Cláudia Filipa Reis Galinha Loureiro

Licenciada em Engenharia Biológica

Monitoring and modelling of membrane bioreactors for wastewater treatment incorporating 2D fluorescence spectroscopy

Dissertação para obtenção do Grau de Doutor em
Engenharia Química e Bioquímica,
Especialidade em Engenharia Bioquímica

Orientador: Maria da Ascensão Miranda Reis, Professora Associada com Agregação, FCT-UNL
Co-orientador: João Goulão Crespo, Professor Catedrático, FCT-UNL
Co-orientador: Gilda Carvalho, Investigadora de Pós-doutoramento, IBET e REQUIMTE

Júri:

Presidente: Doutora Maria Rosa Santos de Paiva

Vogais: Doutor Torove Leiknes
Doutor João Paulo Serejo Goulão Crespo
Doutora Maria da Ascensão Carvalho Fernandes Miranda Reis
Doutor João Pedro Martins de Almeida Lopes
Doutora Gilda de Sousa Carvalho Oehmen
Doutor António Manuel Pedro Martins

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março 2012

Monitoring and modelling of membrane bioreactors for wastewater treatment incorporating 2D fluorescence spectroscopy

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer aos orientadores deste meu trabalho. À São e ao João Paulo pela oportunidade que me deram em trabalhar com os dois. Agradeço imenso o trabalho que me propuseram, e o apoio, optimismo e rigor com que me orientaram ao longo destes anos. À Gilda, mais do que minha orientadora, sempre ao meu lado, desde o primeiro dia; o meu muitíssimo obrigado.

Gostaria também de agradecer às pessoas que me ajudaram a desenvolver o presente trabalho. À Carla Portugal, não só pela ajuda com a fluorescência, mas também pelo indiscutível bom humor. Ao Rui Oliveira, que me iniciou na modelação e aconselhou quando foi preciso. Ao Giuseppe Gugliemi, pela sua imprescindível colaboração neste trabalho, e dedicação, mesmo em tempos mais conturbados.

Agradeço ainda todo o apoio científico e pessoal que recebi dos membros do BioEng, os que estão e os que estiveram. Em especial, à Rita Ricardo, companheira desde o início desta nossa jornada, e com quem aprendi imenso, à Filipa Pardelha e à Rita Moita pela companhia e apoio diários que me têm dado, à Andreia Teixeira pelos momentos de boa disposição, à Ana Lanham, pelo entusiasmo contagiante e motivação constante.

Aos meus pais, que sempre me incentivaram e apoiaram, com infinita paciência. À minha irmã, que muito cedo plantou em mim o "bichinho das ciências".

E finalmente, ao Bruno, que esteve sempre a meu lado e me apoiou incondicionalmente.

Ao Guilherme

Apoio financeiro pela Fundação para a Ciência e Tecnologia, através da bolsa SFRH/BD/30253/2006.

RESUMO

Os biorreactores de membranas (MBRs, 'membrane bioreactors') para o tratamento de águas residuais combinam o processo de lamas activadas com um passo de filtração para obtenção de um efluente limpo, livre de sólidos. Os MBRs representam uma tecnologia em expansão no tratamento de águas residuais sobretudo devido ao reduzido espaço que requerem e à elevada qualidade do efluente obtido. No entanto, a colmatação das membranas pode reduzir o desempenho do MBR. Por este motivo, no presente trabalho, pretendeu-se estudar a monitorização dos MBRs, com o objectivo de minimizar o número de parâmetros de monitorização necessários para descrever o desempenho do processo e obter uma monitorização em tempo real com recurso mínimo a técnicas laboratoriais demoradas. Para este fim, estudou-se a aplicabilidade da fluorescência bidimensional em meios biológicos complexos, tais como as lamas activadas utilizadas para o tratamento de águas residuais. A fluorescência bidimensional mostrou ser uma técnica abrangente, capaz de recolher informação relevante sobre o estado do sistema em tempo real. Devido à complexidade da informação contida nos espectros de fluorescência, usaram-se técnicas de estatística multivariada, tais como análise de componentes principais e projecção de estruturas latentes (PLS, 'projection to latent structures'), para extrair a informação dos espectros e correlacioná-la com parâmetros de operação e de desempenho do MBR. O uso de modelos estatísticos permitiu a previsão de parâmetros chave para o desempenho do MBR usando somente dados de processo impostos ou facilmente adquiríveis em tempo real. Adicionalmente, a modelação estatística foi combinada com um modelo mecanístico, numa estrutura híbrida, de forma a melhorar a previsão mecanística. Este estudo demonstrou ser possível usar modelos PLS para incorporar dados de fluorescência obtidos em tempo real, de modo a melhorar a previsão mecanística sem requerer análises laboratoriais adicionais.

Palavras-chave: Biorreactor de membranas; tratamento de águas residuais; monitorização; fluorescência bidimensional; modelação estatística multivariada; modelação híbrida.

ABSTRACT

Membrane bioreactors (MBRs) for wastewater treatment combine an activated sludge process with a filtration step for solids separation. The application of MBRs for wastewater treatment is growing worldwide due to their compactness and high effluent quality. However, membrane fouling, mostly associated to biologic products, can reduce MBR performance. Therefore, the present study aimed at improving the monitoring of MBRs with simultaneous reduction of the analytical effort. Regarding this objective, the applicability of 2D fluorescence spectroscopy as a monitoring tool in complex biological media, such as activated sludge systems for wastewater treatment, was evaluated. It was shown that 2D fluorescence spectroscopy is a comprehensive technique, able to assess the system status at real-time. Due to the complexity of fluorescence interactions in 2D fluorescence spectroscopy, multivariate statistical analysis, such as principal components analysis (PCA) and projection to latent structures (PLS), was used to extract the information contained in fluorescence spectra and correlate it with operating and performance parameters of an MBR. Through this modelling approach, it was possible to predict key performance parameters of an MBR based only on on-line monitoring data (including 2D fluorescence) or in combination with few additional imposed operating parameters. Additionally, a modelling hybrid approach was developed to improve the predictions of a mechanistic model for MBR performance. It was found that PLS models can be used to incorporate on-line data from 2D fluorescence spectroscopy in order to improve the mechanistic prediction without additional laboratory analysis.

Keywords: Membrane bioreactor; wastewater treatment; monitoring; 2D fluorescence spectroscopy; multivariate statistical modelling; hybrid modelling

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ASM	Activated sludge models
ASM	Activated sludge models
ASM3e	Extension of activated sludge model number 3
BOD_5	5-day biological oxygen demand at five days
BSA	Bovine serum albumin
CAS	Conventional activated sludge
CLSM	Confocal laser scanning microscopy
C*n*	n^{nd} compression component of fluorescence
COD	Chemical oxygen demand
CODfw	Chemical oxygen demand after filtration in wastewater
CODsp	Soluble chemical oxygen demand, after flocculation, in the permeate
CODsw	Soluble chemical oxygen demand, after flocculation, in wastewater
CODtp	Total chemical oxygen demand in the permeate
CODtw	Total chemical oxygen demand in wastewater
CST	Capillary suction time
DO	Dissolved oxygen
EEMs	Excitation-emission matrices
EPS	Extracellular polymeric substances
ESEM	Environmental scanning electron microscopy
FTIR	Fourier transform infrared spectroscopy
FT-NIR	Fourier transform near infrared
HP-SEC	High performance size exclusion chromatography
HRT	Hydraulic retention time
IPW	Iterative predictor weighting
ISE	Iterative stepwise elimination
IWA	International Water Association
Jp	Permeate flux
LV	Latent variables
MLSS	Mixed liquor suspended solids
MLSSb	Mixed liquor suspended solids, acquired on-line, in the biological tank
NH4p	Ammonia in the permeate
NH4w	Ammonia in wastewater
NMR	Nuclear magnetic resonance
NO2p	Nitrite in the permeate
NO2w	Nitrite in wastewater
NO3p	Nitrate in the permeate
NO3w	Nitrate in wastewater
Norgp	Organic nitrogen in the permeate
Norgw	Organic nitrogen in wastewater
PARAFAC	Parallel factor analysis
PC	Principal component
PCA	Principal component analysis

PLS	Projection to latent structures
PO4p	Phosphate in the permeate
PO4w	Phosphate in wastewater
Ptp	Total phosphorus in the permeate
Ptw	Total phosphorus in wastewater
RMSEP	Root mean square error of prediction
SEC	Size exclusion chromatography
SEM	Scanning electron microscopy
SMP	soluble microbial products
SRT	Solid retention time
S_S	Readily biodegradable substrate
SVI	Sludge volume index
T	Temperature
TMP	Transmembrane pressure (mbar)
TOC	Total organic carbon
TSSs	Total suspended solids in the sludge
TSSw	Total suspended solids in wastewater
Vslg/d	Volume of sludge purged per day
VSSs	Volatile suspended in the sludge
VSSw	Volatile suspended solids in wastewater
WWTP	Wastewater treatment plant
X_{STO}	Storage compounds
$Y_{H,S}$	Heterotrophic direct growth on readily biodegradable substrate
$Y_{H,STO}$	Heterotrophic growth on stored substrates
Y_{STO}	Conversion of readily biodegradable substrate into storage compounds
μ_{A20}	maximum growth rate of autotrophs
$\mu_{H,S20}$	maximum growth rate of heterotrophs on readily biodegradable substrate
$\mu_{H,STO20}$	maximum growth rate of heterotrophs on storage compounds

Chapter

1

INTRODUCTION

1.1. MOTIVATION AND WORK OBJECTIVES

Membrane bioreactor (MBR) technology is increasingly applied in wastewater treatment plants, mainly due to their small footprint and high effluent quality. The increasing demand for MBR technology requires the development of adequate monitoring and control techniques, particularly in view of the high operational costs associated to membrane fouling. Extracellular polymeric substances (EPS) are recognised as major fouling agents in MBRs. EPS and cells may deposit and/or adsorb at the membrane surface and within the pores causing fouling. Therefore, monitoring and control during MBR operation is essential. Additionally, adequate and easy operating monitoring tools for MBR will further increase confidence and acceptance of MBR technology, which has been pointed out as one of the main factors influencing the MBR market (Judd, 2006).

Monitoring the performance of an MBR usually involves a high number of time-consuming off-line analytical techniques, due to the complexity of the media. Therefore, MBR technology would greatly benefit from real-time monitoring, able to assess the system status, without further off-line analytic measurements. Furthermore, the development of comprehensive new on-line tools for MBRs monitoring could be used to support immediate control actions.

2D fluorescence spectroscopy is a sensitive, on-line, non-invasive technique that can provide rapid information about the composition of complex biological media. Biological wastewater treatment media contain high quantities of natural fluorophores, such as amino acids (e.g. tyrosine, tryptophan and phenylalanine), vitamins, coenzymes and humic compounds. Furthermore, EPS are composed by large amounts of proteins and aromatic organic substances, making fluorescence a powerful technique to monitor their production in MBRs. Thus, the excitation-emission matrices (EEMs) obtained by scanning a range of spectra wavelengths can cover a wide diversity of natural fluorophores, capturing the physiological activity of a biological system as a fingerprint. 2D fluorescence is thus a promising technique for MBR monitoring, able to capture fingerprinting information about the state of the biological system, including EPS. However, it generates a large amount of complex data that requires mathematical analysis to be fully interpreted.

Multivariate statistical modelling is able to correlate large sets of data, integrating different types of monitoring data, for prediction of performance parameters. The present work is focused on the improvement of MBRs monitoring, through the combination of a comprehensive technique, such as the 2D fluorescence spectroscopy, with the appropriated mathematical tools to extract quantitative information from EEMs.

The main objectives of this thesis can then be defined as: i) monitoring the performance of MBRs for domestic wastewater treatment with minimal off-line analytical effort, ii) use of 2D fluorescence spectroscopy to monitor MBRs under a large range of operating conditions and iii) modelling the performance of a MBR operated at pilot scale for domestic wastewater treatment.

1.2. RESEARCH STRATEGY AND THESIS OUTLINE

The strategy followed in this PhD project involved an initial analysis of the possible interdependencies between common monitoring parameters assessed for a MBR for domestic wastewater treatment in an attempt to reduce redundant analytical parameters. Additionally, 2D fluorescence spectroscopy was investigated regarding the information captured from the main fluorophores, the interferences present in biological wastewater treatment systems and the feasibility of extraction of quantitative information from

fluorescence spectra. Multivariate statistical tools, based on principal component analysis and projection to latent structures, were selected to model, alone or in combination with a simple mechanistic model, key performance parameters of a MBR for domestic wastewater treatment. The final purpose of this work is the development of an effortless and real-time monitoring technique for biological wastewater treatment systems, such as MBRs. The approach used is supported in the belief that 2D fluorescence spectroscopy can be used as a fingerprinting monitoring technique of a MBR and that multivariate statistical analysis are able to extract the quantitative information enclosed in the fluorescence spectra, improving the overall acceptance of MBRs in the wastewater treatment market.

This Thesis is divided into seven chapters following the work performed during this PhD project. Each chapter includes an introduction with a short review of the specific state of the art related with each chapter subject, describes the materials and methods used in that chapter, and discusses the results and main conclusions obtained. The methodology used in each individual chapter is detailed in the context of the respective subject. The work performed during this PhD has resulted in four scientific articles, presented in Chapters 3, 4, 5 and 6, respectively. The articles related to Chapters 3 and 4 are already published in peer reviewed international journal, and the articles related to Chapters 5 and 6 were recently submitted for publication.

Chapter 1 describes the motivation for this PhD project and defines the work objectives. Additionally, Chapter 2 presents the actual state of the art as a context for the work developed in this thesis.

In Chapter 3, 2D fluorescence data obtained from the influent and the permeate of a MBR operated for the treatment of domestic wastewater were successfully modelled using projection to latent structures (PLS) to monitor variations in the influent and effluent total chemical oxygen demand (COD), an indicator of biological performance of the system. However, this approach was not valid for other performance parameters of the MBR system (such as influent and effluent ammonia and phosphorus), which is usually characterised through a high number of analytical and operating parameters. Principal component analysis (PCA) was thus used to find possible correlations between these parameters, in an attempt to reduce the analytical effort required for full MBR

characterisation and to reduce the time frame necessary to obtain monitoring results. This approach alone could not provide robust enough correlations to enable the elimination of parameters for process description. Additionally, it was hypothesised that the information captured by 2D fluorescence spectroscopy could replace some of the analytical and operating parameters, since this technique was able to successfully describe influent and effluent total COD. It was then proposed that a combined modelling of 2D fluorescence data and selected performance/operating parameters should be further explored for efficient MBR monitoring aiming at real-time process control.

In Chapter 4, 2D fluorescence spectroscopy was further investigated in view of the numerous contributions of different compounds in fluorescence spectra. In this chapter it is investigated the occurrence of interference effects (such as quenching and inner filter effects) due to the presence of multiple species in complex biological media, such as natural water matrices, wastewaters and activated sludge. It is shown that the response of fluorescence to a large range of interferences does not represent a problem but a source of information if adequate mathematical tools are used. A statistical multivariate analysis based in a combination of principal component analysis (through the use of PARAFAC function) and PLS modelling is proposed to extract relevant information from 2D fluorescence data. This chapter demonstrates the potential of using 2D fluorescence spectroscopy as a status fingerprint, and how statistical multivariate data analysis can be used to correlate EEMs with selected performance parameters for monitoring of biological systems.

Chapter 5 presents the development of the multivariate statistically-based models, previously defined, for monitoring several key performance parameters of a MBR for wastewater treatment of domestic effluent. PLS modelling was used to integrate 2D fluorescence data, after compression through a PARAFAC function, with operation and analytical data to describe a MBR fouling indicator (transmembrane pressure, TMP), five descriptors of the effluent quality (total COD, soluble COD, nitrite and nitrate concentration, total nitrogen and total phosphorus in the permeate) and the biomass concentration in the bioreactor (MLSS). This study investigated the correlations between inputs and outputs, either through multilinear PLS or by including quadratic and interaction terms of the compressed 2D fluorescence matrices in PLS modelling.

This work demonstrates the applicability of 2D fluorescence and statistically-based models to simultaneously monitor multiple key MBR performance parameters with minimal analytical effort.

In Chapter 6, the MBR performance was modelled using a hybrid approach based on the activated sludge model number 3 (ASM3) combined with PLS to predict the residuals of the ASM. The objective of the modelling strategy used in this chapter was to improve the prediction ability of a plain and easy to implement ASM, with minimal additional monitoring effort. Hybrid models were developed to predict three MBR performance parameters: MLSS, COD in the permeate and nitrite and nitrate concentration in the permeate. PLS modelling of ASM residuals was performed using three different input strategies: 1) analytic and operating data; 2) operating data plus 2D fluorescence spectroscopy; 3) all the data. With the first input strategy, PLS modelling was used to investigate what type of information is missing in the ASM modelling. In the second input strategy, the incorporation of updated data from 2D fluorescence spectroscopy aimed the improvement of the ASM prediction at real-time. Finally, the third input strategy incorporated all the collected data in an attempt to find the best prediction possible of the outputs. In this study demonstrates that 2D fluorescence spectroscopy is a comprehensive monitoring tool, able to capture on-line the required information to complement, through hybrid modelling, the mechanistic information described by an ASM.

In Chapter 7, the main results obtained in this PhD project are summarised, and the main conclusions are discussed. Some possible challenges and suggestions for future research are also presented.

Chapter 2

STATE OF THE ART

2.1. BIOLOGICAL WASTEWATER TREATMENT

Domestic wastewaters are mainly composed by organic compounds that were not completely metabolised in the human body. These compounds, that are rich in nutrients (C, N and P), if discharged in large quantities directly in natural water bodies, are responsible for water pollution and for the well-known phenomena of eutrophication, destroying the natural ecosystems. Additionally, domestic wastewaters are high in pathogenic microorganisms that can contaminate a body of water and spread infectious diseases throughout the population. Therefore, sewage treatment requires reduction of both organic and inorganic nutrients concentrations of the wastewaters and disinfection prior to discharge.

The development of the activated sludge processes for wastewater treatment resulted from the common observation that wastewaters would clarify if left long enough in contact with air. In 1914, Arden and Lockett discovered that the sludge sediment from previous wastewater treatments could be used to supplement fresh untreated wastewater and thus, accelerate the treatment process (Seviour and Nielsen, 2010). The so called activated sludge is composed by a consortium of microorganisms (mainly bacteria and protozoa) that together can degrade and consume the organic and inorganic nutrient present in the domestic wastewaters.

Originally, wastewater treatment systems were conceived to remove only carbonaceous material and ammonia, which is toxic to fish, from domestic wastewaters. This treatment aimed at producing a treated effluent with low levels of organic carbon and

suspended solids, safe for discharge into natural water bodies. Following narrower requirements for nitrogen compounds content in treated wastewaters, treatment plants were progressively modified. Indeed, in the last 3 decades, a significant effort has been done to update plants to remove both nitrogen and phosphorus in addition to organic carbon (Seviour and Nielsen, 2010). Biological processes are preferred for their versatility, low cost and environmental sustainability.

2.1.1 Biological removal of nutrients

In the presence of oxygen, the carbon present in the organic matter is consumed by heterotrophic microorganisms, for biomass growth, with the release of carbon dioxide:

$$C+O_2 \rightarrow \text{biomass} + CO_2$$

Additionally, ammonia is oxidised by autotrophic microorganisms to nitrite and nitrate in the presence of oxygen, in a process called nitrification:

$$NH_4^+ + O_2 \rightarrow NO_2^-$$
$$NO_2^- + O_2 \rightarrow NO_3^-$$

Conversely, in the absence of oxygen other heterotrophic microorganisms (so called denitrifying bacteria) can reduce the NO_3^- sequentially to nitrogen gas (N_2), which is harmless and disperses into the atmosphere:

$$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

In wastewater treatment plants, excess phosphorus is commonly removed by chemical precipitation or biologically, in a process called enhanced biological phosphorus removal (EBPR), by polyphosphate accumulating organisms.

2.1.2. Conventional activated sludge systems

Nowadays, the activated sludge process most used and well studied for domestic wastewater treatment is the conventional activated sludge (CAS) system. The basic design for conventional plants consists in an aerobic reactor with either submerged

diffusers or mechanical surface agitators to provide mixing and aeration, followed by a clarifier where the mixed liquor is separated into sludge and liquid supernatant, by gravity. The settled sludge is then partially recycled to inoculate the incoming raw wastewater. This process allows heterotrophic microorganisms to be in contact with nutrients in the bulk liquid which, in the presence of oxygen, rapidly oxidise the organic compounds, with release of carbon dioxide. Simultaneously, other important communities of microorganisms present in conventional activated sludge systems are able to oxidise compounds like ammonia to nitrite and nitrate (autotrophic microorganisms).

Biological denitrification is achieved in activated sludge systems with the incorporation of an anoxic zone, either preceding or following the aerobic zone. These processes for nitrogen removal are respectively called pre-denitrification and post-denitrification.

Additionally, despite activated sludge processes being able by themselves to eliminate part of the pathogenic microorganisms, disinfection of wastewater in CAS systems is usually performed after biological treatment by chlorine, ultraviolet light or ozone treatment of the clarified effluent.

2.2. MEMBRANE BIOREACTORS

In some wastewater treatment plants the clarifier used in the conventional systems is replaced by membranes, which are very effective to separate the mixed liquor into solids (sludge) and liquid phase. In these processes, called membrane bioreactors (MBRs), the final effluent has lower suspended solids (and then lower turbidity) than the effluent from clarifiers, and also higher effluent quality, since the membranes enable the retention of pathogens.

With the use of membranes to retain suspended solids, MBRs do not depend on the settle ability of the biomass, and thus, all solids have the same residence time. MBRs can thus be operated at high biomass concentration, with hydraulic retention time (HRT) defined independently of the solids retention time (SRT) (Judd, 2006). Therefore, unlike the conventional activated sludge systems, MBRs can easily retain slow growing organisms with poor settling ability, like nitrifying bacteria (Seviour and Nielsen, 2010).

The high quality of the effluent permeate from a MBR permits the direct application of this technology when advanced treatments are required, such as for bathing water, sensitive discharge bodies or water reuse. Additionally, due to the elimination of settlers and due to the possibility of operation with higher biomass concentration in low operating volumes, MBRs have smaller footprints than CAS, which can be valuable when a compact system is needed, such as in areas with high population density.

Despite its advantages, the application of MBR technology for wastewater treatment is still conditioned by the inevitable membrane fouling, by the operational high costs (mostly associated to the aeration of the membrane) (Judd, 2008) and by the complex control systems required (Lesjean et al., 2011).

2.2.1. Configurations of membrane bioreactors

Membrane bioreactors can be generally classified as side-stream or submerged based, according with the membranes placement and operation (Figure 2.1). In side-stream MBRs, membranes are placed externally to the biological reactor and the mixed liquor is pumped into the membrane module. In the membrane module, a permeate stream is generated and the concentrated sludge recycled to the bioreactor. In this configuration, membrane filtration occurs as a typical cross-flow process.

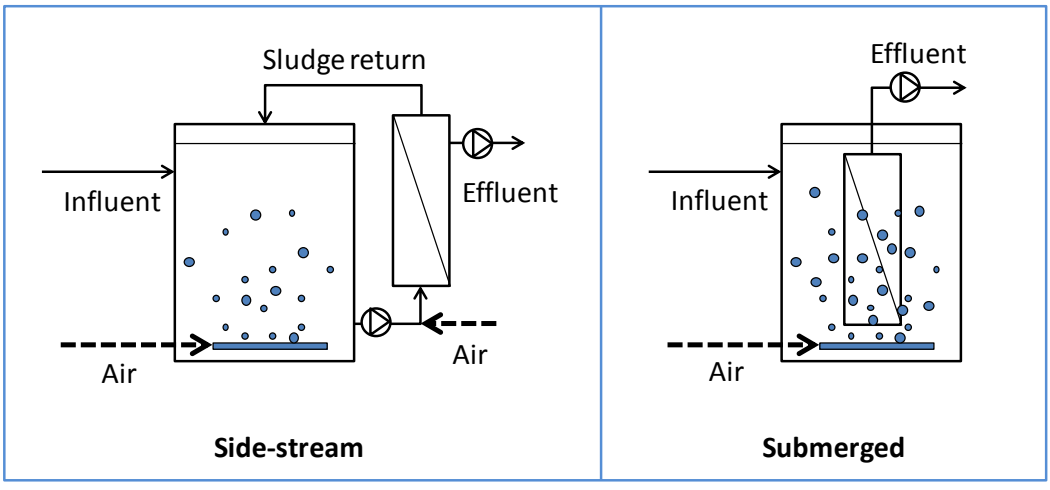

Figure 2.1. Side-stream and submerged MBR configurations.

In submerged MBRs, the membrane module is directly immersed in the mixed liquor, in the aerated bioreactor. This operating strategy was first introduced by Yamamoto et al.

(1989) with the objective of reducing energy consumption associated with the recirculation pump in the side-stream configuration. The submerged MBR configuration corresponds to a dead-end filtration, with shear stress generated by the air bubble flow.

Besides the membrane module placement, the biological compartment of an MBR, like in CAS systems, can be operated in a wide range of bioreactor configurations to achieve specific nutrient removal (e.g. pre- or post-denitrification).

2.2.2. Filtration process in MBRs

Membrane processes depend on both the driving force applied and the membrane characteristics, such as structure and material (Mulder, 1997). Pressure-driven membrane filtration processes are commonly used in water and wastewater treatment systems, and can be divided in four general types, according to the molecular weight cut-off and transmembrane pressure applied: reverse osmosis, nanofiltration, ultrafiltration and microfiltration (Figure 2.2).

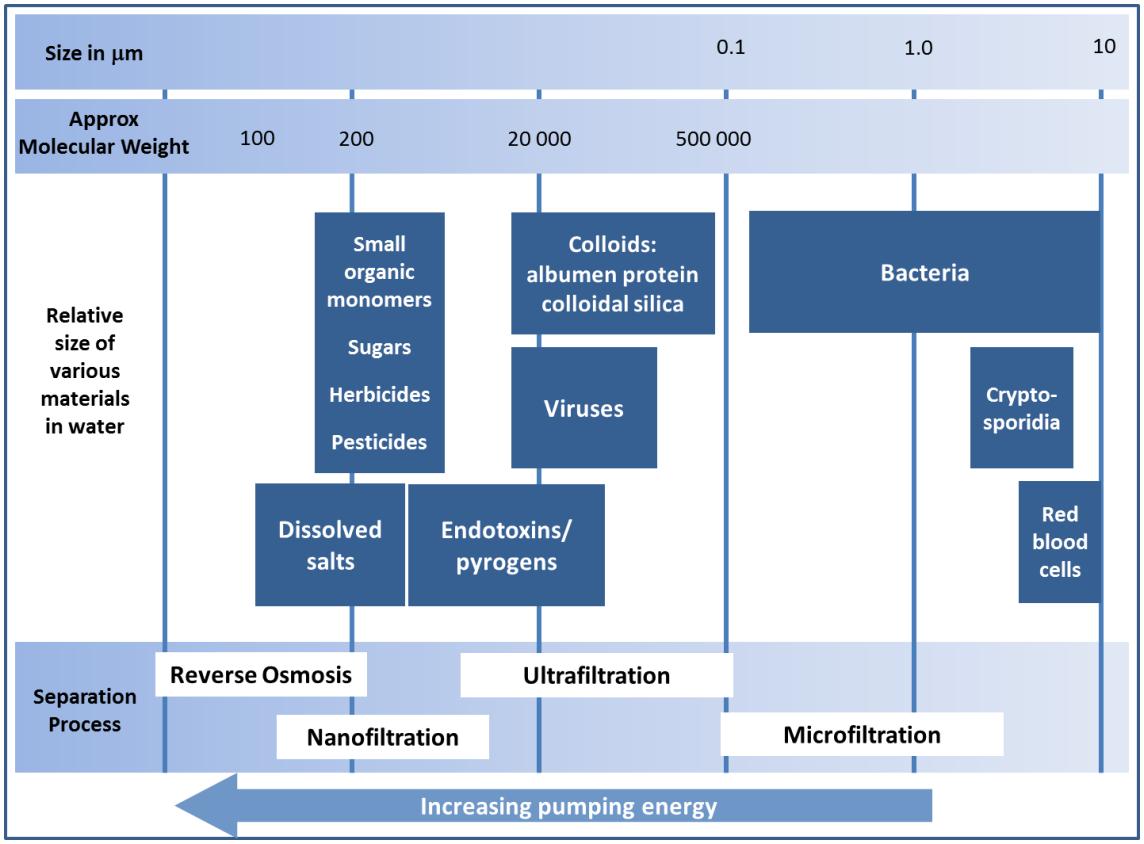

Figure 2.2. Classification of membranes and characteristic retained compounds, based on membrane separation ranges and particle sizes, respectively (adapted from Judd (2006)).

In MBRs for wastewater treatment, either tight microfiltration or loose ultrafiltration membranes are typically used. Since the filtration process determines which components of the mixed liquor can be permeated, the effluent quality of an MBR is also dependent on the membrane used. In fact, polysaccharides, proteins and peptides, colloids, bacteria or virus will cross or be retained by the membrane in accordance with their sizes and the membrane properties (Figure 2.2).

Other membrane characteristics essential for the performance of an MBR are the membrane material and the module configuration. Concerning material, membranes applied in MBRs for domestic wastewater treatment are usually polymeric, with a negative surface charge and hydrophilic. Concerning configuration, the filtration module in MBRs is normally composed by either flat-sheet membranes or hollow-fibre membranes. While submerged MBRs hollow-fibre systems are operated with an outside-inside flux direction, in side-stream configurations hollow-fibre membranes can be operated with either an outside-inside or an inside-outside flux direction. However, in the later flux operation mode, the hollow-fibre membranes used have higher internal diameter (also called tubular membranes) (Judd, 2006).

Fouling

Despite the evident advantages of MBRs for wastewater treatment, the main problem associated with their use is membrane fouling. Membrane fouling refers to the progressive deposition and/or adsorption of material on the membrane surface, or within the membrane structure, resulting in a gradual loss of membrane permeability.

Mixed liquors are a complex mixture of variable amounts of particulate, colloidal and dissolved fractions, all of which containing potential foulants. In MBRs, fouling development is mainly due to the adsorption of colloidal material at the membrane surface and deposition of flocs (composed by cells and organic and inorganic compounds). This type of fouling is usually the predominant fouling component in MBRs, though it is easily removed by physical means (Meng et al., 2009).

Intrapore adsorption occurs in MBRs mainly due to the adsorption of organic matter present in the mixed liquor, either colloidal or soluble. These organic compounds result both from incoming wastewaters (such as humic compounds) and from microbial

activity (such as polysaccharides and proteins). Despite their diverse origin, these compounds are generally termed extracellular polymeric substances (EPS) and are classified into bound EPS, when they are attached to biomass to form flocs, and soluble EPS (also called soluble microbial products, SMP), when freely suspended.

Additionally, at the surface of membranes there is also the adhesion and growth of microorganisms, forming a biofilm, and the precipitation of some salts. The attached deposit of cells, cell debris and suspended materials, are not truly discernible and all together form a more or less bound fouling layer at the membrane surface. Indeed, all mechanisms of fouling can occur simultaneously forming complex forms of fouling with increase difficulty of cleaning (Poele and van der Graaf, 2005).

According to Meng et al. (2009), membrane fouling can be classified in three types of fouling, based on their persistence to cleaning: removable fouling, irremovable fouling and irreversible fouling. Removable fouling is the fouling easily removed by physical cleaning (e.g. backwash), whereas the elimination of irremovable fouling is only achieved with chemical cleaning. Removable fouling is caused by loosely attached compounds and, in general, corresponds to the cake layer. Irremovable fouling is attributed to pore adsorption and strongly attached foulants. Irreversible fouling is a permanent fouling that cannot be removed by cleaning, and will eventually lead to the need of replacing the membranes.

Fouling control is, then, essential in MBRs to maintain high membrane permeability and minimise operational costs associated with the system performance and membranes lifespan. To minimise fouling, the major aspects to take into account during the operation of an MBR are: i) the influent characteristics; ii) the microbial population and activity; iii) the permeate flux imposed; iv) the shear at the membranes surface; v) and membrane cleaning strategies. Concerning the influent wastewater, the concentration, composition, size and degradability of the organic matter can have a direct impact in MBR fouling. In fact, fouling due to natural organic matter, mainly humic compounds, is a well-known problem in the filtration of natural waters, due to their ability to adsorb on membranes (Jones and O'Melia, 2000; Lee et al., 2004; Yuan and Zydney, 1999). Furthermore, the degradation of organic matter by microbial population can result in

several intermediate compounds and microbial sub-products which can act as fouling agents as well.

Regarding the MBR operating conditions, the imposed permeate flux is probably the most important parameter in fouling control. Thus, membrane bioreactors are usually operated at a low permeate flux (below critical flux) to avoid the deposition of particles on the membrane surface and keep permeability for a long operating time. The concept of critical flux was introduced by Field et al. (1995) and it is defined as the flux where the forces linked to filtration pressure and shearing forces are balanced. In MBR systems, the critical flux value depends on the characteristics of the membrane (e.g. pore diameter and material), characteristics of the mixed liquor, shear forces at the membrane surface and temperature. The application of shear forces at the surface of the membranes is a current operational strategy to hamper fouling formation, either through the use of coarse air bubbles in submerged configurations or through increased crossflow velocity (with or without addition of air) in side-stream configurations (Judd, 2006).

Additionally to the strategies used to prevent fouling, membrane cleaning, either physical or chemical, is an essential step in MBRs operation and membrane maintenance. The physical cleaning strategies predominantly used in MBRs are relaxation and backwash. In relaxation, permeation is interrupted for short periods of time, while continuing to scour the membranes with coarse air bubbles, to allow the detachment of the fouling agents from membranes surface. In backwash, the permeate flux is periodically inverted for few seconds to compel the release of loosely bound compounds from membrane pores. While backwash might be more effective, its application is limited to robust membranes (usually hollow-fibres) (Judd, 2006). These two techniques may be used in combination, and backwash may be enhanced by adding air to the backwash flow, increasing the shear (Judd, 2006). Chemical cleaning is the strongest form of cleaning and it is used when fouling cannot be removed physically. Chemical cleaning is generally carried out with mineral or organic acids, sodium hydroxide or, more usual in MBRs, sodium hypochlorite, and can be performed either *in situ* or *ex situ* (Judd, 2006). However, chemical cleaning has a negative impact on the lifespan of the membrane due to the effect of free radicals.

2.3. MONITORING

2.3.1. Conventional monitoring of wastewater treatment plants

In wastewater treatment systems, monitoring of the biological process and of key effluent parameters is essential to achieve conformity with quality and safety requirements. Usual characterisation of influent wastewater in activated sludge systems includes the assessment of suspended solids, 5-day biochemical oxygen demand (BOD_5), total organic carbon (TOC), chemical oxygen demand (COD), nitrogen (as ammonia, total nitrogen, organic nitrogen, nitrate and nitrite), phosphorus (as orthophosphate and total phosphorus), sulphate, alkalinity, greases and coliform bacteria (as an indicator of pathogenic organisms). BOD_5, COD and TOC are used to assess the global organic content of wastewaters, regardless of their composition. However, while BOD_5 is a standard method used to evaluate the oxygen necessary for the oxidation of the biodegradable organic compounds (during 5 days), COD assesses all organic compounds able to be oxidised, without distinction on biodegradability. Nevertheless, fractionation of wastewater COD into readily biodegradable, slowly biodegradable or inert is also actually performed through respirometric batch tests for biodegradability assessment (Sperandio and Etienne, 2000). In addition, TOC is an easy and rapid measurement of the organic carbon, but does not give information on their oxidation state. Although these analytic techniques are used in wastewater treatment plants for a better characterisation of organic matter, frequent determinations for monitoring purpose are usually performed only by COD or TOC, which are quicker and easier techniques then BOD_5.

Additionally to wastewater characterisation, the following water quality parameters of the effluent are typically monitored: suspended solids, coliform bacteria, nutrients (mainly organic compounds, nitrogen, phosphorus and sulphate), pH, and toxic compounds. However, in the last decades, due to higher environmental requirements and increased knowledge on the chemistry and microbiology of wastewater treatment, additional and more specific monitoring analysis of wastewater treatment systems are being claimed (Metcalf and Eddy, 1991).

2.3.2. Monitoring membrane bioreactors

Membrane bioreactors for wastewater treatment, like conventional activated sludge systems, require monitoring of both influent and effluent streams, as well as frequent

assessment of the biological activity. Moreover, tight monitoring of MBR systems is critical, not only because of the quality of the final treated effluent, but also due to the fouling potential of the complex biological media. Indeed, several techniques are being developed and tested for the evaluation and characterisation of fouling and fouling agents in MBRs.

Total COD measurements of wastewater and permeate are easy to perform but not sufficient to characterise the organic compounds fed to the MBR, that remain in the permeate after treatment, in terms of biodegradability and composition. Additionally, in MBRs, large colloidal organic compounds are retained by the membrane and can either be returned to the bioreactor bulk liquor or deposited on the membrane causing fouling.

Several studies pointed natural organic matter from influent wastewater as a major fouling agent (Jones and O'Melia, 2000; Yuan and Zydney, 1999). Furthermore, the biological performance of an MBR also depends on the characteristics of the influent organic carbon, which includes soluble simple organic molecules with low molecular weight, and colloidal organic compounds present in the soluble COD. A physical-chemical method to distinguish between solutes and colloids in the soluble fraction of COD (usually obtained by filtration) was proposed by Mamais et al. (1993). This method uses zinc sulphate to flocculate colloidal material in a wastewater sample before filtration, and thus obtain a soluble COD without the colloids.

Transmembrane pressure (TMP), defined as the difference of pressure between the feed and the permeate sides of the membrane, and permeate flux are typically monitored on-line at full-scale plant. MBRs for wastewater treatment are usually operated at constant permeate flow (Drews, 2010), and thus, TMP is used as a fouling indicator parameter. However, TMP increase does not distinguish the dominant fouling mechanism or the upcoming of extreme fouling (observed as a TMP jump). Therefore, various methods and techniques to evaluate and monitor fouling with more detail were developed.

To assess the physical properties of fouling and potential foulants, some authors developed filtration tests to determine the filtration resistances of fouling (Chu and Li, 2005) and of suspended solids, solutes and colloids in mixed liquor (Bouhabila et al., 2001). Additionally, the use of filtration devices has also been used to characterise the

fouling potential of mixed liquor as reversible or irreversible (Huyskens et al., 2008). Capillary suction time (CST) and sludge volume index (SVI) were also used in the past as sludge characteristics that would be directly linked to its fouling potential, although they have been shown to be poor measures of filterability (Drews, 2010).

Still concerning physical characteristics of fouling, the exact location of fouling formation on the membrane module (knowing exactly which membrane or location of the module affected) and at the membrane level (at the surface or inside the pores) was also previously studied through model-based strategies. These strategies, aiming at the recognition of the underlying mechanism of fouling, included either the estimation of permeate flow distribution in membrane modules (Wicaksana et al., 2009) or the use of filtration models to describe filtration and fouling mechanisms (Drews et al., 2009).

Concerning chemical analysis of fouling components, several techniques have been reported for characterisation of both the foulants, effectively deposited on the membrane, and potential fouling agents found in the bulk liquor. EPS, mainly composed of proteins, polysaccharides, lipids, nucleic acids and humic compounds, were previously found to play a major role in membrane fouling of MBRs (Judd, 2008; Le-Clech et al., 2006; Meng et al., 2009). Besides being the most abundant EPS components, proteins and polysaccharides are usually assumed to be the EPS major contributors for fouling (Drews, 2010), therefore, the evaluation of EPS concentration relies almost exclusively on their measurement. Photometric methods are generally used to assess proteins (Lowry et al., 1951) and polysaccharides (Dubois et al., 1956). Due to the large importance of proteins and polysaccharides in MBRs, Mehrez et al. (2007) developed an automated version of Lowry and Dubois assays using an at-line sensor to monitor soluble EPS. Despite the large use of photometric methods for determination of both proteins and polysaccharides, their application has several interferences, thus, correction methods are also often applied. For protein assessment, (Frolund et al., 1995) proposed a modified Lowry method to subtract humic compounds contribution, while the interferences in the Dubois method for polysaccharide determination can be corrected with the concentrations of nitrate and nitrite in the sample (Drews et al., 2007).

Soluble EPS (also referred as SMP) is usually obtained by centrifugation (e.g. Sperandio and Espinosa, 2008) or filtration (e.g. Rosenberger et al., 2006) of the mixed liquor, whereas several methods for the extraction of bound EPS have been applied, without consensus. The most common methods for bound EPS extraction found in literature include the cation exchange resin method (Frolund et al., 1996; Rosenberger et al., 2006) and heating methods (Ng and Ng, 2010; Wang et al., 2009a), however, several other methods have been used, such as formaldehyde and NaOH methods or combinations of them (Liu and Fang, 2002).

Despite the general recognition of EPS as major fouling agents, there is still no consensus about the role played by each of the EPS fractions and by their components (proteins and polysaccharides) on fouling and on factors governing their occurrence (Drews, 2010). However, great variety of samples preparation, analytical methods, system configurations and operating conditions are simultaneously used, which could be partially responsible for the discrepancy in results (Drews, 2010).

In addition, for determination of EPS concentrations (either soluble or bound) some authors measured the organic carbon (e.g. Dong and Jiang, 2009; Lyko et al., 2008) or the COD (e.g. Jiang et al., 2008) of EPS samples, in substitution or in combination with the proteins and polysaccharides assessment. EPS quantification through the measurements of organic carbon and COD do not distinguish the EPS composition, however they were found to be an alternative to more complex and costly measurements of EPS components to characterise sludge (Lyko et al., 2008).

Chromatographic methods have been applied to characterise the fouling agents after a separation step. Size exclusion chromatography (SEC) has been used for characterisation of EPS constituents based on molecular size, with different detectors, such as UV (e.g. Her et al., 2003; Lyko et al., 2008; Rosenberger et al., 2006), organic carbon (e.g. Her et al., 2003; Rosenberger et al., 2006) or fluorescence (Her et al., 2003). In general, the chromatograms obtained were able to separate peaks with distinguished molecular weight, corresponding to polysaccharides, some proteins, colloids and humic compounds.

In addition, characterisation of membrane organic fouling has also been performed by Fourier transform infrared (FTIR) spectroscopy (e.g. Kimura et al., 2009; Meng et al., 2007; Wang et al., 2009b) and by ^{13}C-nuclear magnetic resonance (NMR) spectroscopy (e.g. Kimura et al., 2009; Meng et al., 2011), to identify functional groups of organic molecules generally accepted as fouling agents. Further evaluation of the proteins present in EPS, by their identification, has recently been adapted through gel electrophoresis (Kuhn et al., 2011; Silva et al., 2011).

Microscopic structural properties of membrane fouling have been also assessed using confocal laser scanning microscopy (CLSM) (Bjorkoy and Fiksdal, 2009; Chu and Li, 2005; Meng et al., 2007; Ng and Ng, 2010) and scanning electron microscopy (SEM) (Chu and Li, 2005; Meng et al., 2007). These techniques assessed the presence of bacterial cells attached to membrane and their interaction with biopolymer foulants. In this context, Le-Clech et al (2007) compared environmental scanning electron microscopy (ESEM), confocal laser scanning microscopy and direct observation of fouling. These authors concluded that, despite CLSM ability to differentiate between the different types of foulants, the direct observation of fouling (using a specially designed microscope-based installation (Li et al., 1998) appeared to be the most promising technique for direct and *in situ* observation of MBR fouling. However, its use is limited to optically accessible systems such as dilute suspensions or single fibres (Drews, 2010).

In addition to the monitoring techniques targeting specific compounds in the MBR, fingerprinting methods, able to assess the system status, have also been studied for characterisation of such complex biological systems, e.g. Fourier transform near infrared (FT-NIR) (Reed et al., 2011) and 2D fluorescence spectroscopy (Wolf et al., 2007; Wolf et al., 2001). Indeed, spectroscopic methods can be used as fingerprint techniques providing large sets of data (the spectra) from which meaningful information could be extracted (Pons et al., 2004).

Two-dimensional fluorescence spectroscopy

The use of fluorescence spectroscopy in such complex systems explores the natural fluorescence of several compounds typically present on biological media. When different compounds, with different excitation/emission wavelengths, are present, it is

possible to assess all compounds by simultaneously scanning a range of excitation and emission wavelengths. Natural fluorophores include proteins, and a large range of other organic compounds, as well as some vitamins, co-factors and NADH. Additionally, fluorescence spectra are sensitive to several other factors, such as the presence of salts, temperature and medium turbidity (Lakowicz, 1983).

Further analysis of the current state of the art on the application of fluorescence spectroscopy to biological systems is explored in Chapter 4, as well as the advantages and disadvantages of the application of 2D fluorescence spectroscopy as a monitoring tool in wastewater treatment systems.

In view of the complexity of MBR systems for wastewater treatment and the wide range of monitoring techniques available, the development of mathematic methods to integrate and correlate different (and disperse) types of information would greatly benefit the comprehension and monitoring of these systems. Therefore, some modelling strategies have been used, such as mechanistic modelling of MBR performance. However, non-mechanistic modelling tools have greater potential to integrate, analyse and correlate different types of information.

2.4. MODELLING MBRS

2.4.1. Activated sludge model

The activated sludge system has been widely studied in the past, resulting in deep understanding of the kinetics of the main heterotrophic and autotrophic biological processes, which set the basis for the development of mechanistic models. These kinetic models have been synthesised in four activated sludge models (ASM) by the International Water Association (IWA) Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment (Henze et al., 2000).

The first activated sludge model published, ASM1, was developed to model biological treatment for organic carbon removal, nitrification and denitrification. This model is able to predict oxygen demand and sludge production in activated sludge systems.

ASM2 was developed later to incorporate phosphorus removal from wastewater, and ASM2d to account for the ability of phosphorus-accumulating organisms to use cell internal substrates for denitrification.

ASM3 does not include phosphorus removal, but it was established to addresses problems found in the first model ASM1, such as the inclusion of internal cell storage compounds in heterotrophs (shifting the focus from hydrolysis to the storage of organic substrates) and the replacement of the death-regeneration concept by the growth-endogenous respiration model.

Due to the similarity between the biological processes occurring in conventional activated sludge systems and in MBRs, ASM can be used for modelling the biological removal of nutrients in a MBR. Nevertheless, some differences between the two systems have to be taken into account when applying an ASM to model the biological performance of an MBR, for instance, the elimination of sludge in the effluent with the replacement of the clarifier by a membrane, and the additional supply of oxygen for aeration of membrane modules.

According with Fenu et al. (2010), the application of the ASMs to the MBR processes can be divided into unmodified and modified ASMs in view of the adaptations made to the ASM to fit an MBR. The expression unmodified or plain encloses ASM applications with modelling objectives similar to those originally stipulated for the CAS systems (process design, effluent characterisation, oxygen demand and sludge production) and without modification of the ASM model structure. However, it includes slight modifications to the biokinetic processes. Modified ASMs extend the original ASM in terms of biokinetic process models, namely the so-called SMP and EPS models. While unmodified ASMs are usually more readily applicable in practice, modified ASMs are developed in academic work to improve process understanding.

Despite the applicability of plain ASM to model the MBR biological performance, such models do not link the biological process of an MBR with membrane fouling. Therefore, modifications to the ASM have been developed to include models for EPS (both bound and soluble) and SMP (defined here as the soluble microbial products resulting from cell lysis) description in MBRs (Jiang et al., 2008; Laspidou and

Rittmann, 2002). Despite the good predictions obtained in these studies, they are complex, involve the calibration of several model parameters and require frequent input data about the feed characteristics.

2.4.2. Statistically-based tools

Principal component analysis

The objective of principal component analysis (PCA) is to substitute the representation of objects in its initial representation into the new Principal Component coordinate space. PCA results in a new co-ordinate system with reduced noise and lower dimensionality through decomposition of a data matrix, X, into a "structure" part plus a "noise" part.

$$X=CP^{T}+E$$

C are the scores matrix, and have as many rows as the original data matrix, P^{T} are the loadings matrix, and have as many columns as the original data matrix. The scores matrix can be seen as the representation of the initial data in the new and reduced co-ordinate system, composed by the new components (Principal Components, PC), while the loadings describe the 'distance' between the initial co-ordinate system and the PC system. The E matrix contains unexplained data variance, such as co-linearity and noise.

Principal component analysis can be used either as a qualitative data analysis tool, through the analysis of loadings plots, or for compression of the number of parameters needed to describe spectroscopic data. PARAFAC (parallel factor analysis) function can be used as a principal component analysis computational tool, to find the scores matrices for 2D fluorescence spectra, which include the most relevant information from original data, but with reduced dimension (Andersson and Bro, 2000; Bro, 1997).

Projection to latent structures regression

Projection to latent structures (PLS) is a non-parametric model that reveals linear relations between the data, by maximising the covariance between the input matrix X, and the output Y. This technique combines features from the principal component analysis and multiple linear regression, and aims at the prediction of dependent variables by decomposing iteratively both the X and Y matrices into reduced orthogonal

factors, termed latent variables. Therefore, PLS regression differs from traditional multivariate regression due to elimination of redundancy in the input and output data. PLS is considered a simple but powerful predictive modelling technique due to its ability to handle co-linearity between variables, data noise and missing data (Wold et al., 2001).

In PLS regression methodology, the input matrix, X, is decomposed as the product of the matrix of scores (T) and the matrix of loadings (P^T) that minimises the residuals E.

$$X = TP^T + E$$

Similarly, the product of T and C^T estimates the output Y, where C is the Y-weights matrix and F is the error term.

$$Y = TC^T + F$$

Therefore, the scores matrix, T, enclosing the new variables obtained from the latent variables, is estimated in a way that is simultaneously able to describe the original X and is a good predictor of Y. The weights matrix W (defined by a linear correlation of T with the original data, T=XW) quantifies the relation between X and Y, therefore, it can be used to identify the important variables to the output. Finally, a multivariate regression model is obtained as follows:

$$Y = XWC^T + F = XB + F$$

where the regression coefficients are given by:

$$B = WC^T$$

Since the linear PLS model finds a new data arrangement, it is possible to determine and interpret the contribution of the input parameters to the model. A large numerical value of B is highly correlated with Y and similar profiles of B-values provide the same contribution to the prediction (Wold et al., 2001).

Using PLS modelling is possible to correlate the information contained in 2D fluorescence spectra with a quantitative output parameter, and thus extract qualitative information from fluorescence spectra.

Chapter

3

REAL-TIME MONITORING OF MEMBRANE BIOREACTORS WITH 2D FLUORESCENCE DATA AND STATISTICALLY-BASED MODELS

SUMMARY

The application of membrane bioreactors (MBR) for wastewater treatment is growing worldwide due to their compactness and high effluent quality. However, membrane fouling, mostly associated to biological products, can reduce MBR performance. Therefore, it is important to monitor MBRs as close to real-time as possible to accelerate control actions for maximal biological and membrane performance. 2D fluorescence spectroscopy is a promising on-line tool to simultaneously monitor wastewater treatment efficiency and the formation of potential biological fouling agents. In this study, 2D fluorescence data obtained from the wastewater and the permeate of a MBR was successfully modelled using projection to latent structures (PLS) to monitor variations in the influent and effluent total chemical oxygen demand (COD). Analysis of the results also indicated that humic acids and proteins highly contributed to the measured COD in both streams. Nevertheless, this approach was not valid for other performance parameters of the MBR system (such as influent and effluent ammonia and phosphorus), which is usually characterised through a high number of analytical and operating parameters. Principal component analysis (PCA) was thus used to find possible correlations between these parameters, in an attempt to reduce the analytical effort required for full MBR characterisation and to reduce the time frame necessary to obtain monitoring results. The 3 first principal components, capturing 57% of the variance, indicated and confirmed expected relationships between the assessed parameters. However, this approach alone could not provide robust enough correlations to enable the elimination of parameters for process description (PCA loadings ≤ 0.5). Nevertheless, it is possible that the information captured by 2D fluorescence spectroscopy could replace some of the analytical and operating parameters, since this technique was able to successfully describe influent and effluent total COD. It is thus proposed that combined modelling of 2D fluorescence data and selected performance/operating parameters should be further explored for efficient MBR monitoring aiming at rapid process control.

*Published as: Galinha, C.F., Carvalho, G., Portugal, C.A.M., Guglielmi, G., Oliveira, R., Crespo, J.G. and Reis, M.A.M., 2011. Real-time monitoring of membrane bioreactors with 2D-fluorescence data and statistically based models. Water Science and Technology, **63**, 1381-1388.*[1]

[1]Reproduced with the authorization of the editor and subjected to the copyrights imposed.

3.1. INTRODUCTION

Membrane bioreactors (MBR) are increasingly applied in wastewater treatment plants, mainly due to their small footprint and high effluent quality. The increasing demand for MBR technology requires the development of adequate monitoring and control techniques, particularly in view of the high operational costs associated to membrane fouling resulting from the adhesion of cells and cell products. Monitoring of MBR performance involves a high number of off-line and time-consuming analytical techniques, regarding the complexity of the media. Therefore, MBR technology would greatly benefit from real-time monitoring techniques that could be used to support immediate control actions.

2D fluorescence spectroscopy is an on-line and non-destructive technique that can quickly provide information about the composition of complex biological media and consequently be used as a real-time monitoring tool. Wastewater media contain high quantities of natural fluorophores, such as aminoacids (e.g. tyrosine, tryptophan and phenylalanine), vitamins, coenzymes and aromatic organic matter in general. Furthermore, extracellular polymeric substances (EPS) containing large amounts of proteins, are the major fouling agent of MBRs. Thus, fluorescence is a good candidate technique for MBR monitoring, able to capture fingerprinting information on the state of the biological media, including EPS.

Fluorescence spectroscopy, using selected excitation/emission wavelengths, was first explored by Li et al. (1991) and Li and Humphrey (1991) for monitoring cell growth and activity in biological reactors. Later, other studies have applied 2D fluorescence spectroscopy to monitor water and wastewater treatment processes (Her et al., 2003; Kimura et al., 2009; Lee et al., 2006; Wang et al., 2009b), and the need to extract deeper, quantitative information from 2D fluorescence spectra obtained from high complex media guided other authors to use multivariate statistical tools (Boehl et al., 2003; Ganzlin et al., 2007; Surribas et al., 2006; Teixeira et al., 2009; Wolf et al., 2001).

In wastewater treatment plants (WWTPs), media are highly complex, generally composed by a wide variety of molecular species (fluorophores and non-fluorophores) which may promote mutual interference effects on the fluorescence signal. Consequently fluorescence excitation-emission matrices (EEMs) obtained from these

systems contain highly embedded information. In this study, the information contained in fluorescence spectra was assessed using projection to latent structures (PLS) aiming at establishing correlations with performance parameters of a MBR.

PLS is a linear regression method that maximises the covariance between input data and output performance parameters. PLS models generate a set of latent variables that explain the maximum variance in the output variables, combining in a single step data decomposition and correlation with predicted outputs. PLS were used in this work to predict quality parameters (such as chemical oxygen demand, COD) of wastewater and permeate based only on 2D fluorescence data. Moreover, the importance of specific areas of the fluorescence matrices was investigated for the prediction of each output.

In MBRs for wastewater treatment the relationships between operating and performance parameters are abundant and complex, essentially due to the interdependency between operating parameters, biological performance and membrane performance. This complexity was assessed in this work by principal component analysis (PCA) (Rencher, 2002) of analytical parameters conventionally used to monitor MBRs for wastewater treatment.

PCA is often used to examine interrelationships between a large number of variables and to explain these variables in terms of their common underlying dimensions (Hair et al., 1998). In PCA, the input matrix is described as a linear correlation between scores and loadings that minimises the residuals. The best linear combination of variables is determined in order to capture the maximum variance in data. So, the first principal component (PC1) can be seen as the best summary of linear relationships present within data. The second principal component (PC2) is identified as orthogonal to the first one, and aims to find the best relationship for the remaining variance. The following PCs are all orthogonal between each other. The interpretation of the contribution of each variable in the PCA is quantified by the loadings matrix. The loadings represent the degree of correlation between the variables and the principal components. Therefore similar values represent variables with a high correspondence as well as higher values of loadings point to a more representative variable (Jackson, 2003). In this work, correlations across data were analysed based on loadings resultant from PCA applied to a comprehensive set of MBR operating and performance parameters.

The aim of this work was the development of a strategy for real-time monitoring of a MBR through the use of an on-line method based on 2D fluorescence spectroscopy data and PLS modelling. Alternatively, this study also evaluated the reduction of the number of MBR monitoring parameters by multivariate data analysis that would enable the elimination of redundant analysis.

3.2. METHODS

3.2.1. Membrane bioreactor

The pilot scale MBR was operated to treat domestic wastewater and monitored with 2D fluorescence spectroscopy. It was located at the wastewater treatment plant of Lavis, Italy and consisted of a biological anoxic/aerobic tank followed by a separate tank with a submerged hollow fibre system (GE Zenon ZW500d) with 0.04 μm membrane pore size (Figure 3.1). This pilot MBR was monitored with 2D fluorescence for a period of 10 months, when it was operated under controlled permeate flux. During this period, operational changes were programmed and imposed in the permeate flux and solids retention time. Temperature changed due to seasons' weather, hydraulic retention time (HRT) and dissolved oxygen (DO) changed due to other operating and control experiments.

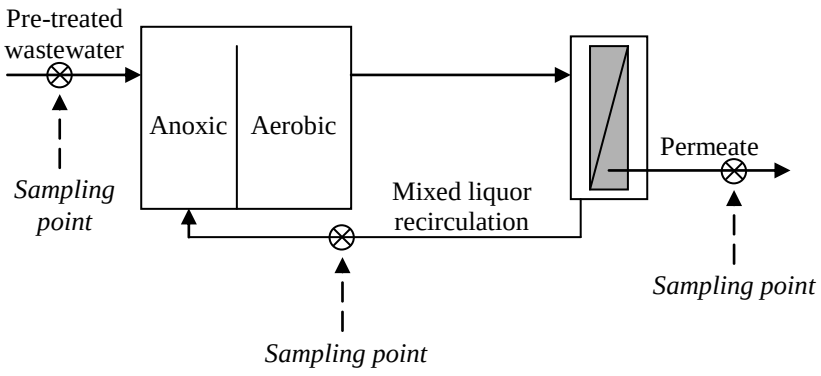

Figure 3.1. Representation of the pilot MBR located in at the Lavis wastewater treatment plant (Trento, Italy).

3.2.2. 2D fluorescence spectroscopy

Fluorescence EEMs were acquired with a fluorescence spectrophotometer Varian Cary Eclipse coupled to a fluorescence optical fibre bundle probe. Fluorescence spectra were generated in a range of 250 to 700 nm (excitation) and 260 to 710 nm (emission), with

an excitation wavelength incrementing step of 10 nm. Fluorescence spectra were obtained using excitation and emission slits of 10 nm and a scan speed of 3000 nm/min.

3.2.3. Data collection

2D fluorescence spectra were acquired in the wastewater feed, in the bulk activated sludge and in the permeate at the same time that samples were collected for further analysis of wastewater, permeate and mixed liquor (Table 3.1). Also transmembrane pressure (TMP), temperature and dissolved oxygen were measured on-line. All these data collected together with selected operating parameters – permeate flux, hydraulic retention time and sludge retention time – were used in this work to find correlations by multivariate analysis.

3.2.4. Multivariate data analysis

Before multivariate data analysis, all data was normalised by subtracting the respective average values and dividing by their standard deviations. The 2D fluorescence spectroscopy measurements were acquired and plotted in excitation-emission matrices (EEMs) where each value of fluorescence intensity corresponds to a pair of excitation/emission wavelengths, totalising 5490 model input variables. The mathematical models were obtained through PLS regression, using EEMs of wastewater and permeate as model inputs and total COD in wastewater and permeate as outputs, respectively. Data from 146 observations obtained throughout the 10 months of operation were used for PLS modelling. These 146 observations were divided randomly into a training set (75% of the observations, which were used to calibrate the model) and a validation set (25% of the observations, which were used to validate the final model). The PLS models thus obtained are a linear correlation of the 5490 fluorescence inputs to predict total COD in the wastewater and in the permeate accordingly. Model fitting to the experimental data was assessed by the training and validation correlation coefficients (R2) and root mean square error of prediction (RMSEP), calculated as the squared root of the sum of the squared differences between predicted and experimental values.

Table 3.1. Monitoring parameters and respective range of values assessed during MBR operation.

	Abbreviation	Range		
		Min.	Max.	average
Process and Performance Parameters:				
Total chemical oxygen demand in wastewater (mg L^{-1})	CODtw	151	1862	524
Soluble chemical oxygen demand in wastewater obtained by sample flocculation (mg L^{-1})	CODsw	17	410	108
Soluble chemical oxygen demand in wastewater obtained by sample filtration (mg L^{-1})	CODfw	46	467	155
Ammonia in wastewater (mg N L^{-1})	NH4w	9.1	88.5	34.2
Nitrite in wastewater (mg N L^{-1})	NO2w	0.01	9.98	1.08
Nitrate in wastewater (mg N L^{-1})	NO3w	0.1	26.8	4.4
Organic nitrogen in wastewater (mg L^{-1})	Norgw	0.1	89.0	19.6
Phosphate in wastewater (mg P L^{-1})	PO4w	0.10	7.35	3.26
Total phosphorus in wastewater (mg L^{-1})	Ptw	1.2	33.0	9.5
Total suspended solids in wastewater (mg L^{-1})	TSSw	50	2100	316
Volatile suspended solids in wastewater (mg L^{-1})	VSSw	30	2071	264
Total chemical oxygen demand in permeate (mg L^{-1})	CODtp	10	77	36
Soluble chemical oxygen demand in permeate obtained by sample flocculation (mg L^{-1})	CODsp	10	67	28
Ammonia in permeate (mg N L^{-1})	NH4p	0.1	19.3	2.4
Nitrite in permeate (mg N L^{-1})	NO2p	0.01	0.84	0.13
Nitrate in permeate (mg N L^{-1})	NO3p	0.1	37.2	13.6
Organic nitrogen in permeate (mg L^{-1})	Norgp	0.1	6.0	1.4
Phosphate in permeate (mg P L^{-1})	PO4p	0.10	3.67	1.31
Total phosphorus in permeate (mg L^{-1})	Ptp	0.1	4.4	1.7
Total suspended solids in membrane reactor (mg L^{-1})	TSSs	5900	13200	9121
Volatile suspended solids in membrane reactor (mg L^{-1})	VSSs	4266	9491	6694
Mixed liquor suspended solids acquired on-line in the biological reactor (g L-1)	MLSSb	4.60	8.70	6.97
Transmembrane pressure (mbar)	TMP	100	378	195
Operating Parameters:				
Permeate flux (m^3 h^{-1})	Jp	0.80	3.00	1.80
Hydraulic retention time (h)	HRT	5.30	19.86	9.71
Dissolved oxygen (mg L^{-1})	DO	0.10	7.37	1.64
Sludge wastage (m^3 day^{-1})	Vslg/d	0.00	1.29	0.57
Temperature (ºC)	T	9.0	26.9	20.7

PCA was applied to all data, except fluorescence matrices, and the loadings obtained for the first 3 PCs were analysed in search for correlations within data. Both PLS models and PCA were implemented in Matlab according to nPLS and PARAFAC functions (Andersson and Bro, 2000), respectively.

3.3. RESULTS AND DISCUSSION

3.3.1. PLS models based on 2D fluorescence spectroscopy

Multivariate linear models were obtained to predict total COD in permeate and total COD in wastewater using fluorescence data from permeate and wastewater, respectively. PLS models were performed using the values of emission intensity for each pair of excitation/emission wavelength as inputs (total of 5490 inputs) and respective COD as output. Predicted results of training and testing data sub-sets were plotted against the experimental values obtained from the pilot MBR (Figure 3.2 and 3.3). The PLS model to predict total COD in the permeate was generated using 3 latent variables and has a good fitting for both training and validation sets (R^2 of 0.88 and 0.92, respectively) and RMSEP of 5.4 mg L^{-1}, corresponding to a mean error of 15%. The PLS model to predict total COD in wastewater made with 9 latent variables gave an overall lower correlation (R^2 of 0.95 and 0.60 for training and validation data, respectively) with a RMSEP of 175 mg L^{-1} corresponding to a mean error of 33%.

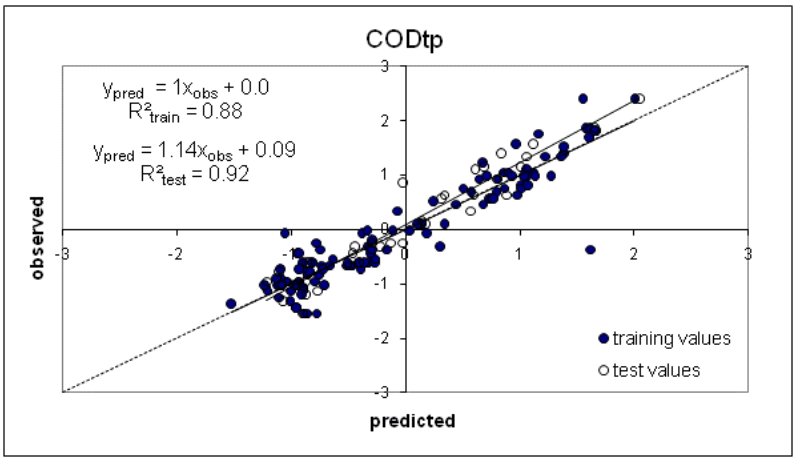

Figure 3.2. Prediction of total permeate COD based on 2D fluorescence spectra acquired in the permeate.

The establishment of these equations demonstrates that 2D fluorescence spectroscopy is able to describe the variation of total COD in wastewater and permeate media. Furthermore, the coefficients of the multilinear regression were analysed in order to identify the 2D fluorescence spectral regions that specifically correlate with total COD.

Thus, if the model coefficients are regarded as the weight of each input (in this case each position in the fluorescence matrices) it is possible to identify the excitation/emission pairs that may have a higher contribution to the description of total COD. In Figures 3.4 and 3.5 these coefficients are plotted as a function of their position in the fluorescence spectrum.

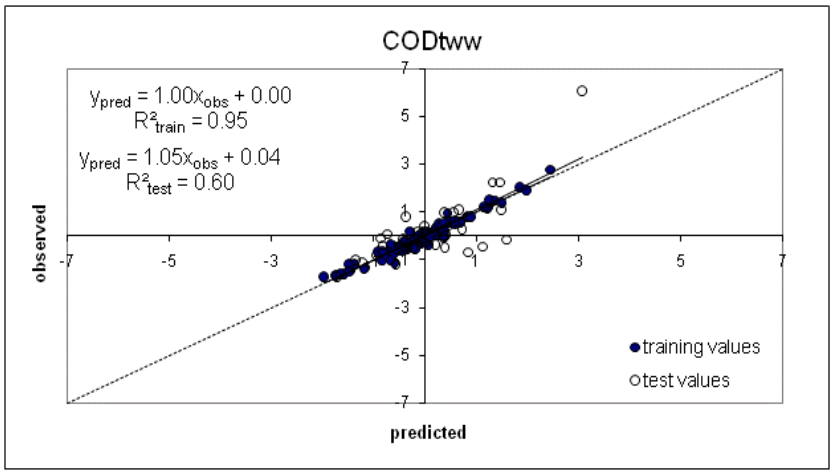

Figure 3.3. Prediction of total wastewater COD based on 2D fluorescence spectra acquired in the wastewater.

The coefficients for total COD prediction in wastewater (Figure 3.4) show that COD is mostly expressed by variations in two spectral regions at approximately 280/340 nm and 360/430 nm of excitation/emission, where proteins and humic compounds have higher emission, respectively. Thus, fluorescence response due to the presence of proteins and humic compounds appears to be the most relevant to predict COD in wastewater. In addition, the two diagonal lines in fluorescence matrices, resultant from light scattering, which is commonly correlated with media turbidity, suspended solid matter and high concentrations of solutes (Harms et al., 2002; Rinnan et al., 2005), also appear to be of high importance. This result is not a surprise when taking into account that some compounds contributing to total COD in domestic wastewaters (particulate COD) are associated with media turbidity.

In the permeate (Figure 3.5), total COD seems to be strongly described by the humic compounds region of fluorescence spectra and also by light scattering. This result was expected since proteins are generally larger than humic compounds and thus are preferentially retained by the membrane; therefore, the permeate COD can be mostly predicted by fluorescence signal of humic compounds, as opposed to wastewater COD

where proteins also have an important contribution. Light scattering in permeate EEMs is also lower than in wastewater (data not shown), likely due to the absence of turbidity and to reduced concentrations of solutes and colloids. However, its contribution to COD prediction is also important as revealed by the analysis of contributions given by the PLS coefficients (Figure 3.5).

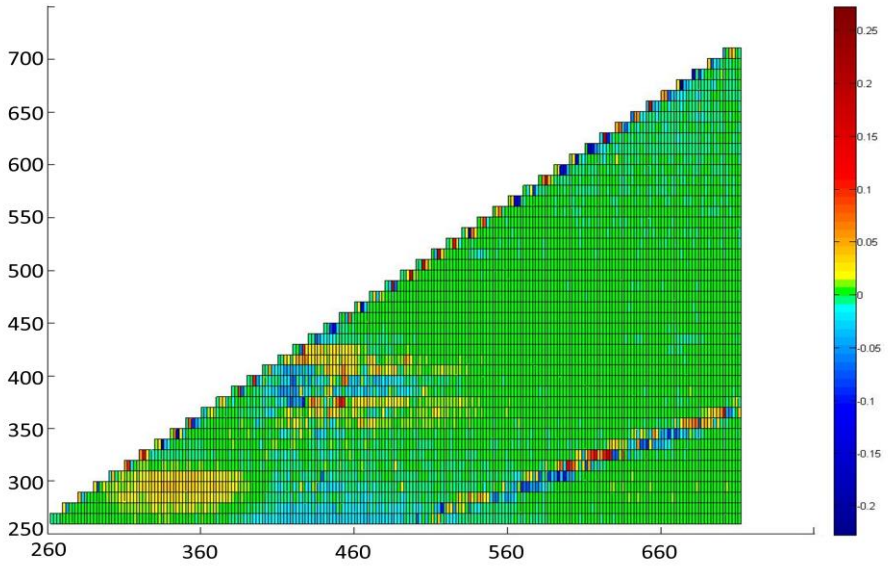

Figure 3.4. Coefficient values of the PLS model for prediction of total COD in wastewater plotted in function of the respective variable position in fluorescence matrices.

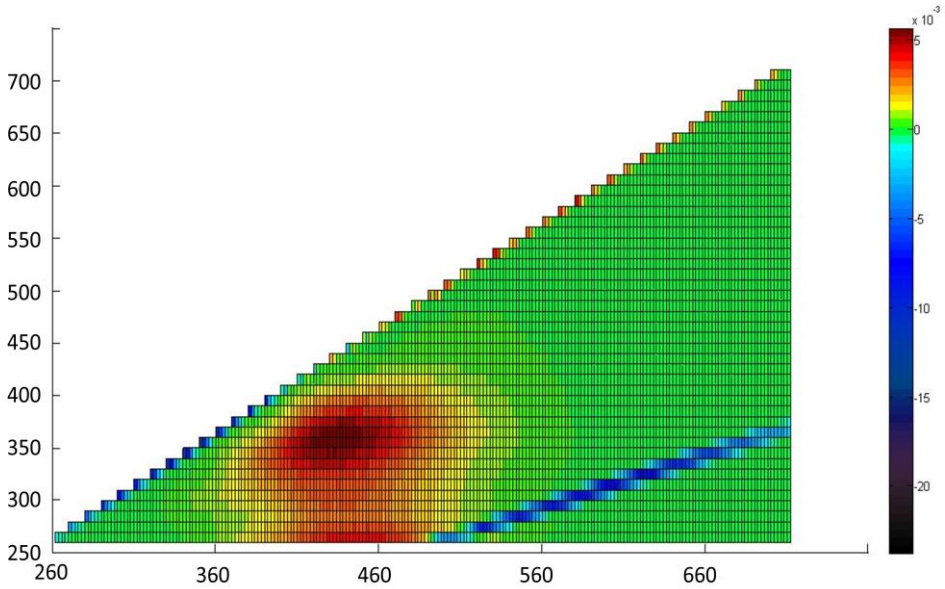

Figure 3.5. Coefficient values of the PLS model for prediction of total COD in permeate plotted in function of the respective variable position in fluorescence matrices.

Despite the fact that some areas of the fluorescence spectra have higher importance to predict total COD in permeate and in wastewater, the complete matrices have a role in modelling and it is not possible to arbitrary remove these inputs without compromising output prediction.

2D fluorescence spectroscopy showed to be rich in information about media composition (enough to predict the variation of total COD) and therefore can be a useful tool in real-time monitoring of MBRs. However, the fluorescence EEMs could not provide sufficient information to predict other parameters, such as influent and effluent ammonia and phosphorus (results not shown). Therefore, the accurate quantitative monitoring of MBR performance will still require the acquisition of at least some process variables.

3.3.2. Principal component analysis of operational and performance parameters

In an attempt to reduce the analytical effort required for full characterisation and to reduce the time frame necessary to obtain monitoring results, PCA was used to study the complex correlations between a comprehensive set of operating and performance parameters (Table 3.1). Using PCA applied simultaneously to all data, it is possible to understand which parameters are strongly correlated and therefore reduce the number of parameters used to monitor MBRs performance.

In PCA, a new system of axes is generated based in the maximum variance captured in order to reduce redundancy and noise, and describe data more accurately. In this study, PCA was applied to all analytical data acquired together with the operating conditions used to obtain each observation. Loadings quantify the difference between the new system of axes and the original one. Therefore, the variables that are correlated have a similar value of loading or are in the same line that crosses the axes intersection, since the direction of the PC is identical. Otherwise, the uncorrelated variables have an orthogonal relationship and appear in a perpendicular direction to the considered variable. The strongest correlation is obtained when loadings have the value of one.

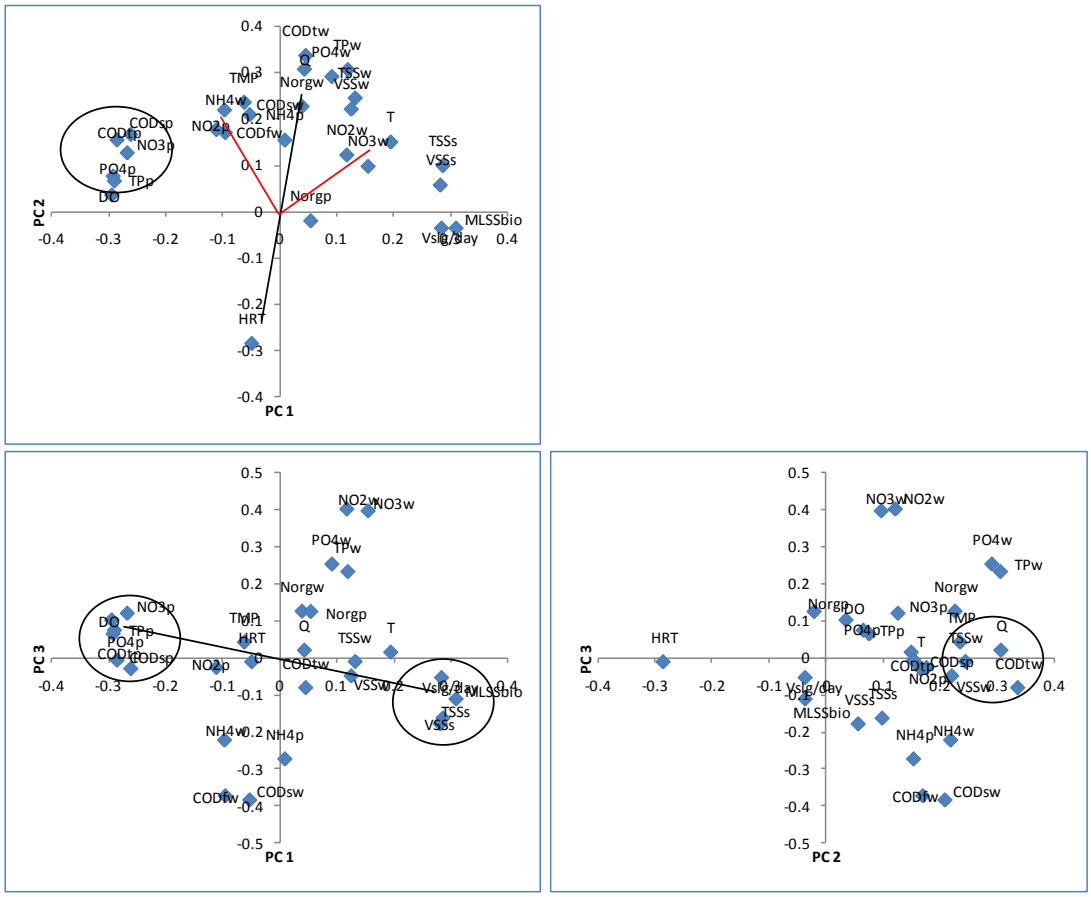

Figure 3.6. Loadings plots for PC1, PC2 and PC3 obtained from PCA of the process and operating parameters used to characterise the pilot MBR.

Figure 3.6 illustrates the loadings of the PC1, PC2 and PC3 for operating and performance MBR data. PC1 captured 28.9% of variance, PC2 19.6% and PC3 8.5%, totalising 57% of variance for the first 3 principal components. The subsequent PCs captured progressively lower variance percentages, thus only the loadings of the first 3 PCs were analysed. Loadings of PC1 and PC2 show that transmembrane pressure (TMP) belongs to the same cluster as ammonia (NH4w), soluble COD by flocculation (CODsw) and soluble COD by filtration (CODfw) in wastewater, and therefore is positively correlated with these wastewater parameters. Also shows that ammonia in the permeate (NH4p) decreases when the hydraulic retention time (HRT) increases. Loadings of PC2 and PC3 show that TMP is positively dependent of the permeate flux (Jp) and the suspended solids in wastewater (total and volatile, respectively TSSw and VSSw). It is also noticeable, looking to the loading plots, that PC1 captures essentially the variation in permeate quality in function of the biomass concentration: permeate quality parameters (CODtp, CODsp, NO3p, PO4p and Ptp) are inversely correlated with the amount of sludge in the MBR (mixed liquor suspended solids in biological

compartment, MLSSbio, and volatile and total suspended solids in sludge, VSSs and TSSs respectively). Additionally, PC2 captures variations due to the flux (note that both permeate flux, Jp, and hydraulic retention time, HRT, were plotted since the MBR volume changed during the study).

It is also evident that permeate quality parameters vary inversely with temperature (T), which is consistent with the higher activity of sludge expected and observed at higher temperature (during the summer period). Moreover, since aeration was maintained constant, the increase of dissolved oxygen (DO) means that less oxygen is consumed and therefore lower sludge activity and lower COD and nutrients removal, as reflected by the inverse correlation of DO with T and positive correlation with permeate quality parameters.

Other expected correlations are evident in the loadings plots such as the proximity between soluble COD by filtration and by flocculation in wastewater, or between MLSSbio, TSSs, VSSs and sludge wastage (Vslg/d) since the first three are different measurements of suspended solids and sludge wastage were performed in order to keep specific solid retention times.

These results confirm the complexity and abundance of correlations across data and show that process variables are not fully described by the first 3 PCs of the PCA, which only capture 57% of total variance. Additionally, loadings have low values (≤ 0.5), meaning that the monitoring parameters are not fully described by any of the three PCs. Therefore, and despite all the relationships found, it is not possible to identify strong correlations that could reduce the number of analytical parameters needed to make a complete characterisation of the MBR performance.

3.4. CONCLUSIONS

In this study, 2D fluorescence spectroscopy data was combined with PLS modelling to predict performance parameters of an MBR, aiming at developing a strategy for real-time monitoring. Using only this technique, it was possible to describe total COD in both the influent wastewater and the permeate of a MBR. However, this approach was not successful in accurately predicting other performance parameters, suggesting that 2D fluorescence spectroscopy cannot totally replace conventional MBR monitoring.

Alternatively, a set of MBR process and operating parameters were analysed by PCA in order to find correlations between them and thus reduce redundancy and analytical time necessary for monitoring. This multivariate analysis tool revealed some linear degree of correlation between certain parameters. However, these correlations were not strong enough (loadings≤0.5) to reduce the number of parameters needed to describe the process. Additionally, the PCA possibly did not cover all the existing correlations, since only a low level of variance (57%) was captured using the most contributing PCs.

Overall, it was found that correlations across the MBR data are abundant and that relationships between operating parameters and performance variables are complex and interdependent. However, it is possible that the information captured by 2D fluorescence spectroscopy could replace some of the analytical and operating parameters, since this technique was able to successfully describe some MBR influent and effluent quality indicators (total COD). It is thus proposed that combined modelling of 2D fluorescence data and selected performance/operating parameters should be further explored for efficient MBR monitoring aiming at rapid process control.

Chapter 4

2D FLUORESCENCE AS A FINGERPRINTING TOOL FOR MONITORING WASTEWATER TREATMENT SYSTEMS

SUMMARY

The use of 2D fluorescence for monitoring complex biological systems requires a careful assessment of the effect of chemical species present, which may be fluorescent and/or may interfere with the fluorescence response of target fluorophores. Given the complexity of fluorescence data (excitation emission matrices – EEMs), the challenge is how to recover the information embedded into those EEMs that can be quantitatively related with the observed performance of the biological processes under study. In this chapter it is shows clearly that interference effects (such as quenching and inner filter effects) occur due to the presence of multiple species in complex biological media, such as natural water matrices, wastewaters and activated sludge. A statistical multivariate analysis is proposed to recover quantitative information from 2D fluorescence data, correlating EEMs with the observed performance. A selected case study is discussed, where 2D fluorescence spectra obtained from the effluent of a membrane bioreactor were compressed based on principal component analysis and successfully correlated with the effluent chemical oxygen demand, using projection to latent structures modelling. This chapter demonstrates the potential of using 2D fluorescence spectroscopy as a status fingerprint. Additionally, it is shown how statistical multivariate data analysis can be used to correlate EEMs with selected performance parameters for monitoring of biological systems.

*Published as: Galinha, C.F., Carvalho, G., Portugal, C.A.M., Guglielmi, G., Reis, M.A.M. and Crespo, J.G., 2011. Two-dimensional fluorescence as a fingerprinting tool for monitoring wastewater treatment systems. Journal of Chemical Technology and Biotechnology, **86**, 985-992.*[1]

[1] Reproduced with the authorization of the editor and subjected to the copyrights imposed.

4.1. INTRODUCTION

Fluorescence spectroscopy is a highly sensitive and non-invasive technique that can be used *in situ* and on-line through an optical fibre probe. Biological systems contain many natural fluorophores, such as amino acids (tryptophan, tyrosine and phenylalanine), vitamins, coenzymes and aromatic organic matter in general that can be detected by fluorescence spectroscopy, regardless if they are intra- or extra-cellular. Therefore, this technique has great potential for real-time monitoring of biotechnological systems, as was previously demonstrated by e.g. Li et al. (1991) and Li and Humphrey (1991). In those studies, the fluorescence signal obtained at specific excitation wavelengths was used to measure key fluorophores in biological media. A broader approach consists of scanning a sample with simultaneous variation of the excitation and emission wavelengths, i.e., two-dimensional (2D) fluorescence. The resulting excitation-emission matrices (EEMs) capture information about the presence of multiple substrate components and microbial products, and thus can be seen as a fingerprint of the biological system (Marose et al., 1998).

The 2D fluorescence signal is complex and subject to interferences due to the presence of multiple molecules in the system. Changes in the fluorescence signal (such as decrease in fluorescence intensity or change of the emission band shape) may be caused by quenching and/or by inner filter effects. Fluorescence quenching affects the fluorescence characteristics of target fluorophores present in the media, and results from the interaction of the fluorophore in the excited state with a (fluorescent or non-fluorescent) quencher, which interferes with the de-excitation process. The extent of the quenching effect will depend on the type and concentration of the quencher (Valeur, 2002). On the other hand, inner filter promoters interfere with the light pathway, e.g. by absorbing light either in the same absorbance region or the same emission region as the fluorophores (Valeur, 2002).

Although interference effects hamper the recovery of quantitative information about specific fluorophores, they are an extremely rich source of information if adequate mathematical tools are used to analyse the data. In fact, changes in spectra caused by interference effects can increase the richness of information contained in 2D fluorescence spectroscopy results, providing an overall fingerprint of the physiological

state of a biological system. Additionally, EEMs are able to elicit information on the performance of bioreactors under specific process environments, since the fluorescence response is also sensitive to the environmental conditions (pH, ionic strength and salt composition). The challenge is how to recover and use all of this embedded information for systems monitoring and develop appropriate models that correlate input (the fluorescence EEMs) with output data (performance parameters), for process control.

2D fluorescence has been applied for monitoring surface water composition and wastewater systems (Her et al., 2003; Kimura et al., 2009; Lee et al., 2006; Tang et al., 2010; Wang et al., 2009b). However, most of those studies did not use fluorescence on-line; the samples were extracted and/or fractioned before fluorescence analysis, and the interferences existing in a complex media were not considered. Recently, Kobbero et al. (2008) assessed the quenching effects observed in activated sludge samples, and concluded that individual components could not be quantified by direct observation of 2D fluorescence peak intensity. However, no attempt to deconvolute the matrices through modelling techniques was made.

This study aims at demonstrating the applicability of 2D fluorescence for real-time monitoring of biological wastewater treatment systems, by regarding the EEMs as fingerprints that may be mathematically correlated with representative performance parameters. Similar approaches have been previously employed by Scheper and co-workers (Boehl et al., 2003; Ganzlin et al., 2007; Marose et al., 1998; Surribas et al., 2006) as well as by our research group (Teixeira et al., 2009; Wolf et al., 2007; Wolf et al., 2001), to monitor other biotechnological systems with better defined media and/or biological composition.

The sensitivity of this technique to water and wastewater compositions was assessed in the present study, as well as the occurrence of fluorescence interferences in these media. Multivariate statistical modelling, using principal component analysis (PCA) (implemented with PARAFAC function) and multiple linear regression (done by projection to latent structures – PLS), were used to correlate the interlinked information contained in EEMs from a wastewater treatment system with a selected performance indicator (chemical oxygen demand of the treated effluent). Data from a membrane bioreactor (MBR) for wastewater treatment was selected as a case study to illustrate this

approach and exemplify how the information obtained by 2D fluorescence spectroscopy can be correlated with performance parameters.

The modelling approach described in this study could be used in the development of expert-control systems based on the modelling approach described. Such control systems may be operated together with on-line 2D fluorescence monitoring, using optical fibres located in defined positions of the biological process, without the need to sample the system periodically and run all the time-consuming and laborious analytical procedures. Therefore, real-time control and operator decision-making may be ensured.

4.2. MATERIALS AND METHODS

4.2.1. Water and wastewater samples and aqueous solutions used for 2D fluorescence analysis

Surface water was collected from the river Tagus, in Portugal, spring water was collected in Alenquer, Portugal, and domestic wastewater was collected at the entrance of the wastewater treatment plant (WWTP) of Mutela, located in the region of Lisbon (Almada), Portugal. All samples were collected within a 5 week period during autumn. The wastewater sample was previously centrifuged at 9000 rpm, 10 min at 10 ºC. The samples were filtered using a filter with 0.20 μm pore diameter to remove microorganisms before fluorescence scanning.

Solutions of commercial humic acids (Sigma, USA) in deionised water were used to confirm the excitation/emission regions typical from these compounds. Excitation-emission maps of a model protein, bovine serum albumin (BSA, Fluka, Switzerland), were obtained in the presence and absence of commercial humic acids (10 mg L^{-1}), in order to test fluorescence interferences caused by the simultaneous presence of humic acids and proteins in media. BSA was employed at concentrations of 10 and 50 mg L^{-1} for these tests. Subsequently, 2D fluorescence maps were obtained for different water samples with and without addition of BSA at the same final concentrations (10 and 50 mg L^{-1}) to assess the interference of natural humic acids in protein fluorescence.

4.2.2. Membrane bioreactor set-up and operation

Pre-screened wastewater, activated sludge and permeate samples were collected from a pilot MBR located in the WWTP of Lavis (Trento), in Italy. The system consisted on an anoxic/aerobic compartment (4.7 m^3 denitrification, 8.7 m^3 nitrification) and a separate membrane tank (1.5 m3), in which a hydrophilized PVDF ultrafiltration membrane was immersed (GE Zenon ZW500d hollow fibre module; 0.04 μm; 100 m^2). The plant was operated at an average SRT of ~ 22 days; the MLSS in the biological tank was around 7.5 kg m^{-3} whereas the MLSS content in the membrane compartment ranged between 10 and 11 kg m^{-3}, with a recirculation ratio of ~ 2.5 from the membrane tank to the anoxic one. Samples were taken simultaneously from the feed, the activated sludge recirculation and the permeate, and were immediately analysed by 2D fluorescence spectroscopy by immersion of an optical fibre probe in a stirred beaker. In order to avoid sedimentation of suspended solids present in wastewater and activated sludge samples, identical stirring conditions were applied to all the samples (wastewater, activated sludge and permeate) during the acquisition of the fluorescence spectra. In addition, chemical oxygen demand (COD) was determined in the permeate samples according to the APHA Standard Methods. COD was selected as a performance indicator since it is the principal parameter used to characterise the effluent's quality in WWTPs used for carbon removal.

4.2.3. Acquisition of 2D fluorescence spectra

The 2D fluorescence spectra of humic and BSA solutions, surface water, spring water and domestic wastewater from Mutela were obtained with a fluorometer computer-interface Perkin-Elmer LS 50B. The excitation source was a pulsed Xenon UV lamp, and the detector was a gated photomultiplier. Reflection grating monochromators were used on both excitation and emission sides. The scanning speed was 1500 nm/min; excitation and emission slits were 10 nm. Fluorescence spectra were generated in a range of 200 to 600 nm (excitation) and 225 to 625 nm (emission), with an excitation wavelength incrementing step of 10 nm.

The 2D fluorescence spectra obtained from the MBR in the WWTP of Lavis were acquired with a fluorescence spectrophotometer Varian Cary Eclipse equipped with excitation and emission monochromators and coupled to a fluorescence optical fibre

bundle probe. The optical fibre bundle is constituted by a total of 294 randomized optical fibres (147 excitation and 147 emission) each with a diameter of 200 µm and a length of 2 m. Fluorescence spectra were generated in an excitation wavelength range of 250 to 700 nm and an emission wavelength range between 260 and 710 nm, with an excitation wavelength incrementing step of 10 nm. Fluorescence spectra were obtained using excitation and emission slits of 10 nm and a scan speed of 3000 nm/min.

4.2.4. Mathematical data deconvolution

The 2D fluorescence spectroscopy measurements result in excitation-emission matrices (EEMs) where each value of fluorescence intensity corresponds to each pair of excitation/emission wavelengths, totalising more than 5 thousand model input variables. In order to reduce the number of variables to introduce in mathematical models, EEMs were vectorised and compacted. Compression was done to decrease the number of variables, co-linearity and noise using principal component analysis (PCA). PCA was applied using three principal components, corresponding to > 99.5% of variance captured. Projection to latent structures (PLS) were then used to maximise the co-variance between the compacted fluorescence maps of permeate, measured during the operation of the pilot MBR, and COD in the permeate. The PLS model was developed using 146 different measurements of 2D fluorescence spectroscopy and COD in the MBR permeate effluent, obtained throughout 10 months of operation. These 146 observations were divided randomly into a training set (75% of the observations, which were used to calibrate the model) and a validation set (25% of the observations, which were used to validate the final model). A model was thus obtained for prediction of the permeate COD through a linear correlation of the three new fluorescence input variables (PC1, PC2 and PC3) resulting from EEM compression.

4.3. RESULTS AND DISCUSSION

4.3.1. Discriminating water samples of different nature with 2D fluorescence spectroscopy

Three different types of real water samples were analysed by 2D fluorescence spectroscopy: spring water, surface water and domestic wastewater. Their spectra showed significant differences between them (Figure 4.1). The spring water sample had few components in solution, which is reflected in the low complexity of the EEM, displaying no detectable fluorescence peaks of protein or humic-like compounds in their

corresponding regions of the contour maps (at Ex/Em of 280 / 320-350 nm and Ex/Em of 320-340 / 420-455 nm regions, respectively). The surface water sample had fluorescence signal in the humic compounds region, but no detectable proteins. It is also interesting to note the similarity between the surface water fingerprint (Figure 4.1b) and the humic acid aqueous solution fingerprint (see Figure 4.4b), which reflects the presence of humic compounds in this natural water, as could be anticipated. Domestic wastewater showed high fluorescence peaks in the protein and the humic-like compounds regions, as expected from such a complex biological matrix. These results clearly show that it is possible to qualitatively distinguish water samples with different nature by comparing the profile and intensities of their fluorescence matrices.

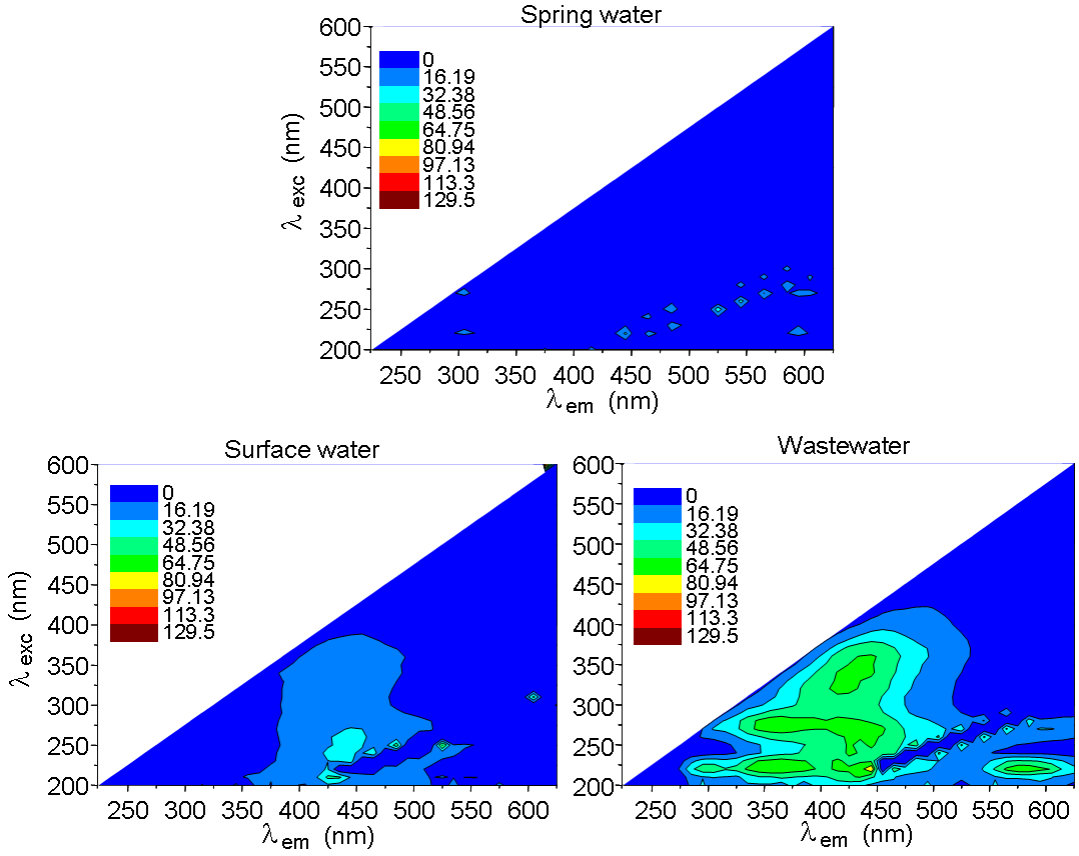

Figure 4.1. 2D fluorescence spectra of (a) spring water collected in Alenquer, Portugal; (b) surface water collected from the river Tagus, Portugal; and (c) domestic wastewater collected at the entrance of a WWTP in Almada, Portugal.

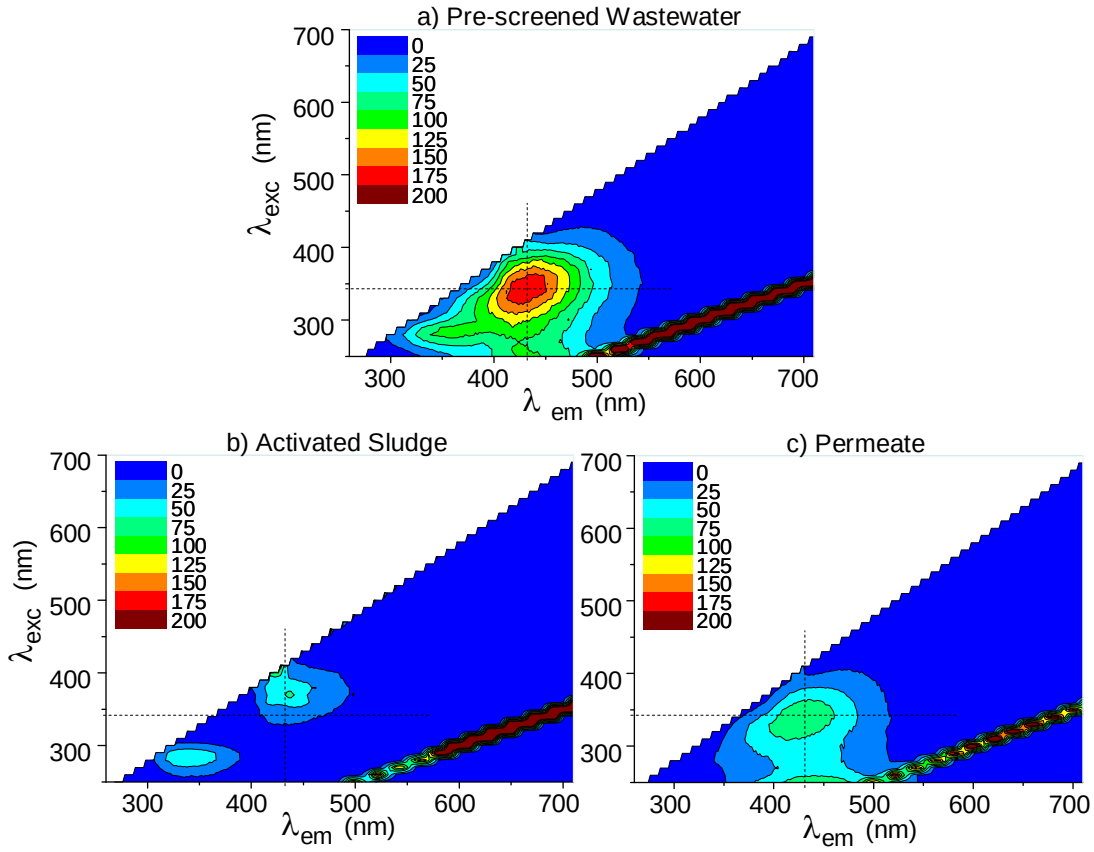

Figure 4.2. 2D fluorescence spectra of (a) influent pre-screened wastewater, (b) activated sludge and (c) effluent permeate collected in the WWTP of Lavis in Italy at the 7th day of operation.

Additionally, samples collected from different sampling points of an MBR system (influent wastewater, activated sludge and permeate effluent) were also analysed by 2D fluorescence spectroscopy. Figures 4.2 and 4.3 show the results obtained for the three sampling points for two different days of operation. The obtained EEMs show significant differences between the three types of samples, corresponding to their different characteristics, which confirm the capability of this technique to qualitatively distinguish samples collected in different treatment phases of a WWTP. Although fluorescence spectra have a consistent profile for each sampling point, small differences can be observed between different days. In Figures 4.2 and 4.3 these differences are likely related to changes observed in the wastewater composition (e.g. total COD in wastewater changed from 970 mg L^{-1} on day 7 (Figure 4.2) to 377 mg L^{-1} on day 23 (Figure 4.3)) and operating conditions, such as hydraulic retention time (11.6 and 7.9 hours, respectively) and average temperature (16.2 and 13.1 ºC, respectively), which affect the system conditions and process performance. From these results it can be seen

that 2D fluorescence measurements of influent wastewater, activated sludge and effluent permeate can be used to monitor changes occurring through MBR operation.

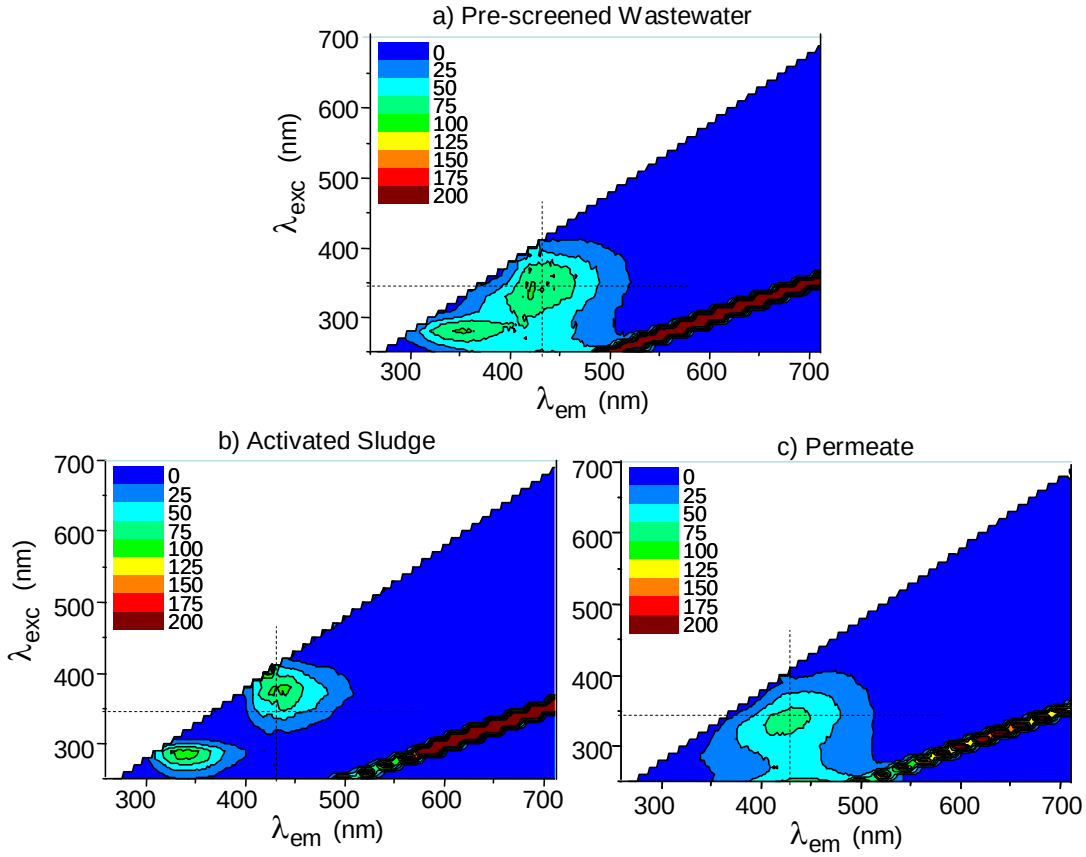

Figure 4.3. 2D fluorescence spectra of (a) influent pre-screened wastewater, (b) activated sludge and (c) effluent permeate collected in the WWTP of Lavis in Italy at the 23rd day of operation.

The influent wastewater samples had high fluorescence intensity in the excitation/emission regions of humic-like compounds and proteins, as observed for the wastewater analysed in the first assay (Figure 4.1). Permeate samples had lower fluorescence intensity in both regions when compared with the influent wastewater, particularly in the protein region. This observation suggests that small molecules like small peptides, amino acids and small humic compounds can cross the membranes while the larger molecules (such as proteins) are preferentially retained, which agrees with previously published results (Teychene et al., 2008). Activated sludge samples showed peaks both in the protein and humic compounds region; however, when compared with wastewater and permeate fluorescence peaks, their intensity was lower than expected, especially in a media with such a high cell concentration and consequently with high amounts of fluorophores. The relatively low fluorescence

intensity in activated sludge samples was likely due to the high cell density in the media. High concentrations of cells and other suspended solids (high turbidity), and the colour of the liquid phase itself, increase light scattering and media absorbance. Highly concentrated complex multi-component matrices, such as in the present study, can also lead to inner filter effects. Indeed, other substances present in the mixture may compete with the fluorophores for the incident light or reabsorb emitted light (Valeur, 2002). The overall contribution of these effects explains the reduced fluorescence intensity observed for activated sludge samples. Moreover, the humic-like compounds peak revealed a shift in the maximum excitation wavelength (see Figures 4.2 and 4.3 *versus* Figure 4.4).

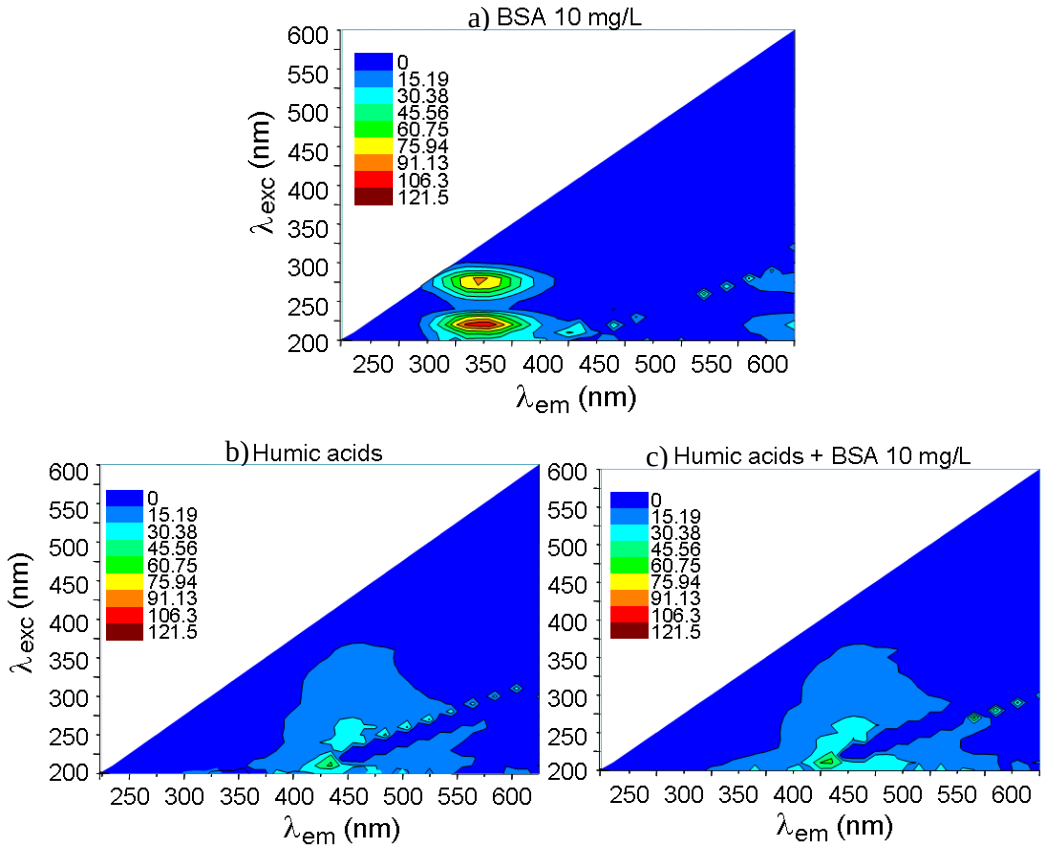

Figure 4.4. 2D fluorescence spectra of (a) a BSA solution with a concentration of 10 mg L^{-1}, (b) a commercial humic acid solution (10 mg L^{-1}) and (c) a mixed solution of humic acid and BSA, both at 10 mg L^{-1}.

This apparent peak shift could be due to a change in composition of the humic compounds fraction, since organic compounds present in the wastewater are degraded, whereas other organics resulting from biological activity are generated. Additionally, fluorescence interference effects may also have a role on this peak shift, since

quenching and inner filters effects can interfere with fluorescence intensity in this specific region of the spectra, masking part of the humic substances peak.

The analysis carried out in this study demonstrates the high sensitivity of the 2D fluorescence technique, beyond the direct response to variations in the fluorophores concentrations.

4.3.2. Assessment of interference effects on protein fluorescence spectra

Proteins are one of the key compounds that can be monitored using fluorescence techniques in biological processes such as wastewater treatment systems. However, these systems also possess high amounts of numerous organic and inorganic compounds (e.g. the humic compounds present in the wastewater, oxygen, nitrates) that have been shown to interfere with protein fluorescence spectra (Kobbero et al., 2008; Lakowicz, 1983; Ricci and Nesta, 1976). In the present study, quenching and inner filter effects on protein fluorescence spectra were investigated using a commercial humic acid and a standard protein, bovine serum albumin (BSA). Figures 4.4a and b show the 2D fluorescence spectra obtained with two independent solutions of BSA and humic acid, both at a concentration of 10 mg L^{-1} (Figures 4.4a and b), displaying the expectable peaks in the protein and in the humic-like compounds regions, respectively. However, when humic acids and BSA were both combined at the same final concentration of 10 mg L^{-1} in a mixed solution, the protein peak was no longer visible (Figure 4.4c). Protein and humic acid peaks do not overlap in their emission spectra and consequently this cannot be the cause for protein signal disappearance. To investigate the reason for this fact, a UV/visible light absorption spectrum of the humic acids was obtained using a spectrophotometer (data not shown), which showed that humic compounds absorb in a range of wavelengths that includes the region of protein absorption (competing with proteins by excitation light) and in the region of protein emission, absorbing the light emitted by proteins. These effects may explain the disappearance of the protein peak in Figure 4.4c.

Table 4.1. Decrease in fluorescence intensity at λ_{exc} = 280 nm and λ_{em} = 345 nm (maximum emission of BSA) of a standard BSA solution in comparison with solutions prepared with either commercial humic acids (at 10 mg L^{-1}), or real water/wastewater samples.

	Fluorescence Intensity Decrease	
	BSA 10 mg L^{-1}	BSA 50 mg L^{-1}
Humic acids	95 %	96 %
Surface water	52 %	57 %
Spring water	33 %	31 %
Wastewater	37 %	58 %

In order to assess fluorescence interference effects in natural water sources and wastewater samples, BSA solutions were prepared using domestic wastewater, as well as surface and spring water. Two different concentrations of BSA (10 and 50 mg L^{-1}) were tested in order to understand if the interference effects are constant for each type of water, independently of the concentration of protein. BSA solutions with the same concentration, prepared in deionised water, were used for comparison purposes. The results showed that the presence of other components in water and wastewater samples reduced the protein fluorescence emission of the obtained EEMs, as exemplified in Figure 4.5. The fluorescence intensity values at the 280/345 nm excitation/emission pair in these tests are shown in Table 4.1. The degree of interference of water/wastewater samples on the BSA fluorescence emission was lower than a concentrated humic acids solution, but still induced a decrease in the protein peak intensity of 31-58% (despite the additional contribution of the proteins present in the wastewater to this peak). These interferences were likely due to the presence of humic-like compounds in the water matrices, as revealed by their fluorescence spectra (Figure 4.1). The extent of interference varied in these different matrices (Table 4.1), which was probably due to differences in the concentration and composition of the humic-like compounds. Indeed, the surface water spectrum showed higher fluorescence intensity than the spring water spectrum in the humic compounds region, and a different contour shape than the wastewater spectrum (Figure 4.1). Correspondingly, different interference levels were observed for all three water types (Table 4.1). Moreover, the results obtained with wastewater showed that the percent fluorescence decrease was dependent on the protein concentration, reflecting the high complexity of the interference phenomena occurring in such media.

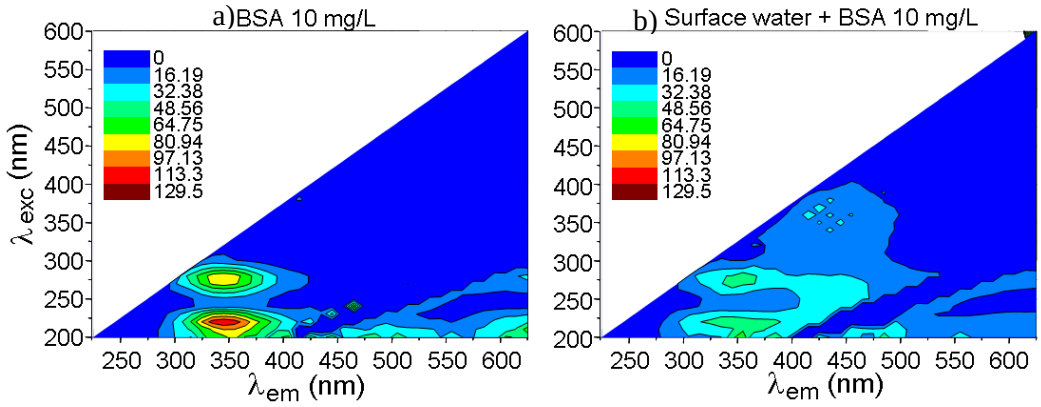

Figure 4.5. Example of fluorescence interference effects in surface water: EEM of a BSA solution (10 mg L^{-1}) in (a) deionised water and (b) prepared with surface water.

Additionally to the humic-like compounds, there are other substances with quenching properties that can be present in wastewater systems, such as dissolved oxygen and heavy metals (Lakowicz, 1983), which were not investigated in this study. However, it is clear that the complexity of interferences on the fluorescence signal limits the simple and direct quantitative measurement of specific fluorophores in complex biological systems, such as wastewater treatment systems. Nevertheless, fluorescence EEMs may be used as fingerprints, which can be regarded as extremely rich, although complex, sources of information. Actually, since fluorescence spectroscopy is highly sensitive to the composition of biological media and to the environmental conditions (T, pH, ionic strength), these effects should not be regarded as a problem but, on the contrary, as a source of information. The challenge is to develop methodologies that allow the extraction of this embedded information, correlating 2D fluorescence data with selected performance indicators.

4.3.3. Use of multivariate statistical analysis to extract quantitative information from EEMs

To extract deeper, quantitative information from 2D fluorescence spectra, previous studies used either an approach based on pattern recognition, i. e. artificial neural networks (ANN) (Wolf et al., 2007; Wolf et al., 2001) or PLS modelling (Morel et al., 2004) to deconvolute EEMs from mixed culture bioreactors. In the present study, PLS modelling was chosen over ANN to describe the data analysed since the multivariate linear PLS models are less complex than non-linear ANN models, and therefore minimise over-fitting of the data points (Chan et al., 2006). Previous compression of fluorescence spectra with PCA is an important step, since it translates fluorescence

EEM information into a lower number of new orthogonal variables. These new variables describe data contained in the spectra while eliminating noise, and are each one a combination of the information present in different regions of the EEMs.

2D fluorescence data of a MBR for municipal wastewater treatment was selected to illustrate this approach. (2D) fluorescence spectroscopy has been previously applied for monitoring MBR (Tang et al., 2010; Teychene et al., 2008) aiming at the detection of single specific compounds. In this study, the complete EEMs obtained in a MBR were used as fingerprints, which were correlated with a selected performance indicator (permeate COD) through multivariate statistical analysis.

A PLS model was developed to predict COD in the permeate, as a function of the compressed permeate EEMs (PC1, PC2 and PC3).

$$[COD] = - 0.198\ PC1 - 0.887\ PC2 + 0.198\ PC3 \qquad \text{(Equation 4.1)}$$

The model obtained (Equation 4.1) shows good fitting for both training and validation sets of data, as can be seen from Figure 4.6 where experimentally observed values are plotted against predicted values. Regression coefficients (R^2) were 0.86 and 0.89, respectively for training and validation, with slopes of approximately 1, and a root mean square error of prediction (RMSEP) of 6.0 mg COD L^{-1}, which is approx. 8% of the maximum measured COD. The COD variance captured by the three compressed inputs was 86%. This value shows that the model was able to predict the experimental COD values while excluding noise.

This example demonstrates that multivariate statistical models can adequately correlate EEM information to predict physical parameters associated with the system performance.

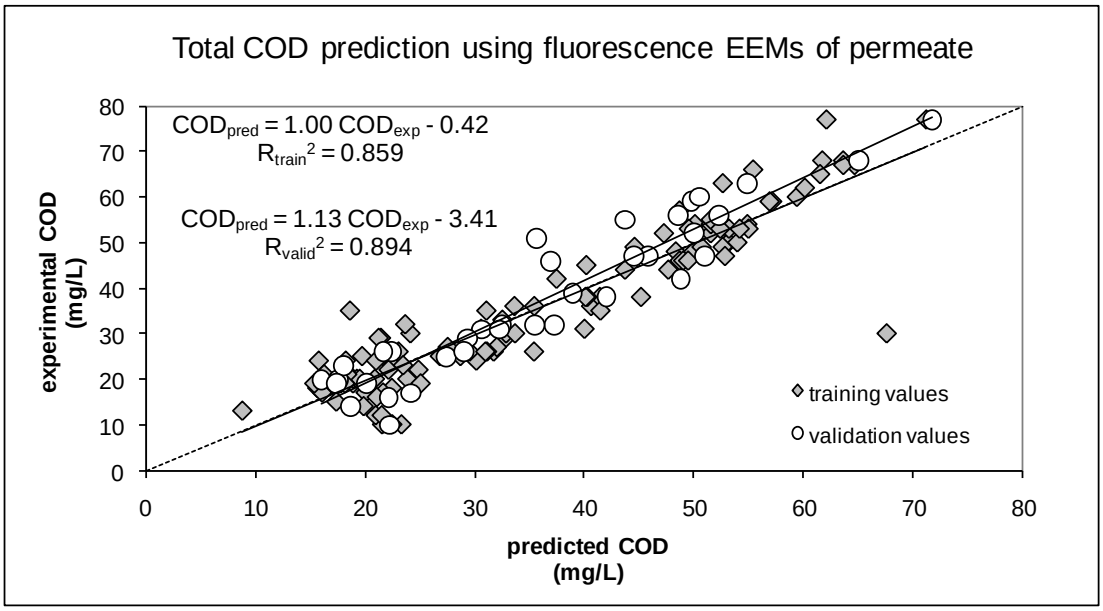

Figure 4.6. Observed values of COD in the permeate represented *vs* the corresponding values predicted by a PLS model developed with permeate EEM data.

4.4. CONCLUSIONS

Two-dimensional fluorescence can be applied for monitoring of biological systems due to the ability of this technique to distinguish matrices with different compositions. 2D fluorescence data can be acquired in multiple locations of the system (off-line or on-line, according to specific needs), and time-programmed with the help of an optical "switch-box". Considering that the acquisition of a complete fluorescence map takes only a few minutes (depending on the number of data points aimed), this tool can be used as an on-line, non-invasive, real-time monitoring technique. The challenge, however, remains on the ability of integrating this information in quantitative models, where fluorescence data are related with relevant process performance parameters determined independently.

To solve this problem, the EEMs can be vectorised and introduced as input information in multivariate statistical models, namely using PLS tools. Through this procedure, 2D fluorescence input data (that characterise influent, effluent and biological media) can be correlated with relevant output variables (e.g. effluent quality).

The approach presented in this paper has high potential for several reasons: 1) 2D fluorescence monitoring of biological systems is simple and fast; other traditional characterisation methods, which are laborious and slow, may be totally replaced by this

technique; 2) The multivariate statistical model developed may be continuously improved with new data; 3) This modelling approach could be implemented within expert-control systems; the data captured by the fluorescence fingerprints could then be used to rapidly evaluate the system status and support a real-time adjustment of the operating conditions.

Chapter 5

MULTIVARIATE STATISTICALLY-BASED MODELLING OF A MEMBRANE BIOREACTOR FOR WASTEWATER TREATMENT USING 2D FLUORESCENCE MONITORING DATA

SUMMARY

This work presents the development of multivariate statistically-based models for monitoring several key performance parameters of membrane bioreactors (MBR) for wastewater treatment. This non-mechanistic approach enabled the deconvolution of 2D fluorescence spectroscopy data, a powerful technique that has previously been shown to capture important information regarding the status of MBRs. Projection to latent structure (PLS) modelling was used to integrate 2D fluorescence data, after compression through parallel factor analysis (PARAFAC), with operation and analytical data to describe a MBR fouling indicator (transmembrane pressure, TMP), five descriptors of the effluent quality (total COD, soluble COD, concentration of nitrite and nitrate, total nitrogen and total phosphorus in the permeate) and the biomass concentration in the bioreactor (MLSS). A multilinear correlation was successfully established for TMP, CODtp and CODsp, whereas the optimised models for the remaining outputs included quadratic and interaction terms of the compressed 2D fluorescence matrices. Additionally, revealing correlations between inputs and outputs were found when analysing the coefficients of the optimised multilinear models. This work demonstrates the applicability of 2D fluorescence and statistically-based models to simultaneously monitor multiple key MBR performance parameters with minimal analytical effort. This is a promising approach to facilitate the implementation of MBR technology for wastewater treatment.

*Published as: Galinha, C.F., Carvalho, G., Portugal, C.A.M., Guglielmi, G., Reis, M.A.M. and Crespo, J.G., 2012. Multivariate statistically-based modelling of a membrane bioreactor for wastewater treatment using 2D fluorescence monitoring data. Water Research, **46**, 3623-3636.*[1]

[1] Reproduced with the authorization of the editor and subjected to the copyrights imposed.

5.1. INTRODUCTION

Despite the increasing acceptance of membrane bioreactors (MBR) for wastewater treatment, membrane fouling is still a major obstacle during MBR operation. The adsorption of microbial cells and organic compounds (colloids and solutes) on the membrane surface and porous structure decreases permeability, leading to frequent membrane cleaning, reducing membrane life span and increasing operating costs.

Several studies have found that extracellular polymeric substances (EPS) play a major role in membrane fouling of MBRs (Judd, 2008; Le-Clech et al., 2006; Meng et al., 2009). EPS are organic compounds that result from microbial metabolism and that can be solubilised in the bulk liquid (soluble EPS) or remain on the cells external surface (bound EPS). EPS are mainly composed of proteins, polysaccharides, lipids, nucleic acids and humic compounds. Indeed, activated sludge is a highly complex matrix composed by a great diversity of microorganisms and microbial products, and also by wastewater organic and inorganic components.

Despite the multiple studies on fouling agents and mechanisms carried out in the latest years, fouling formation and the conditions that lead to it are far from being completely understood. A large variety of techniques has been employed to investigate organic fouling, including Fourier transform infrared (FTIR) spectroscopy, solid state ^{13}C-nuclear magnetic resonance (NMR) and high performance size exclusion chromatography (HP-SEC), which cover a broad range of molecules (Meng et al., 2009). However, even though these comprehensive techniques can provide detailed information about the fouling agents, they cannot be used for real-time monitoring due to their technical requirements and limitations. Simple, real-time and *in situ* monitoring of the fouling agents evolution in the reactor is desirable to improve MBR operation, avoiding damage due to unpredictable membrane fouling or module clogging.

Real-time monitoring techniques are also important to partially or totally replace the numerous and time-consuming analytical techniques used to monitor the biological activity performance for control purposes. Real-time monitoring information can thus be used for optimisation and control of the operating parameters, to maximise process performance and minimise fouling. Furthermore, simultaneous assessment of the

wastewater influent, biological media and permeate effluent allows inferring about the relationship between process performance and different operating and environmental conditions.

Fluorescence spectroscopy is a highly sensitive and non-invasive technique that can be applied *in situ* and on-line using an optical fibre probe without disturbing the biological system. Moreover, all microorganisms produce natural fluorophores, such as amino acids (tryptophan, tyrosine and phenylalanine), vitamins, coenzymes and aromatic organic matter in general that can be detected by fluorescence spectroscopy, regardless if they are intra- or extra-cellular. Wastewater and EPS released by cells also comprise fluorescent molecules, such as proteins and humic compounds. Therefore, fluorescence spectroscopy can be used to provide rapid information about components in MBR media that are relevant for biological and membrane performance.

In complex systems, two dimensional (2D) fluorescence can be used by varying simultaneously the excitation and emission wavelengths, detecting at the same time the presence of a wide diversity of natural fluorophores (Marose et al., 1998). The presence, concentration and interaction between several fluorescent and non-fluorescent components in complex biological media, result in different fluorescence spectra patterns. The fluorescence excitation-emission matrices (EEMs) obtained throughout time may, thus, reflect the physiological activity of a biological system.

Nevertheless, as shown in Chapter 4, fluorescence spectroscopy is susceptible to several interferences, which can prevent the direct measurement of fluorophores in complex media such as wastewater and activated sludge. Previous studies have applied 2D fluorescence spectroscopy to monitor water and wastewater treatment processes either by characterising the organic matter after chromatographic fractioning (Her et al., 2003; Lee et al., 2006) or by comparing the location and intensity of fluorescence spectra peaks of EPS extracts (Kimura et al., 2009; Wang et al., 2009b). While these methodologies can provide information on specific fluorescent compounds present in media, EEMs of the whole sample may contain additional information about the whole complex media that should not be excluded *a priori*.

To rapidly extract the full contextual information contained in spectroscopic data, a non-mechanistic approach can be followed as used in previous studies (Wolf et al., 2007; Wolf et al., 2001; Wolf et al., 2005). Non-mechanistic models are able to correlate large sets of data (including 2D fluorescence spectra and other performance parameters) and thus extract hidden information, disclosing unobvious relationships between different variables. Thereby, 2D fluorescence spectroscopy has the ability to detect mutual interferences of fluorophores with the surrounding medium, which represents an advantage over other analytical methodologies that have been applied so far for MBR monitoring.

A mathematical approach based on projection to latent structures (PLS) was proposed in Chapter 4 for the application of 2D fluorescence data for monitoring wastewater treatment systems. This methodology comprises compression of the EEMs through parallel factor analysis (PARAFAC) to reduce data redundancy and eliminate noise, and PLS modelling to correlate the compressed information with performance parameters. In the present study, this methodology was extended to the combined application of 2D fluorescence and process data (operating and analytical parameters) to predict biological and membrane performance parameters of a pilot MBR plant operated for domestic wastewater treatment.

Fouling control and the complexity of operating MBR plants were previously pointed out as the major challenges for the implementation of MBR technology (Judd, 2005; 2008; Lesjean et al., 2011). The combination of PLS modelling with 2D fluorescence spectra can result in the development of an on-line monitoring tool with minimal analytical effort. These models can be invaluable decision supporting tools to optimise on-site MBR operation.

5.2. MATERIALS AND METHODS

5.2.1. Membrane bioreactor set-up and operation

The pilot scale MBR was located in the wastewater treatment plant of Lavis (Trento, Italy). The MBR was fed with domestic wastewater collected after fine screens (2 mm) and sand/oil removal. The MBR system consisted of a biological tank with a 4.7 m^3 anoxic compartment and a 8.7 m^3 aerobic compartment for denitrification and nitrification, respectively, followed by a tank (1.5 m^3), in which a hydrophilised PVDF

ultrafiltration membrane was immersed (GE Zenon ZW500d hollow fibre module with 0.04 μm pore size and 100 m^2 area). The permeation regime consisted of alternate relaxation (1 min) and suction phases (9 min). The MBR plant was operated with a solids retention time (SRT) between 60 and 15 days; the mixed liquor suspended solids (MLSS) in the biological tank ranged between 4.6 and 8.7 kg m^{-3} whereas the MLSS content in the membrane tank ranged between 6.0 and 9.6 kg m^{-3}, with a recirculation ratio (between the recirculation flow and permeate flow) of ~ 2.5 from the membrane tank to the anoxic compartment.

Table 5.1. Operating and analytical data collected during the 10 months of MBR operation, and input and output parameters used in PLS modelling.

