Universidade do Minho
Escola de Engenharia

Ana Maria da Silva Paulo

Monitoring of Biological Wastewater Treatment Processes using Indirect Spectroscopic Techniques

Mestrado Gestão Ambiental

Trabalho efectuado sob a orientação do
Doutor Eugénio Manuel de Faria Campos Ferreira
Co-Orientadora
Doutora Ana Maria Antunes Dias

Setembro de 2008

Agradecimentos

Gostaria de expressar o meu sincero agradecimento a todas as pessoas que contribuíram para a realização desta dissertação, em particular:

Ao professor Eugénio Ferreira, pela oportunidade que me foi dada de integrar este projecto, pelo incentivo, apoio e intervenção atenta nos momentos chave do trabalho;

À Ana Dias, por todo o trabalho desenvolvido em conjunto, pela amizade, pelos conhecimentos partilhados e dedicação, pela postura crítica e capacidade de trabalho excepcionais;

Ao João Lopes e Mafalda Sarraguça, da Faculdade de Farmácia da Universidade do Porto, por toda a dedicação e apoio, pela partilha de conhecimentos essencial ao enriquecimento deste trabalho;

À Daniela Mesquita, pela amizade, apoio, participação interessada e crítica;

À professora Madalena Alves, pelo apoio e incentivos sempre presentes;

A todos os meus colegas do Laboratório de Biotecnologia Ambiental, especialmente à Ana Júlia Cavaleiro, Andreia Salvador, Ângela Abreu, Catarina Oliveira, Diana Sousa, Frank Van der Zee, Joana Alves e José Carlos Costa, pela extraordinária capacidade de partilha de conhecimentos, de interajuda, pelo acompanhamento e amizade;

À minha família e amigos, que acreditam em todos os meus projectos e me apoiam sempre incondicionalmente.

Monitoring of Biological Wastewater Treatment Processes using Indirect Spectroscopic Techniques

Abstract

Real-time monitoring can enhance the performance of biological wastewater treatment processes by preventing incidents that can lead to the imbalance of the system and eventually to the total loss of biological activity. For this purpose, *in-situ* monitoring techniques should not require sample pre-treatment and chemicals addition. Nowadays automation is still limited by poor sensor performance and high maintenance costs. Hence, further investigation is required in order to achieve new developments in monitoring techniques.

Spectroscopic methods together with chemometrics are being presented as a powerful tool for process monitoring and control, since they can be fast, non-destructive and without the use of chemicals.

In this work, UV-Visible and Near-Infrared (NIR) spectroscopy were used to monitor an activated sludge process using immersion probes connected to the respective spectrophotometers through optical fibbers. During two monitoring periods changes were induced to the system to test the ability of both probes in detecting them.

While UV-Visible spectroscopy showed to be suitable for on-line monitoring, by detecting chemical oxygen demand (COD) variations in the effluent and identifying different nitrification status, NIR range also demonstrated potentialities, however, due to several experimental constrains, the results were not conclusive.

Partial least squares (PLS) regression was performed for the prediction of COD, nitrate and total suspended solids (TSS) concentrations in the effluent using immersible UV-Visible probe and off-line spectra acquisition. The best results were obtained for the *in-situ* technique. The root mean squared error of cross validation (RMSECV) obtained for the estimative of each parameter was 15.4 mg O_2/L for COD, 19.0 mg N-NO_3^-/L for nitrate and 35.3 mg/L for TSS.

In-situ UV-Visible range proved to be valuable for the monitoring and control of biological wastewater treatment processes, although some improvements identified in this work are still needed to overcome its limitations.

Monitorização de Processos Biológicos de Tratamento de Águas Residuais com Técnicas de Espectroscopia Indirectas

Resumo

A monitorização em tempo real permite melhorar o desempenho dos processos de tratamento biológico de efluentes através da prevenção de incidentes que levam ao desequilíbrio do sistema e eventual perda da actividade biológica. Entre outras vantagens, as técnicas de monitorização *in-situ* podem ainda evitar a necessidade de efectuar o pré-tratamento da amostra e a adição de produtos químicos. Actualmente, a automação encontra-se ainda limitada pelo fraco desempenho e elevados custos de manutenção dos sensores. Deste modo, torna-se necessária mais investigação de modo a desenvolver novas técnicas de monitorização. Os métodos espectroscópicos aliados à quimiometria têm sido apresentados como técnicas com uma enorme potencialidade para a monitorização e controlo de processos, uma vez que podem ser rápidos, não destrutivos e não utilizam produtos químicos. Neste trabalho, as regiões do UV-Visível e do Infra-Vermelho Próximo (Near Infra-Red – NIR) foram usadas na monitorização de um processo de lamas activadas através do uso de sondas ligadas, por fibras ópticas, aos respectivos espectrofotómetros. Durante dois períodos de monitorização foram induzidas alterações ao sistema de forma a testar a capacidade de ambas as sondas na detecção destas variações. A espectroscopia do UV-Visível demonstrou ser adequada para a monitorização em linha, ao detectar variações de carência química de oxigénio (CQO) no efluente e ao identificar diferentes fases do processo de nitrificação. Apesar de a espectroscopia na gama NIR ter mostrado potencialidades, devido a diversas limitações experimentais os resultados foram inconclusivos. A técnica de regressão dos mínimos quadrados parciais (partial least squares – PLS) foi utilizada com o objectivo de prever as concentrações da CQO, nitrato e sólidos suspensos totais (SST) no efluente através da sonda de UV-Visível submersível e da aquisição de espectros em diferido. Os melhores resultados foram obtidos com a técnica *in-situ*. O erro médio de validação (RMSECV) obtido para a estimativa de cada um dos parâmetros foi de 15.4 mg O_2/L para a CQO, 19.0 mg $N\text{-}NO_3^-$/L para o nitrato e 35.3 mg/L para os SST.

A técnica UV-Visível *in-situ* demonstrou ser adequada para a monitorização e controlo de processos de tratamento biológico de efluentes, apesar de ter sido identificada neste trabalho a necessidade de melhorias, por forma a ultrapassar as suas limitações.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANNs	Artificial Neural Networks
AOTF	Acousto-Optic Tunable Filter
ASTM	American Society of Testing and Materials
COD	Chemical Oxygen Demand
CCD's	Charged Coupled Devices
BOD	Biological Oxygen Demand
DCOD	Dissolved Chemical Oxygen Demand
DOC	Dissolved Organic Carbon
F/M	Food-to-Microorganism ratio
HPLC	High Performance Liquid Chromatography
HRT	Hydraulic Retention Time
InGaAs	Indium Gallium Arsenide
KHP	Potassium Hydrogen Phthalate
LOO	Leave-one-out
LV	Latent Variables
MIR	Mid Infrared
MLSS	Mixed Liquor Suspended Solids
MLVSS	Mixed Liquor Volatile Suspended Solids
NIR	Near-Infrared
N-Kj	Kjeldahl Nitrogen
N_{org}	Organic Nitrogen
OLR	Organic Loading Rate
PAT	Process Analytical Technology
PbS	Lead sulphide

PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
Q_{in}	Inflow
RMSEP	Root Mean Square Error for Prediction
RMSECV	Root Mean Square Error for Cross Validation
SBR	Sequencing Batch Reactor
SEP	Standard Error of Prediction
SNV	Standard Normal Variate
SRT	Solids Retention Time
TOC	Total Organic Carbon
TS	Total Solids
TSS	Total Suspended Solids
UV	Ultraviolet
VS	Volatile Solids
VSS	Volatile Suspended Solids
WW	Wastewater
WWTP	Wastewater Treatment Plant

1 INTRODUCTION

1.1 Context and Motivation

The recent trends on environmental protection indicate that, in the immediate future, regulators in Europe will increase their demands towards wastewater treatment activities. In fact, a number of European regulatory measures and recommendations, such as the 91-271 EEC Directive already exist with the objective of preventing adverse effects of urban and industrial wastewater discharges on the environment. Hence, the development of wastewater monitoring tools has been an object of growing concern, and the search for a more complete knowledge of the treatment processes is considered an important path towards a higher efficiency.

In order to comply with these regulations and to prevent possible incidents related to the spatial and time dependent variability of wastewater composition, on-line monitoring is clearly pointed out as a solution (Bourgeois *et al.*, 2001). However, the available wastewater quality monitoring technologies have several drawbacks concerning the control of the treatment systems (Lourenço *et al.*, 2006).

Traditional wastewater (WW) characterization uses aggregate parameters like biological oxygen demand (BOD), chemical oxygen demand (COD) and total organic carbon (TOC), to diagnose the WW treatment status. The analytical used to measure these parameters are cumbersome and time consuming, what makes them difficult to adapt to real-time control, since sampling, sample pre-treatment and chemicals addition are needed in most of the cases. Hence, novel techniques or improved tools are required to meet the actual WW quality standards.

Among the potential candidates for the development and application of on-line measurements, spectroscopy can lead to very interesting results, since it can be the basis for non-invasive and non-destructive measuring systems (Pons *et al.*, 2004).

When using spectroscopic methods, the characteristic transmission, absorption, fluorescence spectrum and vibrational properties of chemical species are measured in order

to determine its concentration or identity (Bourgeois *et al*., 2001). Spectroscopic techniques like UV-Visible, Infrared (IR, mid or Near) and Fluorescence have already been tested *in-situ*.

Indirect chemometric models are used in wastewater for correlating the concentrations of the required parameters to spectral information, since direct models cannot be used when there is no linear relation between these parameters as required by the Beer- Lambert law. This fact is due to the kind of parameters monitored in wastewater, since there is a strong correlation between them, like in the case of COD with soluble COD (SCOD) and total suspended solids TSS (Langergraber *et al.*, 2004a).

Principal Component Analysis (PCA) and Partial Least Squares (PLS) are multivariate statistical projection methods that can be used to make data easier to understand by extracting relevant information and modelling it. These tools are usually used to deal with large amounts of data, such as spectral data. PCA and PLS make use of data directly collected from the process to build an empirical model, providing graphical tools easy to apply and to interpret (Aguado and Rosen, 2007), making them very useful for real-time control and monitoring.

UV-Visible spectroscopic techniques for *in-situ* monitoring of a wastewater treatment plant (WWTP) have proven to be possible but also limited as they require calibration procedures and are very much dependent on matrix stability to achieve good correlations. This is due to the complex matrix present in the biological wastewater treatment processes, which results in a mixture of different organic and inorganic compounds, together with colloidal and suspended matter, making the identification and determination of a single compound or determinant very difficult (Pons *et al.*, 2004). It is nowadays accepted that on-site calibrations have to be performed in order to obtain a correct description of the system media (Langergraber *et al*. 2004b, Rieger *et al.*, 2006; Maribas *et al.*, 2008).

The objective of obtaining information related to the monitoring of a lab scale activated sludge process for the development of a model suitable for future control is the basis of this work.

Two spectroscopic ranges were selected to perform the *in-situ* monitoring, namely the UV-Visible and Near-Infrared (NIR) range. The first choice is in agreement with the fact that UV

Visible spectroscopy has already proved to be an adequate technique for application in wastewater monitoring and it can be suitable for control purposes. However, this technique has some problems, once it is limited to the detection of compounds that can absorb in the UV or visible part of the electromagnetic radiation and can have other disadvantages such as signal saturation and necessity of dilution (Vaillant *et al.*, 2002). NIR spectroscopy, even though is not so usually applied for biological wastewater monitoring, has a great application for quality control in food and pharmaceutical industry (Blanco and Villarroya, 2002). This technique has several advantages related to the detection of chemical and physical properties and to its flexibility in terms of equipments and measuring modes (Reich, 2005). Even though there are several interesting works using NIR spectroscopy in wastewater monitoring (Stephens and Walker, 2002; Hansson *et al.*, 2003; Holm-Nielsen *et al.*, 2006), the technique is still underdeveloped. More research is necessary for a better understanding of its applications, limitations and advantages when compared to other methods (Dias et *al.*, 2008). Hence, NIR application to the monitoring of an activated sludge systems constitutes, by itself, an interesting opportunity of research, as suggested by the study performed by Dias *et al.* (2008).

1.2 Objectives

In-situ spectra acquisition with immersible probes together with off-line parameters analysis was used to study an activated sludge process with the main objective of contributing to the future development of an on-line real-time monitoring system.

The following specific objectives were defined:

- Monitor a lab scale activated sludge system *in-situ* using two different probes: UV-Visible and NIR immersible probes;
- Perform off-line monitoring of parameters such as COD, TSS, Kjeldahl nitrogen, NH_4^+, NO_3^-, NO_2^- in the effluent and mixed liquor volatile suspended solids (MLVSS) in the reactor;

- Induce imbalances to the process and analyse the spectral information obtained before and after the disturbance by using PCA and other statistical methods, evaluating the suitability of both immersion probes for real-time monitoring and detection of changes in the biological process;
- Relate the spectral information with the off-line analysed parameters;
- Construct PLS models for the prediction of parameters such as COD, TSS and nitrate.

1.3 Activated Sludge Systems

The use of microorganisms to degrade different kind of effluents, removing contaminants from wastewater by assimilating them, is effective and widespread. When considering biological wastewater treatment for a particular application it is important to have information about the wastewater composition and discharge requirements. So, with proper analysis and environmental control, almost all wastewaters containing biodegradable constituents can be treated biologically (Metcalf and Eddy, 2003).

The principal biological processes used for wastewater treatment can be divided into two main categories: suspended growth and attached growth (or biofilm) processes.

In suspended growth processes, the microorganisms are maintained in liquid suspension by appropriate mixing methods, and these systems can be performed in the presence of oxygen (aerobic) or in its absence (anaerobic, anoxic). The most common suspended biological process used for municipal and industrial wastewater treatment is the activated sludge process. The production of a very active mass of microscopic organisms capable of stabilizing waste under aerobic conditions is the basis for its designation (Rittmann and McCarty, 2001). Over the last 30 years numerous activated-sludge processes have been developed for the removal of organic material (BOD) and for nitrification. According to basic reactor configurations these processes can be grouped as: plug-flow, complete-mix and sequentially operated systems (Metcalf and Eddy, 2003). The complete-mix system became

the favorite of design engineers, since it is the simplest system to analyze (Rittmann and McCarty, 2001).

A complete-mix activated sludge process is based on a continuous-flow stirred-tank reactor (aeration tank), where a relatively large number of microorganisms is in contact with dissolved oxygen, carbonaceous and nitrogenous wastes (Gerardi, 2002). In the aeration tank time is provided for mixing and aerating the influent wastewater with the suspended microorganisms, generally referred to as mixed liquor suspended solids (MLSS) or mixed liquor volatile suspended solids (MLVSS). This guarantees that the organic load, microbial suspension and oxygen demand are uniform in the aeration tank (Metcalf and Eddy, 2003). Usually mechanical equipment is used for mixing and to improve the transfer of oxygen into the process. The mixed liquor then flows to a clarifier where the microbial suspension is settled and thickened. The settled biomass is returned to the aeration tank to continue biodegradation of the influent organic material (Figure 1).

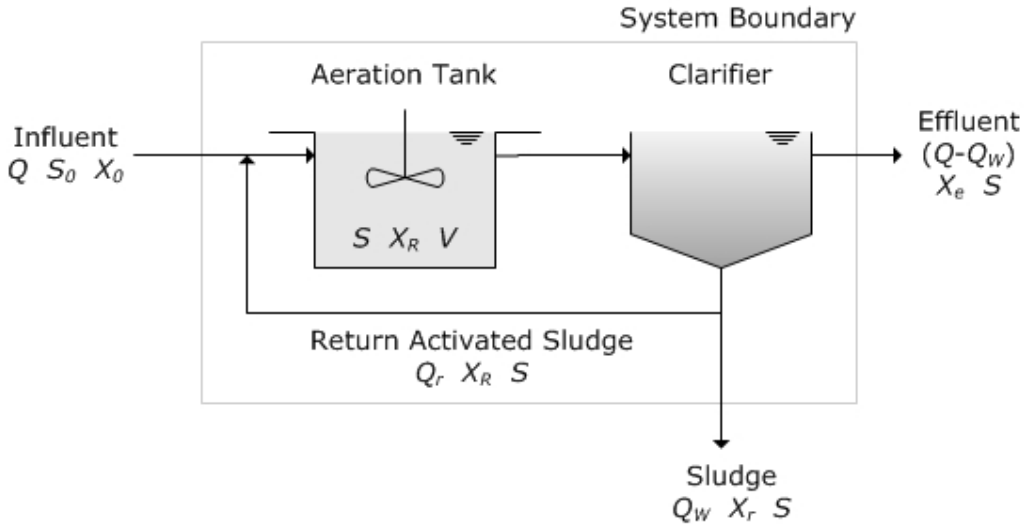


Figure 1. Schematic diagram of an activated sludge process. Legend: Q - flowrate of influent; Q_W - waste sludge flowrate; Q_r - flowrate in return line from clarifier; V - volume of aeration tank; S_0 - influent soluble substrate concentration; S - effluent soluble substrate concentration; X_0 -concentration of biomass in influent; X_R - concentration of biomass in return line from clarifier; X_r - concentration of biomass in sludge drain; X_e - concentration of biomass in effluent (Metcalf and Eddy, 2003).

A large diversity of microorganisms can be found in the activated sludge process and a true ecosystem develops inside the aeration tank. Carbonaceous BOD removal in activated sludge processes is accomplished by aerobic heterotrophic microorganisms which are able to obtain energy and carbon from organic compounds. Nitrate production from ammonium is possible when strict aerobic autotrophs are present, using minerals and inorganic compounds to grow and reproduce, thus reducing nitrogenous BOD of wastes (Gerardi, 2002).

The main microorganisms responsible for most, if not all, nitrification in activated sludge process belong to the genera *Nitrosomonas* and *Nitrobacter*, which oxidize ammonium to nitrite and then to nitrate, respectively, in a two-step process, as follows:

$$2NH_4^+ + 3O_2 \rightarrow 2NO_2^- + 4H^+ + 2H_2O \quad (1)$$

$$2NO_2^- + O_2 \rightarrow 2NO_3^- \quad (2)$$

Total oxidation reaction:

$$NH_4^+ + 2O_2 \rightarrow NO_3^- + 2H^+ + H_2O \quad (3)$$

Reaction (2) is usually very fast and nitrite concentration in the effluent of a WWTP is very low and around 0.1 mg/L (Rieger *et al.,* 2004).

All biological nitrogen-removal processes include an aerobic zone in which biological nitrification occurs but, to satisfy a total nitrogen discharge requirement, the wastewater treatment system must nitrify and denitrify, preventing eutrophication by avoiding the emission of inorganic nitrogen forms to water bodies (Metcalf and Eddy, 2003).

Denitrification is the biological reduction of nitrite to nitric oxide, nitrous oxide and nitrogen gas, as follows:

$$NO_3^-(aq) \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2(g) \quad (4)$$

1.4 UV-Visible Spectroscopy

1.4.1 Fundamentals

Spectroscopic processes rely on the fact that electromagnetic radiation interacts with atoms and molecules in discrete ways to produce characteristic absorption or emission profiles (Burgess, 2007).

Electromagnetic radiation is a type of energy that is transmitted through space, taking many forms: visible light is the most easily recognized, but it also includes X-rays, ultraviolet radiation, radio waves and microwave radiation. The visible region constitutes a small part of the electromagnetic spectrum, when compared to other spectral regions (Figure 2). The various types of radiation can be defined in terms of their wave frequency (Thomas *et al.*, 1996).

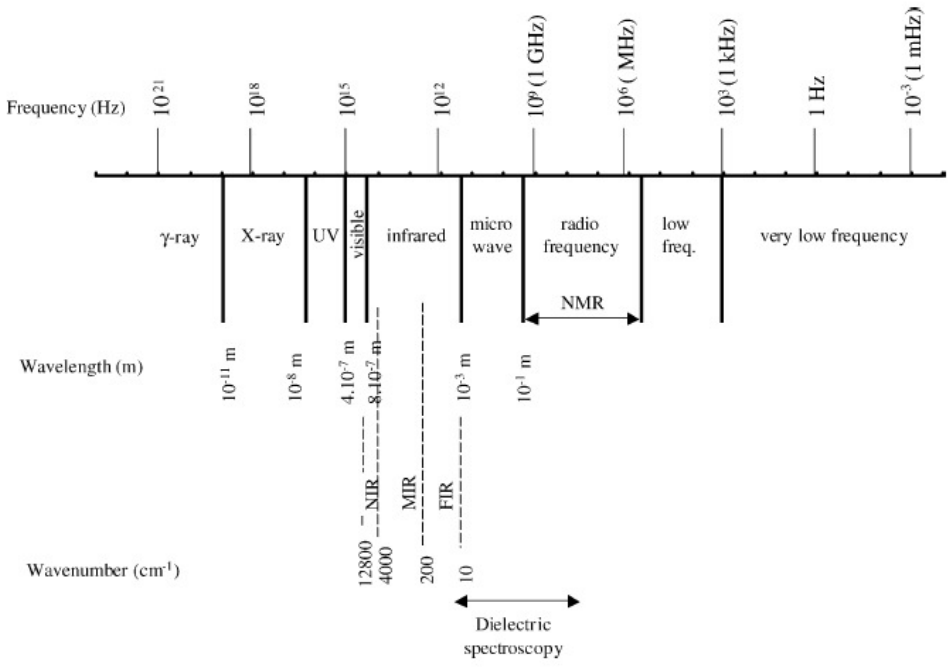


Figure 2. Classification of the different spectral regions (Pons *et al.*, 2004).

The interaction of a photon with the electron cloud of a particular molecule causes the promotion of an electron from the ground to an excited state (Figure 3). The difference in the molecular energy levels, E_2-E_1, will correspond exactly to the photon energy.

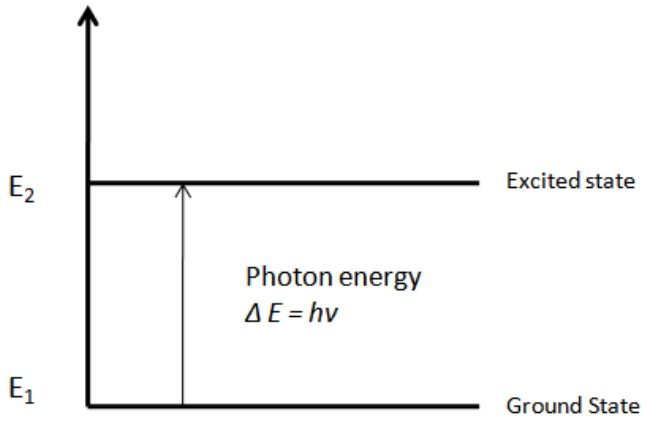


Figure 3. Photon capture by a molecule (Burgess, 2007).

The interaction between the photon and the electron cloud of matter is specific and discrete, being quantized and the energies associated with them related to the type of transition involved. The wavelength of each absorption is dependent on the difference between the energy levels. Hence, some transitions require less energy and consequently appear at longer wavelengths.

If a molecule is only capable of a single electronic transition it will yield a sharp single spectral line, but molecular spectra are not solely derived from single electronic transitions between the ground and excited states. Quantized transitions do occur between vibrational states within each electronic state and between rotation sublevels. Electronic transitions occur at higher energies (ultraviolet) than vibrational (infrared) or rotational ones (microwave). Hence, the molecular spectra observed in the UV-Visible-NIR region are a combination of different transitions (Burgess, 2007).

Electronic transitions related to the UV-Visible spectroscopy are only possible when the molecule involved in the absorption process has a chromophore (Table 1). Chromophores are the basic building blocks of spectra and are associated with molecular structure and the types of transition between molecular orbitals. Chromophores are characterized by the existence of electrons liable to absorb a given radiation, the energy of which corresponds exactly to that required for electron excitation (Thomas *et al.*, 1996).

Table 1. Examples of molecules with chromophores for UV absorption and respective absorption band (adapted from Workman and Springsteen, 1998)

Chromophore	Absorption band (nm)
Nitriles (R – C ≡ N)	160
Alcohols (R – OH)	180 (170-200)
Amines, primary (R – NH_2)	190 (200-220)
Nitrites (R – NO_2)	271
Azo group (R – N ≡ N – R)	340

There are three types of ground state molecular orbitals:

- Sigma (σ) bonding,
- Pi (π) bonding,
- Non-bonding (n),

and two types of excited state:

- Sigma star (σ^*) antibonding,
- Pi star (π^*) antibonding,

from which transitions are observed in the UV region (Figure 4). These four transitions yield different values for ΔE, and, hence, wavelength (Burgess, 2007).

Possible electronic transitions of π, σ, and n electrons are:

- $\sigma \rightarrow \sigma^*$ Transitions

An electron in a bonding σ orbital is excited to the corresponding antibonding orbital. The energy required is large. For example, methane (which has only C-H bonds and can only undergo $\sigma \rightarrow \sigma^*$ transitions) shows a maximum absorbance at 125 nm. Maxima absorption due to $\sigma \rightarrow \sigma^*$ transitions are not seen in typical UV-Visible spectra (200–700 nm).

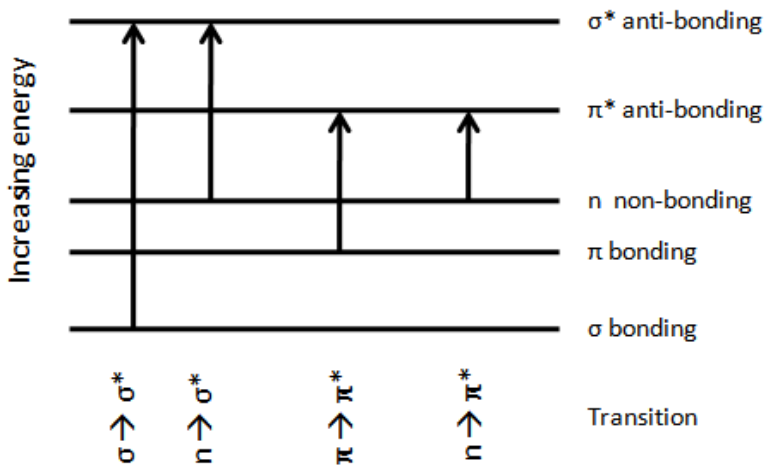


Figure 4. Transitions between molecular orbitals (Burgess, 2007).

- $n \rightarrow \sigma^*$ Transitions

Saturated compounds containing atoms with lone pairs (non-bonding electrons) are capable of $n \rightarrow \sigma^*$ transitions. These transitions usually need less energy than $\sigma \rightarrow \sigma^*$ transitions. They can be initiated by light whose wavelength is in the range 150–250 nm. The number of organic functional groups with $n \rightarrow \sigma^*$ peaks in the UV region is small.

- $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ Transitions

Most absorption spectroscopy of organic compounds is based on transitions of n or π electrons to the π^* excited state. This is because the absorption peaks for these transitions fall in an experimentally convenient region of the spectrum (200–700 nm). These transitions need an unsaturated group in the molecule to provide the π electrons.

Since only $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions are possible in the UV-Visible spectral range, only non-saturated organic compounds or ions, which contain a chromophoric group, can absorb directly radiation in this spectral region and thus be detected (Thomas *et al.*, 1996).

Hence, saturated bonds present in saturated hydrocarbons (oils, fuel), carbohydrates (sugar) and almost all mineral species, except oxyanions like nitrate and nitrite, are not able to absorb in UV-Visible region (Thomas *et al.*, 1999; Pons *et al.*, 2004).

UV-Visible spectroscopic techniques used for quantifying purposes are based on Beer-Lambert law. According to the Beer-Lambert law for a single wavelength and a single component, the following relation is valid:

$$A = \varepsilon bc \qquad (5)$$

where

A – Absorbance (A.U.); ε - Molar absorptivity ($mol^{-1}.cm^{-1}$); b - Path length of the cell in which the sample is contained (cm); c - Concentration of the absorber ($mol.dm^{-3}$).

Therefore, for a given wavelength and a single component, absorbance is a linear function of the concentration of the component.

However, this equation is based on a number of assumptions, including:

- Radiation is perfectly monochromatic;
- There are no uncompensated losses due to scattering or reflection;
- Radiation beam strikes the cuvette at normal incidence;
- There are no molecular interactions between the absorber and other molecules in solution;
- Temperature remains constant.

These assumptions are not always met and cause deviations from ideal Beer-Lambert law behavior, like in the case of water and wastewater UV-Visible spectra (Burgess, 2007).

The chemical nature and concentration of absorbent dissolved components together with the physical characteristics and concentration of heterogeneous material are the two phenomena responsible for the shape of the UV-Visible spectrum of a water sample. Consequently, direct spectroscopy involves two main phenomena: the chemical absorption mechanism, explained by the Beer-Lambert law, and the scattering effect and its associated diffusion, related to the suspended solids and colloids (Thomas *et al.*, 1996).

Since anthropogenic and natural organic compounds contain chromophoric groups, associated to the unstable (oxydizable), condensed state or organic matter, these can be detected by UV-Visible spectrophotometry (Thomas *et al.*, 1996). UV region concentrates a part of the relevant spectral information that can be used for wastewater characterization, as shown by Figure 5.

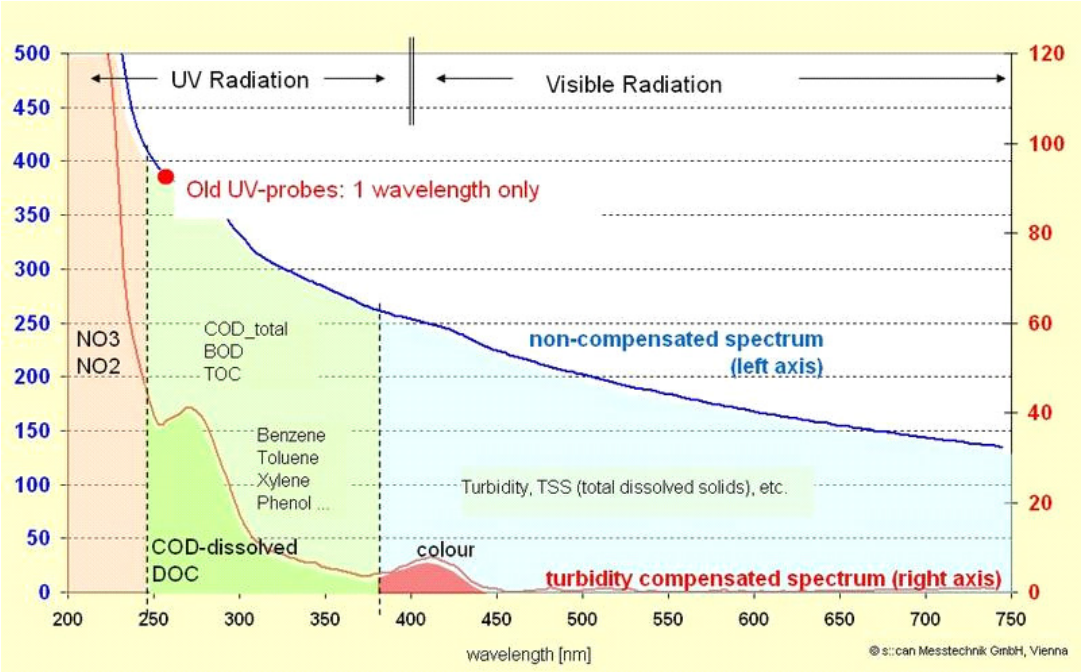


Figure 5. Detection of different wastewater monitoring parameters in the UV-Visible spectral range (s::scan Messtechnik GmbH, Vienna, Austria).

1.4.2 Instrumentation

The general arrangement of an UV-Visible spectrometer and its usual components are presented in Figure 6.

Two radiation sources are generally used in UV-Visible spectrometers which together cover the range from 200-800 nm. For measurements below 320 nm a deuterium or a hydrogen lamp at low pressure is used for emitting a continuous spectrum. If a tungsten halogen lamp is used to emit below 400 nm, special filters are often included in the optical path, to reduce the stray radiation. For measurements above 320 nm compact tungsten halogen sources in

quartz envelope are often used. This type of source is used in the wavelength range of 350–2500 nm. Tungsten/halogen lamps are very efficient, and their output extends well into the ultraviolet region.

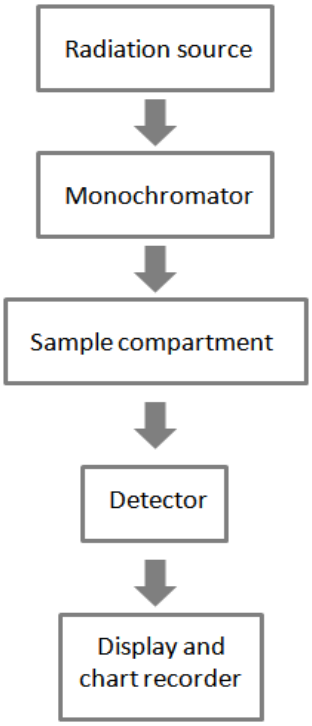


Figure 6. Basic construction of a spectrophotometer (Thomas, 1996).

Wavelength selectors are needed to guarantee a monochromatic radiation, since a narrow bandwidth is required in order to enhance the sensitivity of the absorbance measurements.

As sample containers cuvettes are usually used and must be made of a material which is transparent to the radiation concerned – silica or quartz for the UV-Visible region and glass or plastic to the visible region.

Since cuvettes are only feasible for off-line and at-line measurements, new materials like optical fibers connected to immersible probes can be more suitable for on-line spectroscopic analysis. Optical fibers are, along with mirrors and windows, passive optical

components of great interest for use in several different applications and also as optical data communication links (Sporea and Sporea, 2005).

A variety of detectors is available for UV-Visible measurements. High-performance instruments utilize photomultiplier tube technology from the ultraviolet into the visible region. The more common detectors are given bellow with the useful operating ranges indicated (Table 2).

Table 2. Different types of UV-Visible detectors and useful working ranges in nanometers (adapted from Workman and Springsteen, 1998)

Detector Type	Useful working range (nm)
Silicon photodiode	350-1100
Photomultiplier tubes	160-1100
CCD's (charge coupled devices)	180-1100
Photodiode arrays	180-1100

1.4.3 Applications

UV-Visible spectroscopy is a mature analytical technique, basis of several established applications. Although it's obvious utility, this technique is still poorly exploited in several fields (Thomas, 2007). However, it is not a novelty the study of UV-Visible spectroscopy as an alternative and rapid method to obtain information about the quality of water and wastewater.

The main application of the technique is to correlate the UV-Visible response (e.g. absorbance) to the parameter to be estimated (Thomas *et al.*, 1996). Considering only UV spectroscopy, the 200–300 nm range has been considered particularly interesting for this purpose (Wu *et al.*, 2006).

Using the absorbance at 254 nm, correlations were performed for COD (Mrkva, 1975) and TOC (Dobbs *et al.*, 1972), for municipal and industrial wastewaters. Since this technique can be very sensitive to turbidity, a second wavelength can be used as a correction (Wu *et al.*,

2006). This was performed by Matsché and Stumwöhrer (1996), where COD and TOC were determined using the absorbance at 254 together with the absorbance at 350 nm, for TSS correction. More recently, absorbance at 254 nm was also used to estimate dissolved chemical oxygen demand (DCOD), COD, ammonia and turbidity in a municipal wastewater, being this information associated to synchronous fluorescence spectroscopy results, for fingerprinting purposes (Wu *et al.*, 2006).

Even though it is very interesting to use a fast and simple UV measurement at one or two wavelengths instead of a usual COD or BOD measurement, frequent calibration should be assured to guarantee good results (Thomas *et al.*, 1993). Moreover, a univariate approach is based on the fact that the organic pollution present in effluent has a peak of maximum absorbance. However, this value can vary, depending on the matrix composition (Fogelman *et al.*, 2006).

The increasing computational power observed during the last years allowed a shift towards a multiple wavelength approach (Fogelman *et al.*, 2006). Even though the equipment needed can be more complex, results are more robust (Thomas *et al.*, 1993). In fact, a multiwavelength approach can achieve better results when compared to the use of single-wavelength procedures, mostly for monitoring effluents characterized by constant variations in composition (Rieger *et al.*, 2004; Langergraber *et al.*, 2004a).

Different mathematical procedures have been used for UV spectral processing (Vaillant *et al.*, 2002). Using the spectral range of 205–330 nm and a deconvolution method for the determination of dissolved organic carbon (DOC), COD, TOC, BOD, TSS, and nitrate Thomas *et al.* (1996) demonstrated that it is possible to obtain very good correlations for all of the referred parameters, with the purpose of improving WWTP control. El Khorassani *et al.* (1999) also concluded that using a deterministic deconvolution method and the UV spectral range it is possible to achieve good calibration results to determine COD, TOC, TSS, nitrate and chromium IV, present in different industrial wastewaters. Escalas *et al.* (2003) used a modified UV deconvolution method to estimate DOC of raw and diluted samples from a WWTP.

The influence of turbidity in COD quantification in grey and sewage effluents using the UV range and artificial neural networks (ANNs) was investigated by Fogelman *et al.* (2006). The best results were obtained between 190 and 350 nm, when comparing to the range between 200 and 350nm. Moreover, the authors concluded that for grey waters, better correlations were achieved without sample filtration only when turbidity was not higher than 150 NTU.

Considering that nitrate has a maximum absorbance between 200 and 220 nm, Karlsson *et al.* (1995) used UV-Visible spectroscopy together with PLS multivariate calibration for the determination of nitrate concentration between 0.5 and 13.7 mg/L. Samples from three different WWTPs were collected during a period of more than one year. Correlation coefficients (R^2) for the PLS calibration, for several raw spectra pre-treatments, were always very high and close to unity. Also with the concern of determining total nitrogen present in wastewaters Ferree and Shannon (2001) studied the use of a second derivative method for the determination of nitrate and total nitrogen, by oxidizing all nitrogenous compounds to nitrate by auto-claving. A correlation coefficient of 0.99 was achieved, even though the results were only suitable for determination of concentrations of $N\text{-}NO_3^-$ between 0.1 and 3 mg/L. These examples show that UV spectroscopy can be considered an alternative method for nitrate monitoring without the use of hazardous reagents (e.g. cadmium reduction technique) or expensive equipments (e.g. ion chromatography) (Ferree and Shannon, 2001).

All the previous applications needed sampling for off-line spectral analysis, suitable only for in-line WWTP monitoring. Meanwhile, new developments were achieved by constructing submersible equipments which can perform a spectra analysis directly in liquid media. The use of this type of *in-situ* spectrometers for the determination of several parameters in the effluent of a WWTP, such as COD, TSS, nitrate and nitrite, has been successfully applied using the UV spectra range 200-400 nm (Rieger *et al.*, 2004). The same *in-situ* spectrometer was used for UV-Visible range acquisition in the determination of COD, filtered COD and nitrate to monitor a paper mill WWTP (Langergraber *et al.*, 2004a) and for quantifying filtered COD, TSS and nitrate values, for control of a pilot-scale sequencing batch reactor (Langergraber *et al.*, 2004b). More recently, Maribas *et al.* (2008) used a submersible UV-Visible spectrophotometer to monitor the rapid changes in total COD and TSS, testing three

different places in a WWTP pre-treatment unit. For one of the chosen locations, better correlations were achieved by performing a local calibration and new calibrations were needed every time sudden composition variation occurred. This work demonstrated that rapid changes difficult calibration procedures, even though it was still possible to achieve a good qualitative monitoring. Thus, the results show that it is not easy to take into account large variations in the wastewater matrix as also stated by Rieger *et al.* (2006), while studying different calibration approaches for six WWTP, using a UV-Visible spectrometer.

In Table 3 relevant results are presented concerning some of the studies above referred, mainly the ones focused on a multiwavelength and/or multiparametric approach.

Table 3. Concentration ranges and correlation coefficients for each of the determined parameters obtained in several studies using UV or UV-Visible spectroscopy

Reference and method	**Parameters (Concentration range and R^2 correlation coefficient)**
Thomas *et al*. (1996) UV spectrophotometry and deconvolution method	0 < COD < 500 mg O_2/L (R^2 = 0.940) 0 < DOC < 120 mg/L (R^2 = 0.917) 0 < TOC < 150 mg/L (R^2 = 0.963) 0 < BOD < 250 mg O_2/L (R^2= 0.905) 0 < NO_3^- < 15 mg/L (R^2 = 0.992) 0 < TSS < 350 mg/L (R^2 = 0.938)
El Khorassani *et al.* (1999) UV spectrophotometry and deconvolution method	0 < COD < 150 mg O_2/L (R^2 = 0.89) 0 < TOC < 60 mg/L (R^2 = 0.91) 0 < NO_3^- < 50 mg/L (R^2 = 0.99) 0 < TSS < 100 mg/L (R^2 = 0.77) 0 < Cr IV < 300 mg/L (R^2 = 0.96)
Rieger *et al.* (2004) Submersible UV spectrometer and	17.4 < COD < 21.4 mg O_2/L (R^2 = 0.905) 2.5 < DOC < 17.5 mg/L (R^2 = 0.382)

multivariate calibration based on PLS regression	$0 < N\text{-}NO_2^- < 3.5$ mg/L ($R^2 = 0.993$) $4 < N\text{-}NO_3^- < 17.5$ mg/L ($R^2 = 0.978$) $0 < TSS < 25$ mg/L ($R^2 = 0.848$)
Langergraber *et al.* (2004a) Submersible UV/VIS spectrometer and multivariate calibration based on PLS regression	$700 < COD_{influent} < 1600$ mg O_2/L (COD_{total} $R^2 = 0.95$; $COD_{filtered}$ $R^2 = 0.95$) $75 < COD_{effluent} < 175$ mg O_2/L (COD_{total} $R^2 = 0.90$; $COD_{filtered}$ $R^2 = 0.91$) $0 < N\text{-}NO_3^-{}_{effluent} < 5$ mg/L ($R^2 = 0.87$)
Langergraber *et al.* (2004b) Submersible UV/VIS spectrometer and multivariate calibration based on PLS regression	$20 < COD_{filtered} < 540$ mg O_2/L ($R^2 = 0.90$) $0 < N\text{-}NO_3^- < 25$ mg/L ($R^2 = 0.98$) $0 < TSS < 15$ g/L ($R^2 = 0.995$)
Fogelman *et al.* (2006) UV–Vis spectroscopy (190–350 nm) and artificial neural networks (ANNs)	Unfiltered grey water $120 < COD < 420$ mg O_2/L ($R^2 = 0.93$) Filtered grey water $80 < COD < 350$ mg O_2/L ($R^2 = 0.81$) Unfiltered raw sewage effluent $400 < COD < 1200$ mgO_2/L ($R = 0.726$) Filtered grey water $200 < COD < 500$ mg O_2/L ($R^2 = 0.88$)

UV-Visible spectrophotometry has several advantages when compared to other techniques (Thomas *et al.*, 1993; Winiarski *et al.*, 1995; Thomas *et al.*, 1997; Langergraber *et al.*, 2004a; Pons *et al.*, 2004; Vargas and Buitrón, 2006; Lourenço *et al.*, 2006), since traditional methods are inadequate for real-time monitoring of water quality (Bourgeois *et al.*, 2001; Fogelman *et al.*, 2006; Vargas and Buitrón, 2006). But optical techniques based on UV-Visible radiation also have some drawbacks that need to be overcome for better acceptance and implementation in wastewater treatment plants (Bourgeois *et al.*, 2001, Vaillant *et al.*, 2002; Pons *et al.*, 2004; Wu *et al.*, 2006). Table 4 resumes the main advantages and drawbacks of using UV-Visible spectroscopy for WW monitoring and control.

Table 4. Advantages and disadvantages of UV-Visible spectroscopy

Advantages	Disadvantages
Fast and simple technique	
No sample pre-treatment is required	Sensitive to turbidity
No chemicals addition	No detection of compounds with saturated bonds
No wastes production	Fouling
Small portable equipment	Signal saturation and necessity of sample dilution
Measurement of several parameters using only one sample	Calibration stability
Short analysis time	
Non-destructive	
Non-invasive	
Low maintenance cost	
Useful for on-line monitoring	
Different types of applications	

1.5 NIR Spectroscopy

1.5.1 Fundamentals

Near Infrared region was discovered in 1800 by Sir William Herschel being defined as "beyond the red". This spectral range interfaces the visible and infrared portions of the electromagnetic spectrum and there is a big controversy related to the definition of its limits (Figure 7). The American Society of Testing and Materials (ASTM) defined the NIR region of the electromagnetic spectrum between 780-2526 nm (12,820 cm^{-1} – 3959 cm^{-1}) (Reich, 2005).

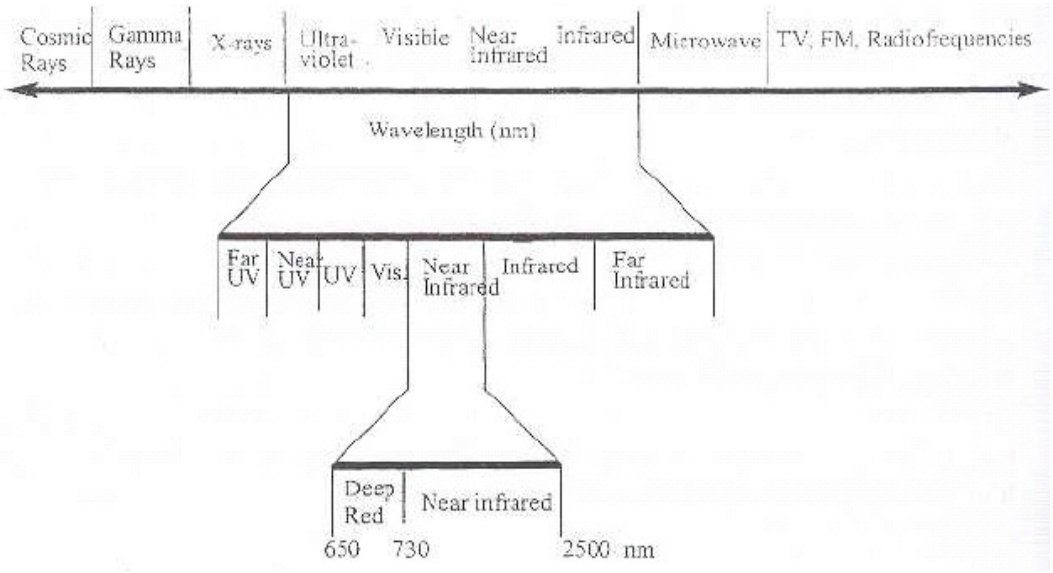


Figure 7. Position of NIR region in the electromagnetic spectrum (Raghavachari, 2001).

NIR spectroscopy employs a determined photon energy higher than necessary to promote molecules only to their lowest excited vibrational states (through a fundamental vibrational transition), but not too high for the molecules electron excitation (except for some rare earth compounds) (Pasquini, 2003).

The vibrations in a molecule can be described using the harmonic oscillator model, where only the transitions between consecutive energy levels in a molecule that cause a change in dipole moment are possible. However, this model cannot explain the behavior of actual

molecules, which resemble the model of an anharmonic oscillator. In an anharmonic oscillator energy levels are not equally spaced and transitions occur between non-contiguous vibrational states. These vibrational states yield absorption bands known as overtones, which are much less likely than the fundamental transitions (Blanco and Villarroya, 2002).

The NIR reflecting spectra are characterized by overtone and combination bands of fundamental vibrations occurring in the mid infrared (Büning-Pfaue, 2003), as presented in Figure 8. The band of the first overtone is 10-100 times weaker than that for the fundamental frequency, depending on the particular bond. These bands appear between 780 nm and 2000 nm, depending on the overtone order and the bond nature and strength. In polyatomic molecules, combination bands appear between 1900 and 2500 nm, when two or more vibrational modes interact and cause simultaneous energy changes (Blanco and Villarroya, 2002). The intensity of a given absorption band is associated with the magnitude of the dipole change during the displacement of atoms in a vibration and with its degree of anharmonicity (Pasquini, 2003).

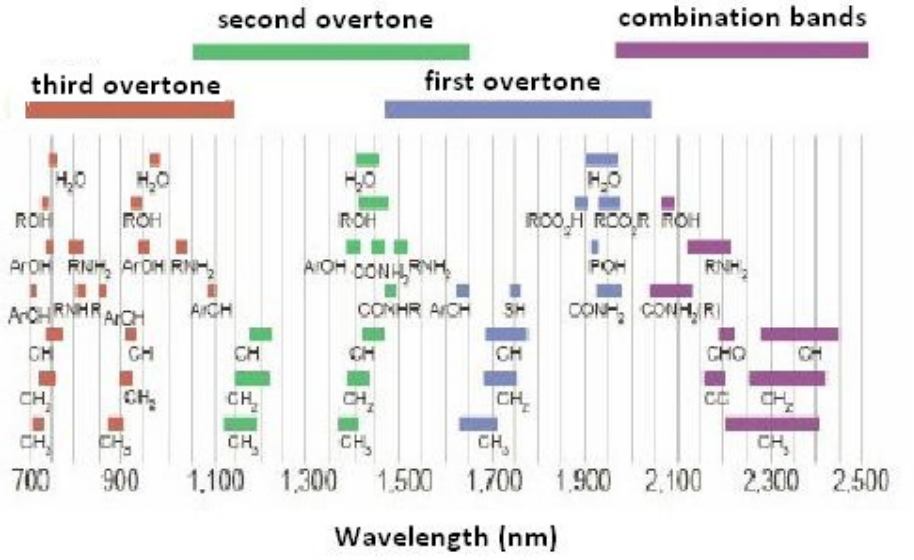


Figure 8. Principal analytic bands and location in NIR spectrum (according to Páscoa, 2006).

Since hydrogen is the lightest atom and, therefore, exhibits the largest vibrations and the greatest deviations from harmonic behavior (Blanco and Villarroya, 2002), NIR spectroscopy

is essentially useful to detect components that contain hydrogen atoms bound to a heteroatom (e.g. C–H for aliphatics and aromatics, N–H for proteins, O–H for alcohols and water) (Pons *et al.*, 2004). Due to this characteristic absorption, molecules containing bonds such as C=O, C–C and C–Cl give origin to much weaker or even absent bands which are difficult to detect by NIR spectral range (Blanco and Villarroya, 2002). All organic bonds have absorption bands in the NIR region, whereas minerals may only be detected in organic complexes and chelates or indirectly by their effect on hydrogen bonds (Büning-Pfaue, 2003).

The complexity of NIR absorption spectra of a substance or sample is due to the fact that in a given wavelength range, some frequencies will be absorbed, others will not and others will be only partially absorbed. Thus, a NIR spectrum is the result of a selective response that depends on the match between radiation energy and the energy difference between two vibrational levels (Pasquini, 2003).

NIR absorption bands are typically broad, overlapping, with poor baseline resolution and of much lower amplitude than those found in Mid Infrared (MIR), resulting from the convolution of the fundamental MIR vibrations. These characteristics make spectral information interpretation difficult and the use of chemometrics compulsory (Vaidyanathan *et al.*, 2001; Mark, 2001; Reich, 2005). However, if the fact that NIR absorptions are generally 10–100 times weaker than the fundamental bands of MIR can be an advantage it can also be a disadvantage, since it makes the method less suitable for detection of minor components, present in small concentrations (Pons *et al.*, 2004). Low absorption coefficient of NIR radiation makes it appropriate for measurements in solids and turbid liquids, since it allows direct analysis of strong absorbing or even high scattering samples (Reich, 2005).

In NIR spectroscopy the analytical signal is dependent on the chemical and physical properties (temperature, viscosity, turbidity, refractive index, ionic strength, etc.) of the sample (Pons *et al.*, 2004), what can be a positive or a negative aspect having into account the sensitivity of NIR spectroscopy (Reich, 2005).

1.5.2 Instrumentation

Spectrophotometers used to record NIR spectra are essentially identical to those employed in other regions of the electromagnetic spectrum. NIR equipment can incorporate a variety of devices (Figure 9), depending on the characteristics of the sample and the particular analytical conditions and needs (such as speed, sample complexity and environmental conditions) what makes this technique quite flexible (Blanco and Villarroya, 2002).

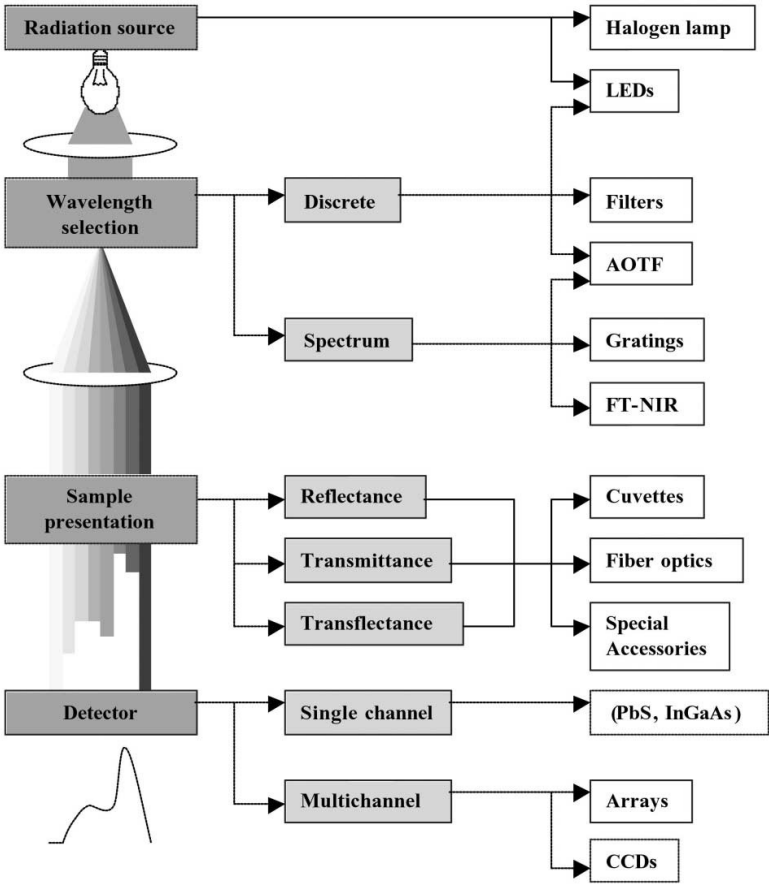


Figure 9. Principal features of NIR spectroscopy equipment (Blanco and Villarroya, 2002).

Samples can be analyzed directly by using cuvettes or, e.g., optical probes. The introduction of fiber optics contributed to an enormous expansion of NIR spectroscopy for remote measurements, as the case of food products or for real-time monitoring (Büning-Pfaue, 2003).

The light source is usually a tungsten halogen lamp, since it is small and rugged (Reich, 2005). Detector types include silicon, lead sulfide (PbS) and indium gallium arsenide (InGaAs), being able to impart a very high signal-to-noise ratio for NIR measurements, which will partially compensate for the lower intensities of NIR absorption bands (Pasquini, 2003).

In multi-channel detectors several detection elements are arranged in rows (diode arrays) or planes (charged coupled devices - CCD's) in order to record many wavelengths at once, so as to increase the speed at which spectral information can be acquired (Blanco and Villarroya, 2002).

Discrete-wavelength spectrophotometers can be used, irradiating only a few wavelengths, which makes them suitable only for applications with analytes that absorb in specific spectral zones. Diffraction grating, interferometer, diode-array or acousto-optic tunable filter (AOTF)-based instruments can provide full spectral coverage.

1.5.3 Measuring modes

NIR spectroscopy can be used to analyze samples with different types of physical properties. The most appropriate measuring mode will be dictated by the optical properties of the samples.

Transparent materials are usually analyzed by measuring the transmittance (Figure 10 (a)), as similarly to conventional UV-Visible spectroscopy. These samples can be measured in glass/quartz cuvettes with typical optical paths varying from 1 to 50 mm (Pasquini, 2003).

Transflectance is a special way to obtain transmittance measurement (Figure 10 (b)) and is frequently applied when optical bundle probes are employed. The difference relatively to a simple transmittance measurement is in doubling the optical path as the radiation beam passes twice through the sample (Pasquini, 2003).

In diffuse reflectance measurement of solid samples (Figure 10 (c)) the change in signal intensity is caused by scattering and absorbance by solid granules.

In the interactance mode there is a higher probability of interaction between the incident beam and the sample (Figure 10 (d)), thus, the emerging beam will contain more information regarding the sample constituents, reflecting better the actual composition of the sample.

Transmittance measurement of dense solid samples (Figure 10 (e)) has proved to be useful for quantitative determination of certain compounds in pharmaceutical tablets since, the longer optical path resulting from internal scattering, can provide information which is better correlated with the average sample content than the surface dominated diffuse reflectance signal.

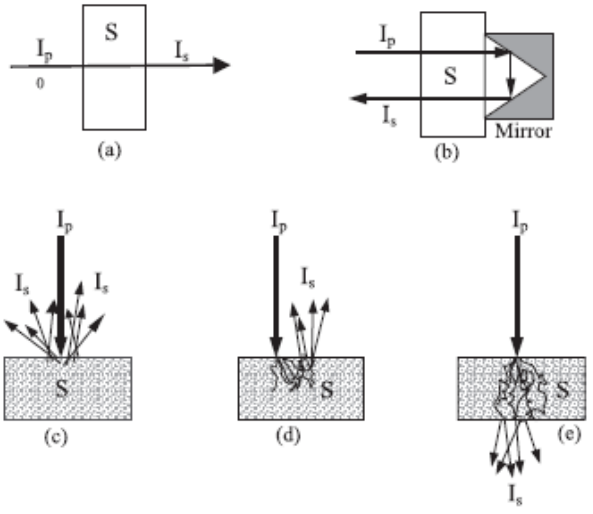


Figure 10. Modes of measurement employed in NIR spectroscopy. (a) transmittance; (b) transflectance; (c) diffuse reflectance; (d) interactance and (e) transmittance through scattering medium (Pasquini, 2003).

1.5.4 Applications

NIR first applications are dated from the mid-1960 at the U.S.A., when Karl Norris initiated his work searching for new methods to determine the moisture content of agricultural products, becoming one of the pioneers in this field (Nicolaï *et al.*, 2007).

After NIR potentialities were demonstrated in practical terms it didn't take too long for this technique to meet a fast development, provided mainly by instrumental (spectrophotometer) and chemometrics improvements.

Its application in food industry is still one of the most relevant, being used for several different purposes such as determination of oil content in hazelnut (Franco *et al.*, 2006) and adulteration (Ozen and Mauer, 2002); cheese quality control assessment (Čurda and Kukačková, 2004); analysis of fat, protein and casein in milk (Laporte and Paquin, 1999); determination of soluble solid content of tea (Li *et al.*, 2007) and assessment of fruit and vegetable quality (Nicolaï *et al.*, 2007). NIR spectroscopy has been extensively applied in pharmaceutical analysis, both qualitatively and quantitatively, namely in process analytical technology (PAT), defined as systems for real-time monitoring and control (Lopes *et al.*, 2004). Analysis of tablets, capsules, film coating and packaging, are only some of the possible applications in pharmaceuticals industry processes (Reich, 2005), being also helpful for counterfeit drug detection (Rodionova *et al.*, 2005). Applications in petroleum and fuel industry are also known, such as determination of components in crude oils (Aske *et al.*, 2001; Balabin *et al.*, 2007). The latest success of this technology in many areas of application is related with its ability to incorporate the benefits resulting from the development of correlated fields such as chemometrics, new materials for optical components, new sensors and sensor arrays, microcomputers and micro-electronics. This partnership will allow continuous and fruitful research and development in NIR applications (Pasquini, 2003).

While industrial processes are usually very well controlled, matrix composition and compounds concentration in a bioprocess can be subject of several variations. However, NIR range is gaining importance in the area of biotechnology, being considered an optical technique that has several features for cost-effective real-time bioprocess monitoring (Vaidyanathan and McNeil, 1998). Vaidyanathan *et al.* (2003) studied the influence of morphology of mycelial biomass in NIR spectra. Many examples of applications to suspended cultures using off-line or at-line measurements for bioprocess real-time monitoring have been described (Vaidyanathan *et al.*, 2001; Arnold *et al.*, 2002; Giavasis *et al.*, 2003).

Despite all the mentioned applications, the use of NIR technology in environmental processes is still underdeveloped, especially due to the limitations resulting from the complexity of the matrix (Pons *et al.*, 2004). However, the search for more suitable techniques in environmental monitoring and control using NIR range has been the subject of several studies. Regarding anaerobic processes, the monitoring of the dynamics of biogas production using NIR spectroscopy and electronic gas sensors has been performed by Nordberg *et al.* (2000). Concerning volatile fatty acids production, the NIR model developed for acetate achieved a root mean square error for prediction (RMSEP) of 0.20 g/l within the range of 0.14-1.72 g/l. Hansson *et al.* (2003) monitored an anaerobic digester using NIR spectroscopy and PCA. A PLS regression was also achieved for the prediction of propionate concentration (concentration range of 0.1-3.6 g/l, RMSEP of 0.53 g/l and correlation coefficient of 0.85), being considered suitable for real-time monitoring. COD, total solids (TS) and volatile solids (VS) concentrations were predicted by at-line NIR spectroscopy and image analysis for monitoring an industrial anaerobic process, being the best predictions obtained for VS (R^2 = 0.91), followed by TS (R^2 = 0.81) and COD (R^2 = 0.77) (Holm-Nielsen *et al.*, 2006). Stephens and Walker (2002) tested a NIR-Visible spectroscopic method to rapidly evaluate BOD_5 in a wastewater treatment plant. Transmission spectral measurements were performed in wastewater samples and PLS and principal component regression (PCR) models were used for BOD_5 prediction. Sousa *et al.* (2007) proposed a method based on NIR reflectance of seston for determination of COD in domestic wastewater. *In-situ* NIR spectroscopy was also used to monitor a lab scale activated sludge system, showing promising results for NIR applications in this type of biological processes (Dias *et al.*, 2008). In this work a PCA was performed, showing that it is possible to extract relevant qualitative information for process monitoring. More recently, Páscoa *et al.* (2008) monitored a sequential batch reactor (SBR) on-line for aerobic treatment of dairy residues, using a NIR *in-situ* transflectance probe, performing calibration models for TS, TSS and COD using PLS regression.

Some of the advantages exhibited by this method are also shared by other spectroscopic techniques, such as UV-Visible spectroscopy (Laporte and Paquin, 1999; Arnold *et al.*, 2000; Blanco and Villarroya, 2002; Büning-Pfaue, 2003; Chen *et al.*, 2004a; Pons *et al.*, 2004; Reich, 2005; Franco *et al.*, 2006; Qu *et al.*, 2008), but others can be adressed only to NIR, due to the intrinsic particularities characterizing this spectroscopic range (Blanco and Villarroya, 2002; Büning-Pfaue, 2003; Reich, 2005; Rodionova *et al.*, 2005; Blanco *et al.*, 2006). On the other hand, some disadvantages should be considered, in order to achieve new developments and decrease the technique's limitations (Vaidyanathan, 2001; Blanco and Villarroya, 2002; Büning-Pfaue, 2003; Reich, 2005; Franco *et al.*, 2006; Uddin *et al.*, 2006). Table 5 resumes the main advantages and disadvantages of NIR range.

Table 5. Common advantages between NIR and UV-Visible spectroscopy and NIR particular advantages and disadvantages

Common advantages	Particular advantages	Disadvantages
Fast and simple technique	Suitable for highly-scattering and strongly absorbing matrices (culture media)	Chemometrics use is compulsory
No sample pre-treatment	Determination of chemical and physical properties simultaneously (e.g. temperature, density, viscosity or particle size)	Laborious calibration procedures
Non-destructive	Records spectra for solid and liquid samples	Reference method is obligatory
No chemicals addition	NIR instruments have a very high signal to noise ratio (typically 10000:1)	Not very sensitive to minor components
No wastes production		Physical properties can interfere with chemical characteristics
More economical		
Suitable for on-line monitoring (optical fibers)		
Simultaneous analysis of several components		

1.6 Chemometrics

Spectroscopic techniques can deliver a large amount of data when several spectra, with several wavelengths, are recorded in order to have as much information as possible related to a process. Considering bioprocess applications, the matrix complexity can difficult compounds identification. Therefore, data-reduction techniques such as chemometric tools are essential to rapidly extract the relevant information, presenting the data in a more clear way.

According to the International Chemometrics Society "Chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods". Chemometrics can not only be used to design or select optimal measurement procedures and experiments, but also to provide the maximum relevant information by analyzing all sorts of data, having many different possible applications. Chemometrics has developed considerably in the past decades, being this related to the advances in intelligent instruments, laboratory automation, powerful computers and user-friendly software (Einax *et al.*, 1997).

Regarding environmental analytical methods, chemometrics is becoming a very important tool to understand the complexity of pollutants pathways and effects, since multicomponent and multielement analytical methods available produce an enormous flood of data (Einax *et al.*, 1997). As a consequence, it is often necessary to model results that are more difficult to obtain with results more easily obtained to better understand the process (Geladi *et al.*, 1999).

There are several different chemometric tools available and its use is defined by the nature of the data to be analyzed and the final purpose (mainly correlation or tendencies identification).

PCA is the multivariate statistical method most frequently used for environmental data analysis (Christensen *et al.*, 2004; Peré-Trepat *et al.*, 2004; Nake *et al.*, 2005). Considering wastewater monitoring it has already been applied in some cases (Winiarski *et al.*, 1995;

Rosen and Lennox, 2001; Miettinen *et al.*, 2004; Lourenço *et al.*, 2006; Aguado and Rosen, 2007).

For quantitative analysis PLS regression is frequently performed. PLS regression has been applied for several purposes in wastewater monitoring, namely for *in-situ* monitoring of a WWTP using UV-Visible spectroscopy (Langergraber *et al.*, 2004a, Rieger *et al.*, 2004; Rieger *et al.*, 2006) and *in-situ* SBR monitoring applications (Langergraber *et al.*, 2004b); Vargas and Buitrón, 2006).

PCA and PLS application in NIR spectroscopy is very usual, since these kind of statistical tools are essential for NIR spectral information extraction (Heikka *et al.* 1997; Aske *et al.*, 2001; Vaidyanathan *et al.*, 2003; Qu *et al.*, 2007). NIR spectroscopy applications in the environmental field using PLS regression were also recently reported in the literature (Holm-Nielsen *et al.*, 2006; Sousa *et al.* 2007).

A more detailed description of these and other chemometric tools used for this study will be presented in what follows.

1.6.1 Spectral preprocessing

Spectral preprocessing techniques are used to remove irrelevant information from raw data, making processing more feasible. This step is quite usual in defects removal (Büning-Pfaue, 2003) and to eliminate interferences that cause spectral variations during its acquisition, which can disturb precision of prediction models (Chen *et al.*, 2004a). Since there are many different types of mathematical methods for pre-treatment, it is necessary to understand which method can be the most suitable.

Several types of the pre-processing methods used for this study are described as follows.

1.6.1.1 Derivatives - Savitzky-Golay Algorithm

Derivative computation is used to remove baseline shifting and to help in the resolution of overlapping peaks, being the first and second derivatives the most commonly used (Büning-Pfaue, 2003). However, derivative spectra of order two are most popular as they can correct for both additive and multiplicative effects. Derivation is accomplished after filter application to remove noise and spectra incongruence. This is usually calculated according to a discrete form of filter namely the Savitzky–Golay algorithm (Savitzky and Golay, 1964). The Savitzky–Golay filter is a simplified least squares-fit convolution for smoothing and computing derivatives of a set of consecutive values (a spectrum) (Chen *et al.*, 2004b). These filters are essentially local functions that are applied to each spectrum and to use them it is necessary to specify some parameters, which define the local function resolution.

The convolution can be understood as a weighted moving average filter with weighting given as a polynomial of a certain degree. The weight coefficients when applied to a signal perform a polynomial least-squares fit within the filter window. This polynomial is designed to preserve higher moments within the data and to reduce the bias introduced by the filter. This filter can be applied to any consecutive data when the points of the data are at a fixed and uniform interval along the chosen abscissa, and the curves formed by graphing the points must be continuous and more or less smooth. The filter consists in the determination of a sequence of steps:

- Definition of the filter's order;
- Definition of the dimension of the filter (window's dimension);
- Obtain the coefficients according to the tabled values and divide them by a constant which depends on the order and size of the filter window (Páscoa, 2006).

1.6.1.2 Mean-Centering

Mean centering is usually applied to remove constant background contributions, which are considered of little interest for data variance interpretation. However, the use of this pre-treatment must be carefully analyzed given that data mean centering can result in a loss of quantitative information that may be important in environmental studies (Peré-Trepat *et al.*, 2004).

Mean centering is commonly applied on any multivariate calibration model, which involves calculating the average spectrum of all the spectra in the training set and then subtracting the result from each spectrum. In addition, the mean concentration value for each constituent is calculated and subtracted from the concentrations of every sample (Franco *et al.*, 2006). This operation ensures that results will be interpretable in terms of variation around the mean, being recommended for all practical applications (Nicolaï *et al.*, 2007).

The mean-centering operation can be represented by equation 6:

$$\boldsymbol{x}_{i,cent} = \boldsymbol{x}_i - \bar{\boldsymbol{x}} \qquad (6)$$

Where

$\boldsymbol{x}_{i,cent}$ - centered value of $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}$ - mean value of variable $\boldsymbol{x}$, the original feature.

The mean of the variable x after centering is zero point on the new $\boldsymbol{x}_{i,cent}$ - axis. Mean centering is only possible if the variances of the different features are similar. Otherwise autoscaling is necessary (Einax *et al.*, 1997).

1.6.1.3 Standard Normal Variate

Standard normal variate (SNV) was developed by Barnes *et al.* (1989). This method is used for centering and scaling individual spectra. Processing according to SNV is given by equation (9) (Naes *et al.*, 2002):

$$\hat{x}_{ik} = \frac{x_{ik} - m_i}{s_i} \quad (7)$$

where,

x_{ik} – spectra measurement at the kth wavelength for the ith sample;

m_i – mean of the k spectral measurements for sample i;

s_i – standard deviation of the same k measurements.

This kind of pre-treatment is used in many spectroscopic applications. SNV makes an additive and multiplicative adjustment. It is performed without a reference spectrum, improving predicting precision but not simplifying the model. SNV standardizes each spectrum using only the data from that spectrum and not using the mean spectrum of any set. This method cannot reduce the influence of water (in the case of NIR spectra) neither the systematic interference.

1.6.2 Principal Component Analysis

One problem regarding multivariate data is that its sheer volume makes it difficult to see patterns and relationships. The amount of intensity measurements obtained in a spectrum is directly related to the amount of values that a correlation matrix would have. There is frequently some correlation between the variables which makes some information redundant.

Principal Component Analysis (PCA) is aimed at finding and interpreting hidden complex and possibly causally determined relationships between features in a data set. Correlating features are converted to factors which are non correlated (Einax *et al.*, 1997). PCA can reduce the amount of data when there is correlation present. However it is not a useful technique if the variables are uncorrelated (Miller and Miller, 2000).

The method can be mathematically explained as follows (Carvalho *et al.*, 2006):

Considering a matrix $\boldsymbol{X}$, where each row in the matrix represents the spectrum at one point in time and each column represents the absorbance at a given wavelength, the central point of PCA is to reduce the original data matrix $\boldsymbol{X}$ to factor loadings and factor scores.

This matrix $\boldsymbol{X}$ (dimensions $\boldsymbol{I} \times \boldsymbol{J}$) can be decomposed into a product of two matrices:

$$\boldsymbol{X} = \boldsymbol{T} \times \boldsymbol{P} + \boldsymbol{E} \qquad (8)$$

The $\boldsymbol{T}$ matrix contains the scores of $\boldsymbol{I}$ objects on $\boldsymbol{K}$ principal components. The $\boldsymbol{P}$ matrix is a square matrix and contains the loadings of $\boldsymbol{J}$ variables on the $\boldsymbol{K}$ principal components. $\boldsymbol{E}$ is the error matrix. If the original data matrix is dimension $\boldsymbol{I} \times \boldsymbol{J}$, no more than $\boldsymbol{J}$ principal components can be calculated if $\boldsymbol{J} \leq \boldsymbol{I}$. PC1 (Principal Component #1) represents the direction in the data, containing the largest variation. PC2 (Principal Component #2) is orthogonal to PC1 and represents the direction of the largest residual variation around PC1 and so on. These will contain less and less variation and therefore less information. Hence, when significant correlation occurs, the number of useful PCs is much less that the number of original variables. It is often found that PC1 and PC2 account between them for most of the variation on the data set. As result, the data can be represented only in two dimensions instead of the original *n* (Miller and Miller, 2000). The first scores vector and the first loadings vector are often called the eigenvectors of the first principal component. Each successive component is characterized by a pair of eigenvectors.

The information provided by the monitoring algorithm can be used to classify the current operational state of the process, being able to identify different locations in a multivariate space which describe different operational states. Thus, a cluster represents similar process behavior and different clusters usually characterize different operational states. Many clustering techniques are available and can be applied either to the original data or to the scores from a multivariate projection method (Aguado and Rosen, 2007).

Once two or more groups have been identified by using PCA, it may be possible to explain the differences between them in terms of chemical structure. The ultimate objective is to identify and give a physical interpretation to the principal components. For this reason,

principal components are sometimes referred to as latent (i.e. hidden) variables (Miller and Miller, 2000).

1.6.3 Outlier Detection

Outliers may derive from all sorts of mistakes or problems, such as interface errors, sensor malfunctions, fouling or bad sample. But an outlier is not necessarily an erroneous observation, but merely an observation that is different from the rest and may possibly have a strong influence on the results. One of the reasons for an observation to be considered an outlier is when a sample, either a calibration or a prediction sample, belongs to another population than the "normal" samples (Naes *et al.*, 2002). A sample may be outlying according to the $\boldsymbol{x}$ - variables only, to the $\boldsymbol{y}$ - variables only, or to both. Alternatively, the leverage of a spectrum may be calculated as the distance to the centre of all spectra relative to the variability in its particular direction. If the leverage exceeds a certain threshold value the spectrum may be considered as an outlier. In practice, however, only those outliers which have an effect on the regression model are to be removed. Excessive pruning of the data set for outliers should be avoided because important information can be rejected (Nicolaï *et al.*, 2007).

1.6.4 Residuals Statistics (Q)

Residuals statistics (Q) is used to analyze the deviation between each sample and the model. The vector for prediction error for each sample is given by $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{x}_i \boldsymbol{b}_{PLS}$. The Q statistic for each sample i is:

$$\boldsymbol{Q}_i = \boldsymbol{e}_i \boldsymbol{e}_i^{\boldsymbol{t}} \qquad (9)$$

The confidence level for Q statistic is estimated through X^2 distribution if residues are independent and normally distributed (Páscoa, 2006).

The samples with a Q statistic value higher than the estimated confidence level are considered outliers.

1.6.5 Partial Least Squares

To establish a relation between a spectrum and a chemical property it is necessary to perform a model calibration. The achievement of an optimal calibration is an essential step for applying spectroscopic techniques. The calibration emerges from previous analysis of a lot of samples, which must be representative of the expected variability in unknown samples (Franco *et al.*, 2006).

Partial Least Squares (PLS) regression is a major regression technique for multivariate data used for calibration (Carvalho *et al.*, 2006). PLS has been applied to many fields in science with great success. One important feature of PLS is that it takes into account errors in both the concentration estimates and spectra. In a manner similar to PCA, PLS extracts linear combinations of essential features which model the original data. But, compared with PCA method, PLS can also model the dependence of the two data sets. In contrast with other methods of multivariate data analysis the PLS algorithm is an iterative algorithm which makes it possible to treat data with more features that objects. This type of model is well suited for modelling and simulating environmental relationships or for multivariate calibration (Einax *et al.*, 1997). The four steps for the application of PLS (Carvalho *et al.*, 2006), are described as follows:

1. A calibration design is built with a training set;

2. The optimum number of PLS components is selected using cross-validation;

3. The prediction capacity of this model is assessed with an additional group of samples called a test set;

4. The model is then applied to predict the concentration profiles during the reaction period.

There are different algorithms to calculate PLS, namely NIPALS algorithm (Rosen, 2001). Given a matrix of cause data, X (of size $m \times n_x$, where m is the number of observations and n_x

is the number of cause variables), and effect data, Y (of size $m \times n_y$, where n_y is the number of effect variables), a factor of the cause data, t_k (length m), and effect data, u_k (length m), is evaluated, such that

$$X = \sum_{k=1}^{np<nx} t_k p_k^T + E \quad (10)$$

$$Y = \sum_{k=1}^{np<nx} u_k q_k^T + F \quad (11)$$

The t_k and the u_k vectors are selected to maximize the covariance between each pair (t_k, u_k). The matrix E and F are errors and p_k and u_k are referred to as loading vectors. Linear regression is performed between the p_k and the u_k vectors, to produce the inner relationship:

$$u_k = b_k t_k + \varepsilon_k \quad (12)$$

where b_k is a regression coefficient, and ε_k refers to the prediction error.

It is important to determine how many significant PLS components should be used to obtain an accurate and robust model. A model accuracy assessment is accomplished by using validation procedures. These procedures can also avoid over fitting. Cross-validation is usually employed for this purpose. In leave-one-out cross validation, one sample is removed from the dataset, and a calibration model is constructed for the remaining subset. The removed samples are then used to calculate the prediction residual. The process is repeated with other subsets until every sample has been left out once and in the end the variance of all prediction residuals is estimated. After selecting the proper number of latent variables, the robustness of the model is finally tested by analyzing the predictive ability of a model created on part of a dataset (training set) and evaluating how well it predicts the remainder of the data (test set).

Once the PLS model has been applied to the training set, validated using the test set and demonstrated to have good predictive abilities, it can be applied to new datasets where the concentration profiles are unknown.

1.6.6 Bootstrapping – Wavenumber selection

Regarding spectroscopic applications, usually not all the spectrum is equally relevant. It is possible to select the wavenumber regions that contain the required information, in order to improve the model robustness and prediction ability. Bootstrapping can be used for this purpose. Bootstrapping is a statistical method that generates a set of samples by sampling with replacement from the original data set. A typically large number of "new" data sets are generated, each one with the same size of the original data set (Wehrens and Van der Linden, 1997).

From each bootstrap sample the statistical parameter of interest is calculated. This yields an ensemble of estimates that is used to obtain, for example, the mean, standard error, or confidence intervals for different model statistical parameters (Wehrens and Van der Linden, 1997). In spectral data, this procedure can be used to assess the statistical significance and the standard deviation of regression coefficients (confidence intervals). If the interval encloses the zero value then that corresponding wavenumber is discarded. After bootstrapping, the group of wavelengths which was not discarded is used to perform a final PLS regression.

In the regression context two types of bootstrapping methods can be used: bootstrap objects (BO) and bootstrap residuals (BR). In BO the new data sets are generated by randomly drawing objects (original variables) with replacement, while BR performs the resembling of the uncertainty estimates (residuals) from the regression model (Faber, 2002).

1.6.7 Model Accuracy

The prediction error of a calibration model is defined as the root mean square error for cross validation (RMSECV) when cross validation is used (Nicolaï *et al.*, 2007):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i)^2}{\boldsymbol{n_p}}} \qquad (13)$$

with $\boldsymbol{n_p}$ the number of validated objects, and $\hat{\boldsymbol{y}}_i$ and $\boldsymbol{y_i}$ the predicted and measured value of the ith observation in the test set, respectively. This value gives the average uncertainty that can be expected for predictions of future samples. The number of latent variables in the calibration model is typically determined as that which minimizes the RMSECV. In some publications the standard error of prediction (SEP) is reported:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i - \boldsymbol{b})^2}{\boldsymbol{n_p}}} \qquad (14)$$

with b the model bias.

Another useful statistic is the R^2 value. It essentially represents the proportion of explained variance of the response variable in de calibration (R^2_c) or validation (R^2_v) set.

1.6.8 Model robustness

Calibration models are called robust when the prediction accuracy is relatively insensitive towards unknown changes of external factors. The main factors which may affect model performance are (Wang *et al.*, 1991): (i) the calibration model developed on one instrument is transported to another instrument that produces instrumental responses that differ from the responses obtained on the first instrument; (ii) the instrumental responses measured on

a single instrument drift because of temperature fluctuations, electronic drift, and changes in wavelength or detector stability over time; and (iii) the samples belong to different batches. Clearly model robustness increases at the expense of model accuracy.

2. Materials and Methods

2.1 Activated sludge system

A complete-mix activated-sludge reactor with both populations of microorganisms present in suspension (heterotrophic microorganisms and nitrifying bacteria), was monitored considering only the nitrification process. By promoting disturbances to the system it was possible to follow the consequent phenomena with spectroscopic techniques.

The lab scale system used in this work is based on a 25 L total volume tank with 17 L of suspended biomass, followed by a 2.5 L cylindrical settler (Figure 11).

The system was fed with a synthetic wastewater based on peptone and meat extract as carbon sources. Every two days new feed solution was prepared according to Marquéz *et al.* (2004) (Table 6). The synthetic wastewater solution was kept in the fridge to avoid degradation inside the reservoir. Tap water was used to prepare the feed solution and was pumped to dilute twice this solution before entering the system.

To maintain pH values between 7.2 and 7.5 the system was controlled with a pH meter and control pump (Model BL 7916 – BL 7917, Hanna Instruments), by pumping a NaOH solution into the system to increase pH. When nitrification process was significantly disturbed a HCl diluted solution was dosed to the reactor to decrease the pH.

The complete mix inside the reactor was guaranteed by a continuous inflow of air bubbles by using an air diffuser which covered its bottom. This system maintained the dissolved oxygen above 7 mg O_2/L being the concentration in the reactor measured by a TriOxmatic 690 dissolved oxygen probe (WTW) connected to an Oxi 296 R/RS monitor (WTW). Sludge recirculation from the settler to the reactor was guaranteed by an air pump. Reactor was inoculated with activated sludge from Frossos WWTP (Braga), diluted with feed solution.

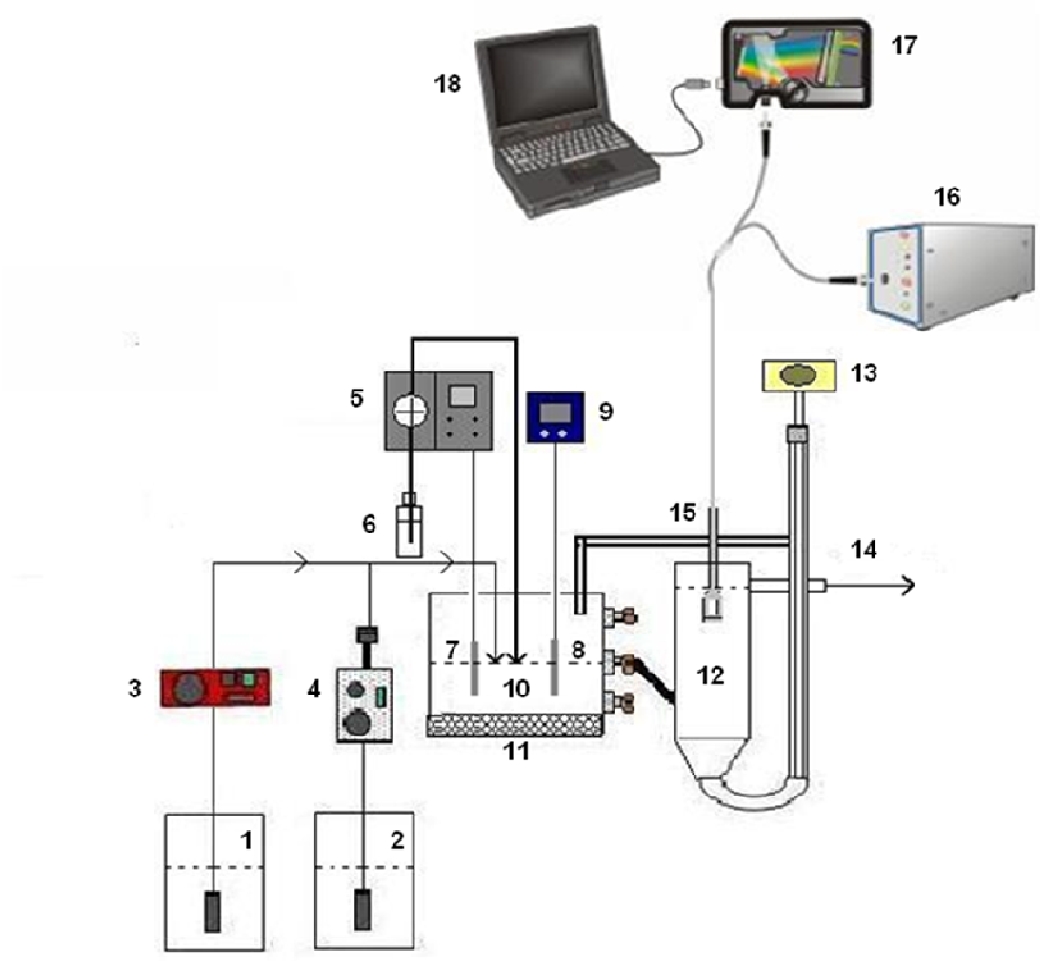


Figure 11. Activated sludge system layout. Legend: 1 – concentrated synthetic wastewater; 2 – tap water; 3 – peristaltic pump; 4 – metering pump; 5 – pH meter and control pump; 6 – base/acid solution; 7 – pH sensor; 8 – dissolved oxygen sensor; 9 - dissolved oxygen monitor; 10 – aerated completed mix reactor; 11 – aeration system; 12 – settler; 13 – air pump; 14 – effluent; 15 – immersible probe; 16 – light source; 17 – spectrometer; 18 – computer.

Table 6. Composition of concentrated synthetic wastewater

Component	Mass concentration (mg/L)
Peptone	1900
Meat extract	110
Urea	30
K_2HPO_4	28
NaCl	7
$CaCl_2.2H_2O$	4
$MgSO_4.7H_2O$	2

2.2 UV-Visible and NIR equipments

UV-Visible and NIR spectra were acquired with immersion probes in the ranges from 230 to 700 nm and 900 to 1700 nm, respectively. The operating mode is the same in both cases. Light travels from the light source through a lens near the end of the probe. The light is then transmitted through the sample compartment to a second-surface mirror. The light reflects and travels back through the sample compartment a second time and is then focused by the lens onto the read fibber and through the spectrometer. The returned beam is sent to the portable detectors connected to a computer, allowing the immediate spectra visualization and acquisition.

The optical path was twice the value of the mechanical gap of the probes. It is adaptable and depends on the characteristics of the reactor content. In this work the optical path was equal to 1 cm, for both probes.

The Ocean Optics USB4000 portable dispersive UV-Visible equipment was used to spectra acquisition in the UV-Visible range. The instrument is a high-performance 3648-element linear CCD-array detector equipped with a grating for the UV range, connected to a DH-2000 deuterium tungsten halogen light source that combines the continuous spectrum of deuterium and tungsten halogen light sources in a single optical path. The immersion probe

is connected to the light source and to the spectrometer by means of two TP300-UV-Visible solarization-resistant optical fibers.

Ocean Optics NIR 512 portable dispersive NIR equipment including a PDA cooling detector was used for spectra acquisition in the NIR range. A transflectance probe (Ocean Optics/T300RT) is connected through optical fibers (OceanOptics/QP400-2-VISNIR) to a light source (Stellarnet/SL1) and to the NIR spectrometer.

In both cases spectra were acquired using the OOIBase32/Ocean Optics software. This software allows the configuration of certain parameters like the integration time, average spectra, filter type (to avoid noise mostly when low integration times are used) and the temperature of the detector in the case of the NIR equipment.

The MatLab version 6.5 Release 13 (The Mathworks, Inc) was used for data treatment, calibration and validation of the chemometric models. The chemometric functions included in the PLS MatLab Toolbox (PLS Toolbox, Eigenvector Research, Inc) were used to generate the PCA model.

2.3 *In-situ* process monitoring

For *in-situ* monitoring the probes were immersed in the settler at the same time, acquiring spectra simultaneously. The spectra acquisition was performed during not more than 45 minutes, every monitoring day (2-3 times a week). The amount of spectra acquired varied according to the integration time, which was selected to give an average of 15 spectra in the settler, in each monitoring day.

The usual procedure performed for both probes is described in Appendix A.

2.4 Off-line process monitoring

Several analytical parameters were monitored off-line during activated sludge process operation:

- COD in the influent and in the effluent (2-3 times a week);
- TSS and volatile suspended solids (VSS) in the reactor and in the effluent (2-3 times a week);
- Kjeldahl nitrogen in the influent and in the effluent (once a week);
- Nitrate and nitrite concentration in the effluent (2-3 times a week);
- Ammonium in the effluent (2-3 times a week).

COD, TSS, VSS, Kjeldahl nitrogen and nitrite were analyzed according to Standard Methods for the Examination of Water and Wastewater (20^{th} Edition).

Samples were collected from the settler and were analyzed after pre-treatment (centrifuged and filtered) and with no pre-treatment (TSS and VSS determination). Feed samples were analyzed after being pre-treated. Samples were collected from the reactor for solids analysis.

In order to obtain a comparison between *in-situ* and off-line UV-Visible spectra acquisition, samples from the settler without pre-treatment were analyzed in a quartz cell with 1 cm path length with a Jasco V-560 UV-VIS spectrophotometer.

2.4.1 COD determination

COD determination was based on a colorimetric method, in closed reflux, according to method 5220 D from Standard Methods.

For COD calibration potassium hydrogen phthalate (KHP) solutions were prepared with the following concentrations in terms of COD: 20, 50, 100, 200, 400, 600, 800 and 1000 mg O_2/L.

Solutions were prepared in triplicate. Digester was connected to reach the temperature of 148ºC. 2.5 mL of sample or standard solution was added to a digestion tube, where 1.5 mL of digestion solution and 3.5 mL of sulfuric acid solution were carefully added after. The tubes were carefully mixed and digested at 148ºC for 2 hours. After digestion, the tubes were taken from the digester to cool down and reach room temperature. The absorbance at 620 nm was measured for each tube (standard, sample or blank).

With the absorbance measurements from standard solutions a calibration curve was calculated, with a correlation coefficient close to unit, and used for sample's COD calculation.

COD samples from the activated sludge process were analyzed immediately after being collected and pre-treated. Thus, only soluble COD was measured. COD analysis was performed in duplicate in the inlet and in triplicate in the outlet.

Reagents preparation is presented in Appendix A.

2.4.2 TSS and VSS determination

TSS and VSS were determined according to methods 2540 D and 2540 E from Standard Methods, respectively.

Whatman glass-fiber filter disks type AP40 were washed in a filtration apparatus with distilled water. The filter disks were transferred to an aluminum weighting dish and ignited at 550ºC during 30 min in a muffle furnace. The disk and aluminum dish were cooled in a desiccator and then weighted (m_1).

Homogeneous samples were taken from the reactor (V = 5 mL) and from the settler (V = 40 mL). These samples were filtered in a filtration apparatus and the glass-fiber disk, aluminum dish and residue retained on the filter (set) were dried at 105ºC during one day. After, the set was cooled in a desiccator and then weighed (m_2).

TSS were calculated according to the following equation:

$$TSS\ (g/L) = \frac{(m_2 - m_1)}{V} \times 1000 \quad (15)$$

After TSS determination, VSS were determined. The residue from later procedure was ignited at 550ºC in a muffle furnace during 2 hours (m_3).

The following equation was used to calculate VSS:

$$VSS\ (g/L) = \frac{m_2 - m_3}{V} \times 1000 \quad (16)$$

TSS and VSS were analyzed in triplicate.

2.4.3 Kjeldahl nitrogen determination

Kjeldahl nitrogen was determined according to Macro-Kjeldahl method 4500-N_{org}-B, from Standard Methods. Kjeldahl method was used for determining organic nitrogen (N_{org}) and ammonium in the feed and effluent samples.

A detailed description of reagents preparation and digestion and distillation procedure is presented in Appendix A.

In each digestion tube 10 mL of concentrated H_2SO_4 were added to 10 mL of sample or ultrapure water (blank), followed by the addition of one Kjeltab with selenium. The tubes were carefully mixed and digested at 400ºC connected to a system for vapors aspiration.

An automatic distillation system (Tecator Kjeltec 1026) was used to rapidly distillate all the samples and blanks after digestion. After distillation each solution is titrated with a solution of H_2SO_4 (0.025 mol/L) until the color changes from green to grey.

N-Kjeldahl concentration was calculated according to the following equation:

$$N\text{-}Kj\ (mg/L) = \frac{V_{acid} - V_{acidblank}}{V_{sample}} \times 14.01 \times 1000 \times 2 \times C_{H2SO4} \quad (17)$$

where,

$V_{acid\ sample}$ – Volume of acid used for sample titration;

$V_{acid\ blank}$ – Volume of acid used for blank titration;

V_{sample} – Initial sample volume (in this case 10 mL);

C_{H2SO4} – Concentration of acid solution used for titration in mol/L.

Samples for Kjeldahl nitrogen analysis were pre-treated and previously acidified and frozen for future analysis.

2.4.4 Nitrate determination

Nitrate was determined by using high performance liquid chromatography (HPLC) equipment (Jasco, Japan) with automatic injection.

HPLC operational conditions are presented in Appendix A.

For calibration curve a stock solution of 100 mg $N\text{-}NO_3^-$/L was used to prepare a 50 mg $N\text{-}NO_3^-$/L to be used for subsequent dilutions. KNO_3 was previously dried at 105ºC during 24h for stock solution preparation.

To avoid interferences resulting from the presence of nitrite, samples, standards and blank solutions were diluted on a ratio of 1:1 with a sulfamic acid solution (0.05 mol/L).

Nitrate calibration curve was performed using the following concentrations: 0, 1, 2.5, 5, 10, 15, 20 and 25 mg $N\text{-}NO_3^-$/L. Correlation coefficient was close to unit.

Samples for nitrate analysis were pre-treated, acidified with HCl and frozen for analysis.

A software (Varian Star Workstation) was used to integrate the resulting peaks for standards and samples.

2.4.5 Nitrite determination

Nitrite was determined with Griess-Hosvay method, similar to the colorimetric method 4500-NO_2^--B, from Standard Methods.

100 µl of sulfanilamide reagent were added to 5 mL of sample, standard or blank and mixed. After 2 minutes 100 µl NEDD reagent was added and mixed. After 10 minutes absorbance at 543 nm was recorded in a Jasco V-560 UV-Visible spectrophotometer with a quartz cell with 1 cm path length.

For calibration curve a stock solution of 50 mg N-NO_2^-/L was used to prepare a 0.5 mg N-NO_2^-/L to be used for subsequent dilutions. $NaNO_2$ was previously dried at 105 ºC during 1h for stock solution preparation.

Nitrite calibration curve was performed using the following concentrations: 0, 0.05, 0.1, 0.15, 0.2, 0.25 and 0.5 mg N-NO_2^-/L. Correlation coefficient was close to unit.

Samples for nitrite analysis were pre-treated, acidified with HCl and frozen for future analysis.

Reagents preparation is presented in Appendix A.

2.4.6 Ammonium determination

Ammonium was determined according to Nessler's method.

Reagents

- Nessler reagent (commercial reagent);

For calibration curve a stock solution of 1 g N-NH_4^+/L was used to prepare a 10 mg N-NH_4^+/L to be used for subsequent dilutions. NH_4Cl was previously dried at 105 ºC during 1h for stock solution preparation.

Ammonium calibration curve was performed using the following concentrations: 0, 1, 2, 3, 5, 10 and 15 mg N-NH_4^+/L. Correlation coefficient was close to the unity.

0.25 mL of Nessler reagent was added to standards, blank and samples. After 15 minutes absorbance at 425 nm was recorded in a Jasco V-560 UV-Visible spectrophotometer with a quartz cell with 1 cm path length.

Samples for ammonium analysis were pre-treated and analyzed after sample collection.

3 RESULTS AND DISCUSSION

The main objective of this work was to detect different kind of variations related to the operation of a lab scale activated sludge system using UV-Visible and NIR immersion probes able to detect those changes through on-line monitoring. Most of the disturbances were induced to the system. Even though the main purpose of a wastewater treatment system is a good stability and efficiency in terms of pollutants degradation, for this study this was not always the objective. Many situations that are usually avoided in a real activated sludge process were promoted in this work as a source of important information regarding what can and cannot be detected by *in-situ* spectroscopy.

Motivated by some observations made during this work and information collected from the literature, it was also found important to optimize the spectra acquisition process which included comprehending how the spectra can change during a long acquisition period without cleaning, while immersed in the settler.

3.1 Synthetic wastewaters study

On-line monitoring is of great interest for fault prevention (e.g. toxic pollutants inlet, variations in influent composition or in certain compounds concentration, etc.) hence, it is important to know if the immersible probes can detect when something different is present in the influent of the treatment system. It is known that biological wastewater treatment processes like activated sludge process can be quite sensitive to changes in the influent, which can disturb the process so dramatically that can lead to a total loss of the biological activity.

Focusing on inlet composition and concentration variations, a test was performed to investigate which information can be obtained by the two probes when immersed in solutions of different composition and concentration.

The selection of the different solutions intends to accomplish two main objectives: to compare the results between the two probes in terms of distinction between the solutions and to compare the ability of the probes to distinguish different concentrations of the same solution.

For that purpose, three different stock solutions were prepared with COD concentrations between 1 and 1.3 g O_2/L, with a composition according to Table 7. Dilutions from the stock solution were prepared. Three spectra were acquired for the stock solutions. The dilutions were prepared in triplicate to analyze the reproducibility of each probe. Measured COD concentrations of the tested solutions are presented in Table 8.

Table 7. Composition of the studied solutions

Solution	**Composition**
# 1	Peptone, meat extract, urea, K_2HPO_4, NaCl, $CaCl_2.2H_2O$ and $MgSO_4.6H_2O$
# 2	Glucose, meat extract, urea, K_2HPO_4, NaCl, $CaCl_2.2H_2O$ and $MgSO_4.6H_2O$
# 3	Skim milk

Table 8. Average COD concentrations of stock and diluted solutions

Solution	**COD (mg O_2/L)**		
Solution #1		1187.9	
1 A	1057.7	1051.4	1057.7
1 B	802.8	810.7	826.4
1 C	585.6	585.6	588.7
Solution #2		1074.5	
2 A	988.4	982.1	975.8
2 B	747.7	751.6	751.6
2 C	532.0	535.2	535.2
Solution #3		1313.2	
3 A	1183.7	1208.9	1196.3
3 B	924.9	924.9	917.0
3 C	667.5	664.3	665.4

The first solution was already tested as a synthetic effluent used for the study related to the removal of organic pollutants on an activated sludge system similar to the one used in this work (Márquez *et al.*, 2004). This feed solution consists in a very nutritive medium which avoids biomass growth limiting problems. Its composition based on peptone and meat extract, as carbon sources, is also rich in nitrogen and should be detected by both spectroscopic techniques.

Since glucose cannot be directly detected by UV-Visible spectroscopy it was found interesting to study if NIR spectroscopy could be more effective in the detection of a solution with glucose as the main constituent. With this in mind, instead of peptone, glucose was added to a solution with the same compounds present in solution #1. Meat extract and urea can be detected by UV-Visible probe but since these were added in very small amounts it was expected a low absorbance spectrum for these solutions using UV-Visible probe, probably close to its detection limit.

Milk proteins (e.g. casein) can be detected by UV-Visible spectroscopy, however, sugar present in milk (e.g. lactose) is not possible to detect directly by UV-Visible spectroscopy, being an important constituent of milk. The idea was to investigate if an effluent of a dairy industry could be better detected by the NIR or the UV-Visible probe, knowing that NIR as been studied as a alternative method for determination of main constituents of milk, such as fats, proteins and lactose (Laporte and Paquin, 1999; Šašić and Ozaki, 2001). In this case the content of fat is expected to be low.

In Figure 12 UV-Visible and NIR raw spectra of the solutions above described are presented. By analyzing the different spectra is already possible to detect the main differences between the UV-Visible and the NIR spectra. While for UV-Visible a variation in the composition can be visually detected by a change of the spectra's shape, giving already some information, in NIR spectra the changes are very difficult to be noticed with naked eye. An expressive shift in the baseline is observed for skim milk solution in UV-Visible spectra and for peptone in NIR spectra. In the first case, the baseline shifts are suggested to be due to the milk solution's characteristic turbidity and its natural decrease along the dilutions, what could be detected by the visible part of the spectrum.

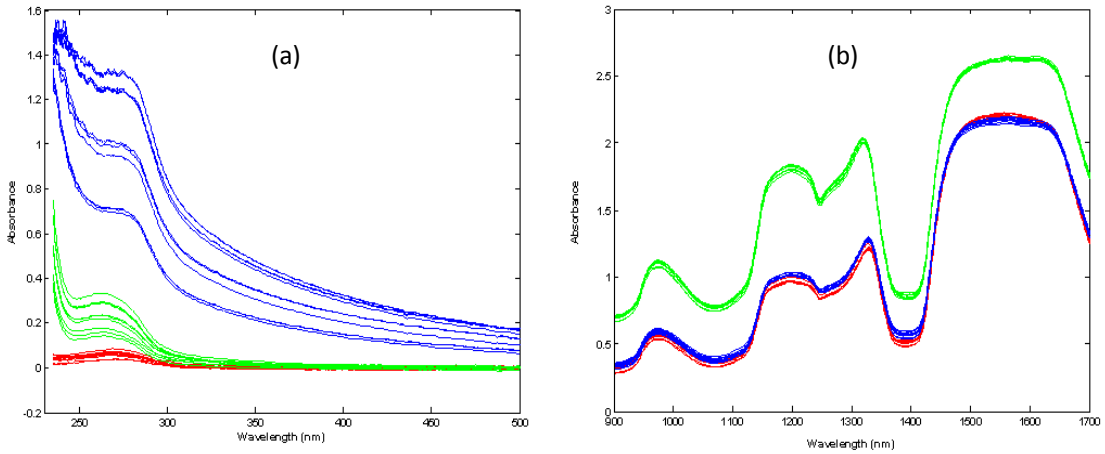


Figure 12. UV-Visible (a) and NIR (b) raw spectra acquired for all the measured solutions. Green line – peptone; red line – glucose; blue line – skim milk.

Several spectral ranges and raw spectra pre-treatments were selected and studied to investigate which combination could achieve the best clustering results with PCA. For NIR probe the studied ranges were: 900-1700 nm, 1000-1600 nm and 1100-1400 nm. For UV-Visible the following ranges were investigated: 235-500 nm, 250-380 nm and 270-310 nm. In the case of the NIR spectra analysis, the objective was to compare the entire spectral range with smaller ranges, where some differences between the three solutions could be observed with naked eye. The selection proposed for UV-Visible is related to the fact that UV region concentrates a great amount of information, regarding a wastewater spectrum (Figure 5), being interesting to compare two different ranges inside de UV region to the entire UV-Visible region in order to analyze the amount and relevance of the information that each specific range can represent.

Table 9 shows the optimal ranges and pre-treatments selected for NIR and UV-Visible spectra.

Table 9. Selected spectral ranges and pre-treatments for solutions study by PCA

Probe	PCA (solutions)	Spectral range (nm)	Pre-treatment
NIR	#1, #2, #3	900 - 1700	Savitzky-Golay (15,2,1) and MNCN
	#1	900 - 1700	Savitzky-Golay (15,2,2), SNV and MNCN
	#2	900 - 1700	Savitzky-Golay (15,2,2), SNV and MNCN
	#3	900 - 1700	Savitzky-Golay (15,2,1) and MNCN
UV-Visible	#1, #2, #3	235 - 500	SNV and MNCN
	#1	270 - 310	Savitzky-Golay (15,2,1), SNV and MNCN
	#2	270 - 310	Savitzky-Golay (15,2,1) and MNCN
	#3	270 – 310	Savitzky-Golay (15,2,1) and MNCN

Legend: SNV – Standard Normal Variate; MNCN - Mean Centering.

The entire NIR spectral range was selected for the study, being the results quite approximated to the ones obtained for the 1000-1600 nm range. For UV-Visible probe the best results were achieved using the ranges 235-500 nm and 250-380 nm for solutions composition differentiation, being the results here presented relative to the first range selected. The selection of the range 270-310 nm was crucial for solution #2 concentration's differentiation. For solution #1 and solution #3 good results were achieved with any of the spectral ranges, being selected the score plot for spectral range of 270-310 nm for results presentation.

Analyzing the PCA score-plots (Figure 13) it is possible to identify the differences between the solutions using two principal components, for both probes. In both cases a clear distinction is observed with the formation of three independent clusters in the score plot. It is also possible to observe that UV-Visible probe can already differentiate between the dilutions of skim milk, what could be expected by the visual indication of spectra variations (Figure 12).

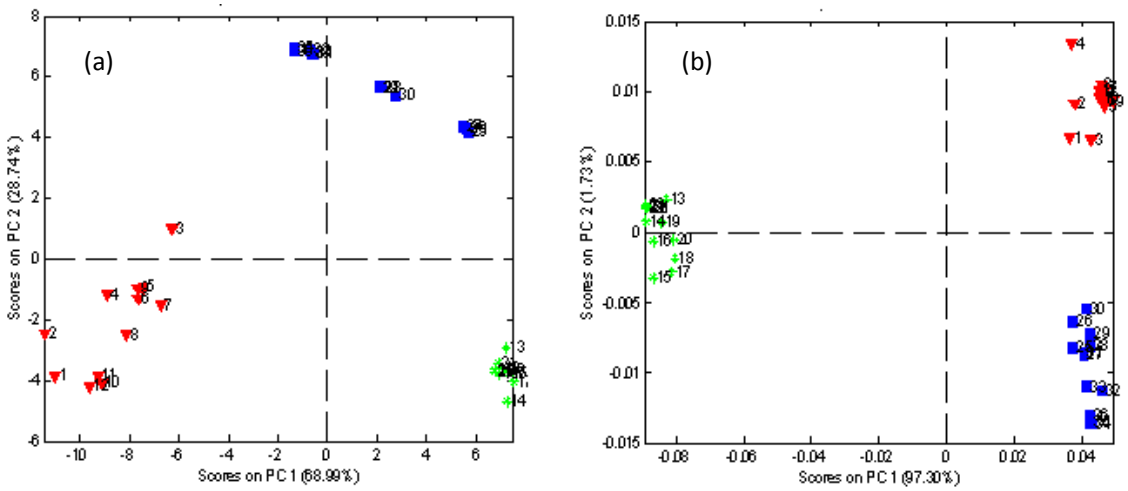


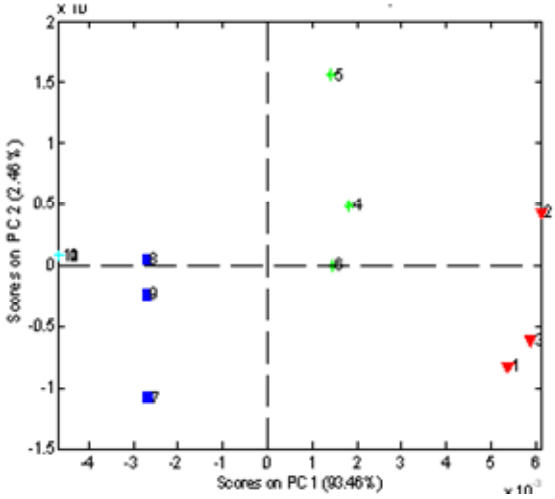
Figure 13. Score plots representing the two principal components used to differentiate among the different feed solutions. Results obtained with the UV-Visible (a) and NIR (b) probes for solutions #1 (green ♦), #2 (red ▼) and #3 (blue ■).

For this first study it was not essential that all stock solutions had exactly the same concentration in terms of COD in order to be compared, since a PCA using two principal components can only describe the highest variance contained in the data. In fact, with both probes, what is mainly revealed by the score-plots is the expected composition difference that exists between the different solutions. Additional principal components would be necessary to achieve a more complete analysis of the existing differences between the three solutions.

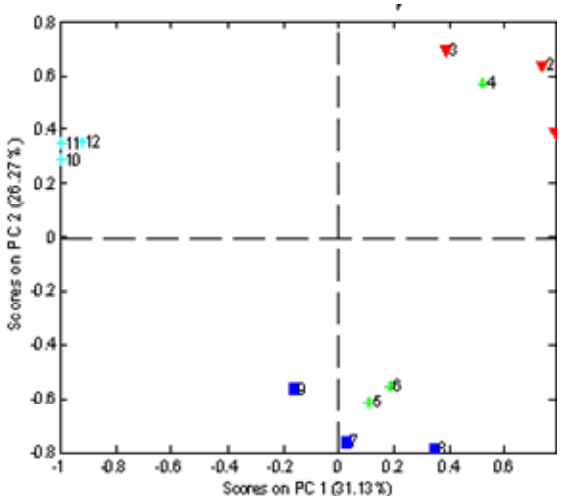
For solution #1 analysis (Figure 14, a, b), it is possible to notice that UV-Visible probe can detect more efficiently the different concentrations, when compared to the NIR probe. While PC1 accounts for the variation of concentration, PC2 accounts for the variation between replicates.

For solution #2, the UV-Visible range is better than the NIR range detecting the different concentrations. The NIR results suggest that this probe doesn't detect so well the lower concentrations. Regarding solution 2B, there are some reproducibility problems (Figure 14, c, d).

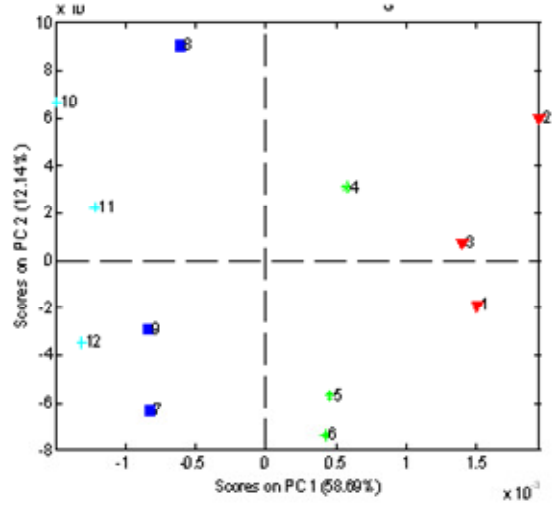
For solution #3 also better results can be achieved by using the UV-Visible probe. By analyzing the score plot from the NIR range it is possible to suggest that this probe has more difficulties in differentiating solutions with lower concentration (Figure 14, e, f) as in the previous case.



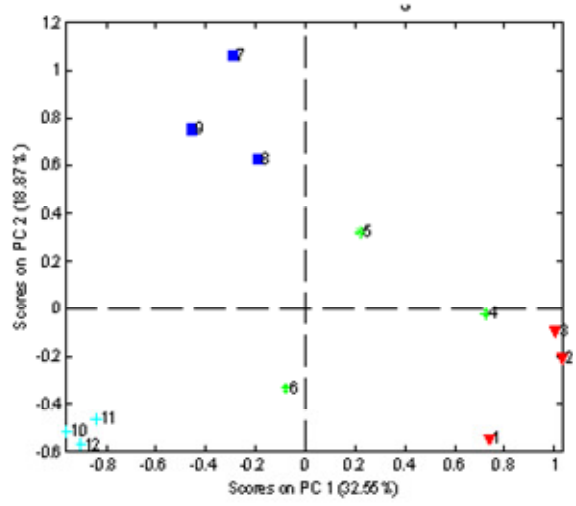
(a)



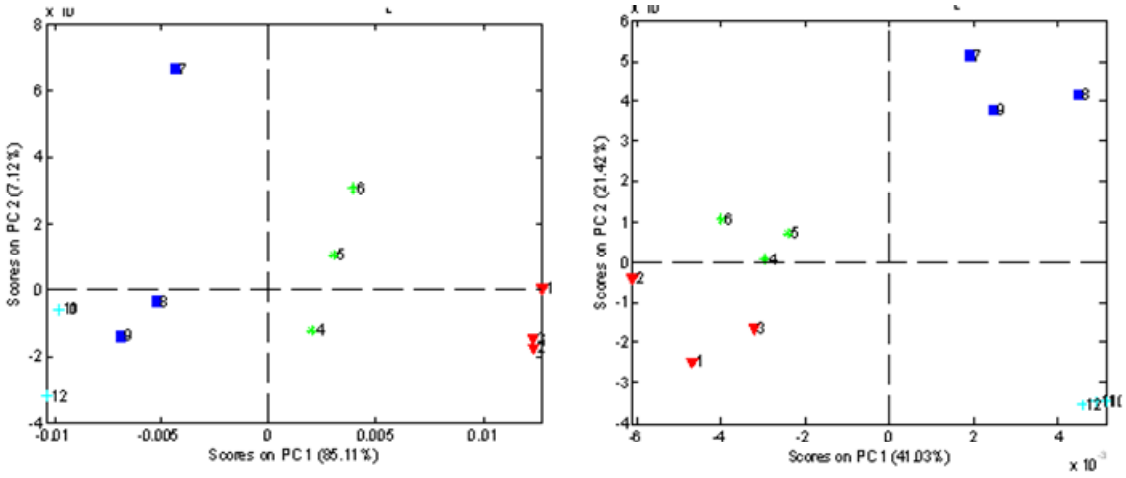
(b)



(c)



(d)



(e) (f)

Figure 14. Score plots representing the two principal components used to discriminate among the different concentrations in the solutions with peptone (a,b), glucose (c,d) and skim milk (e,f). Results obtained with the UV-Visible (a, c, e) and with the NIR (b, d, f) probe. Stock solution samples (light blue +); samples A (blue ■); samples B (green ♦) and samples C (red ▼).

Since it was possible to use the entire UV-Visible range and also the region between 270 and 310 nm to separate the different concentrations of solution #3, this suggests that turbidity was not the most important characteristic of these solutions, being the most determinant amount of information detected by the UV range. Some conclusions can be drawn from this study, namely:

- Both probes can distinguish between different solution's composition;
- UV-visible probe demonstrated to detect more efficiently different concentrations within the studied synthetic solutions, being suggested a difficulty of NIR probe in detecting lower concentrations;
- Even though solution with glucose (solution #2) was not supposed to be better detected using the UV-Visible probe, this was possible due to the fact that the solution had small concentrations of meat extract and urea, which

could be detected by the UV region of the spectra. This means that UV-Visible probe can detect very small concentrations of organic compounds in solution, what is an advantage;

- Skim milk dilutions (solution #3) differentiation was better detected by the UV-Visible probe;

- UV-Visible probe is probably more sensitive to the absence or presence of compounds, what can explain its ability to distinguish between different concentrations of the same solution composition. It is suggested that in this spectral range "concentration" effect can be easier detected;

- In NIR spectroscopy, the different properties of the synthetic solutions, physical and chemical, can cause more variability and difficult, somehow, the differentiation between them, making this spectroscopic technique more sensitive to possible interferences;

- The combination between the selected spectral ranges and the pre-treatment methods can be determinant in terms of final results.

Solution #1 presents the necessary characteristics to be used for monitoring the activated sludge process, since it has the needed composition to be used as a synthetic effluent and can be detected by the probes. The possibility of following up the biological process more easily using UV-Visible direct spectra observation (by using a UV-visible detectable composition) could also be interesting for direct comparison with the NIR probe. These are the main reasons for the choice of solution #1 composition as feed solution for the system studied in this work.

3.2 Location of the *in-situ* monitoring probes

During the first monitoring period several studies were performed in order to better understand were the probes should be placed for accurate monitoring. The acquisition of spectra inside the reactor was performed as it could be interesting to monitor the process in real-time, avoiding the need to wait enough time (at least the residence time) for the effects of the biological reaction to be noticed in the outlet, every time a disturbance was applied to the process. The high concentration of biomass (> 1.5 g MLVSS/L) and the continuous bubbling inside the reactor limited strongly the information that could be acquired with the probes when immersed in the reactor. Hence, this possibility was discarded. Since the settler offered optimal conditions for *in-situ* monitoring with no bubbles formation and no high suspended solids in solution (< 100 mg TSS/L) this location was selected for the monitoring.

Spectra from the influent were acquired every time the feed solution was changed, in order to have a better insight of the spectra of the feed, mainly by UV-Visible spectra direct observation. Since the feed composition and concentration were not often modified along this work, its continuous monitoring was not essential for the study. By comparison with the UV-Visible effluent spectra, it was possible to notice that the compounds leaving the system were different from the initial ones (Figure 15), showing that degradation was taking place, as expected. Langergraber *et al.* (2004a) also compared the spectra from the influent and the effluent of an activated sludge treatment of a paper mill wastewater treatment plant, to search for an indication of biological degradation.

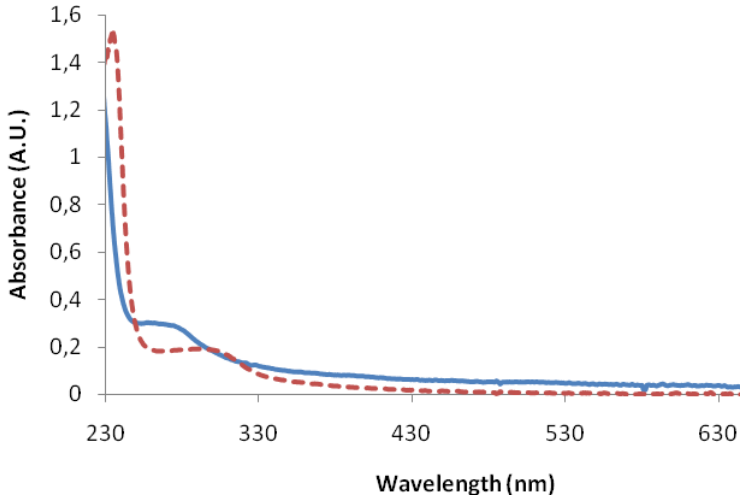


Figure 15. Comparison between influent and effluent spectra from monitoring period I. Continuous line - feed; dashed line – effluent.

3.3 Influence of fouling

Taking into account that one of the main problems associated with *in-situ* measurements is related to the accumulation of solids in the probe's sample window, being responsible for interferences, probes fouling was taken into consideration. Some observations made along this work, related to the detection of changes in the spectra after monitoring during long periods, made it clear the necessity of optimizing the spectra acquisition process.

A test was conducted focusing on exposure time of the probes in the settler, during a stationary period of the activated sludge process, to guarantee that no changes in the process could be the source of the obtained results. For both probes the same procedure was performed, which is explained as follows:

- The probes were immersed in the settler during an entire night and in the morning spectra were acquired;
- The probes were cleaned and immersed again in the settler. New spectra were acquired;

- During a space time of three hours spectra were acquired after periods of one hour, without any cleaning procedure between them;
- The probes were cleaned;
- In the evening spectra were acquired during one hour.

Table 10 shows the tasks performed after overnight probe's immersion till the evening of the same day, showing the moment when the probes were manually cleaned and spectra were acquired.

Table 10. Description of spectra acquisition and cleaning procedure moments

Time	Procedure	Monitoring Period
Probes immersed during night		-
9:30	Spectra acquisition	I
10:00	Probe cleaning and spectra acquisition	II
10:00-11:00	Spectra acquisition	III
11:00-12:00	Spectra acquisition	IV
12:00-13:00	Spectra acquisition	V
18:30	Probe cleaning and spectra acquisition	VI
18:30-19:30	Spectra acquisition	VII

Different results were obtained for each probe (Figure 16 and Figure 17) by performing PCA of the acquired spectra.

Different pre-treatments methods were applied to the UV-Visible and NIR spectra: Savitzky-Golay (15, 2, 2) method, standard normal variate and mean-centering for UV-Visible spectra and Savitzky-Golay (15, 2, 2) method and mean-centering for NIR spectra. For both probes the entire spectra was analyzed. In this case the selection and optimization of smaller spectral ranges for the analysis was not desired since the main objective was to study the influence of fouling taking into account all the acquired information and not only a part, since the changes in the spectra due to deposition of solids would probably affect the entire spectral range.

It is possible to identify clusters of NIR probe spectra throughout the several moments of analysis (Figure 16). The fact that spectra from monitoring period I are apart from the rest indicate that the night period was determinant to influence the spectra shape.

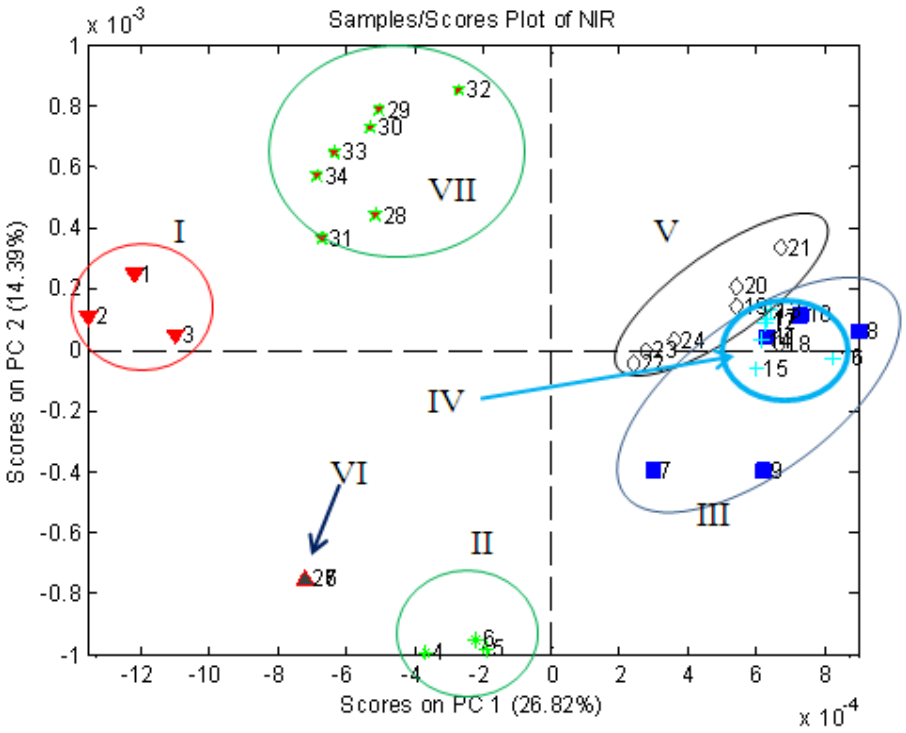


Figure 16. PCA scores plot for NIR spectra throughout the test. Roman numbers identify the different acquisition moments.

The similarity between the spectra acquired after cleaning in the morning and the spectra acquired after cleaning in the afternoon (II and VI, respectively) indicates that after the cleaning procedure the probe has the same conditions, however along with the time immersed in the settler these characteristics seem to change rapidly. After one hour inside the settler the probe is not cleaned and spectra are acquired in intervals of one hour. In this period spectra are alike (III, IV and V) but different from the spectra acquired at 10 h after cleaning (II). Since after 1 h it is possible to detect different spectra, it seems that one hour may be sufficient to promote changes in the spectra due to probe's fouling.

Regarding the UV-Visible probe, it is obvious that an entire night of immersion in the settler affected the conditions in the sample window, since the first spectra are different from the rest (taken after cleaning). However, these changes appear not to be so clear throughout the different monitoring moments of this study, since it is not very obvious the identification of distinct clusters (Figure 17). This may suggest that the night period, being much longer than the day monitoring period, was the main factor affecting the spectra acquisition.

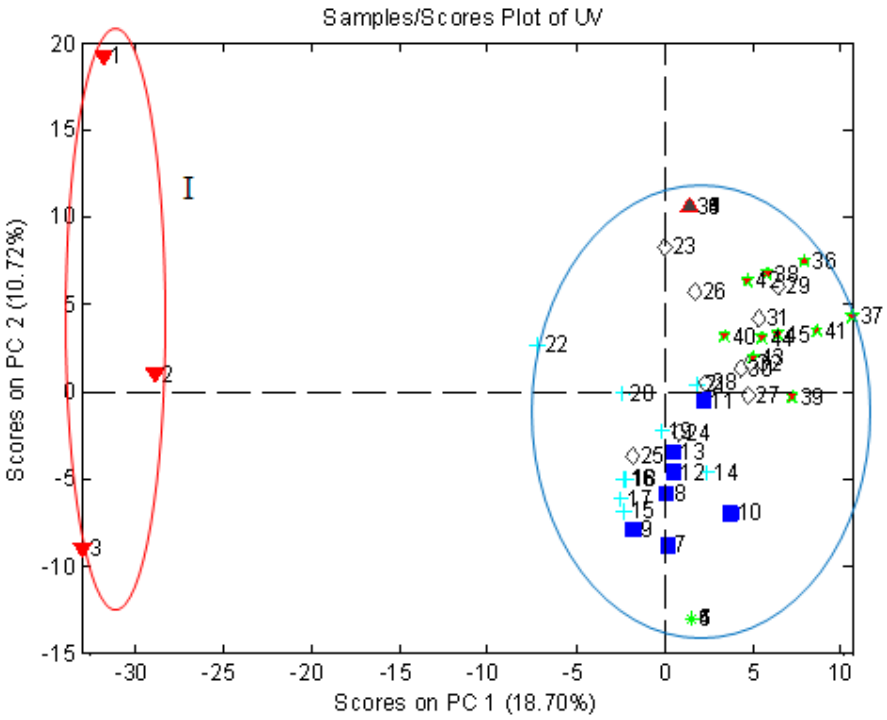


Figure 17. PCA scores plot for UV-Visible spectra throughout the test. Monitoring period I is contained in the left ellipse and the remaining periods are contained in the right ellipse.

This study allowed achieving some conclusions regarding the different behavior of the probes when submitted to the described conditions. It indicates that NIR probe, due to its ability in detecting physical properties in solution, is much more sensitive to small particles accumulation inside the sample window. UV-Visible probe may suffer less influence when immersed for not very long periods of monitoring.

This study was important to understand how the spectra acquisition should be performed during monitoring moments.

In regular monitoring moments the probes were immersed in the settler after being carefully cleaned and the acquisition was performed in the settler during a period of not longer than 45 minutes, so the conditions could be as much identical as possible for comparison between monitoring days. Hence, fouling is not expected to influence that much the monitoring of short periods. The same could not be assured every time a night or a long monitoring period was performed.

Although chemometric tools are essential to eliminate interferences related to scattering due to the presence of solids, they may not be sufficient to remove the influence of solids accumulation characterizing a washout period.

The best option is to clean the immersible probes in a regular basis and preferably with an automatic mechanism, being this solution already applied in studies where the submersible probe is equipped with an auto-cleaning pressurized air system (Langergraber *et al.,* 2004a; Rieger *et al.*, 2004).

3.4 Activated Sludge Process Monitoring

With the main objective of collecting as much data as possible regarding the activated sludge process variations, two different monitoring periods were performed: monitoring period I and monitoring period II.

3.4.1 Off-line monitoring

3.4.1.1 Monitoring Period I

Inflow variations

The first monitoring period was performed during 70 days with a 16 L sludge bulk in the reactor (sludge collected at Frossos WWTP, Braga). A first variation induced to the process

after 21 days of operation was due to a decrease in COD in the feed (COD_{in}) and subsequently in organic loading rate (OLR), as a necessary adjustment in terms of influent flow (by correcting the inflow from the concentrated synthetic wastewater). After this moment, operational conditions were not modified (Table 11). Even though this was not considered a very significant disturbance to the biological process, this variation was monitored *in-situ* with the immersible probes to check if any deviation from the previous days could be detected. This disturbance is designated as disturbance I.

Table 11. Inflow (Q_{in}), hydraulic retention time (HRT), COD_{in}, OLR, COD removal and Kjeldahl nitrogen values obtained during monitoring period I, before and after inflow adjustment

Time period (days)	Q_{in} (L/d)	HRT (h)	COD_{in} (g O_2/L)	OLR (g COD/L.d)	COD removal (%)	$[N–Kj]_{in}$ mg/L
0-21	11.7±0.4	33.6±0.9	0.78±0.12	0.56±0.08	94.3±2.5	125.9±2.6
21-70	13.2±0.6	29.1±1.5	0.95±0.05	0.79±0.06	94.2±3.1	158.1±4.8

Biomass concentration variations

Biomass concentration in the reactor (MLVSS) was subject of several changes throughout this monitoring period. Concentrations between 1.9 and 3.1 g MLVSS/L were present in the reactor during operation, being more stable around 2.8 g MLVSS/L after day 15. Initially the biomass was not purged from the reactor. After day 15, the removal of biomass was carried out with the purpose of improving settleability and control biomass population fluctuations.

To effectively decrease MLVSS concentration and detect all possible variations in the system that could derivate from a disorder in nitrification process and COD degradation, an intensive removal of biomass from the reactor was performed from day 55. An average value of 1.6 g MLVSS/L was achieved. This severe removal of biomass from the system was significant in terms of disturbance to the process and was designated as disturbance II.

COD variations

At the beginning of the experiment some variations were observed in COD_{in} and also in the effluent (COD_{out}) (Figure 18). Till day 21 COD_{in} was decreasing as previously referred, being more stable after this day. COD_{out} was always below 150 mg O_2/L and obtained values varied between 20 mg O_2/L and 120 mg O_2/L. The inflow adjustment applied after day 21 was, as expected, not considered a significant disturbance regarding heterotrophic bacteria activity and, hence, COD removal. Considering that heterotrophic bacteria can be approximately 90 to 97 percent of the bacterial population in the activated sludge process (Gerardi, 2002), the high biomass concentration (between 2-3 g/L) could be able to degrade large amounts of organic matter. This can be explained by the average values of food-to-microorganism ratio (F/M) obtained after day 21 of 0.29 ± 0.03 gCOD/gMLVSS.d or 0.23 ± 0.03 gBOD/gMLVSS.d - if we consider BOD/COD=0.8 for a substrate easily degradable by biological means (Metcalf&Eddy, 2003). This F/M ratio value is not considered high for a complete mix activated sludge process, which can go from 0.2 to 0.6 gBOD/gMLVSS.d, as typical values (Metcalf&Eddy, 2003).

After MLVSS concentration decrease to 1.6 gMLVSS/L an average F/M value of 0.47 ± 0.04 gCOD/gMLVSS.d (0.39 ± 0.01 gBOD/gMLVSS.d) was achieved, not affecting COD removal, which was of 94.4 ± 3.6 % throughout this monitoring period.

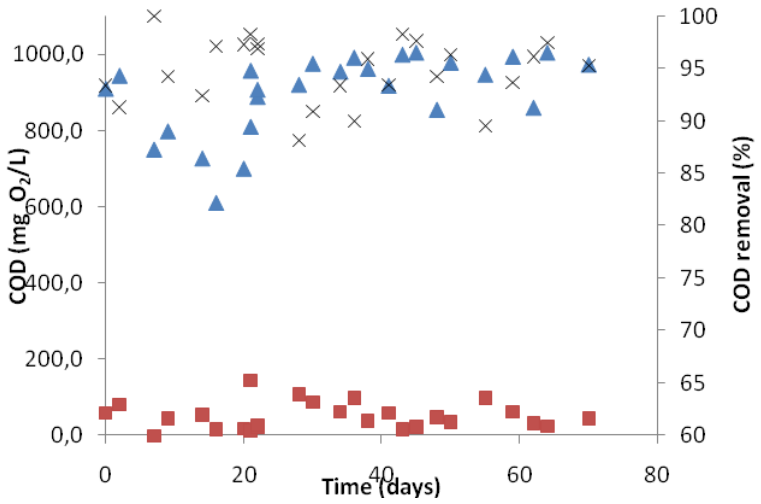


Figure 18. COD efficiency removal and influent and effluent fluctuations during monitoring period I (Legend: ▲ – COD_{in}; ■ – COD_{out}; × - COD removal efficiency).

Nitrogen variations

Kjeldahl nitrogen (organic and ammonium nitrogen) was monitored in the inlet and in the outlet. According to the disturbances applied to the system this parameter also suffered variations.

Inflow adjustment affected inlet concentration of Kjeldahl nitrogen (N-Kj) in a similar way to COD_{in}, since peptone is the main source of carbon and organic nitrogen to the system.

MLVSS decrease in the reactor after day 55 affected tremendously N-NH_4^+ oxidation in the system, what was expected, being all the organic nitrogen only hydrolyzed into ammonium ions but not oxidized to nitrite. As a result, nitrate concentration in the outlet dropped to values close to zero (Figure 19). Nitrifying bacteria population was immediately affected by the biomass purges, being removed from the system. With the lack of ammonia-oxidizing bacteria in the system, organic nitrogen was hydrolyzed to ammonium nitrogen but no nitrite was produced and, consequently, no nitrite oxidation to nitrate occurred. Nitrifying bacteria have a much lower maximum specific growth rate (μ_{nm}) when compared to heterotrophic microorganisms (Metcalf&Eddy, 2003), being this the main reason why these bacteria need high solids retention time (SRT) for good population growth and stability in an activated sludge system.

Nitrite concentration along most of the monitoring period was very close to zero (0.014 ± 0.011 mg $N\text{-}NO_2^-$/L). The second step of nitrification is usually very fast and nitrite concentration is around 0.03 mg $N\text{-}NO_2^-$/L in the outlet of a WWTP (Rieger *et al.*, 2004).

Possible denitrification due to the existence of "dead" zones in the reactor and of biomass retained bellow air diffusing system, nitrogen assimilated by heterotrophic bacteria and/or experimental errors may explain the difficulty in closing nitrogen mass balance.

Kjeldahl nitrogen was used to monitor the process only once a week, since it is a very time consuming analytical technique. To rapidly detect changes in the system, determination of $N\text{-}NH_4^+$ in the effluent was performed using the Nessler method (2-3 times a week).

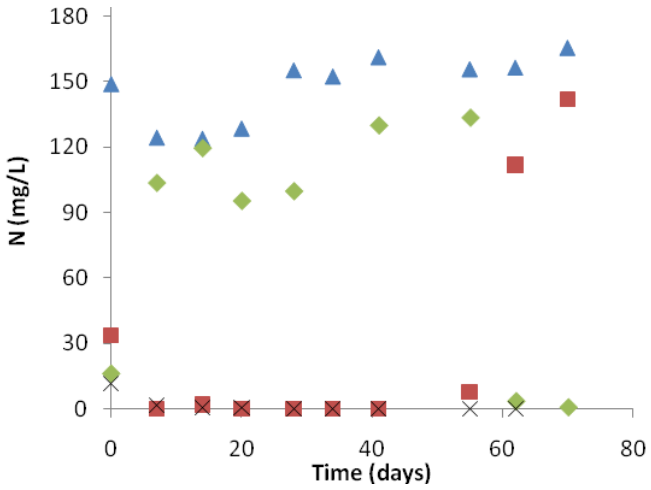


Figure 19. Nitrogen variations during monitoring period I, for the same monitoring days, and N-Kj removal efficiency (Legend: ▲ – $N\text{-}Kj_{in}$; ■ – $N\text{-}Kj_{out\ t}$; ◊ – $N\text{-}NO_3^-$; × - $N\text{-}NO_2^-$).

3.4.1.2 Monitoring Period II

HRT sudden decrease

The second monitoring period was performed during 49 days with a 17 L sludge bulk in the reactor (sludge collected at Frossos WWTP, Braga).

During this monitoring period many parameters were maintained constant throughout the study (Table 12), except for a HRT sudden decrease from 31 h to 15 h, during day 35 – disturbance I (Figure 20). This disturbance was induced to obtain variations in terms of COD concentration in the outlet.

An incident occurred at day 21, when the aeration and mixing stopped for some hours. During that day the probes were immersed in the settler in order to detect any possible variation. Besides TSS increase in the outlet, no other parameters suffered changes after this episode.

Table 12. Average values of several monitoring parameters during the monitoring period II

Time period (days)	Q_{in} (L/d)	HRT (h)	COD_{in} (g O_2/L)	OLR (g COD/L.d)	COD removal (%)	$[N–Kj]_{in}$ mg/L
0-49	13.1±0.4	31.1±1.0	0.10±0.06	0.77±0.06	91.0±3.4	143.2±12.4

Operational conditions were maintained as close as possible to the ones applied during the more stable part of the first monitoring period.

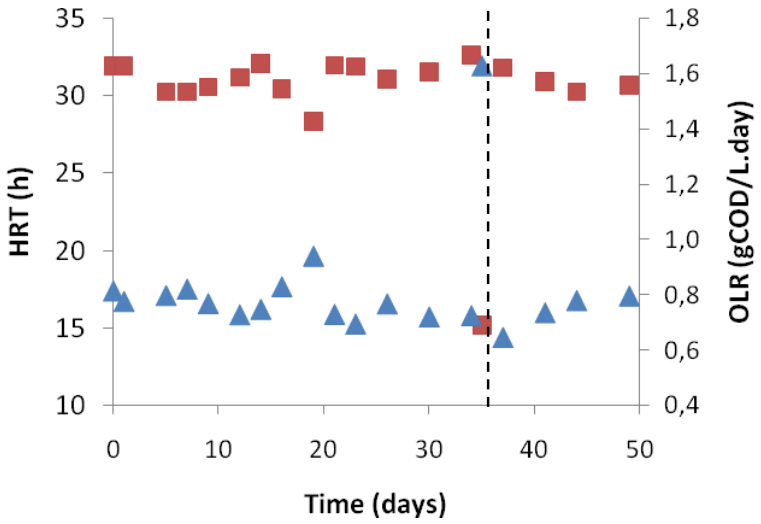


Figure 20. Variations of HRT and OLR during monitoring period II (Legend: ■ – HRT; ▲- OLR; dashed line indicates the moment when the HRT was decreased).

Biomass concentration variations

After one week of operation MLVSS concentration was kept between 1.2 and 1.8 g/L, by weekly biomass removal directly from the reactor.

The sudden HRT decrease affected MLVSS concentration, being a great part of the biomass lost together with scum resulting from peptone higher concentration in the reactor. The final concentration of MLVSS after the disturbance was around 0.5 g/L, which rapidly increased to values above 1.2 g/L after one week of operation.

COD variations

COD degradation was high (91.0 ± 3.4 %), in a similar way to what was found in the monitoring period I (Figure 21). The only moment when COD efficiency was lower (day 35) was due to the MLVSS decrease in the reactor, immediately after the disturbance. COD_{out} was very variable throughout all the monitoring period (between 30 and 160 mg O_2/L), except for day 35 when a value of 620 mg O_2/L was achieved.

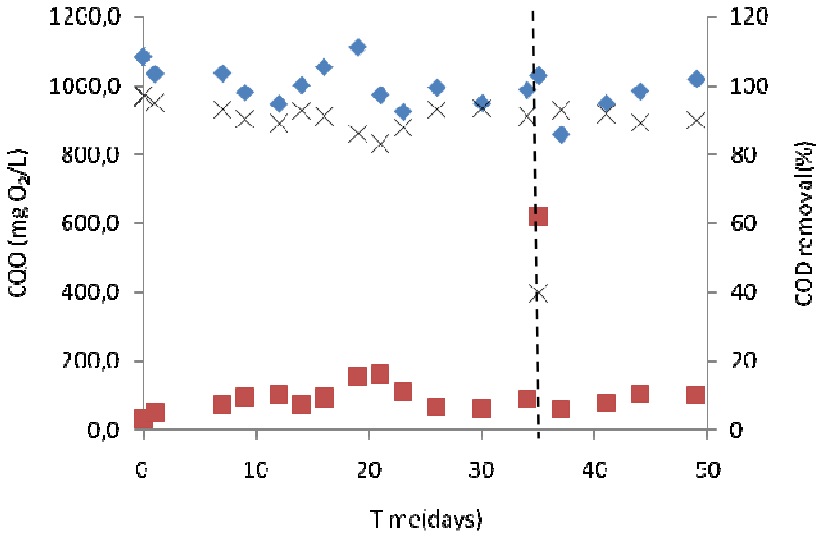


Figure 21. COD efficiency removal and influent and effluent fluctuations during monitoring period II (Legend: ▲ – COD_{in}; ■ – COD_{out}; × - COD removal efficiency. Dashed line indicates the moment when the HRT was decreased).

During days 35 and 36 COD_{out} was monitored in a more frequent basis. Results obtained after the disturbance are presented in Figure 22. It was noticed that after not more than one day the system could rapidly recover from the disturbance in terms of COD degradation.

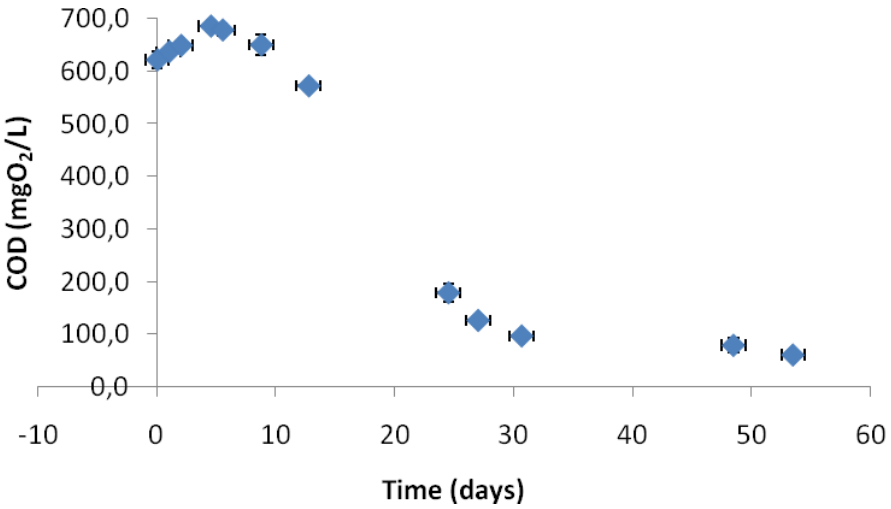


Figure 22. COD concentration variation in the effluent after the disturbance induced to the system.

Nitrogen variations

Kjeldahl nitrogen concentration in the feed was maintained more or less constant throughout this monitoring period (143.2 ± 12.4 mg/L). Regarding the weekly purges from the reactor during this study, nitrification process was not feasible. Nitrate concentrations ranged from 20 mg $N\text{-}NO_3^-$/L to zero concentration, after two weeks of operation, thus, NH_4^+ was not oxidized after this moment, being accumulated in the system. No nitrite was accumulated (Figure 23). Since nitrification process was already inexistent at the moment of the disturbance, no variations were observed in terms of nitrogen forms present in the effluent after the decrease of HRT.

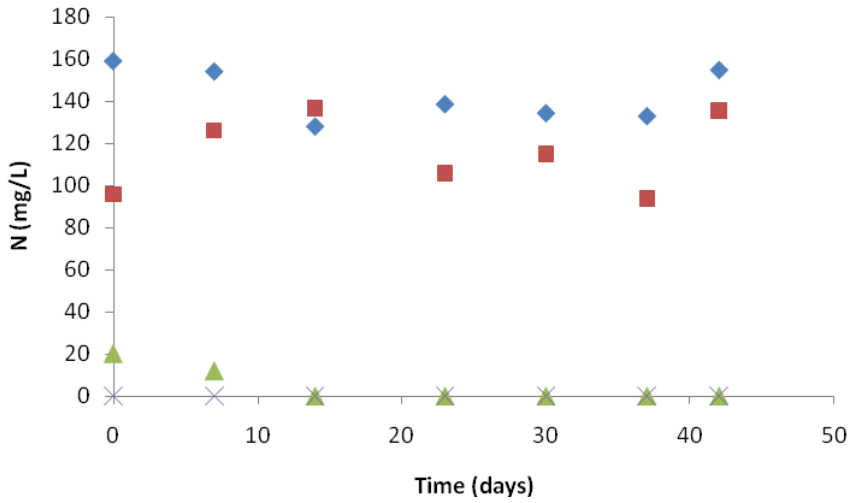


Figure 23. Nitrogen forms variations during monitoring period II, for the same monitoring days, and N-Kj removal efficiency (Legend: ▲ – N-Kj_{in}; ■ – N-$Kj_{out\ t}$; ● – N-NO_3^-; x - N-NO_2^-).

3.4.2 *In-situ* monitoring

During monitoring period I and II the process was monitored in the settler by using UV-Visible and NIR immersible probes, focusing essentially in the detection of the disturbances applied to the process by comparison with the more stable periods. All spectra from both probes were pre-treated and PCA was carried out.

3.4.2.1 UV-Visible *in-situ* monitoring

3.4.2.1.1 Monitoring period I

A PCA was performed to spectra from monitoring period I. A pre-treatment was applied by using a first derivative according to the method of Savitzky-Golay (1984) and mean centering the data, after the filter adjustment. Almost the entire UV-Visible spectra range was used to perform the analysis (230-700 nm). The two PCs describe 88.4 % of the total variance in the spectra (Figure 24).

It is possible to notice that day *0* is apart from the rest of the spectra, being this possibly related to the start-up of the system and, thus, to some instability. Some time is required for the system to acclimate to the new conditions.

Considering the disturbances applied to the process, UV-Visible probe was able to detect both variations: disturbance I – after day 21; disturbance II – during day 55. The PCA shows a cluster formed for days 28 and 30. Since after day 22 no variations in the concentration of COD and/or nitrogen forms in the outlet were noticed after the disturbance, the visible change in spectra form during these days (28 and 30), presenting an extra peak around 350 nm when compared to the previous UV-Visible spectra (Figure 25), was not expected and could not be explained. However, since the variation occurred after the disturbance it is suggested that it could only be due to it. Spectra from days 34 and 36 still have a small peak at 350 nm, meaning that the compound or compounds that appeared in the effluent after the disturbance started to disappear after day 30. Spectra returned to the initial cluster, which corresponds to the stationary stage of the system.

Regarding disturbance II, this was more effective in disturbing the system by disabling the nitrification process. Spectra changed due to the absence of nitrate in the effluent and a different cluster was formed (days 62, 64 and 70). If nitrifiers had enough time and conditions to increase its population, probably the new spectra would form new clusters with time till returning to the initial cluster (stationary system).

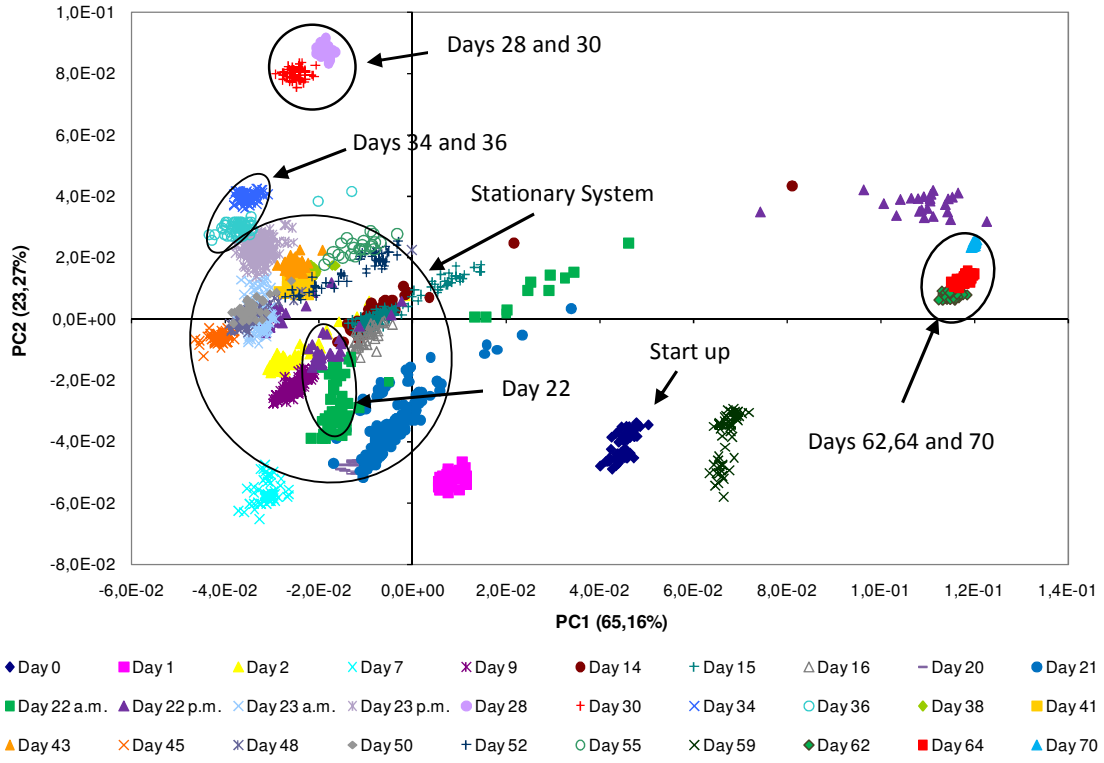


Figure 24. Score-plot representing UV-Visible spectra variations during monitoring period I.

Fouling of the probe's sample window occurred during some monitoring days, due to biomass washout related to bulking problems. Bulking phenomena was related to the presence of filamentous bacteria in the reactor. Its occurrence was noted by the occasional release of concentrated portions of solids from the settled biomass. Thus, some solids were deposited in the probe's sample window and changed the spectra form, through a baseline shift. This occurrence was easily detected by PCA, since spectra tended to describe a line in the score plot. Some examples of this situation are also present in Figure 24. On days 21 and 22 (morning and afternoon) bulking problems were observed and spectra variations allowed to detect it perfectly. Fouling can be a disadvantage if it induces changes in spectra which are not related to biochemical reactions in the process. But when considering short monitoring periods the establishment of a continuous line may indicate solids washout derived from settleability problems.

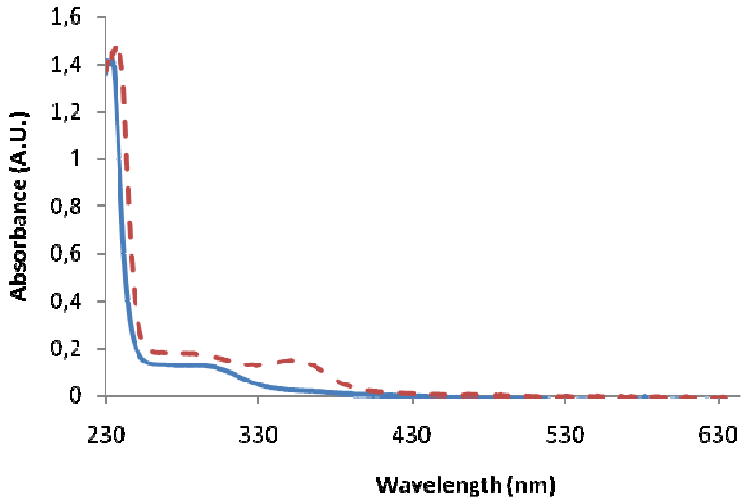


Figure 25. Spectra variation after disturbance I. Continuous line – day 20; dashed line – day 28.

3.4.2.1.2 Monitoring period II

During monitoring period II, a disturbance was induced to the system during day 35, by performing a sudden decrease of the HRT. This perturbation led to an increase of COD in the effluent, which was monitored along day 35 and 36.

A first derivative (Savitzky-Golay method) and mean-centering were performed to the acquired spectra before PC analysis.

In the score plot (Figure 26) it is possible to notice the existence of three clusters. The start-up of the system is clearly identified. After this moment, spectra have similar characteristics till the disturbance. During day 35, spectra have some variations, being apart from the previous.

On day 36, COD values are already close to the usual values found before the disturbance. After day 35 spectra seem to return to their initial characteristics, what was expected since besides COD no other significant changes were detected.

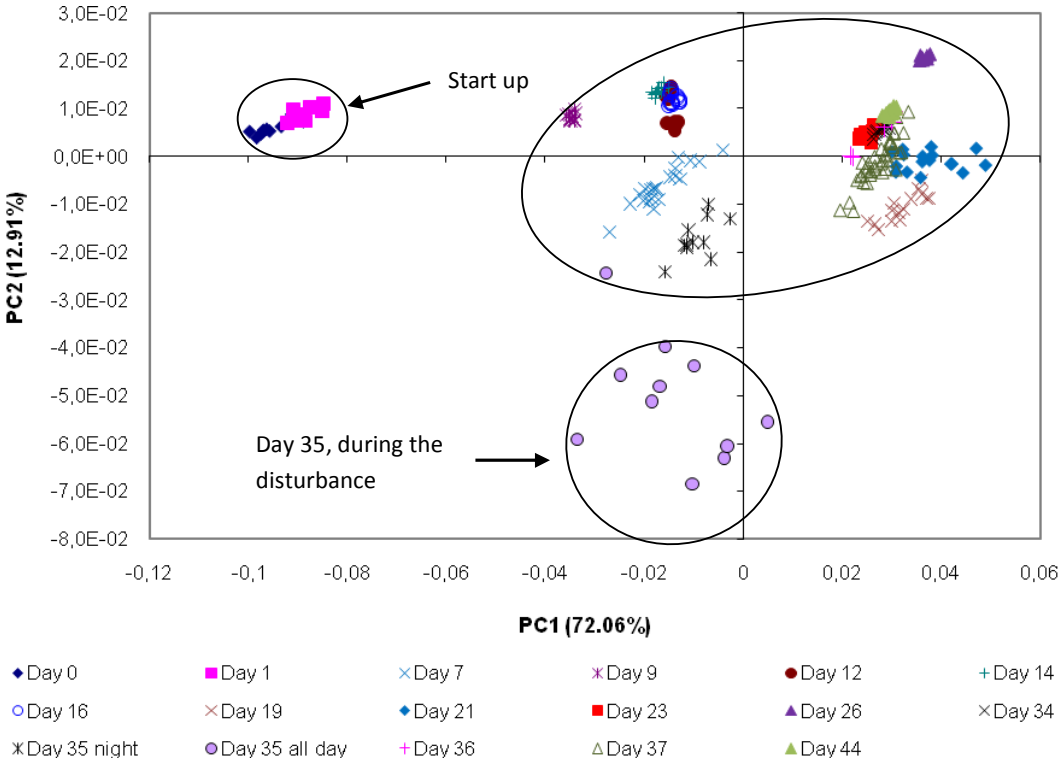


Figure 26. Score-plot representing UV-Visible spectra variations during monitoring period II.

As main conclusions, UV-Visible probe was able to detect a COD increase in the outlet, during the disturbance. These results are in agreement with the ones obtained for the complementary test of the synthetic wastewaters study, showing that the "concentration effect" is efficiently detected by the UV-Visible probe during the monitoring of an activated sludge system.

3.4.2.1.3 Global analysis

Spectra from monitoring period I and monitoring period II were analyzed together with PCA. Spectra were pre-treated by applying a first derivative and mean centering the raw data. With two PCs it was possible to explain 92.5 % of the data (Figure 27).

Two groups were clearly distinguished - Group I and Group II. Group I is characterized by a higher nitrification rate, before disturbance II in the first monitoring period, while Group II is associated to a low or even inexistent nitrification process, after disturbance II of monitoring period I and most part of the second monitoring period.

It was also possible to distinguish perfectly the spectra obtained due to disturbances applied in the different periods - disturbance I of monitoring period I and disturbance I of monitoring period II - which are indicated in the score plot as A and B, respectively.

Some samples corresponding to the monitoring period I are present in group II. This is due to the low nitrification rate promoted after the induction of disturbance II what decreased the levels of nitrate to the similar ranges maintained during the second monitoring period.

By gathering the information related to both monitoring periods it was possible to collect enough information regarding periods with high nitrification and with low or inexistent nitrification process. This allows the possibility of having information about nitrification performance by analyzing some spectra in the settler before checking nitrate concentration in the outlet through an off-line analysis.

3.4.1.1.4 Study of variables relations

Since in monitoring periods I and II it was possible to detect several variations in the measured parameters, it was found interesting to realize how this parameters were related among them, considering both monitoring periods. With that purpose, PCA was applied to the auto-scaled set of nine variables measured in this study (Q_{in}, COD_{in}, COD_{out}, OLR, MLVSS, TSS_{out}, F/M, $pH_{reactor}$ and $N\text{-}NO_3^-{}_{out}$). With two PCs it was possible to explain 71.70 % of the variation in the data.

A biplot was obtained presenting in the same graph the samples and the measured variables (Figure 28).

The major advantage of the biplot representation is the possibility of establishing relations between samples and variables. Moreover it is possible to identify how the different variables are related with each other.

After PCA samples were divided in two main groups – I and II. As considered previously, two groups were distinguished, being the first group composed of spectra related to high nitrate concentration in the effluent and Group II composed of spectra acquired when low nitrate concentration was present in the system outlet.

According to Figure 28 it is also possible to identify how nitrate concentration in the outlet is closely related to the MLVSS concentration and inversely related to pH in the reactor. These results confirm also expected relations, since nitrification is more feasible when MLVSS concentration is high and pH values in the reactor decrease with high nitrification rate, rising when this process is disturbed. Monitored samples were placed along the arrows #2 or #3, depending if they correspond to a high or low nitrification period, respectively.

This study pointed out that it is possible to follow-up variations in the system in an equivalent way to the use of different monitoring parameters.

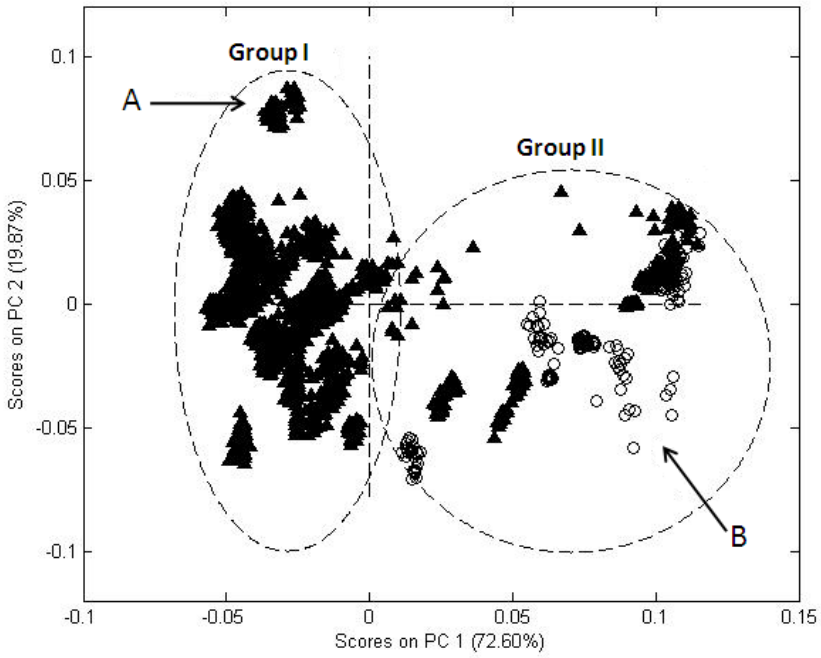


Figure 27. Score plot representing the UV-Visible spectra variation during the monitoring period. Legend: ▲– samples from monitoring period I; o – samples from monitoring period II.

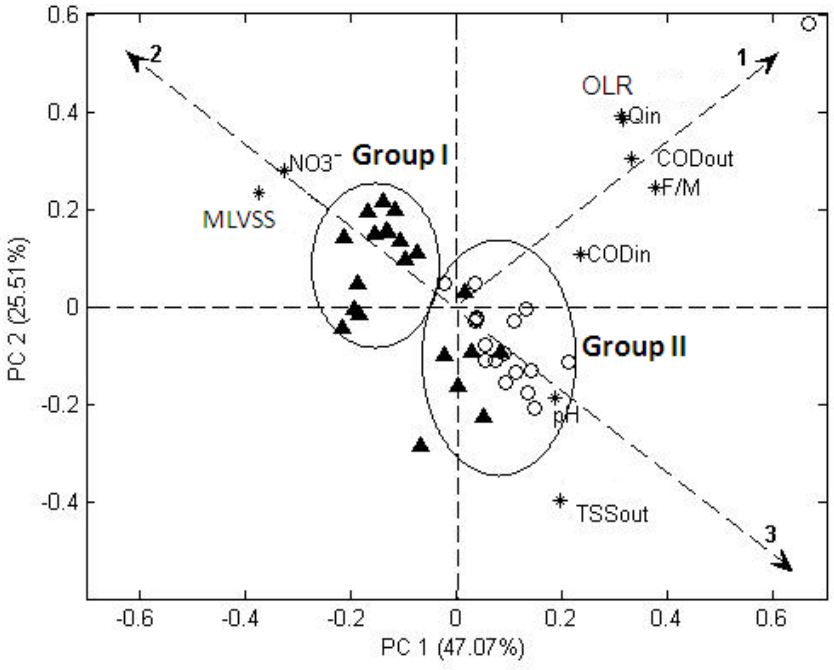


Figure 28. Biplot representing simultaneously the samples and the variables measured during both monitoring periods. Legend: ⁎ - parameters; ▲– samples from monitoring period I; o – samples from monitoring period II.

3.4.1.1.5 Disturbances detection

In order to have a better insight regarding how well the UV-Visible probe can detect a disturbance applied to the system, residuals statistics (Q) was used to analyze the results obtained from PCA. This is a statistical parameter that can easily inform, even a non-expert person, when some deviation is being observed in the monitored system. Monitoring period II was selected for this study.

For residuals statistics (Q) analysis, a PCA was performed using as pre-treatment methods: standard normal variate, first derivative (Savitzky-Golay method) and mean-centering.

From Figure 29, it is possible to notice that the UV-Visible probe can detect four different moments: at the start-up, on day 21 and during day 35.

A line is represented in the plot corresponding to the 95 % confidence range for the measured data, when compared to the model. In this way, all samples bellow this line are considered regular with 95 % of confidence and those located in the upper part of the line are considered to represent a variation to the normal conditions.

This statistical parameter allowed the detection of the system's start up as a different period.

During day 21, the aeration stopped for some hours and no mixing inside the reactor was performed. In order to detect some modification in the process due to this incident, the probes were immersed in the settler for monitoring. The only monitored parameter that changed after the incident was TSS concentration, which increased. Regarding that solids in the settler can modify the spectra shape by rapidly fouling the tip's probes, it is suggested that TSS increase and the exposure time explain the variation detected on day 21.

During day 35 the sudden decrease of HRT induced a rapid increase of COD in the outlet. This was detected by the UV-Visible probe, what could be expected since this probe is able to detect variations in COD concentration. PCA score-plot for UV-Visible monitoring period II (Figure 26), already detected a different cluster for spectra from day 35. As indicated in Figure 29, initially the spectra are different because of the COD change, but when COD starts to decrease spectra start to change its characteristics again, not being detected as

"irregular" by residuals statistics, what is in agreement with PCA for this monitoring period. Fouling could not explain this spectra change, since the change is detected immediately after the probe is immersed in the settler for monitoring.

However, the monitoring proceeded during most of the day, and the probe was kept in the settler till the evening. Again, something that should be expected happened: fouling of the probe's sample window. Once more, and after analyzing all the samples for COD and other parameters determination, this could be the only reason for spectra variation after the COD decrease.

As main conclusion, it was demonstrated that it is possible to detect real-time variations of COD concentration using *in-situ* UV-Visible spectroscopy. This can be considered an important achievement, since a rapid detection of changes in the effluent's quality should be detected as soon as possible.

However, the data must be carefully analyzed now that fouling is clearly identified as an obstacle for monitoring. The cleaning of the tips probe must be performed in a regular basis, although it is not very feasible to do it manually.

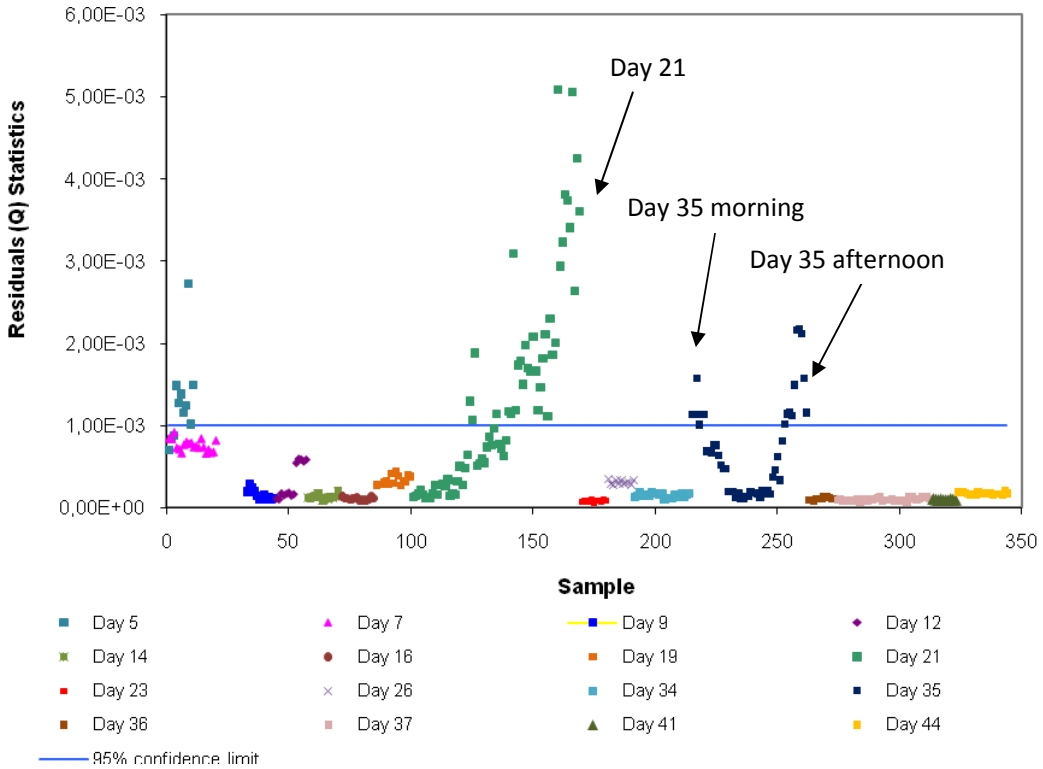


Figure 29. Residuals statistics obtained when PCA is applied to spectra acquired with UV-Visible probe immersed in the settler. The blue line represents the 95 % confidence limit.

3.4.2.2 NIR *in-situ* monitoring

3.4.2.2.1 Monitoring period I

During monitoring period I the NIR probe was damaged. Since the damage only affected the mirror which reflected the radiation back, decreasing only a part of the radiation that arrived to the spectrophotometer, it was still possible to get satisfactory spectra with the probe. Considering that this incident happened after day 28, some days after disturbance I was applied, NIR monitoring was maintained till the end of this monitoring period, since some interesting results were obtained related to disturbance II. The periods before and after this occurrence are referred as period A and B, respectively.

The spectra obtained during monitoring period I were pre-processed using Savitzky-Golay filter (15,2,2) and mean centering and analyzed with PCA, using the entire spectral range (900-1700 nm).

Figure 30 shows how spectra acquired in both periods are contained in separate clusters, confirming that the incident changed the shape of the measured spectra. In this way, data analysis was also divided in two different periods, for monitoring period I. Unless something happens to the system performance, external factors such as those inherent to the spectroscopic equipment must be taken into consideration when different clusters are formed as results show, indicating an obvious change in spectra shape. After the damage, the direct observation of the NIR spectra was not sufficient to detect the incident without the results obtained from chemometric analysis.

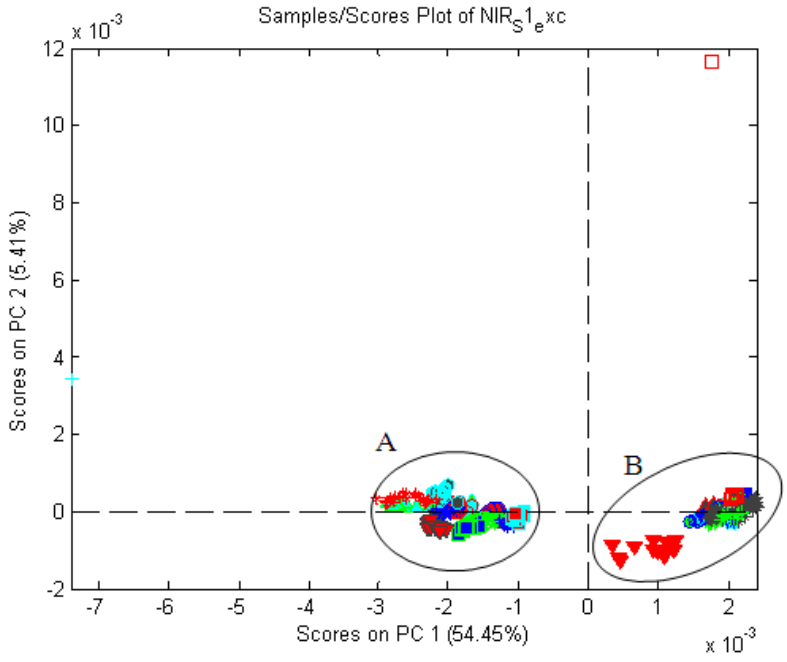


Figure 30. PCA scores plot regarding monitoring period I. A represents the period before NIR probe damage and B represents the period after the incident.

Concerning disturbance I which was applied after day 21 during period A, regarding an inflow adjustment, NIR probe was able to identify the changes in the system (Figure 31). During day 22, spectra were continuously dislocated to form a separate cluster, as indicated by the arrow. At day 23 spectra variation was less and still close to the cluster formed at day

22. At day 28 spectra were already close to the initial location, indicating that the system was returning to its initial equilibrium state.

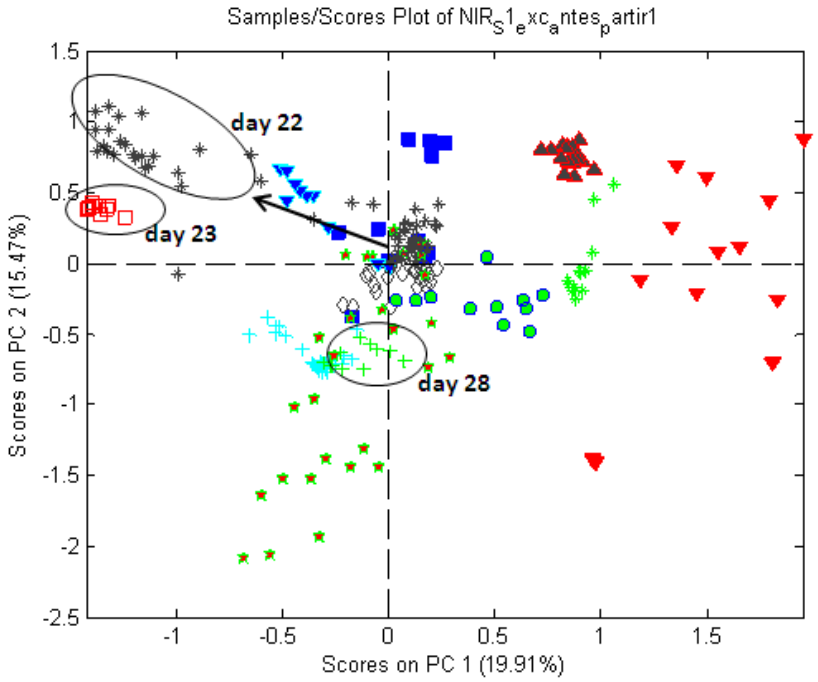


Figure 31. PCA scores plot representing data regarding disturbance I, during monitoring period A, for NIR probe.

During period B, the sudden decrease of MLVSS applied on day 55 (disturbance II) was detected by NIR probe at day 56 (Figure 32). Spectra from days 62 and 64 indicate the return to the initial characteristics. On day 70 some disturbance was detected by the NIR probe, which was not identified. Even though the principal variation after this disturbance was the decrease of nitrate and the increase of ammonium in the outlet, the analysis of all parameters for this particular day suggest that what is detected by NIR probe may not be related to the absence or presence, respectively, of these compounds.

Comparing the results between both probes, it is possible to notice that NIR monitoring suggests that the system returns to its initial conditions, after the disturbance was applied. The same did not happen with UV-Visible probe (Figure 24). This can easily be explained by

the fact that nitrate presence or absence is naturally detected by UV spectral region. Since inorganic compounds are not detected by NIR radiation, nitrate should not be directly detected.

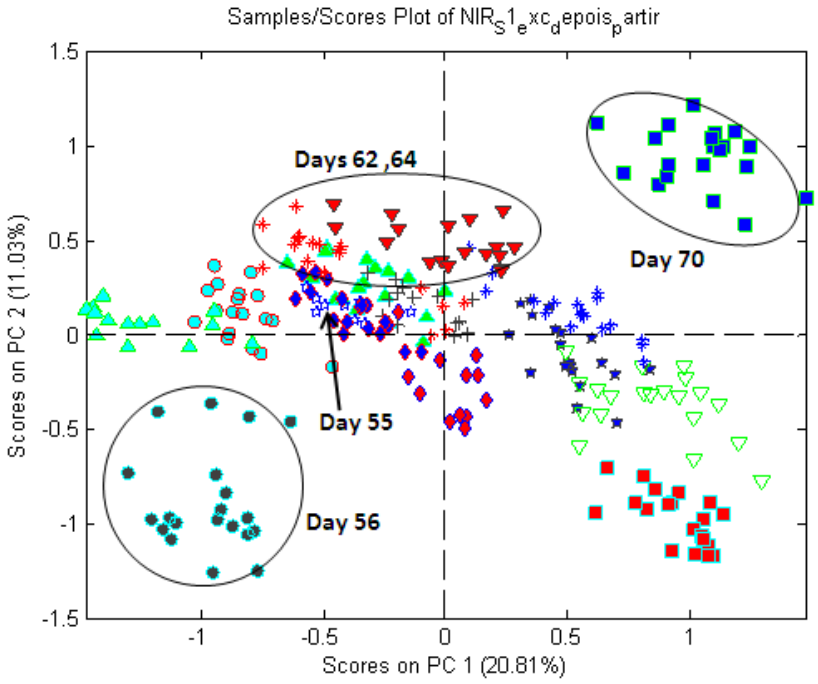


Figure 32. PCA scores plot representing data regarding disturbance II, during monitoring period B, for NIR probe.

3.4.2.2.2 Monitoring II

The damaged tip of the NIR probe was substituted before starting the monitoring period II. In this second period, a disturbance was applied to the system, related to the sudden decrease o HRT during less than a day, during day 35 (night and day). NIR probe was able to detect a disturbance that is coincident to the period when disturbance I is applied and also fouling during day 21, when occurred the stop of mixing and aeration in the reactor.

A first derivative (Savistzky-Golay method) and mean centering were performed before the PCA. Two PCs can explain 83 % of the total variation in the spectra (Figure 33).

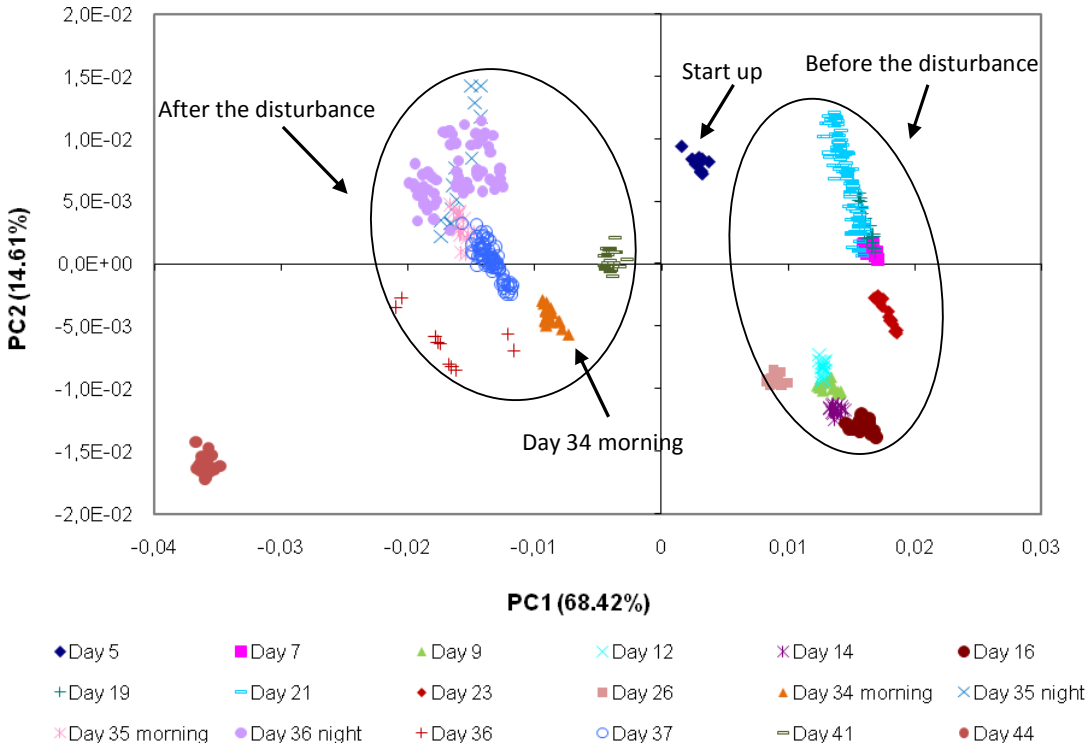


Figure 33. Score-plot representing NIR spectra variations during monitoring period II.

By analyzing the score plot it is possible to notice that spectra before and after the disturbance form two different clusters. On day 5 the spectra were different from the rest, what can be explained by the system's start up. A cluster is formed by the spectra from day 7 till day 26. Another cluster is formed by spectra including day 34 till day 41. Even though the disturbance induced to the system was only applied during day 35, the spectra from day 34 (morning) are already present in the same cluster as the next acquired spectra. This was not expected and by spectra direct observation it was noticed a displacement that was maintained till day 41. Probably the NIR probe was detecting this change and not the disturbance.

Fouling is perfectly noticed every time the spectra form a line in the PCA score-plot, being day 21 a clear example of this type of occurrence, having a similar response to this events as the UV-Visible probe.

Spectra from day 44, the last monitoring day for NIR probe are different from all the rest. This fact was not also possible to explain, since no disturbances in the process were observed by analyzing the monitoring parameters off-line.

3.4.3 Parameters Modelling

Since UV-Visible spectra acquired *in-situ* and off-line (spectrophotometer Jasco) presented satisfactory data for parameters modelling, the accuracy of the different methods was compared for parameters prediction. As for *in-situ* procedures no pre-treatment of the samples was performed for off-line spectral acquisition. NIR spectra could not be used for parameters modelling as it was thought initially, due to limitations encountered along the work.

COD, nitrate concentration and TSS were modelled, by performing PLS regression, selecting the best spectral ranges, the best pre-processing tools and by considering a bootstrap variable selection.

The number of latent variables (LV) was chosen by performing a cross-validation leave-one-out (LOO). By plotting RMSEVC against the number of latent variables it is possible to identify the number of latent variables that is necessary to have a good PLS model. If increasing the number of latent variables will not decrease RMSEVC, then the minimum number of latent variables with the lower RMSEVC should be selected, since considering a higher number of latent variables would probably make the model more complex and less robust.

As observed in score plot presented in Figure 27, it is feasible to gather all the data from monitoring periods I and II, for UV-Visible probe, since samples distribution is according to their nitrate concentration, being the matrix differences not sufficient to separate monitoring period I from period II. This situation was expected, since the activated sludge

system was operated in similar conditions in both periods. Assuming that the same can be applied for off-line analysis, the data from both monitoring periods was used for off-line parameters calibration.

In this analysis two spectral ranges were used: 250-380 nm and 250-500 nm. As previously, the objective of this selection is to compare results when using most of the UV-Visible range or just the UV spectral range, since UV region concentrates the largest part of the information (Figure 5).

After selecting the best spectral range for analysis, the use of the entire spectra or a selection of wavelengths was also considered for PLS regression, by using bootstrap objects or bootstrap residuals as variables selection methods.

A pre-processing procedure was performed for variables selection and for PLS regression. The results will be presented according to the selected pre-processing method for each situation.

The uncertainty of the determination of COD, nitrate and TSS, by the reference methods, was calculated according to the following equation:

$$x = \hat{x} \pm \frac{t_{0.05/2,n-1} s}{\sqrt{n}} \quad (18)$$

Where

x – Real value of the measurement;

$\hat{x}$ – Estimated value of the measurement (average value of the replicates);

s – Standard deviation of the replicates;

$t_{0.05/2,n-1}$ – