Ferreira E.C. Monitoring filamentous bulking and *pinpoint* flocs in a lab-scale activated sludge system using image analysis. *Water Research Conference, 11-14 April, Lisbon, Portugal, 2010.*
- Mesquita D.P., Amaral A.L., Ferreira E.C. Improved image analysis procedures for monitoring activated sludge systems with filamentous bulking. *MicroBiotec, 28-30 November, Vilamoura, Portugal, 2009.*
- Mesquita D.P., Louvet J.N., Potier O., Amaral A.L., Pons M.N., Ferreira E.C. Surveying activated sludge changes during acclimation with artificial wastewater. *XIIème Congrès de la Société Française de Génie des Procédés, 14-16 October, Marseille, France, 2009.*
- Mesquita D.P., Amaral A.L., Ferreira E.C. Surveying a wastewater treatment process through image analysis and partial least squares. *Euroanalysis XV, 06-10 September, PO88-A2, Innsbruck, Austria, 2009.*
- Dias A.M.A., Paulo A., Mesquita D.P., Alves M.M., Lopes J.A., Ferreira E.C. Monitoring an activated sludge reactor using IR spectroscopy. *11th International Conference on Chemometrics in Analytical Chemistry, p. 73, June 30-July 4, Montpellier, France, 2008.*
- Dias A.M.A., Paulo A., Mesquita D.P., Alves M.M., Ferreira E.C. Multivariate monitoring of an activated sludge process for biological treatment of a synthetic wastewater effluent. *10th International Chemical and Biological Engineering Conference, p. 735-736, 04-06 September, Braga, Portugal, 2008.*
- Mesquita D.P., Dias O., Amaral A.L., Ferreira E.C. Monitoring the Operation of Activated Sludge Processes by Image Analysis: Validation of Aggregates and Filaments Assessment with Settling Properties. *Micro '07 - Biotec '07 - XXIII JPG, p. 146, 30 November - 02 December, Lisbon, Portugal, 2007.*