All Data	Average (Stdev)	Predicted Outputs	Inputs Used
Wastewater characteristics:		TMP (Transmembrane pressure)	All data but TMP
CODtw – Total COD (mg L^{-1})	524 (221)		
CODfw – COD after filtration (mg L^{-1})	155 (62)		
CODsw – Soluble COD (mg L^{-1})	108 (58)		
NH4w – Ammonia (mg N L^{-1})	34 (14)	CODtp (Total COD in the permeate)	All data but total and soluble COD in the permeate
NO2w – Nitrite (mg N L^{-1})	1.1 (1.7)		
NO3w – Nitrate (mg N L^{-1})	4.4 (6.1)		
Norgw – Organic N (mg N L^{-1})	19.6 (11.8)		
PO4w – Phosphate (mg P L^{-1})	3.3 (1.6)		
Ptw – Total phosphorus (mg P L^{-1})	9.5 (5.1)	CODsp (Soluble COD in the permeate)	All data but total and soluble COD in the permeate OR Operating parameters and on-line MLSSb
TSSw – Total suspended solids (mg L^{-1})	316 (221)		
VSSw – Volatile suspended solids (mg L^{-1})	264 (199)		
Permeate characteristics:			
CODtp – Total COD (mg L^{-1})	36 (17)		
CODsp – Soluble COD (mg L^{-1})	28 (13)		
NH4p – Ammonia (mg N L^{-1})	2.4 (2.9)	Ntp (Total nitrogen in the permeate)	All data but nitrogen measurements in the permeate
NO2p – Nitrite (mg N L^{-1})	0.13 (0.19)		
NO3p – Nitrate (mg N L^{-1})	13.6 (10.5)		
Norgp – Organic N (mg N L^{-1})	1.39 (0.97)		
PO4p – Phosphate (mg P L^{-1})	1.31 (1.11)	NOxp (Nitrite and nitrate in the permeate)	
Ptp – Total phosphorus (mg P L^{-1})	1.72 (1.25)		
Operating parameters:			
T – Temperature (ºC)	21 (5)		
Vslg/d – Volume of sludge purged per day (mL d^{-1})	573 (359)	Ptp (Total phosphorus in the permeate)	All data but phosphorus measurements in the permeate
HRT – Hydraulic retention time (hour)	10 (3)		
Jp – Permeate flux (L m^{-2} h^{-1})	18.0 (6.2)		
DO – Dissolved oxygen (mg L^{-1})	1.6 (1.5)		
TMP – Transmembrane pressure (mbar)	195 (71)		
Activated Sludge characteristics:		MLSS (Mixed liquor suspended solids on biological tank)	All data but suspended solids of sludge
MLSSb – Mixed liquor suspended solids, acquired on-line in the biological tank (g L^{-1})	7.0 (1.1)		
TSSs – Total suspended solids in sludge recirculation (g L^{-1})	9.1 (1.6)		
VSSs – Volatile suspended in sludge recirculation (g L^{-1})	6.7 (1.2)		

This system was operated under controlled permeate flux for a period of 10 months. During this period, operational changes were imposed in the permeate flux, SRT, temperature (due to seasons' weather), hydraulic retention time (HRT) and dissolved oxygen. The average values of wastewater quality characteristics and operating parameters during the MBR operation are shown in Table 5.1.

5.2.2. Sampling and chemical analysis

Samples taken simultaneously at the influent to the biological tank, the activated sludge recirculation and the permeate, were immediately analysed by 2D fluorescence spectroscopy by immersion of an optical fibre probe in a stirred beaker. In order to avoid sedimentation of suspended solids present in wastewater and activated sludge samples, identical stirring conditions were applied to all samples (wastewater, activated sludge and permeate) during the acquisition of the fluorescence spectra.

Additional laboratory analyses were performed for each sample. Wastewater samples were filtered through a filter with a pore size of 0.45 μm. Three different fractions of chemical oxygen demand (COD) were measured in wastewater samples (total, soluble after filtration and soluble after flocculation), while only total COD and soluble COD after flocculation were determined for permeate samples. The flocculation method used in wastewater and permeate samples is described in Mamais et al. (1993) and deduces the COD contribution of the colloidal particles that normally pass through 0.45 μm filters. COD, ammonia, nitrate, nitrite, total nitrogen, phosphate and total phosphorus were measured in wastewater and permeate samples, whereas total suspended solids (TSS) and volatile suspended solids (VSS) measurements were performed for wastewater and sludge samples. These parameters were analysed according to the APHA Standard Methods (APHA, 1998). Additionally, the following parameters were continuously acquired on-line: transmembrane pressure (TMP), temperature (T), dissolved oxygen (DO) and mixed liquor suspended solids in the biological tank (MLSSb) (see Guglielmi et al. (2007) for equipment details).

5.2.3. 2D fluorescence spectra

2D fluorescence spectra were acquired with a fluorescence spectrophotometer Varian Cary Eclipse equipped with excitation and emission monochromators and coupled to a fluorescence optical fibre bundle probe. The probe comprised 294 randomised optical

fibres (147 excitation and 147 emission), each with a diameter of 200 μm and a length of 2 m. Fluorescence spectra were generated in an excitation wavelength range of 250 to 700 nm (with an incrementing step of 10 nm) and an emission wavelength range between 260 and 710 nm, using excitation and emission slits of 10 nm and a scan speed of 3000 nm min^{-1}.

5.2.4. Development of PLS models

2D fluorescence spectroscopy measurements resulted in excitation-emission matrices (EEMs) with more than 5 thousand values, a value of fluorescence intensity for each pair of excitation/emission wavelengths. Therefore, to reduce the dimension and the number of variables to introduce in the mathematical models, EEMs were vectorised and compacted. Compression was done to decrease the number of variables, co-linearity and noise using PARAFAC (parallel factor analysis), a decomposition method of multi-way data (such as EEMs) that can be considered an extension of the well-known principal component analysis (PCA) (Bro, 1997). PARAFAC was applied using three components for sludge and permeate EEMs, four components for wastewater EEMs and six components when the three types of EEMs where compressed together, which captured more than 99% of variance in all cases. When the EEMs corresponding to the three types of matrices were used in the same model, compression was done simultaneously to eliminate co-linearity and redundancy of information.

Multivariate statistical modelling was used to correlate operating and analytical parameters, including compressed fluorescence data, with selected performance parameters, aiming at process monitoring. Mathematical models were obtained through projection to latent structures (PLS) modelling, where an output is described by linear correlations of the inputs. In this study, selected operating parameters – permeate flux imposed (Jp), DO, HRT and SRT – were used together with the analytical data (obtained either through laboratory analysis or on-line) as inputs for model development (Table 5.1). A total of 130 MBR sampling events or observations were used, which were divided randomly into a training set (75% of the observations, which were used to calibrate the model) and a validation set (25% of the observations, which were used to validate the final model). The same training and validation data sets were used for the different models carried out for an output, in order to simplify the comparison between them. All data used in PLS models were previously normalised by subtracting the

respective average values and dividing by their standard deviations. The biomass concentration (as MLSS), a fouling descriptor parameter (TMP), and five descriptors of the biological treatment efficiency (total COD, soluble COD, concentration of nitrite and nitrate, total nitrogen and total phosphorus in the permeate) were selected as outputs.

Six different statistical parameters were used to compare, select and optimise the best models for each output prediction. These parameters were the variance captured, the root mean square error of prediction (RMSEP), the R^2 coefficients and the slopes between prediction and experimental data for both the training and test sets. The RMSEP was the main criterion used to analyse the prediction ability of each PLS model. It translates the error of the validation set and is calculated using the following equation:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N}(y_{pred} - y_{\exp})^2}{N}}$$

where y_{pred} is the value predicted by the model, y_{exp} is the value experimentally observed, and N is the total number of observations used in the model. Thus, the number of latent variables (LV) used to perform each PLS model was chosen based on the lowest RMSEP.

To assess the importance of 2D fluorescence as a monitoring tool, PLS models were developed for each selected performance parameter (output) with and without the use of fluorescence spectra as inputs. EEMs were combined with operating and analytical data as inputs using four different modelling strategies: in exclusive combination with each type of EEMs (wastewater, sludge or permeate) and in combination with all the EEM types compressed together (only the best model out of these four strategies is presented for each output, for simplicity). In order to find outlier data, the confidence intervals for PLS predictions were defined as 2 times the standard deviation of the output experimental data.

Additionally, when simple, multilinear PLS modelling was not sufficient to predict an output, interaction and quadratic terms of the compressed fluorescence matrix were also introduced as inputs in the PLS models. The models thus obtained are able to describe complex, non-linear interactions:

$$y = a \cdot x_1 + b \cdot x_2 + c \cdot x_3 + ... + d \cdot x_1^2 + e \cdot x_1 x_2 + f \cdot x_1 x_3 + ...$$

Since not all the parameters initially used as inputs are truly correlated with the outputs, the useful predictors were selected for each output by a process of optimisation. Four mathematic methods were used for input elimination: iterative stepwise elimination (ISE) (Boggia et al., 1997), iterative predictor weighting (IPW) (Forina et al., 1999), stepwise elimination (Ryan, 1997), and the Martens uncertainty test (Forina et al., 2004) using the jackknife standard deviations (Duchesne and MacGregor, 2001). The use of different methods for selection of inputs resulted in models with different number of useful inputs, from which the optimal model was selected based on the RMSEP, and on the R^2 coefficients when necessary.

5.3. RESULTS AND DISCUSSION

In the present work, PLS modelling was used to predict seven MBR performance parameters: transmembrane pressure (TMP), total chemical oxygen demand in the permeate (CODtp), soluble chemical oxygen demand in the permeate (CODsp), total nitrogen in the permeate (Ntp), nitrate and nitrite concentration in the permeate (NOxp), total phosphorus in the permeate (Ptp) and mixed liquor suspended solids (MLSS). TMP was chosen as a fouling indicator parameter, while the remaining parameters are key effluent quality characteristics that reflect the biological performance.

Various multivariate regression models were developed to predict each output independently from each other, using as inputs the fluorescence data and/or a set of process data (operating and performance parameters – Table 5.1). Low RMSEP means less error in the prediction of the outputs, thus this criterion was primarily used to identify the best models obtained for each output. Table 5.2 presents the statistical parameters obtained for selected models of TMP, CODtp, CODsp, Ntp, NOxp, MLSS and Ptp.

Table 5.2. Statistical parameters for the selected PLS models.

Outputs	Model #	2D fluorescence data incorporated in PLS	Inputs selection	Number of inputs	LV used in PLS	Variance (%)	RMSEP[a]	R^2_{train}	R^2_{valid}	$Slope_{train}$	$Slope_{valid}$
TMP	1	No	no	27	8	87	39	0.87	0.70	1.00	1.01
	2[b]		yes	7	6	84	35	0.84	0.77	1.00	1.10
	3	Permeate (3 components)	no	22	13	86	36	0.86	0.76	1.00	1.09
	4[b]		yes	11	8	85	34	0.85	0.78	1.00	1.09
TMP (without outliers)	5	No	no	27	15	94	17	0.95	0.94	1.01	1.09
	6[b]		yes	6	3	92	15	0.92	0.96	1.01	1.07
CODtp	7[b]	No	yes	9	3	77	8.4	0.77	0.76	1.00	1.09
	8	Permeate (3 components)	No	29	11	95	4.8	0.95	0.92	1.00	0.99
	9[b]		Yes	2	2	91	4.0	0.91	0.94	1.00	0.95
CODsp	10[b]	No	Yes	8	4	66	6.3	0.66	0.75	1.00	1.19
	11	Permeate (3 components)	No	29	7	92	3.4	0.92	0.92	1.00	0.98
	12		Yes	16	8	93	3.2	0.93	0.93	1.00	0.97
	13[b]		yes (on-line inputs)	7	7	90	2.9	0.90	0.95	1.00	0.91
Ntp	14[b]	No	Yes	5	1	67	5.7	0.67	0.70	1.00	1.03
	15[b]	Permeate (3 components)	Yes	9	9	70	5.6	0.70	0.72	1.00	1.00
	16[b]	Permeate (10 components)	Yes	8	7	70	5.4	0.70	0.73	1.00	1.00
	17	Permeate (3 components + quadratic and interaction terms)	Yes	24	15	78	5.1	0.78	0.77	1.00	0.96
	18[b]	Permeate (10 components + quadratic and interaction terms)	Yes	29	29	87	3.4	0.87	0.89	1.00	1.01
NOxp	19[b]	Permeate (10 components + quadratic and interaction terms)	Yes	31	31	91	4.0	0.91	0.86	1.00	0.99
Ptp	20[b]	All (6 components + quadratic and interaction terms)	Yes	22	5	86	0.48	0.86	0.82	1.00	0.97
MLSS	21[b]	Sludge (10 components + quadratic and interaction terms)	Yes	21	6	89	0.33	0.89	0.90	1.00	0.99

[a] Units: TMP in mbar, MLSS in g L^{-1} and the remaining in mg L^{-1}.
[b] Models represented in Figures 5.1-5.7.

The optimisation of the PLS models, by selection of the useful inputs to predict the output, resulted in a general improvement of fitting (reducing the RMSEP) and it decreased the complexity of the models. The best method for the selection of the useful inputs differed with the output modelled and with the initial inputs used. However, regardless the selection method chosen, optimisation was essential to improve, simplify and reduce noise of the PLS models. Figures 5.1-5.7 summarise the optimised models obtained for the different outputs, where observed data was plotted against predicted data and the confidence intervals of the models were established.

From the optimised models obtained, it is possible to infer about the relationships between the performance parameters modelled and the operating and analytical parameters used as predictors. Nevertheless, some prudence is needed when inferring about the correlation between inputs and outputs. The regression coefficients relative to some of the inputs showed very high standard deviations, meaning that, although these inputs are essential for the outputs prediction, the positive or negative correlation obtained may have or not a mechanistic interpretation related with the system's behaviour.

5.3.1. Modelling of transmembrane pressure

Transmembrane pressure is a parameter usually assessed on-line to monitor the membrane performance of MBRs operated at controlled, constant permeate flux. Since TMP is a fouling indicator, in the present study this parameter was modelled in search of correlations between fouling and operating conditions. Two optimised models were obtained for TMP prediction, one without fluorescence data as input (model #2) and another including the fluorescence spectra obtained in the permeate (model #4) (Table 5.2 and Figure 5.1). Figures 5.1a and 5.1b show the fitting of the models without and with fluorescence data, respectively. Figures 5.1c and 5.1d show the regression parameters of TMP models with the respective standard deviations calculated by the jackknife method (Duchesne and MacGregor, 2001) (results shown are normalised values). The model developed incorporating permeate fluorescence data (model #4) showed that the three components resultant from EEM compression (C1, C2 and C3) could contribute strongly for TMP prediction (Figure 5.1d). However, the overall model

fitting did not improve significantly with the inclusion of fluorescence data as compared to model #2, based only on the operating conditions (Table 5.2).

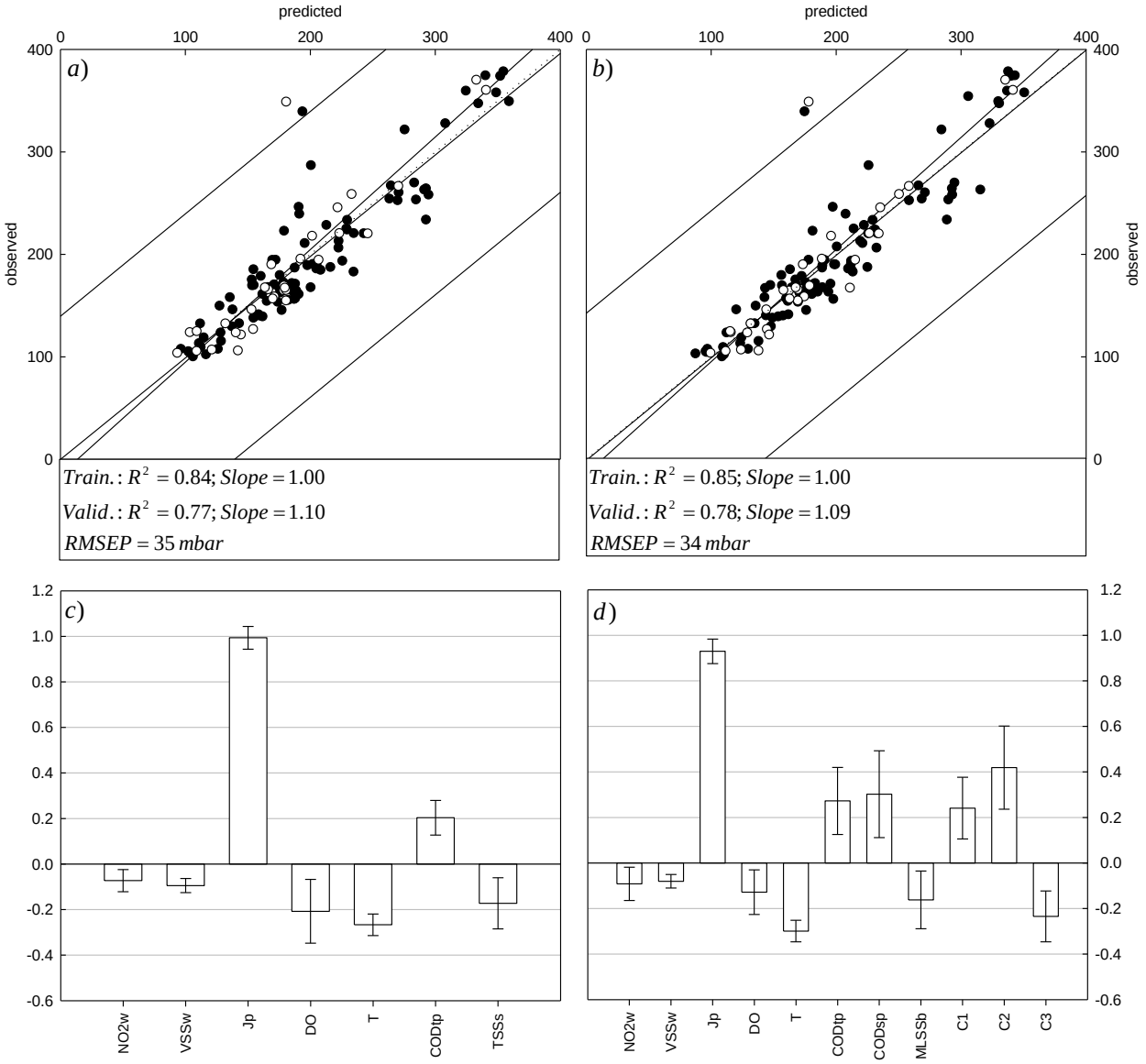

Figure 5.1. TMP prediction: a) without fluorescence data (model #2); b) with permeate fluorescence data (model #4). Regression coefficients of model inputs for TMP prediction: c) without fluorescence data (model #2); b) with permeate fluorescence data (model #4). Closed circles represent training data and open circles represent validation data, both in mbar. Regression coefficients are normalised.

Despite the fact that PLS models carried out in this study did not take into account neither the operating time elapsed nor the time that the membranes were operated at each permeate flux value, satisfactory TMP models were achieved. Even so, a couple of TMP data points were not properly predicted by any of the obtained models and are outside the confidence intervals (Figures 5.1a and 5.1b). A careful look into the

operational history showed that outlier TMP data corresponded to samplings performed after a long time of operation without membrane cleaning (chemical cleaning of membranes was performed immediately after the collection of these data points), where TMP increase was essentially due to fouling accumulation throughout operation time. These two outlier TMP values were in the same range of values occurring in other days (~350 mbar, Figure 5.1a), which were predicted by the models. However, the latter corresponded to situations of higher permeate fluxes and/or removable fouling, which could be reduced by physical means (air scouring) without requiring the permeation interruption and membrane backwashing or chemical washing. This suggests that both optimised TMP models (#2 or #4) could then be used to alert to the need of chemical cleaning through deviations from the model.

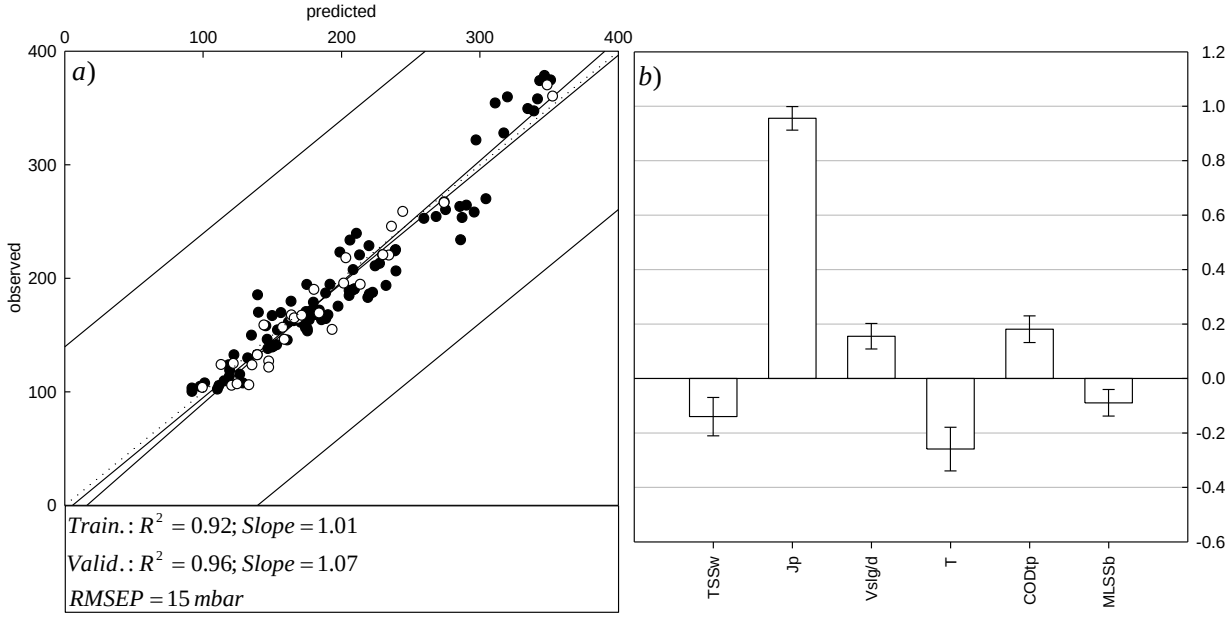

Figure 5.2. TMP prediction without fluorescence and without the outlier data (model #6): a) model fitting; b) regression coefficients of model inputs for TMP prediction. Closed circles represent training data and open circles represent validation data, both in mbar. Regression coefficients are normalised.

The removal of these two data points resulted in an improved model (model #6) (Table 5.2 and Figure 5.2) where the RMSEP decrease from 34.7 mbar to 14.7 mbar. Similarly to models #1-4, the incorporation of 2D fluorescence data did not improve model prediction in the absence of these outliers; indeed, the model optimisation process always resulted on the elimination of fluorescence as input (data not shown).

As expected, the highest contribution to TMP prediction comes from the permeate flux imposed (Jp) for all of the obtained models (Figure 5.1c, 5.1d and 5.2b). In fact, the positive correlation with permeate flux and negative correlation with operating temperature (T), obtained with PLS modelling, is a well-established fact for membrane permeation. Interestingly, the remaining inputs needed to predict TMP in both optimised models obtained with (model #2) and without (model #6) the TMP outliers, reflected the same operating conditions. For instance, biomass concentration correlated negatively with TMP in both models: where in model #2 (Figure 5.1c) the selected input was the total suspended solids in the membrane tank (TSSs), and in model #6 (Figure 5.2b) was the on-line measurement of the mixed liquor suspended solids (MLSSb) and the mean volume of activated sludge purged from the system per day (Vslg/d), which is also correlated (inversely) with the biomass concentration. The negative correlation between TMP and biomass concentration may result from the increased consumption of the organic matter by the higher number of microorganisms at higher suspended solids concentration. This organic matter could include colloids and particles that might contribute to fouling (Rosenberger et al., 2006). Previous studies reported in lieterature pointed out that mixed liquor suspended solids impact on membrane fouling can be either positive, negative or have no impact for high, low or intermediate MLSS values, respectively (Le-Clech et al., 2006; Rosenberger et al., 2005). The negative correlation found in the present study for a low MLSS range (4.6 - 8.7 g L^{-1}) corroborates these findings.

In general, the optimised models obtained for TMP prediction confirmed that mathematical correlations obtained with the statistically-based models developed in this work can be representative of the relationships between the system parameters (operational and/or analytical).

5.3.2. Modelling of chemical oxygen demand (total and soluble) in the permeate

Total and soluble COD in the permeate are two important biological performance indicators. Since particulate COD is retained by the membrane, it is expected that most of the effluent COD was due to either undegraded organic compounds or microbial products. Therefore, the difference between CODtp and CODsp consists of colloids that were able to permeate through the membrane, including extracellular polymeric substances (EPS) produced in the MBR. It is expected that the abundance of this

permeated colloidal fraction is related to the retained colloidal fraction, which is known to be a major fouling agent. Similarly, the soluble COD detected in the permeate is also likely reflective of the total soluble microbial products in the MBR, a part of which is retained in the cake layer formed on the membrane surface, contributing to fouling. Therefore, although total COD in the permeate is mainly in the soluble form (Table 5.1), both performance parameters are of interest in MBR monitoring.

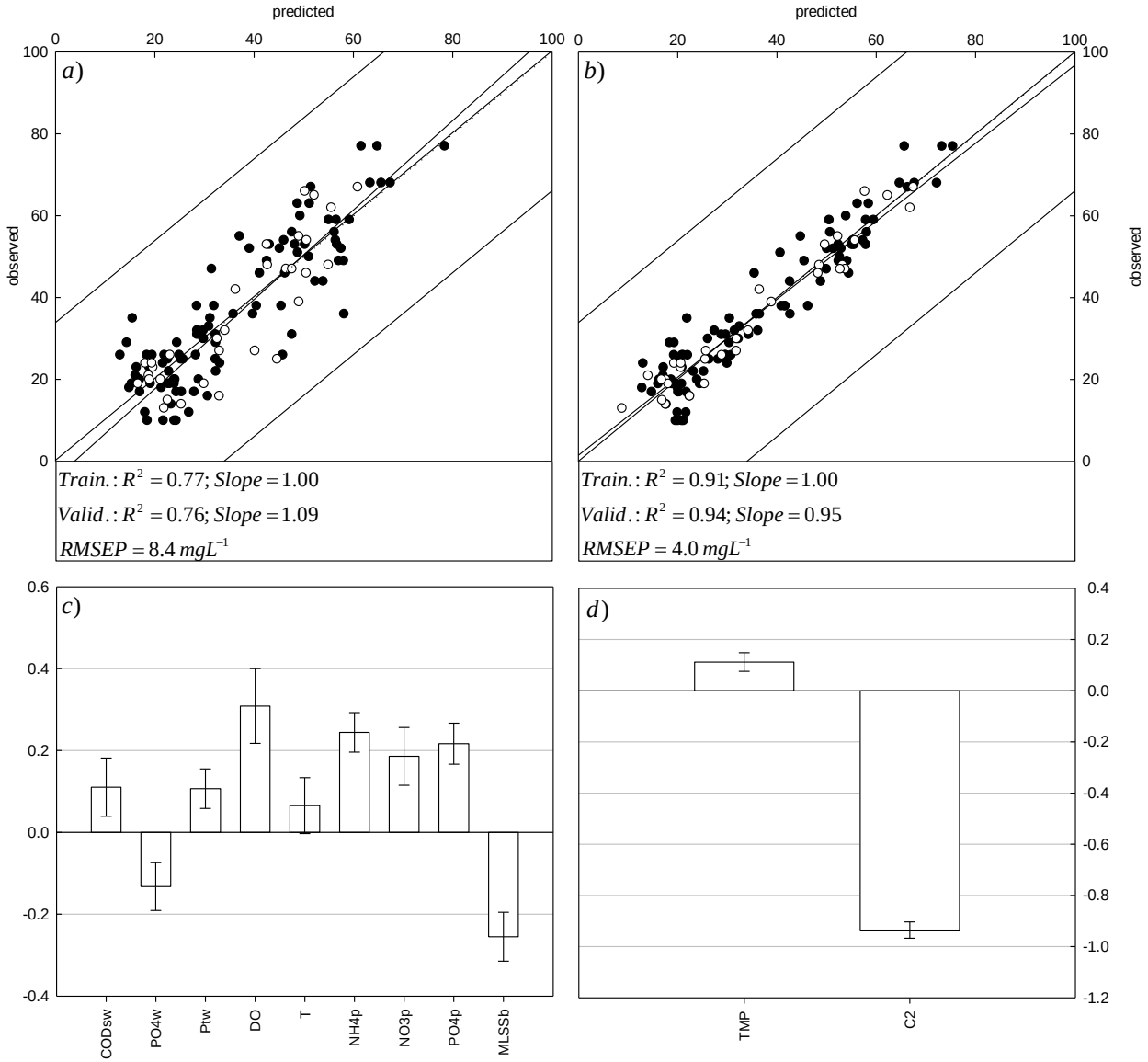

Figure 5.3. Prediction of the total COD in the permeate (CODtp): a) without fluorescence data (model #7); b) with permeate fluorescence data (model #9). Regression coefficients of model inputs for CODtp prediction: c) without fluorescence data (model #7); b) with permeate fluorescence data (model #9). Closed circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients are normalised.

Modelling of the total chemical oxygen demand in the permeate (CODtp) was firstly attempted using operating conditions and analytical parameters as inputs (Table 5.1)

and excluding fluorescence data (Figures 5.3a and 5.3c). The model obtained after optimisation (model #7) described poorly the total COD in the permeate, with fitting correlation coefficients (R^2) of 0.77 and 0.76 for training and validation respectively (see Table 5.2). However, with the addition of permeate fluorescence data as input information (model #9), the model was substantially improved, with R^2 of 0.91 and 0.94 for training and validation, respectively, and the RMSEP decreased from 8.4 to 4.0 mg COD L^{-1}. In the latter model, only two inputs were needed to describe the CODtp after optimisation: the TMP and the second component of compressed permeate fluorescence data (C2) (Figure 5.3d). Thus, the inclusion of fluorescence data as input also improved the model by reducing the number of total inputs needed, simplifying the final model.

Similarly to the total COD in the permeate (CODtp), the inclusion of 2D fluorescence data in PLS models substantially improved the model prediction of chemical oxygen demand in the permeate after precipitation of the colloidal matter or soluble COD (CODsp) (Figure 5.4). The optimised models obtained with fluorescence data (models #11 and 12) captured higher variance, had half of the error and better R^2 coefficients for both training and validation sets (Table 5.2).

Since the best model obtained for CODtp used only permeate fluorescence and TMP as inputs, other initial combination of inputs were attempted in PLS modelling for CODsp prediction (Table 5.1). This approach enabled the development of a model based only on the MLSSb acquired on-line, the imposed operating parameters and permeate fluorescence data. Figures 5.4b and 5.4d represent the optimised model (model #13), which showed a substantially better fitting than model #12, using all the analytical data (Table 5.2).

The high contribution of the fluorescence components to the models developed for both COD forms (CODtp and CODsp) suggests that the majority of the compounds contributing to the COD in the permeate are fluorescent or interfere with the fluorescence signal, and can be accounted for through 2D fluorescence measurements. This finding reinforces the results obtained in the previous studies (Chapter 3 and 4).

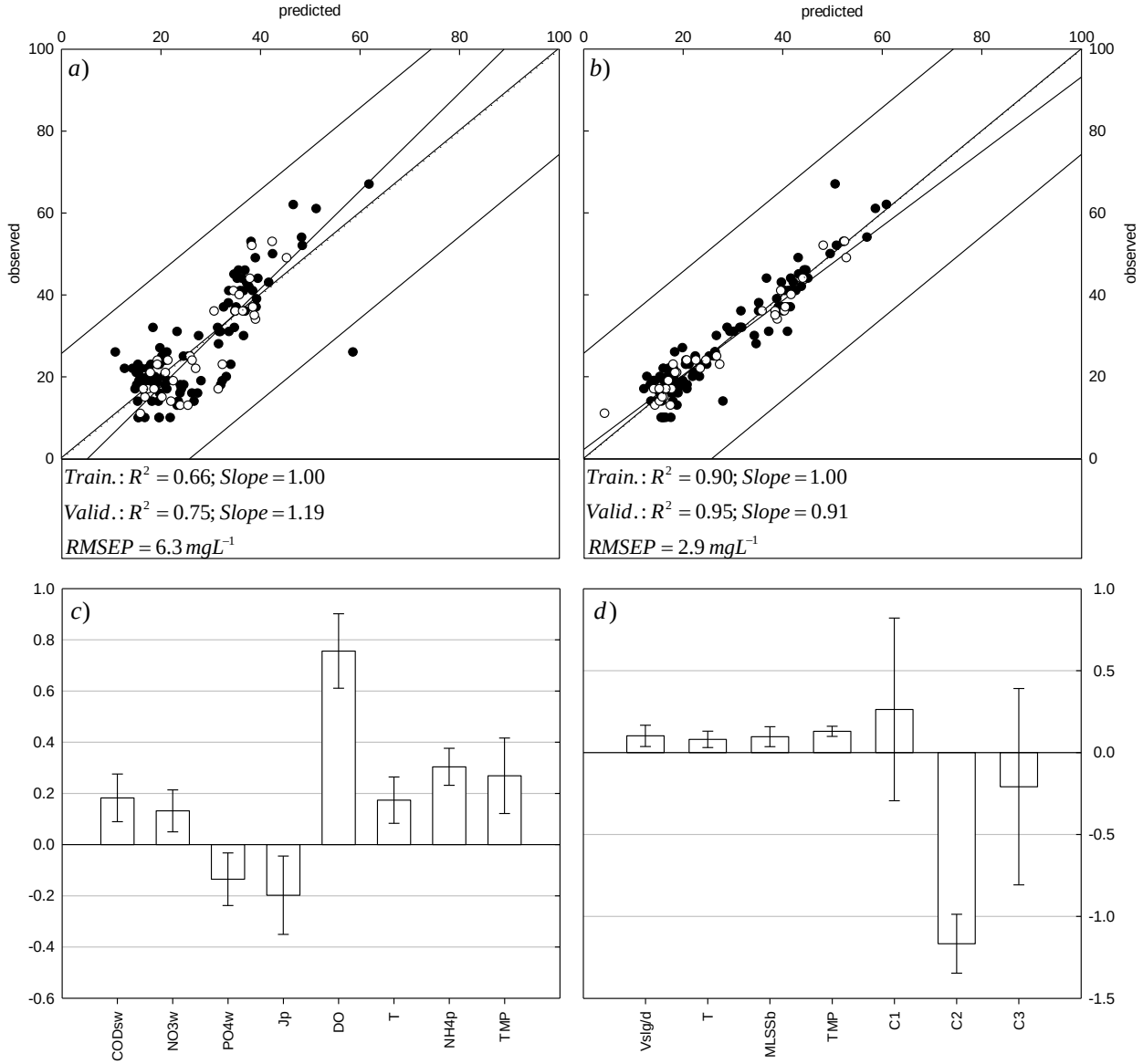

Figure 5.4. Prediction of the soluble COD in the permeate (CODsp): a) without fluorescence (model #10); b) with permeate fluorescence and on-line parameters (model #13). Regression coefficients of model inputs for CODsp prediction: c) without fluorescence (model #10); b) with permeate fluorescence and on-line parameters (model #13). Closed circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients are normalised.

Despite that in the present work 2D fluorescence measurements were performed at-line after sampling, the optical probe can be coupled directly to an MBR to perform on-line measurements (Wolf et al., 2007; Wolf et al., 2001). Furthermore, fluorescence spectra can be easily acquired with only one fluorescence spectrophotometer, using different optical probes in multiple locations of the system and time-programmed with the help of an optical 'switch-box'. Other parameters of the optimised models are also acquired on-line (TMP, temperature (T) and MLSSb) or imposed (sludge purged from the system

(Vslg/d)), meaning that, after calibration of the models developed, the total and soluble COD in the permeate can be monitored on-line, *in situ* and without laboratory work or reagents consumption.

Although modelling of both total and soluble COD in the permeate without using fluorescence as input resulted in models with lower predictive capability, as compared to those incorporating fluorescence data, some relationships are noteworthy (Figures 5.3c and 5.4c). For instance, a strong positive correlation was obtained between the COD in the permeate and the dissolved oxygen in the biological tank (DO). This correlation may result from shear forces promoted by increased aeration that could increase floc disaggregation, and EPS solubilisation, as well as cell lysis. Additionally, both models obtained for total and soluble COD without fluorescence show a relationship with the influent wastewater characteristics (positive correlation between influent and permeate COD), which was expected and might result from incomplete COD removal or increased EPS production at higher influent COD. In the same way, some other characteristics of the permeate (NH_4, NO_3 and PO_4) correlated positively with the total COD, possibly linked by the overall wastewater treatment performance of the MBR.

Another interesting correlation occurred between the permeate flux (Jp) and soluble COD in the permeate, CODsp (Figure 5.4c). This negative correlation with the CODsp (i.e. after precipitation of the colloidal fraction) may be due either to the higher production of EPS at lower hydraulic retention time, or to distinct retention and membrane selectivity due to different fouling mechanisms occurring at different permeate fluxes. Finally, Figure 5.3c shows a negative correlation between the MLSSb and total COD in the permeate, suggesting that the organic content in the permeate is mainly due to unconsumed COD, which would increase with decreasing biomass concentration. However, MLSSb correlated positively with the soluble COD in the permeate (Figure 5.4d), showing that despite the overall decrease in total COD, the soluble fraction increased with the biomass concentration, perhaps meaning it is mostly composed of products resulting from biological activity.

From the COD models achieved with fluorescence it is possible to see not only a strong correlation with the second component of compression of the permeate EEMs, but also

a positive correlation of both total and soluble COD in the permeate with the TMP (Figures 5.3d and 5.4d). This relationship shows a link between membrane fouling and the permeate composition. In model #13 (Figure 5.4d), CODsp also correlated positively with temperature, which can be due to the microbial activity and/or to the increase of membrane permeability at higher temperature.

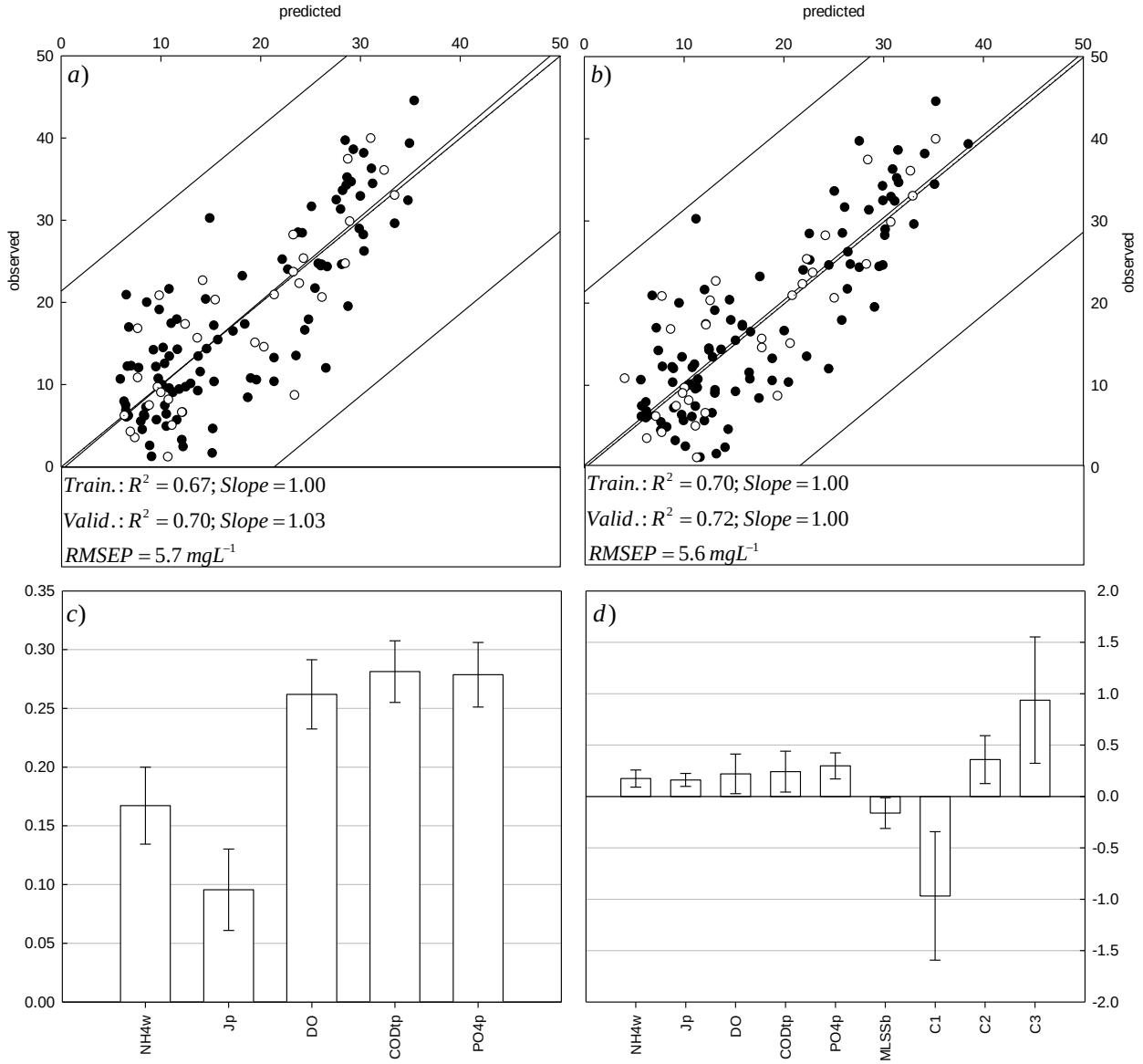

Figure 5.5. Prediction of the total nitrogen in the permeate (Ntp): a) without fluorescence (model #14); b) with 3 components of permeate EEMs compression (model #15). Regression coefficients of model inputs for Ntp prediction: c) without fluorescence (model #14); b) with 3 components of permeate EEMs compression (model #15). Closed circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients are normalised.

5.3.3. Modelling of nitrogen in the permeate

Two different nitrogen parameters were successfully modelled: total nitrogen in the permeate (Ntp) and nitrite and nitrate concentration in the permeate (NOxp). Different data combinations were used to develop PLS models for prediction of the Ntp. Despite the good correlations obtained for both COD measurements in the permeate, modelling of total nitrogen either without or with fluorescence (models #14 and 15, respectively) resulted in weaker correlations (Table 5.2 and Figure 5.5). The best model achieved using fluorescence data incorporated permeate fluorescence matrices compressed into 3 components (C1, C2 and C3).

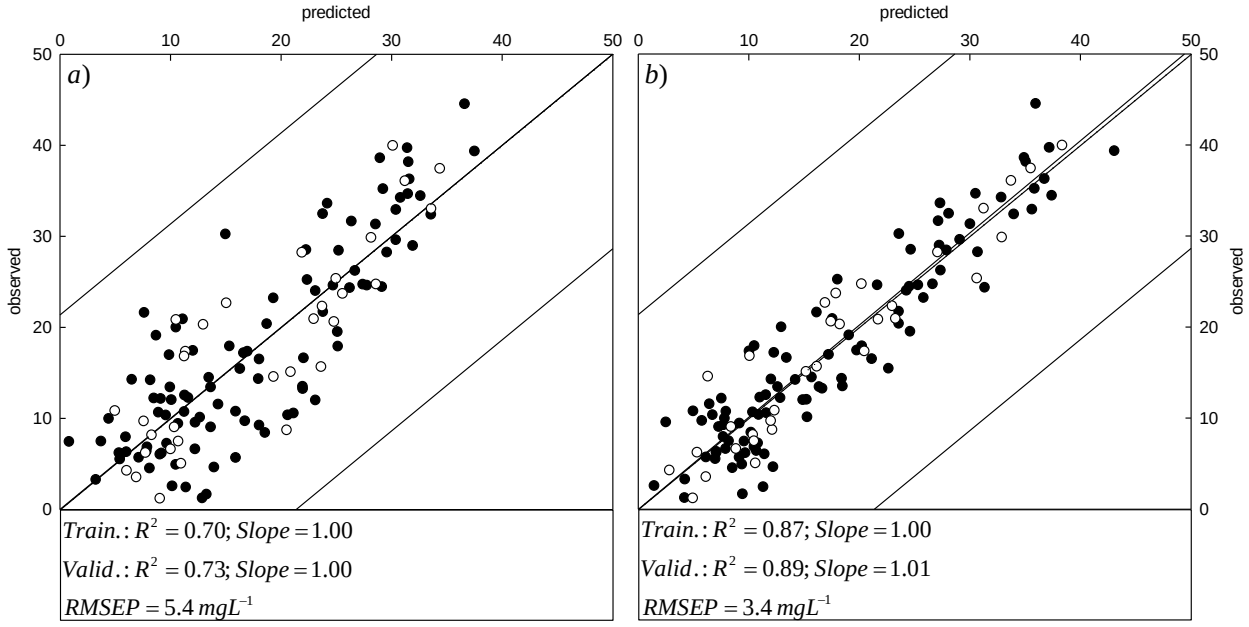

Figure 5.6. Prediction of total nitrogen in the permeate (Ntp): a) with 10 components of permeate EEMs compression (model #16); b) with 10 components of permeate EEMs compression plus their quadratic and interaction terms (model #18). Closed circles represent training data and open circles represent validation data, both in mg L^{-1}.

The models obtained showed that compression of permeate EEMs using only 3 components are likely not enough to include all the information related with nitrogen. Indeed, nitrogen should be captured by 2D fluorescence of amino acids and proteins (as fluorophores), urea (due to colour interferences) and due to ammonia (as a fluorescence quencher). The combined interference of these compounds over the fluorescence spectra is complex and probably not totally captured in the three first PARAFAC components. Therefore, permeate fluorescence was also compressed with 10 components and used together with operating and analytical data to model total nitrogen in the permeate (model #16, Table 5.2, Figure 5.6a). The inclusion of more components in the PLS

models gave additional information for nitrogen prediction, resulting in a slight improvement of the PLS models, with lower RMSEP and higher R^2 coefficients (Table 5.2, models #14-16). However, even using more information from permeate fluorescence, PLS linear correlations did not result in a good and strong model, as R^2 coefficients were below 0.8.

When a simple multilinear approach is not sufficient to model and predict an output, a more complex modelling tool may be considered. Consequently, interaction and quadratic terms (that reflect the crossed interferences between inputs) of the compacted permeate fluorescence matrix (either with 3 and 10 components) were introduced as inputs in the multilinear PLS models, resulting in more complex, non-linear models (models #17 and 18). This modelling approach resulted in better model fitting for total nitrogen prediction (Table 5.2, Figure 5.6b). The best optimised model required fluorescence EEMs compression with 10 components (model #18) and had better R^2 coefficients (0.87 and 0.89 for training and validation, respectively) and a much lower RMSEP, 3.4 mg L^{-1} than the regular multilinear PLS. Improvement of PLS models with the incorporation of quadratic and interaction terms reveals that the complex correlations between the convoluted information enclosed in fluorescence spectra are not linearly correlated with the concentration of total nitrogen in the permeate.

Modelling of nitrite and nitrate concentration in the permeate (NOxp) was also initially attempted using the multilinear PLS modelling with and without fluorescence data. However, a good model prediction (with R^2 coefficients of 0.91 and 0.86 for training and validation, respectively) was only achieved with the inclusion of quadratic and interaction terms of the 10 compression components of the permeate EEMs in the PLS modelling (Table 5.2, Figure 5.7a). Like for total nitrogen, the complex information and crossed interferences present in permeate spectra associated with nitrate and nitrite concentration required the use of non-linear correlations to fully extract that information.

Furthermore, the optimisation of the models obtained for both nitrogen measurements (Ntp and NOxp) with quadratic and interaction terms from fluorescence spectra of the permeate resulted in the elimination of all inputs except the fluorescence data and temperature. Therefore, similarly to the COD measurements in the permeate, 2D

fluorescence spectroscopy can be used as a monitoring tool to predict Ntp and NOxp on-line, *in situ* and without laboratory work or reagents consumption.

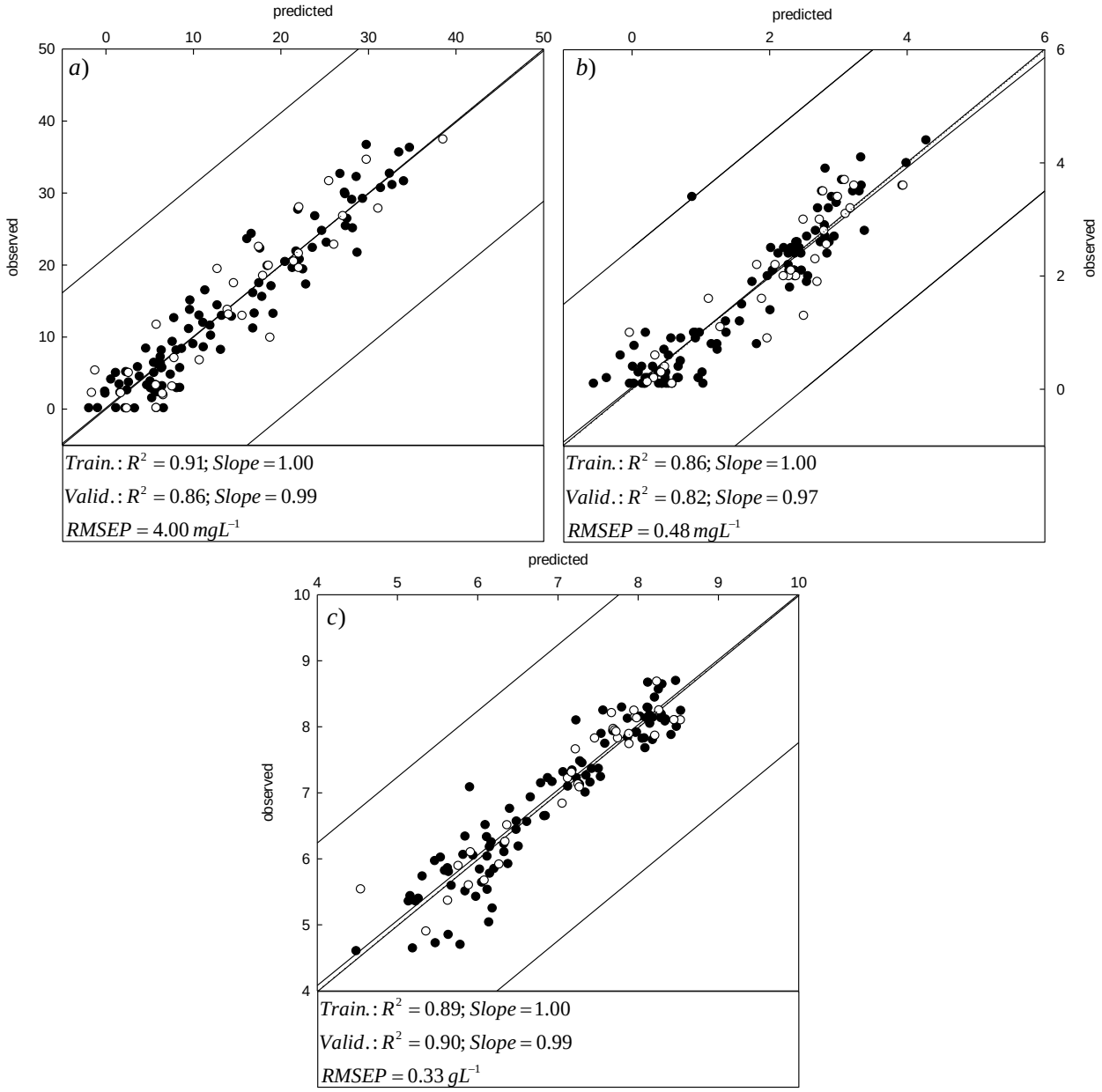

Figure 5.7. Prediction of: a) NOx in the permeate with 10 components of permeate EEMs compression plus their quadratic and interaction terms (model #19); b) total phosphorus in the permeate (Ptp) with 6 components of all EEMs compression plus their quadratic and interaction terms (model #20); c) MLSS with 10 components of sludge EEMs compression plus their quadratic and interaction terms (model #21). Closed circles represent training data and open circles represent validation data, both in mg L^{-1}.

5.3.4. Modelling of total phosphorus in the permeate

Similarly to the previous outputs, to predict total phosphorus in the permeate (Ptp), PLS modelling was attempted using alternatively the wastewater fluorescence matrices,

sludge fluorescence matrices, permeate fluorescence matrices or all matrices combined together. The best obtained model incorporated the information contained in all fluorescence spectra compressed together, but with R^2 coefficients of only 0.74 and 0.72 for training and validation, respectively, and RMSEP of 0.60 mg L^{-1}. Therefore, a more complex, non-linear modelling approach was followed to deconvolute the information from fluorescence matrices. The quadratic and interaction terms of the 6 components resulting from compression of the EEMs of wastewater, sludge and permeate were incorporated as inputs in PLS modelling. With this strategy, a better prediction of total phosphorus in the permeate was achieved, with lower RMSEP (0.48 mg L^{-1}) and better R^2 coefficients (0.86 and 0.82 for training and validation, respectively) (Table 5.2, Figure 5.7b).

Despite the good models obtained for the previous outputs using only permeate EEMs, the best models achieved for Ptp prediction required the three measurements of fluorescence. This suggests that fluorescence from the permeate might not capture the information from all the different phosphorus compounds, but with the addition of fluorescence information from influent wastewater and activated sludge, it was possible to predict total phosphorus concentration in the permeate.

Figure 5.7b shows an outlier value that PLS modelling was not able to predict. This value is in the same range of values occurring in other days and there was nothing remarkable in the operational history. Therefore, the outlier may have occurred due to a variation in the phosphorus concentration in the permeate that is not reflected by any of the inputs parameters (or due to an analytical error when analysing phosphorus off-line).

Although a good fitting was achieved with the incorporation of quadratic and interaction terms, the analysis of the regression coefficients is not feasible due to the complexity and high number of inputs used to model the total phosphorus in the permeate.

5.3.5. Modelling of mixed liquor suspended solids

One of the major concerns in biological wastewater treatment is the sludge production. Therefore, additionally to the permeate quality parameters, the mixed liquor suspended solids (MLSS) in the biological tank was also modelled.

The complexity of the 2D fluorescence spectra acquired from the sludge samples is caused not only by the broad range of compounds present in the medium but also by the high microbial cell density and high turbidity of the medium. Therefore, it was necessary to incorporate the quadratic and interaction terms of 10 compression components of sludge EEMs into PLS modelling to deconvolute the 2D fluorescence spectra of sludge and predict the MLSS. The model then obtained presents a good fitting with R^2 coefficients of 0.89 and 0.90 for training and validation, respectively, and RMSEP of 0.33 g L^{-1} (Table 5.2, Figure 5.7c).

Additionally, the optimised model obtained for MLSS prediction only required as inputs the permeate flux and dissolved oxygen combined with the 2D fluorescence data of the sludge, which are all monitored on-line. In fact, in MBRs for wastewater treatment, the MLSS is already a parameter usually monitored on-line; however the use of 2D fluorescence as a monitoring tool has the potential to incorporate the screening of several performance parameters simultaneously in the same tool.

5.4. CONCLUSIONS

The PLS models obtained prove that the performance of a MBR can be monitored by 2D fluorescence spectroscopy combined with a few other process parameters through statistically-based modelling. Moreover, the correlations between the input and output parameters obtained through this approach are significant and they highlight the relationships between operating and performance variables, which can be used for process control and optimisation.

Three MBR performance parameters were successfully modelled using the a multilinear correlation: TMP, CODtp and CODsp. With the incorporation of 2D fluorescence data in PLS modelling, both CODtp and CODsp predictions were improved.

When multilinear PLS was not sufficient to describe complex data correlations, particularly from the fluorescence matrices, it was possible to incorporate quadratic and interaction terms of the compressed EEMs to improve model prediction. Thereby, Ntp, NOxp, Ptp and MLSS models were achieved with good fitting. The good results obtained with the incorporation of quadratic and interaction terms of EEMs

compression demonstrates the complexity of the information contained in the fluorescence spectra and reinforces the capability of the mathematical tools used to extract the required information.

Furthermore, it was shown that both soluble and total COD in the permeate, both nitrogen parameters in the permeate (Ntp and NOxp) and MLSS may be predicted based fully on on-line data (including 2D fluorescence data) and imposed operating parameters. The use of 2D fluorescence has the additional advantage of integrating the monitoring of several performance parameters of MBR for wastewater treatment.

This work shows that 2D fluorescence spectroscopy deconvoluted by adequate mathematical tools is a promising, simple monitoring technique that can facilitate monitoring and optimisation of MBRs performance.

Chapter 6

DEVELOPMENT OF A HYBRID MODEL STRATEGY FOR MONITORING MEMBRANE BIOREACTORS

SUMMARY

In the present study, the performance of a membrane bioreactor (MBR) was modelled using a hybrid approach based on the activated sludge model number 3 (ASM3) combined with projection to latent structures (PLS) to predict the residuals of the ASM. The application of ASM to MBRs requires frequent re-calibration to adjust the model to variations in influent characteristics, determined through time-consuming analysis and batch tests. Considering this problem, the objective of this study was to improve ASM prediction ability with minimal additional monitoring effort. Hybrid models were developed to predict three MBR performance parameters: mixed liquor suspended solids (MLSS), COD in the permeate (CODp) and nitrite and nitrate concentration in the permeate (NOxp). For PLS modelling of ASM residuals three input strategies were used: 1) analytic and operating data; 2) operating data plus 2D fluorescence spectroscopy; 3) all the data. The first input strategy improved ASM prediction of the three selected outputs, and highlighted the lack of detailed and real-time information from wastewater and operating parameters in the ASM used in this study. In the second input strategy, the incorporation of updated data from 2D fluorescence spectroscopy resulted on better model fitting than in the first input strategy, for all the output parameters studied. Through the hybrid modelling approach it was possible to significantly improve the ASM predictions in real-time using 2D fluorescence measurements and other relevant parameters acquired on-line, without requiring further laboratory analysis. Furthermore, the third input strategy, incorporating all the collected data, did not significantly improve the prediction of the outputs beyond the second strategy. This shows that 2D fluorescence spectroscopy is a comprehensive monitoring tool, able to capture on-line the required information to complement, through hybrid modelling, the mechanistic information described by an ASM.

Submitted as: Galinha, C.F., Carvalho, G., Portugal, C.A.M., Guglielmi, G., Crespo, J.G. and Reis, M.A.M., 2012. Development of a hybrid model strategy for monitoring membrane bioreactors. Journal of Biotechnology.

6.1. INTRODUCTION

Membrane bioreactors (MBR) for wastewater treatment consist in an activated sludge process coupled with a membrane filtration process to separate the treated effluent from the activated sludge. MBRs have been increasingly employed for industrial and municipal wastewater treatment for their numerous advantages in respect to conventional activated sludge systems (CAS), such as compactness, high quality effluent and flexibility in treating problematic wastewaters (e. g. with foaming problems) (Judd, 2008).

Despite the progress achieved in membrane technology in the recent years, the use of a MBR for wastewater treatment still involves high energy consumption (mainly for aeration used for membrane scouring and/or sludge recirculation, depending on the process configuration) and high operational costs due to membrane fouling, which requires tight monitoring and control (Le-Clech et al., 2006; Lesjean et al., 2011; Meng et al., 2009). In order to maintain membrane permeability to a sustainable level, the permeate flux is often carried out with a membrane-relaxation or a backwash mode. Moreover, membrane fouling has to be removed frequently, in a process that involves stopping reactor operation to physically and/or chemically wash the accumulated material. It is desirable that these operation interruptions are as spaced in time as possible in order to maximise the flow of wastewater that is treated. Therefore, further acceptance of MBR technology depends on developing automatic monitoring and control tools that are easy to implement and operate. Mathematical models that accurately describe the MBR performance are necessary tools of such monitoring and control systems.

The activated sludge systems have been studied in detail for many years, resulting in deep knowledge about the main heterotrophic and autotrophic biological processes. Four activated sludge models (ASM) were developed by the International Water Association (IWA) Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment to describe the kinetics and stoichiometry of the different processes occurring in these systems (Henze et al., 2000). The first activated sludge model published, ASM1, was developed to model biological treatment for organic carbon removal, nitrification and denitrification. ASM1 predicts oxygen demand and sludge production in an activated sludge system. Later, ASM2 and ASM2d

were developed to include the phosphorus removal and organic carbon storage, respectively. ASM3 was the last to be developed and, although it does not include phosphorus removal, it addresses some problems found in the first model, ASM1, such as the inclusion of internal cell storage compounds in heterotrophs (shifting the focus from hydrolysis to the storage of organic substrates) and the replacement of the death-regeneration concept by the growth-endogenous respiration model.

Nevertheless, ASM were specifically developed to describe activated sludge processes under the typical operating conditions of the well-known CAS systems. Therefore, in order to apply ASM to model the biological process occurring in MBRs, some specific characteristics of the MBRs, which are different from CAS, must be taken in account: higher biomass concentration and viscosity, with lower biomass production; higher solids retention time; accumulation of influent non-biodegradable solids and microbial products larger than the membrane molecular cut-off; high aeration rates for scouring of the membranes (Fenu et al., 2010; Ng and Kim, 2007).

To date, ASM have been applied to MBRs using two different approaches: the so-called plain or unmodified ASM, where differences between MBRs and CAS were only reflected in the parameters estimated to match model prediction with experimental data; or introducing major modifications by adding state variables to the ASM state vector (e.g. soluble microbial products, which play a major role in MBRs) (Fenu et al., 2010). Despite the overall improved fittings of the latter approach, a simple model is preferable (Fenu et al., 2010), since the addition of new variables implies additional laboratory analyses, some of which still do not have universally accepted analytical methods. An alternative to a complex mechanistic model (such as a modified ASM) could be a hybrid approach to improve ASM for MBR modelling. In fact, combined approaches between mechanistic and non-mechanistic models for chemical and biological processes were previously reported as advantageous either by expanding the mechanistic model to situations beyond its building assumptions, or by extrapolation of the non-mechanistic model to regions lacking calibration data (Duarte and Saraiva, 2003; Ricardo et al., 2012; Thompson and Kramer, 1994).

In the present study, a hybrid model is proposed, where a plain ASM is combined with a statistically-based model to predict the residuals from the mechanistic model. Hybrid

modelling with statistically-based models can use common analytical tools and/or on-line measurements to improve prediction from a plain ASM. Therefore, the selection of the inputs needed to predict residuals with a statistically-based model can, simultaneously, highlight missing information in the mechanistic model.

In the previous studies, 2D fluorescence spectroscopy showed high potential as an on-line monitoring technique able to capture a broad range of information from complex biological media, including biologic wastewater treatment systems. In the present work, 2D fluorescence was used as a tool to improve the predictions obtained with a plain ASM. Through the incorporation of 2D fluorescence measurements in hybrid models, an on-line correction of the ASM could be automatically done, without any additional off-line analytical measurements for ASM re-calibration. This study presents the development of a hybrid model based on ASM3 combined with projection to latent structures (PLS) to predict the residuals of the ASM. The model was developed to predict the performance of an MBR for wastewater treatment, i.e. the ability of the system to remove major macro-pollutants (carbonaceous and nutrients) from wastewater with minimal sludge production.

6.2. MATERIALS AND METHODS

6.2.1. Membrane bioreactor set-up and operation

A pilot scale MBR for domestic wastewater treatment, located in the wastewater treatment plant of Lavis (Trento), in Italy, was monitored for a period of approximately 15 months. The MBR system was composed by a biological tank divided in anoxic and aerobic compartments (4.7 and 8.7 m^3, respectively), followed by a 1.5 m^3 tank where a hydrophilised PVDF ultrafiltration membrane module (GE Zenon ZW500d hollow fibre module with 0.04 μm pore size and 100 m^2 area) was immersed. The excess sludge in the membrane tank was recirculated to the anoxic compartment of the biological tank at a recirculation ratio (ratio between the recirculation flow and permeate flow) ranging between 2 and 3. Membrane permeation was performed by alternating suction and relaxation phases of 9 min and 1 min respectively, under controlled permeate flux. The MBR plant was operated at sludge retention time (SRT) between 15 and 60 days, with mixed liquor suspended solids (MLSS) in the biological tank and influent wastewater characteristics ranging as shown in Table 6.1. During the study period, several operational changes were imposed in the permeate flux, SRT, temperature (due to

seasons' weather), hydraulic retention time (HRT) and dissolved oxygen concentration (DO).

Table 6.1. Range of values of wastewater characteristics and mixed liquor suspended solids during both calibration and validation periods of the ASM.

Parameter	Range of operation in mg/L	
	50 days of ASM calibration	400 days of validation
Wastewater:		
Total COD (CODtw)	165 - 1760	151 - 1862
Soluble COD (CODsw)	24 - 484	17 - 410
COD after filtration (CODfw)	64 - 537	46 - 467
Amonia (NH4w)	19.8 - 63.8	9.1 - 88.5
Nitrite (NO2w)	0.01 - 4.20	0.01 - 9.98
Nitrate (NO3w)	0.1 - 6.0	0.1 - 26.8
Organic nitrogen (Norgw)	7.6 - 80.2	0.1 - 89.0
Phosphate (PO4w)	0.1 - 5.7	0.1 - 7.35
Total phosphorus (Ptw)	2.7 - 20.2	1.2 - 33.0
Total suspended solids (TSSw)	130 - 1700	50 - 2100
Volatile suspended solids (VSSw)	115 - 1370	30 - 2071
Sludge:		
Mixed liquor suspended solids in the biological tank (MLSSb)	5400 - 10200	4266 - 9491

6.2.2. Sampling and chemical analysis

Samples were collected simultaneously in the wastewater fed to the MBR, permeate effluent and mixed liquor recirculation. Suspended solids were measured in wastewater and sludge samples, and influent and effluent characteristics for main macro-pollutants were measured according to the APHA Standard Methods (APHA, 1998). Wastewater samples were filtered through a filter with a pore size of 0.45 μm. Three different fractions of chemical oxygen demand (COD) were measured in wastewater samples (total, soluble after filtration and soluble after flocculation), while only total COD and soluble COD after flocculation were determined for permeate samples. The flocculation method used in wastewater and permeate samples is described in Mamais et al. (1993) and deduces the COD contribution of the colloidal particles that normally pass through 0.45 μm filters. COD, ammonia, nitrate, nitrite, total nitrogen, phosphate and total phosphorus were measured in wastewater and permeate samples, whereas total suspended solids (TSS) and volatile suspended solids (VSS) measurements were performed for wastewater and sludge samples. Additionally, the following parameters were continuously acquired on-line: transmembrane pressure (TMP), temperature (T), dissolved oxygen (DO) and mixed liquor suspended solids in the biological tank (MLSSb).

6.2.3. 2D fluorescence spectra

2D fluorescence measurements were performed immediately after sample collection by immersion of an optical fibre probe in a beaker that was stirred, in order to avoid sedimentation of suspended solids present in wastewater and activated sludge samples. Identical stirring conditions were applied to all the samples (wastewater, activated sludge and permeate) during the acquisition of the fluorescence spectra. Fluorescence spectra were acquired with a fluorescence spectrophotometer Varian Cary Eclipse equipped with excitation and emission monochromators and coupled to a fluorescence optical fibre bundle probe. Fluorescence spectra were generated in an excitation wavelength range of 250 to 700 nm (with an incrementing step of 10 nm) and an emission wavelength range between 260 and 710 nm, using excitation and emission slits of 10 nm and a scan speed of 3000 nm min^{-1}.

6.3. HYBRID MODEL DEVELOPMENT

6.3.1. Development of the ASM

Brief description of the model

An extended version of ASM3 (ASM3e for simplicity) was used as mechanistic model. The approach, originally proposed by Sin et al. (2005a), extends the conventional ASM3, assuming biomass growth to occur on both readily biodegradable substrate (S_S) and storage products (X_{STO}). In terms of biochemical stoichiometry, two different pathways for biomass growth result in three yield coefficients, namely $Y_{H,S}$ (heterotrophic direct growth on S_S), Y_{STO} (conversion of S_S into storage compounds X_{STO}) and $Y_{H,STO}$ (heterotrophic growth on stored substrates X_{STO}). Such stoichiometric coefficients are related to each other by means of the efficiency of oxidative phosphorylation δ (mol_P mol_O^{-1}), according to the metabolic model proposed by van Aalst-van Leeuwen et al. (1997). The pilot plant layout was coded into AQUASIM platform (Reichert, 1994), with the following as input data:

- Influent flowrate and composition;
- Influent COD fractions according to the ASM3e, routinely assessed by means of respirometric batch tests;
- Operational conditions, such as temperature in the biological tank and SRT.

A sensitivity analysis was carried out in order to assess the most suitable parameters for model calibration; given the overall target of the hybrid model in this study, effluent

nitrogen (ammonia and oxidised forms), effluent COD and sludge production were chosen as output functions. The absolute-relative sensitivity method revealed the following to be the most significant parameters to be calibrated:

- maximum growth rates of heterotrophs on both readily biodegradable substrate and storage products ($\mu_{H,S20}$ and $\mu_{H,STO20}$, respectively) and
- maximum growth rate of autotrophs (μ_{A20}).

Model calibration

An initial period of 50 days was considered for the ASM calibration. Data used for calibration included off-line 24h-grab samples of permeate and on-line MLSS concentration. A trial-and-error procedure was used to estimate the biokinetic parameters. As a result of the calibration process, the following values were obtained and subsequently used for model validation:

- $\mu_{H,S20} = 0.89\ d^{-1}$
- $\mu_{H,STO20} = 0.59\ d^{-1}$
- $\mu_{A20} = 0.40\ d^{-1}$

After the initial model calibration, the ASM3e was used throughout a 400-day validation period in order to predict biomass and effluent characteristics. During an ASM validation, an input data vector is supplied to the model in order to obtain the output prediction corresponding to the current system status. Therefore, the application of this model was dependent on the acquisition of updated state data, which, in this study, consisted on the measurement of the wastewater characteristics every two or three days throughout the 400 days period of validation.

6.3.2. Development of PLS models

One hundred and thirty experimental observations, made after the 50 days period of ASM3e calibration, were used to develop the hybrid model. The differences between the ASM predictions and experimental measurements were calculated, and a projection to latent structures (PLS) technique was used to predict these residuals. Therefore, for each output, the hybrid model is the result of the ASM3e prediction plus the PLS prediction of the respective residual.

PLS models were trained using 75% of the experimental data collected, and the remaining 25% were used for validation. Training and validation sets of data were randomly chosen throughout the operation period used for ASM validation. For each output, three different input strategies were used (Figure 6.1): 1) using analytic and operating data (aiming to understand what type of information is missing in the ASM model); 2) using operating data plus fluorescence spectra (aiming at obtaining a model that does not require the acquisition of new analytical data, i.e. an on-line prediction); 3) using all data collected (to achieve the best fitting possible).

Previously to PLS modelling, the excitation-emission matrices (EEMs) obtained through 2D fluorescence spectroscopy were analysed using a PARAFAC (parallel factor analysis) function in order to reduce the number of inputs and remove noise. PARAFAC was used to compress the EEMs of wastewater, sludge and permeate separately, with more than 99% of variance captured in all cases. The parameters resultant from fluorescence compression are referred further in this work as compression components (C1, C2,...). Each of the three types of EEM (wastewater, sludge or permeate) was incorporated into the PLS models either individually or combined altogether, in which case they were also compressed together to eliminate co-linearity and redundancy of information. For simplicity, only the best model out of these four possibilities is presented for each output and modelling strategy.

In PLS models the number of latent variables and the selection of the best models were made based on the lower root mean square error of prediction (RMSEP), or, when the RMSEP results were not conclusive, by the R^2 coefficients and the slopes of the linear correlation between predicted *vs* observed values. The models presented in this work were obtained after selection of useful inputs for each output prediction. Four different mathematic methods were used for input elimination: iterative stepwise elimination (ISE) (Boggia et al., 1997), iterative predictor weighting (IPW) (Forina et al., 1999), stepwise elimination (Ryan, 1997) and by the Martens uncertainty test (Forina et al., 2004) using the jackknife standard deviations (Duchesne and MacGregor, 2001). For each output and each modelling approach the four methods were applied and the best models chosen.

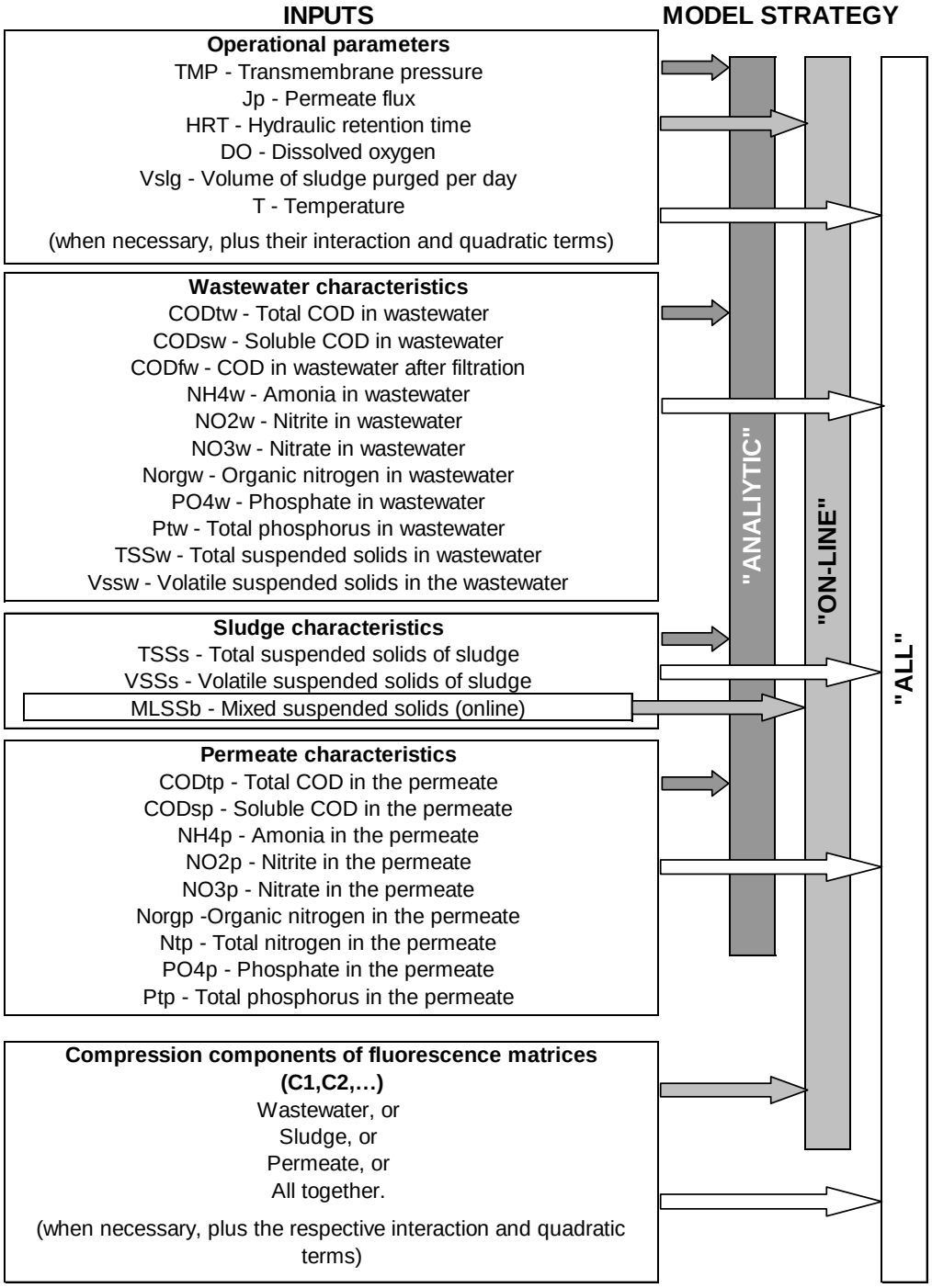

Figure 6.1. Different input combinations used in PLS modelling for prediction of MLSS, CODp and NOxp residuals from ASM.

When a simple PLS multilinear correlation was not sufficient to predict an output, the quadratic and interaction terms of fluorescence compression components were added to PLS models to account for non-linear correlations with fluorescence data. Similarly, the non-linear correlations between the outputs and the operating conditions (temperature (T), transmembrane pressure (TMP), dissolved oxygen (DO), hydraulic retention time

(HRT), permeate flux (Jp) and volume of sludge purged per day (Vslg/d)) were also added by the inclusion of the quadratic and interaction terms of these operating conditions in PLS models, when required to obtain better model prediction.

6.4. RESULTS AND DISCUSSION

6.4.1. Activated sludge models

Mechanistic modifications of ASM aiming at fitting the specific MBR biological processes are usually oriented for research work (requiring additional and laborious bench tests) and hard to implement for practical use. Therefore, for practical applications, a simpler activated sludge model is preferred.

In the present study, an extended version of the ASM3, ASM3e (Sin et al., 2005a), that assumes that biomass growth occurs both on readily biodegradable substrate and on storage products, was applied directly to model the activate sludge performance of an MBR, without any further modification to membrane systems. The model developed aimed at predicting the performance of an MBR for wastewater treatment. Therefore, three outputs descriptive of the MBR sludge production and effluent quality were studied: mixed liquor suspended solids (MLSS), chemical oxygen demand in the permeate (CODp) and nitrite and nitrate in the permeate (NOxp). Figure 6.2 shows the values of MLSS, CODp and NOxp predicted by the ASM3e plotted against the respective observed values for the 130 experimental observations obtained during the ASM validation period, corresponding to those used for PLS modelling. The ASM used did not result in a good prediction for the three outputs, particularly after the initial 50-70 days of the validation period (data not shown). MLSS and NOx prediction had not only low R^2 of validation (of 0.13 and 0.55, respectively) and high RMSEP (1.1 g L^{-1} for MLSS and 8.3 mg N L^{-1} of NOx), but also the slopes of predicted *vs* observed values were clearly different from 1. ASM prediction for COD had the best slope (1.04) of all outputs, showing that the COD model captured the trend of the experimental data, but with high dispersion of values (ASM predicted the same COD value for multiple observation points corresponding to very different COD experimental values).

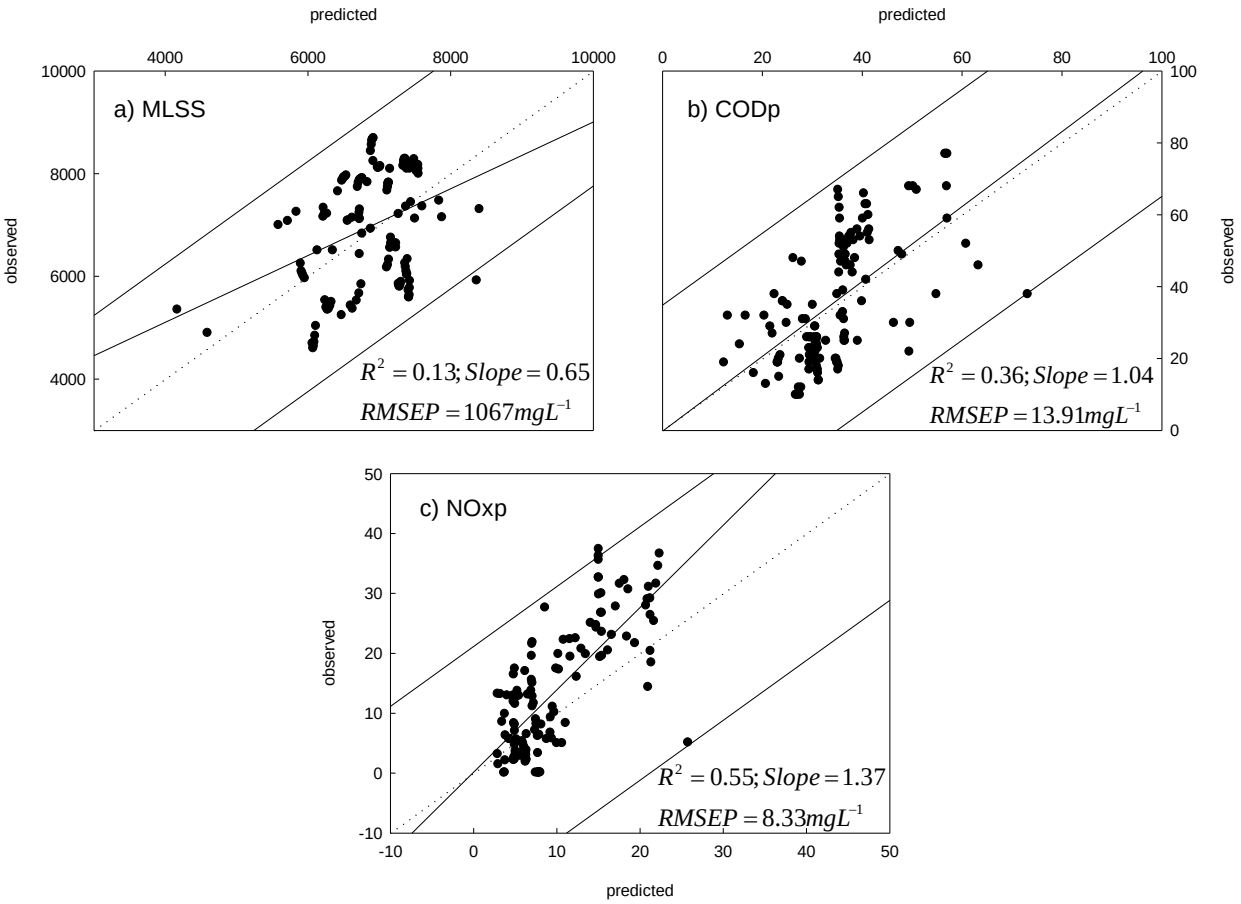

Figure 6.2. ASM3e prediction of a) Mixed liquor suspended solids, b) COD in the permeate and c) Nitrite and nitrate concentration in the permeate. The predicted values are plotted against the observed values for 130 experimental observations obtained throughout the validation period. The units in all axes are in mg L^{-1}.

It should be noted that, in this work, the ASM was used during an extended validation period of more than 400 days, where possible changes on microbial population may affect the biokinetic parameters initially calibrated. After the initial model calibration, ASM may require frequent input of feedwater characterisation through batch tests to re-adjust the model and improve the fitting (Fenu et al., 2010; Sin et al., 2005b). Additionally, despite the extrapolative ability of mechanistic models, the ranges of the parameters that characterise the wastewater during the validation period were not completely overlapped by those in the calibration period (Table 6.1), which also may affect the ASM prediction ability. In this study, the input data consisted only on the conventional feedwater characteristics (Table 6.1), without additional re-calibration data. Therefore, it is clear from the results obtained that the ASM presented required additional input data throughout the validation period, which is one of the well-known shortcomings of ASM (Fenu et al., 2010).

6.4.2. Hybrid Modelling

A hybrid model approach combines mechanistic and statistic-based modelling. In practice, this approach is an alternative to the need of frequently supplying the ASM with new input data which are laborious to acquire and time-consuming. PLS modelling of ASM residuals permits the use of any input data vector for model re-adjustment (on-line data being preferred for practical reasons), as long as it contains valuable information that complements that contained in the mechanistic model. In this study, three different sets of PLS input data were compared: the first, aiming at understanding the pitfalls of the ASM; the second, obtaining a model based only on on-line data to improve the ASM prediction; the third, obtaining the best fitting possible (see Figure 6.1).

Table 6.2. Statistical parameters of selected hybrid models.

Outputs	Model #	Initial inputs in PLS	Number of inputs	LV used in PLS	Variance (%)	RMSEP[a]	R^2_{train}	R^2_{valid}	$Slope_{train}$	$Slope_{valid}$
MLSS	1	no fluorescence, with opx^2	24	3	81	558	0.83	0.76	0.94	0.89
	2	online with $slg10x^2$	27	26	80	522	0.83	0.80	0.91	0.85
	3	all with $all6x^2$	17	17	79	536	0.82	0.78	0.92	0.88
CODp	4	no fluorescence	13	9	70	8.36	0.81	0.77	0.95	1.11
	5	online with $all6x^2$	17	5	76	6.51	0.86	0.86	0.90	0.93
	6	all with perm10	9	5	77	5.63	0.86	0.90	0.92	1.01
NOxp	7	no fluorescence	18	13	58	4.27	0.78	0.84	1.00	0.97
	8	online with $perm10x^2$	30	24	75	3.08	0.87	0.92	0.96	0.93
	9	all with $perm10x^2$	37	34	84	3.20	0.92	0.91	0.98	0.96

opx^2 - interaction and quadratic terms of operational parameters.
$slg10x^2$ - 10 compression components of sludge EEMs, plus their interaction and quadratic terms.
$all6x^2$ - 6 components of all EEMs compressed together, plus their interaction and quadratic terms.
perm10 - 10 compression components of permeate EEMs.
$perm10x^2$ - 10 compression components of permeate EEMs, plus their interaction and quadratic terms.
[a] Root mean square error of prediction in mg L^{-1}.
[b] Latent variables used in PLS modelling.

Regardless the PLS input strategy used, it was found that the hybrid model always resulted in prediction improvements in respect to the plain ASM3e (even for points corresponding to the period immediately following calibration), for the three outputs studied: mixed liquor suspended solids in the biological tank (MLSS), chemical oxygen demand in the permeate (CODp) and nitrite and nitrate concentration in the permeate (NOxp). Table 6.2 summarises the statistical parameters obtained for the selected hybrid models performed for each output.

Prediction of mixed liquor suspended solids

Figure 6.3a shows the best hybrid model obtained for MLSS prediction when using only analytic data as PLS model inputs, and Figure 6.3b shows the normalised regression coefficients of the selected PLS model inputs (model #1). With this input strategy, the ASM prediction of MLSS was significantly improved. From the PLS model inputs selected as useful to predict the MLSS residuals (Figure 6.3b), it was possible to conclude that the ASM model lacks information not only about the influent (wastewater characteristics) but also about the effluent (permeate characteristics). In this model, MLSS residuals were also correlated with some of the operating parameters, mostly in a non-linear way. When incorporating fluorescence data (model #2), the model showed a better fitting to the experimental data (Figure 6.3c). In this model, the selected inputs were the hydraulic retention time (HRT) and dissolved oxygen (DO), besides inputs from 2D fluorescence spectra from the activated sludge (selected from 10 compression components of sludge EEMs plus their interaction and quadratic terms) (Figure 6.3d). Since HRT is an imposed operating parameter, DO is commonly measured on-line and 2D fluorescence spectra can also be easily acquired on-line, this model has the potential to be applied using information exclusively acquired on-line, making possible real-time prediction of mixed liquor suspended solids in the MBR.

Model #3 (Figure 6.3e and f), including both analytic and on-line data, was developed in search for a better model fitting, but it shows that even when all the data collected was incorporated into PLS modelling there was no real improvement in MLSS prediction, when compared with the on-line model (model #2).

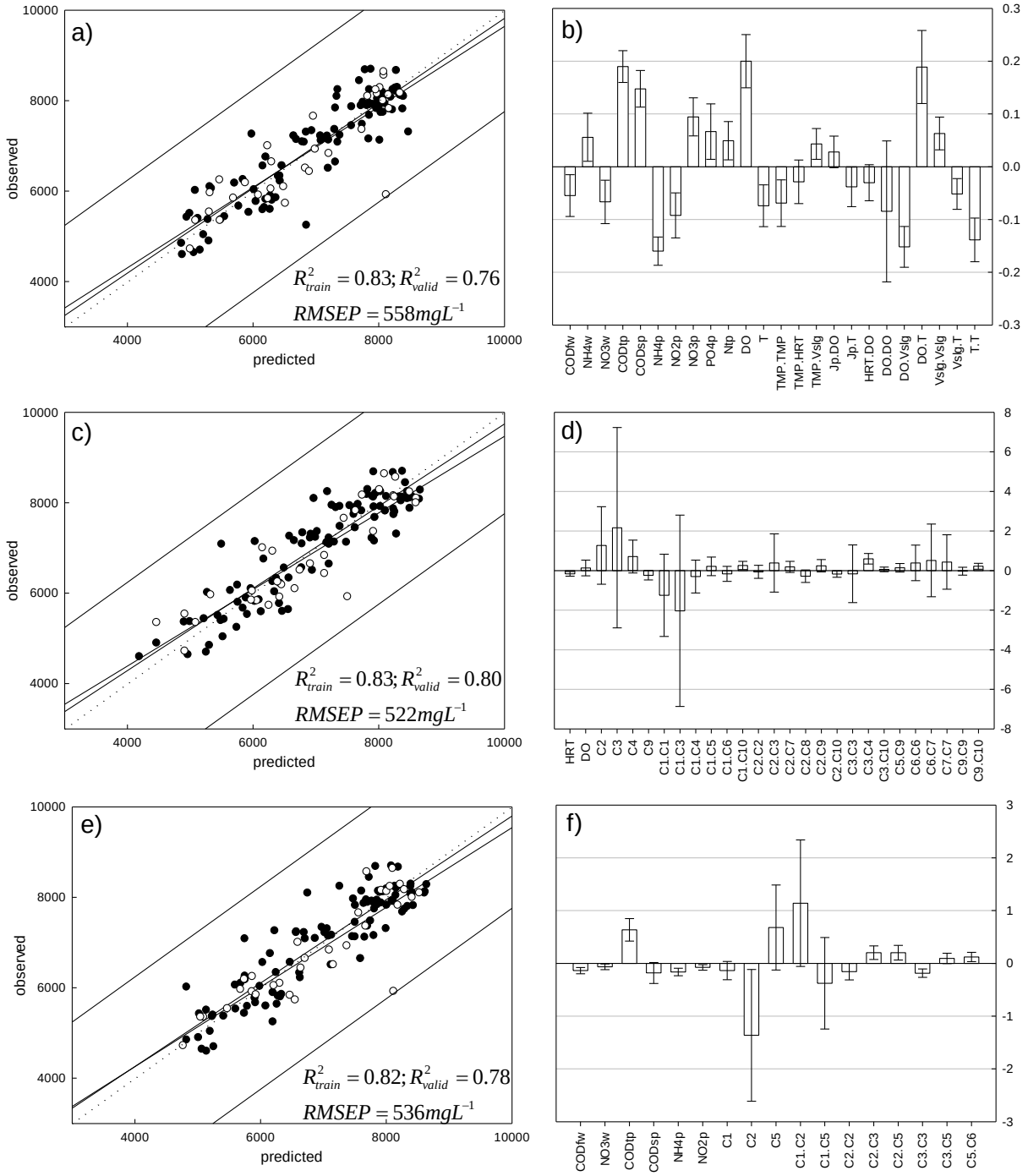

Figure 6.3. MLSS prediction by hybrid models: a-b) model #1, c-d) model #2 and e-f) model #3. Predicted vs observed values respectively for each model (left). Close circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients of model inputs are in normalised units (right).

Prediction of chemical oxygen demand in the permeate

PLS modelling using only analytic data improved considerably the COD prediction (model #4, Figure 6.4a) when compared to the ASM used. In this model, the difference between ASM prediction and experimental data is mainly explained, in a linear correlation, by: i) some operating parameters (transmembrane pressure (TMP), dissolved oxygen (DO), temperature (T) and mixed liquor suspended solids in the biological tank (MLSSb)); ii) almost all wastewater quality parameters, and iii) some of the permeate quality parameters (nitrate (NO3p) and total phosphorus (Ptp)) (Figure 6.4b). This input requirement indicates that the mechanistic model used required additional information in what concerns wastewater characteristics and operating conditions (DO and T) for adequate COD prediction, and lacks crucial information about the membrane performance (which is reflected by TMP). In fact, TMP and DO revealed to be essential in the prediction of COD residuals regardless the input strategy used (Figure 6.4). This gap in the application of ASMs for MBR modelling has been previously acknowledged, and indeed some modified ASM models have included resistance terms to account for fouling development (Jiang, 2007).

The best model achieved to predict COD in the permeate was model #6, which required, besides TMP, DO and T, a few wastewater characteristics and two components of the permeate fluorescence analysis to model the ASM residuals (Figure 6.4e and f). However, on-line adjustment of ASM is possible with fairly good fitting using only TMP, DO and a combination of compression components of the three EEMs obtained together, including some of their quadratic and interaction terms (model #5, Figure 6.4c and d). This model shows that even without acquiring the analytic data concerning wastewater characteristics, this PLS model can use fluorescence data from wastewater, sludge and permeate assuring a good prediction of the COD residuals from the ASM. Indeed, it was previously demonstrated that 2D fluorescence spectroscopy is able to capture the information regarding several key performance parameters of a MBR (Chapter 5), thus this technique likely replaced those additional analytical data.

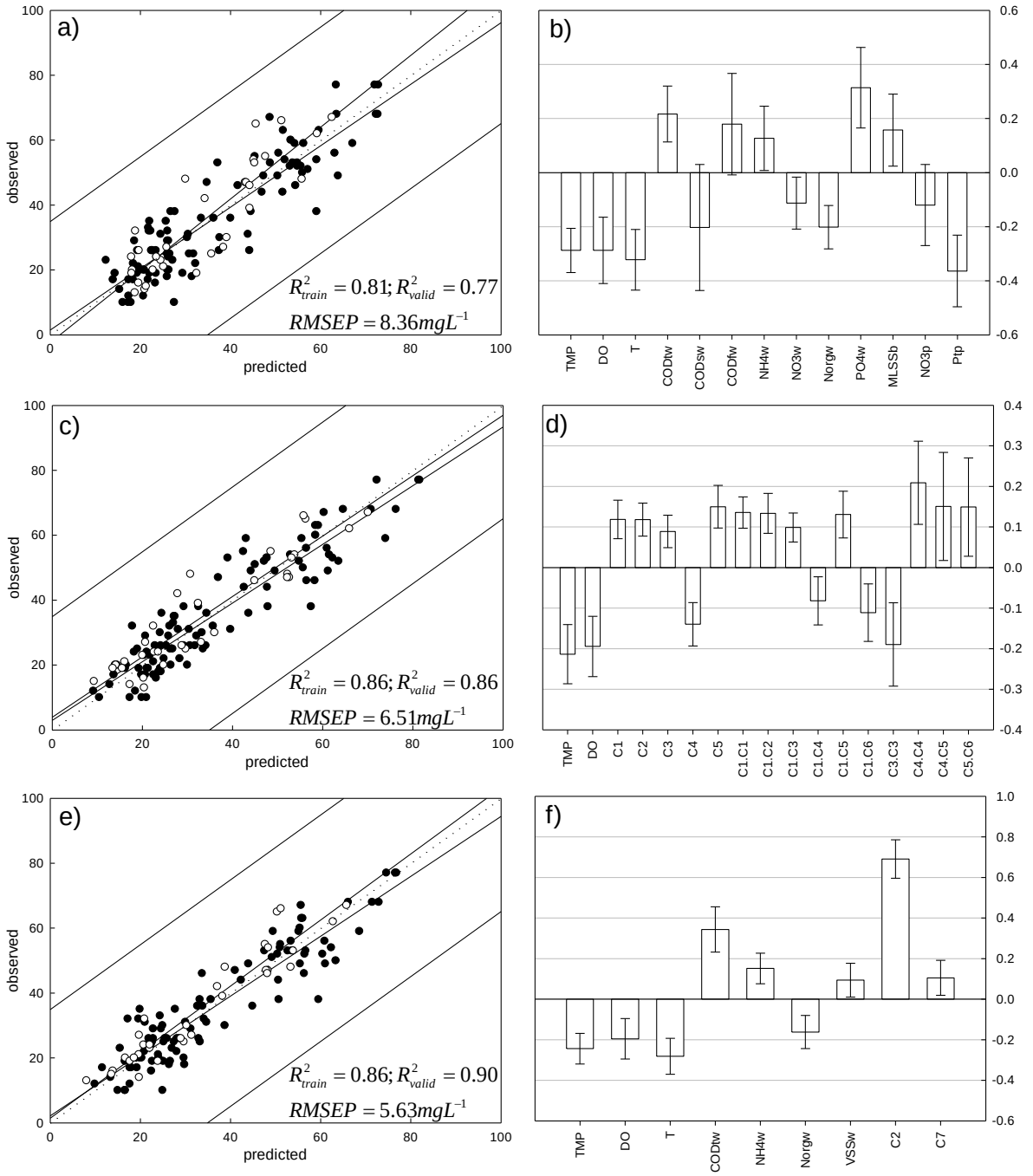

Figure 6.4. COD in the permeate prediction by hybrid models: a-b) model #4, c-d) model #5 and e-f) model #6. Predicted vs observed values respectively for each model (left). Close circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients of model inputs are in normalised units (right).

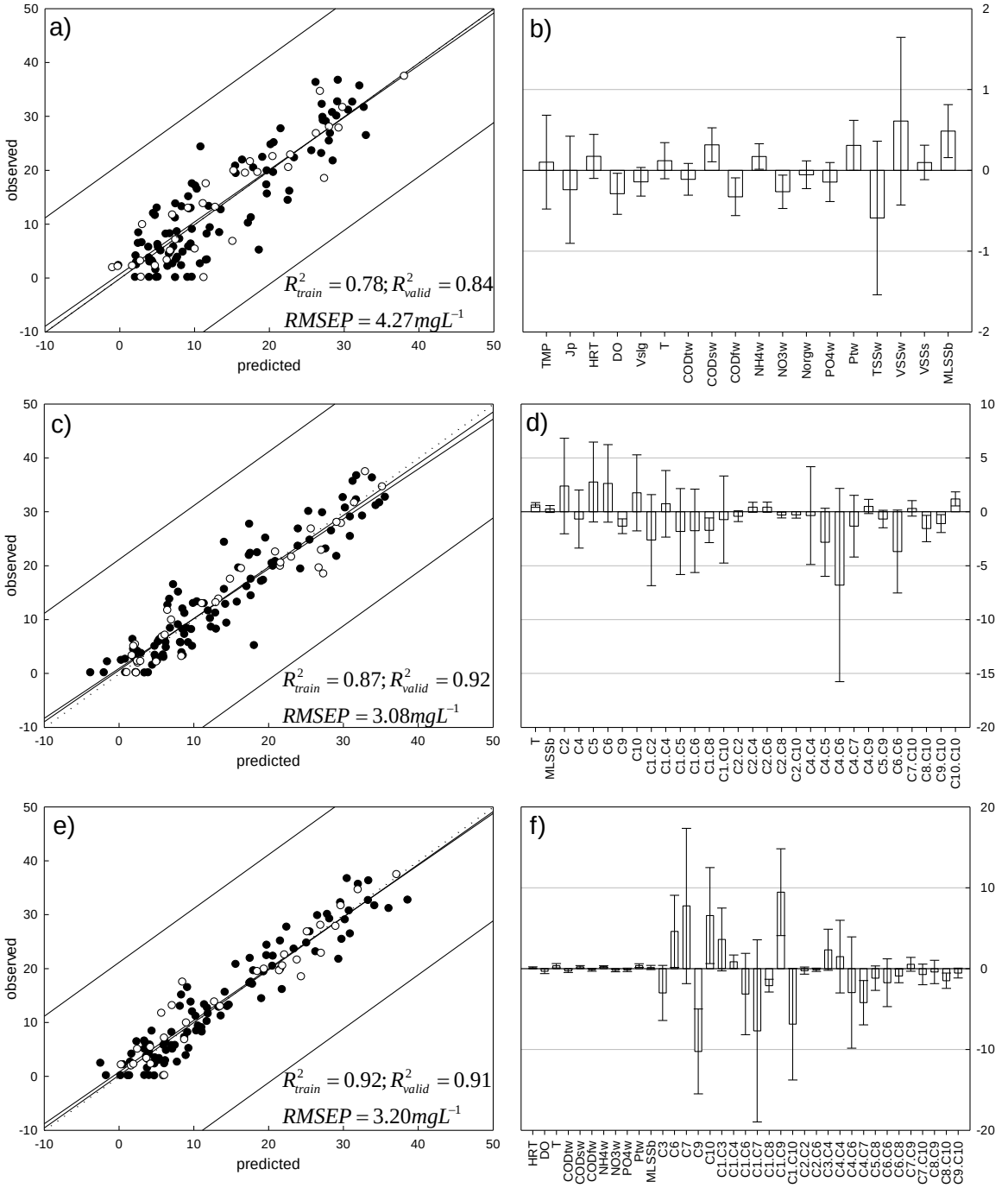

Figure 6.5. NOx in the permeate prediction by models: a-b) model #7, c-d) model #8 and e-f) model #9. Predicted vs observed values respectively for each model (left). Close circles represent training data and open circles represent validation data, both in mg L^{-1}. Regression coefficients of model inputs are in normalised units (right).

Prediction of nitrite and nitrate concentration in the permeate

Figure 6.5a shows the predicted *vs* observed values of nitrite and nitrate concentration in the permeate (NOxp) for the PLS model based purely on analytic data (model #7). With this hybrid strategy, the mechanistic prediction was significantly improved. The selected parameters for this model included the operating conditions and almost all wastewater and sludge analytic parameters, and a multilinear correlation was sufficient to obtain a fairly good fitting (Figure 6.5b). The inputs selection process eliminated all permeate parameters, meaning that these parameters are not useful to predict the residuals of NOxp from the ASM. Interestingly, despite the elimination of permeate characteristics as inputs in the previous model, the best models obtained for NOxp prediction were the ones that incorporated fluorescence data from the permeate (Table 6.2). These results reinforce the assumption that, although 2D fluorescence is not directly correlated with specific quality parameters, it can be used as a fingerprint of the system status, which captures information that indirectly is related to the target quality parameters. Models #8 and #9 gave similar fitting, and both where based on non-linear correlations with the 10 compression components of permeate EEMs (Figure 6.5c-f). The on-line measurements are able to predict NOxp residuals with the best RMSEP (model #8). Nevertheless, the incorporation of wastewater characteristics in model #9 improved the training set fitting (both the R^2 coefficient and the slope, Table 6.2). Through inputs selection analysis, the temperature (T) and mixed liquor suspended solids in the biological tank (MLSSb) were found to be necessary in all PLS models to complement the ASM results for NOxp prediction, meaning that even if these parameters appear to have low contribution (in Figure 6.5d and f) they are essential in the PLS modelling of NOxp residuals.

Overall analysis of the hybrid modelling strategies

Different conclusions can be withdrawn from the three different input strategies used for PLS modelling. With the first input strategy it was shown that the incorporation of analytic and operating data in the hybrid models, through PLS, can substantially improve the prediction of a plain ASM lacking detailed and updated input information about the influent wastewater. Additionally, the selection of useful inputs in these PLS models showed that the hybrid model required information from the sludge and permeate characteristics to improve prediction, reflecting that the ASM kinetic parameters were probably not adjusted throughout the entire validation period.

In the second input strategy, the incorporation of updated data from a comprehensive technique, such as 2D fluorescence (able to assess the system status), not only improved ASM prediction, but also resulted on better model fitting than when only analytic and operating data were used as PLS inputs. Furthermore, the incorporation of 2D fluorescence spectra and operating parameters in hybrid models also resulted on the possibility of on-line correction of the ASM, which could be automatically done without requiring additional laborious analytical measurements for ASM re-calibration. The last input strategy tested, using all data collected, aimed at obtaining the best fitting possible. However, it did not significantly improve model fitting for any of the outputs studied, showing that 2D fluorescence spectroscopy spectra is a powerful technique, able to capture important information about the system, that can be combined with few on-line data to optimise the ASM prediction in real-time.

6.5. CONCLUSIONS

The application of ASM to MBRs requires supplying the model with updated input data that depends on off-line, laborious analysis. However, the present work shows that modelling the residuals of a plain ASM using PLS to correlate them with relevant parameters measured on-line can significantly improve the outputs prediction from ASM in real-time, without requiring further laboratory analysis.

For the three outputs modelled in this study (MLSS, and effluent COD and NOx), the best results were achieved when using inputs from fluorescence spectra acquired either on wastewater influent, sludge or permeate. 2D fluorescence spectroscopy showed to be a comprehensive monitoring tool that can be used in hybrid models, in combination with other on-line parameters to complement the mechanistic information described by an ASM in order to obtain good prediction of key MBR performance indicators.

Chapter 7

CONCLUSIONS AND FUTURE WORK

7.1. FINAL OVERVIEW AND CONCLUSIONS

In this study, it was found that correlations across the conventional MBR monitoring data are abundant and that relationships between operating parameters and performance variables are complex and interdependent. However, the initial approach followed could not provide robust enough correlations to enable the elimination of monitoring parameters for process description, and thus reduce the analytical effort required for full MBR characterisation. Furthermore, it was shown that two-dimensional fluorescence can be applied for monitoring of biological systems due to the ability of this technique to distinguish matrices with different compositions. It was also proven that the complexity of interferences on the fluorescence signal prevents the simple and direct quantitative measurement of specific fluorophores in complex biological systems, such as wastewater treatment systems. However, since fluorescence spectroscopy is highly sensitive to the composition of biological media and to the environmental conditions, these effects were not regarded as a problem but, on the contrary, as a source of information. Therefore, fluorescence EEMs were used as fingerprints, which can be regarded as extremely rich, although complex, sources of information.

The challenge, then, was the integration of such information in quantitative models, where fluorescence data can be related with selected process performance parameters determined independently. Hence, multivariate statistical analysis, combining PARAFAC and PLS, was used to successfully extract relevant information contained in fluorescence EEMs and correlate it to process parameters. The PLS models obtained

proved that the performance of a MBR can be monitored by 2D fluorescence spectroscopy combined with a few other process parameters through statistically-based modelling. Moreover, optimisation of PLS models, through the selection of useful modelling inputs, resulted in the elimination of the redundant input parameters. Therefore, the correlations obtained between the selected inputs and the output parameters are significant and highlight the relationships between operating and performance variables.

Following this multilinear approach, three MBR performance parameters were successfully modelled: TMP, CODtp and CODsp. The incorporation of 2D fluorescence data in PLS modelling proved to be essential for successful prediction of both CODtp and CODsp. On the other hand, when the multilinear PLS was not sufficient to describe the complex data correlations, particularly from the fluorescence matrices, quadratic and interaction terms of the compressed EEMs were incorporated in the PLS model to improve prediction. Thereby, Ntp, NOxp, Ptp and MLSS models were achieved with good fitting. The good results obtained with the incorporation of quadratic and interaction terms of EEMs compression demonstrates the complexity of the information contained in the fluorescence spectra and reinforces the capability of the mathematical tools used to extract the required information. Moreover, it was found that both soluble and total COD in the permeate, both nitrogen parameters in the permeate (Ntp and NOxp) and the MLSS can be predicted based fully on on-line data (including 2D fluorescence data) and imposed operating parameters.

Besides the pure statistical modelling approach, this study demonstrated as well that a hybrid approach, combining a plain ASM with PLS to monitor a MBR, can result in significant improvement of the prediction of MLSS, effluent COD and effluent NOx. Furthermore, the best results achieved with the hybrid modelling strategy were obtained through the incorporation of 2D fluorescence data and other on-line parameters to complement the mechanistic information described by a plain ASM. It was thus shown that PLS, through the integration of relevant monitoring data measured on-line, can be used to adjust the ASM prediction, in real-time, without requiring further laboratory analysis.

Therefore, in this PhD project it was found that multivariate statistical modelling, such as PCA and PLS, can use on-line data from fingerprinting techniques, such as 2D fluorescence spectroscopy, in combination with few additional operating parameters, to predict simultaneously several key performance parameters of an MBR, replacing analytical and time consuming measurements. Furthermore, statistically-based tools and 2D fluorescence spectroscopy proved also to be useful when used in combination with a plain mechanistic model, such as ASM3, for prediction improvement of performance parameters, in real-time, without requiring further laboratory analysis.

2D fluorescence spectroscopy showed to be a powerful and promising monitoring tool for application in MBRs for domestic wastewater treatment. Moreover, after development of the multivariate statistical models they may be continuously updated and improved with new data. 2D fluorescence data may also be acquired in multiple locations of the system (off-line or on-line, according to specific needs), and time-programmed with the help of an optical "switch-box". Considering that the acquisition of a complete fluorescence map takes only a few minutes (depending on the number of data points aimed), this tool can be used as an on-line, non-invasive, real-time monitoring technique.

7.2. SUGGESTIONS FOR FUTURE WORK

The results obtained in the frame of this PhD thesis were a step forward on the development of new on-line monitoring tools for MBRs for domestic wastewater treatment. However, these results can be improved and extended in several ways. The following recommendations for future work can be proposed.

Despite the correlations already found between 2D fluorescence data and performance parameters of an MBR, this same multivariate approach could be used to explore further information contained in fluorescence spectra. With this objective, more specific parameters, such as proteins and polysaccharides composing both soluble and bound EPS, could be evaluated and correlated with fluorescence. The statistical models found would then be useful to monitor these compounds, with reduced analytical effort, during the MBR operation, and infer on their impact on membrane performance.

Furthermore, since 2D fluorescence spectroscopy can assess common fluorophores constituent of EPS, which are accepted as major fouling agents, the acquisition of fluorescence spectra at the membrane surface, *in situ*, should be also considered in future research work.

Some of the PLS models presented required the MLSS data as model input, which in the present work was acquired on-line. However, since MLSS was also successfully modelled, further investigation on the selection of useful inputs would be necessary to fully replace the MLSS on-line probe.

Regarding the study of further correlations of 2D fluorescence with new parameters, it may occur that PLS modelling, with linear or with quadratic and interaction terms, becomes inefficient to extract more complex information from fluorescence spectra. In that case, other non-mechanistic approaches can then be considered, such as artificial neural networks. This alternative approach could be useful to predict e.g., the effluent ammonia, which was not achieved in the present work.

Concerning hybrid modelling, it would be interesting the use of 2D fluorescence with appropriate modelling tools to assess wastewater characteristics (or even the biokinetic parameters), on-line. This information could then be used in an ASM to model and predict the performance of the activated sludge system, without any loss of the mechanistic information.

With the elimination of redundant modelling inputs, the correlations obtained between the selected inputs and output parameters are significant and highlight the relationships between operating and performance variables. These correlations could then be used for process control. Therefore, as the ultimate objective, the modelling approach presented could be implemented within expert-control systems, where the data captured by the fluorescence fingerprints could be used to rapidly evaluate the system status and support a real-time adjustment of the operating conditions.

BIBLIOGRAPHY

Andersson, C.A. and Bro, R., 2000. The N-way Toolbox for MATLAB. Chemometrics and Intelligent Laboratory Systems, **52**, 1-4.

APHA, 1998. Standard methods for the examination of water and wastewater, American Public Health Association, Washington DC.

Bjorkoy, A. and Fiksdal, L., 2009. Characterization of biofouling on hollow fiber membranes using confocal laser scanning microcscopy and image analysis. Desalination, **245**, 474-484.

Boehl, D., Solle, D., Hitzmann, B. and Scheper, T., 2003. Chemometric modelling with two-dimensional fluorescence data for Claviceps purpurea bioprocess characterization. Journal of Biotechnology, **105**, 179-188.

Boggia, R., Forina, M., Fossa, P. and Mosti, L., 1997. Chemometric study and validation strategies in the structure-activity relationships of new cardiotonic agents. Quantitative Structure-Activity Relationships, **16**, 201-213.

Bouhabila, E., Ben Aim, R. and Buisson, H., 2001. Fouling characterisation in membrane bioreactors. Separation and Purification Technology, **22-3**, 123-132.

Bro, R., 1997. PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems, **38**, 149-171.

Chan, Z.S.H., Ngan, H.W., Rad, A.B., David, A.K. and Kasabov, N., 2006. Short-term ANN load forecasting from limited data using generalization learning strategies. Neurocomputing, **70**, 409-419.

Chu, H.P. and Li, X.Y., 2005. Membrane fouling in a membrane bioreactor (MBR): Sludge cake formation and fouling characteristics. Biotechnology and Bioengineering, **90**, 323-331.

Dong, B. and Jiang, S.Y., 2009. Characteristics and behaviors of soluble microbial products in sequencing batch membrane bioreactors at various sludge retention times. Desalination, **243**, 240-250.

Drews, A., 2010. Membrane fouling in membrane bioreactors-Characterisation, contradictions, cause and cures. Journal of Membrane Science, **363**, 1-28.

Drews, A., Arellano-Garcia, H., Schoneberger, J., Schaller, J., Wozny, G. and Kraume, M., 2009. Model-based recognition of fouling mechanisms in membrane bioreactors. Desalination, **236**, 224-233.

Drews, A., Mante, J., Iversen, V., Vocks, M., Lesjean, B. and Kraume, M., 2007. Impact of ambient conditions on SMP elimination and rejection in MBRs. Water Research, **41**, 3850-3858.

Duarte, B.P.M. and Saraiva, P.M., 2003. Hybrid models combining mechanistic models with adaptive regression splines and local stepwise regression. Industrial & Engineering Chemistry Research, **42**, 99-107.

Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F., 1956. Colorimetric Method for Determination of Sugars and Related Substances. Analytical Chemistry, **28**, 350-356.

Duchesne, C. and MacGregor, J.F., 2001. Jackknife and bootstrap methods in the identification of dynamic models. Journal of Process Control, **11**, 553-564.

Fenu, A., Guglielmi, G., Jimenez, J., Sperandio, M., Saroj, D., Lesjean, B., Brepols, C., Thoeye, C. and Nopens, I., 2010. Activated sludge model (ASM) based modelling of membrane bioreactor (MBR) processes: A critical review with special regard to MBR specificities. Water Research, **44**, 4272-4294.

Field, R.W., Wu, D., Howell, J.A. and Gupta, B.B., 1995. Critical Flux Concept for Microfiltration Fouling. Journal of Membrane Science, **100**, 259-272.

Forina, M., Casolino, C. and Millan, C.P., 1999. Iterative predictor weighting (IPW) PLS: A technique for the elimination of useless predictors in regression problems. Journal of Chemometrics, **13**, 165-184.

Forina, M., Lanteri, S., Oliveros, M.C.C. and Millan, C.P., 2004. Selection of useful predictors in multivariate calibration. Analytical and Bioanalytical Chemistry, **380**, 397-418.

Frolund, B., Griebe, T. and Nielsen, P.H., 1995. Enzymatic activity in the activated-sludge floc matrix. Appl Microbiol Biotechnol, **43**, 755-761.

Frolund, B., Palmgren, R., Keiding, K. and Nielsen, P.H., 1996. Extraction of extracellular polymers from activated sludge using a cation exchange resin. Water Research, **30**, 1749-1758.

Ganzlin, M., Marose, S., Lu, X., Hitzmann, B., Scheper, T. and Rinas, U., 2007. In situ multi-wavelength fluorescence spectroscopy as effective tool to simultaneously monitor spore germination, metabolic activity and quantitative protein production in recombinant Aspergillus niger fed-batch cultures. Journal of Biotechnology, **132**, 461-468.

Guglielmi, G., Chiarani, D., Judd, S.J. and Andreottola, G., 2007. Flux criticality and sustainability in a hollow fibre submerged membrane bioreactor for municipal wastewater treatment. Journal of Membrane Science, **289**, 241-248.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.L., 1998. Multivariate data analysis, Prentice-Hall PTR, USA.

Harms, P., Kostov, Y. and Rao, G., 2002. Bioprocess monitoring. Current Opinion in Biotechnology, **13**, 124-127.

Henze, M., Gujer, W., Mino, T. and van Loosdrecht, M., 2000. Activated Sludge Models: ASM1, ASM2, ASM2d and ASM3, Scientific and Technical Report No 9, IWA publishing, London.

Her, N., Amy, G., McKnight, D., Sohn, J. and Yoon, Y.M., 2003. Characterization of DOM as a function of MW by fluorescence EEM and HPLC-SEC using UVA, DOC, and fluorescence detection. Water Research, **37**, 4295-4303.

Huyskens, C., Brauns, E., Van Hoof, E. and De Wever, H., 2008. A new method for the evaluation of the reversible and irreversible fouling propensity of MBR mixed liquor. Journal of Membrane Science, **323**, 185-192.

Jackson, J.E., 2003. A user's guide to principal components, John Wiley and Sons, Inc,, USA.

Jiang, T., 2007. Characterization and modelling of soluble microbial products in membrane bioreactors. PhD Thesis, Ghent University, Belgium, pp. 241.

Jiang, T., Myngheer, S., De Pauw, D.J.W., Spanjers, H., Nopens, I., Kennedy, M.D., Amy, G. and Vanrolleghem, P.A., 2008. Modelling the production and degradation of soluble microbial products (SMP) in membrane bioreactors (MBR). Water Research, **42**, 4955-4964.

Jones, K.L. and O'Melia, C.R., 2000. Protein and humic acid adsorption onto hydrophilic membrane surfaces: effects of pH and ionic strength. Journal of Membrane Science, **165**, 31-46.

Judd, S., 2005. Fouling control in submerged membrane bioreactors. Water Science and Technology, **51**, 27-34.

Judd, S., 2006. The MBR Book: Principles and Applications of Membrane Bioreactors in Water and Wastewater Treatment, Elsevier, Oxford.

Judd, S., 2008. The status of membrane bioreactor technology. Trends in Biotechnology, **26**, 109-116.

Kimura, K., Naruse, T. and Watanabe, Y., 2009. Changes in characteristics of soluble microbial products in membrane bioreactors associated with different solid retention times: Relation to membrane fouling. Water Research, **43**, 1033-1039.

Kobbero, C., Keiding, K., Larsen, K.L. and Nielsen, P.H., 2008. Quenching effects in the application of multi-channel fluorescence in activated sludge suspended solids. Water Research, **42**, 2449-2456.

Kuhn, R., Benndorf, D., Rapp, E., Reichl, U., Palese, L.L. and Pollice, A., 2011. Metaproteome analysis of sewage sludge from membrane bioreactors. Proteomics, **11**, 2738-2744.

Lakowicz, J.R., 1983. Principles of fluorescence spectroscopy, Plenum Press, New York

Laspidou, C.S. and Rittmann, B.E., 2002. Non-steady state modeling of extracellular polymeric substances, soluble microbial products, and active and inert biomass. Water Research, **36**, 1983-1992.

Le-Clech, P., Chen, V. and Fane, T.A.G., 2006. Fouling in membrane bioreactors used in wastewater treatment. Journal of Membrane Science, **284**, 17-53.

Le-Clech, P., Marselina, Y., Ye, Y., Stuetz, R.A. and Chen, V., 2007. Visualisation of polysaccharide fouling on microporous membrane using different characterisation techniques. Journal of Membrane Science, **290**, 36-45.

Lee, N., Amy, G. and Croue, J.P., 2006. Low-pressure membrane (MF/UF) fouling associated with allochthonous versus autochthonous natural organic matter. Water Research, **40**, 2357-2368.

Lee, N.H., Amy, G., Croue, J.P. and Buisson, H., 2004. Identification and understanding of fouling in low-pressure membrane (MF/UF) filtration by natural organic matter (NOM). Water Research, **38**, 4511-4523.

Lesjean, B., Tazi-Pain, A., Thaure, D., Moeslang, H. and Buisson, H., 2011. Ten persistent myths and the realities of membrane bioreactor technology for municipal applications. Water Science and Technology, **63**, 32-39.

Li, H., Fane, A.G., Coster, H.G.L. and Vigneswaran, S., 1998. Direct observation of particle deposition on the membrane surface during crossflow microfiltration. Journal of Membrane Science, **149**, 83-97.

Li, J.K., Asali, E.C. and Humphrey, A.E., 1991. Monitoring Cell Concentration and Activity by Multiple Excitation Fluorometry. Biotechnology Progress, **7**, 21-27.

Li, J.K. and Humphrey, A.E., 1991. Use of Fluorometry for Monitoring and Control of a Bioreactor. Biotechnology and Bioengineering, **37**, 1043-1049.

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

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J., 1951. Protein Measurement with the Folin Phenol Reagent. Journal of Biological Chemistry, **193**, 265-275.

Lyko, S., Wintgens, T., Al-Halbouni, D., Baumgarten, S., Tacke, D., Drensla, K., Janot, A., Dott, W., Pinnekamp, J. and Melin, T., 2008. Long-term monitoring of a full-scale municipal membrane bioreactor - Characterisation of foulants and operational performance. Journal of Membrane Science, **317**, 78-87.

Mamais, D., Jenkins, D. and Pitt, P., 1993. A Rapid Physical-Chemical Method for the Determination of Readily Biodegradable Soluble Cod in Municipal Waste-Water. Water Research, **27**, 195-197.

Marose, S., Lindemann, C. and Scheper, T., 1998. Two-dimensional fluorescence spectroscopy: A new tool for on-line bioprocess monitoring. Biotechnology Progress, **14**, 63-74.

Mehrez, R., Ernst, M. and Jekel, M., 2007. Development of a continuous protein and polysaccharide measurement method by Sequential Injection Analysis for application in membrane bioreactor systems. Water Science and Technology, **56**, 163-171.

Meng, F.G., Chae, S.R., Drews, A., Kraume, M., Shin, H.S. and Yang, F.L., 2009. Recent advances in membrane bioreactors (MBRs): Membrane fouling and membrane material. Water Research, **43**, 1489-1512.

Meng, F.G., Zhang, H.M., Yang, F.L. and Liu, L.F., 2007. Characterization of cake layer in submerged membrane bioreactor. Environmental Science & Technology, **41**, 4065-4070.

Meng, F.G., Zhou, Z.B., Ni, B.J., Zheng, X., Huang, G.C., Jia, X.S., Li, S.Y., Xiong, Y. and Kraume, M., 2011. Characterization of the size-fractionated biomacromolecules: Tracking their role and fate in a membrane bioreactor. Water Research, **45**, 4661-4671.

Metcalf and Eddy, 1991. Wastewater Engineering: Treatment, Disposal and Reuse, McGraw-Hill International Editions, Singapore.

Morel, E., Santamaria, K., Perrier, M., Guiot, S.R. and Tartakovsky, B., 2004. Application of multi-wavelength fluorometry for on-line monitoring of an anaerobic digestion process. Water Research, **38**, 3287-3296.

Mulder, M., 1997. Basic Princibles of Membrane Technology, Kluwer Academic Publishers, London.

Ng, A.N.L. and Kim, A.S., 2007. A mini-review of modeling studies on membrane bioreactor (MBR) treatment for municipal wastewaters. Desalination, **212**, 261-281.

Ng, T.C.A. and Ng, H.Y., 2010. Characterisation of initial fouling in aerobic submerged membrane bioreactors in relation to physico-chemical characteristics under different flux conditions. Water Research, **44**, 2336-2348.

Poele, S.T. and van der Graaf, J., 2005. Enzymatic cleaning in ultrafiltration of wastewater treatment plant effluent. Desalination, **179**, 73-81.

Pons, M.N., Le Bonte, S. and Potier, O., 2004. Spectral analysis and fingerprinting for biomedia characterisation. Journal of Biotechnology, **113**, 211-230.

Reed, J.P., Devlin, D., Esteves, S.R.R., Dinsdale, R. and Guwy, A.J., 2011. Performance parameter prediction for sewage sludge digesters using reflectance FT-NIR spectroscopy. Water Research, **45**, 2463-2472.

Reichert, P., 1994. Aquasim - a Tool for Simulation and Data-Analysis of Aquatic Systems. Water Science and Technology, **30**, 21-30.

Rencher, A.C., 2002. Methods of multivariate analysis John Wiley and Sons, Inc, USA.

Ricardo, A.R., Oliveira, R., Velizarov, S., Reis, M.A.M. and Crespo, J.G., 2012. Hybrid modelling of counterion mass transfer in a membrane-supported biofilm reactor. Biochemical Engineering Journal. DOI:10.1016/j.bej.2011.12.010.

Ricci, R.W. and Nesta, J.M., 1976. Intermolecular and Intramolecular Quenching of Indole Fluorescence by Carbonyl-Compounds. Journal of Physical Chemistry, **80**, 974-980.

Rinnan, A., Booksh, K.S. and Bro, R., 2005. First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. Analytica Chimica Acta, **537**, 349-358.

Rosenberger, S., Evenblij, H., Poele, S.T., Wintgens, T. and Laabs, C., 2005. The importance of liquid phase analyses to understand fouling in membrane assisted activated sludge processes - six case studies of different European research groups. Journal of Membrane Science, **263**, 113-126.

Rosenberger, S., Laabs, C., Lesjean, B., Gnirss, R., Amy, G., Jekel, M. and Schrotter, J.C., 2006. Impact of colloidal and soluble organic material on membrane performance in membrane bioreactors for municipal wastewater treatment. Water Research, **40**, 710-720.

Ryan, T.P., 1997. Modern Regression Methods, John Wiley & Sons, Inc.

Seviour, R. and Nielsen, P.H., 2010. Micobial ecology of activated sludge, IWA Publishing, New York.

Silva, A.F., Carvalho, G., Soares, R., Coelho, A.V. and Crespo, M.T.B., 2011. Step-by-step strategy for separation and identification of extracellular proteins in wastewater treatment systems. Submitted to Applied Microbiology and Biotechnology.

Sin, G., Guisasola, A., De Pauw, D.J.W., Baeza, J.A., Carrera, J. and Vanrolleghem, P.A., 2005a. A new approach for modelling simultaneous storage and growth processes for activated sludge systems under aerobic conditions. Biotechnology and Bioengineering, **92**, 600-613.

Sin, G., Van Hulle, S.W.H., De Pauw, D.J.W., van Griensven, A. and Vanrolleghem, P.A., 2005b. A critical comparison of systematic calibration protocols for activated sludge models: A SWOT analysis. Water Research, **39**, 2459-2474.

Sperandio, M. and Espinosa, M.C., 2008. Modelling an aerobic submerged membrane bioreactor with ASM models on a large range of sludge retention time. Desalination, **231**, 82-90.

Sperandio, M. and Etienne, P., 2000. Estimation of wastewater biodegradable COD fractions by combining respirometric experiments in various So/Xo ratios. Water Research, **34**, 1233-1246.

Surribas, A., Geissler, D., Gierse, A., Scheper, T., Hitzmann, B., Montesinos, J.L. and Valero, F., 2006. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by Pichia pastoris. Journal of Biotechnology, **124**, 412-419.

Tang, S.J., Wang, Z.W., Wu, Z.C. and Zhou, Q., 2010. Role of dissolved organic matters (DOM) in membrane fouling of membrane bioreactors for municipal wastewater treatment. Journal of Hazardous Materials, **178**, 377-384.

Teixeira, A.P., Portugal, C.A.M., Carinhas, N., Dias, J.M.L., Crespo, J.P., Alves, P.M., Carrondo, M.J.T. and Oliveira, R., 2009. In Situ 2D Fluorometry and Chemometric Monitoring of Mammalian Cell Cultures. Biotechnology and Bioengineering, **102**, 1098-1106.

Teychene, B., Guigui, C., Cabassud, C. and Amy, G., 2008. Toward a better identification of foulant species in MBR processes. Desalination, **231**, 27-34.

Thompson, M.L. and Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. Process Systems Engineering, **40**, 1328-1340.

Valeur, B., 2002. Molecular fluorescence: Principles and applications, Wiley-VCH, Germany.

van Aalast - van Leeuwen, M.A., Pot, M.A., vanLoosdrecht, M.C.M. and Heijnen, J.J., 1997. Kinetic modeling of poly(beta-hydroxybutyrate) production and consumption by Paracoccus pantotrophus under dynamic substrate supply. Biotechnology and Bioengineering, **55**, 773-782.

Wang, Z.W., Wu, Z.C. and Tang, S.J., 2009a. Characterization of dissolved organic matter in a submerged membrane bioreactor by using three-dimensional excitation and emission matrix fluorescence spectroscopy. Water Research, **43**, 1533-1540.

Wang, Z.W., Wu, Z.C. and Tang, S.J., 2009b. Extracellular polymeric substances (EPS) properties and their effects on membrane fouling in a submerged membrane bioreactor. Water Research, **43**, 2504-2512.

Wicaksana, F., Fane, A.G. and Law, A.W.K., 2009. The use of Constant Temperature Anemometry for permeate flow distribution measurement in a submerged hollow fibre system. Journal of Membrane Science, **339**, 195-203.

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

Wolf, G., Almeida, J.S., Crespo, J.G. and Reis, M.A.M., 2007. An improved method for two-dimensional fluorescence monitoring of complex bioreactors. Journal of Biotechnology, **128**, 801-812.

Wolf, G., Almeida, J.S., Pinheiro, C., Correia, V., Rodrigues, C., Reis, M.A.M. and Crespo, J.G., 2001. Two-dimensional fluorometry coupled with artificial neural networks: A novel method for on-line monitoring of complex biological processes. Biotechnology and Bioengineering, **72**, 297-306.

Wolf, G., Almeida, J.S., Reis, M.A. and Crespo, J.G., 2005. Modelling of the extractive membrane bioreactor process based on natural fluorescence fingerprints and process operation history. Water Sci Technol, **51**, 51-58.

Yamamoto, K., Hiasa, M., Mahmood, T. and Matsuo, T., 1989. Direct Solid-Liquid Separation Using Hollow Fiber Membrane in an Activated-Sludge Aeration Tank. Water Science and Technology, **21**, 43-54.

Yuan, W. and Zydney, A.L., 1999. Humic acid fouling during microfiltration. Journal of Membrane Science, **157**, 1-12.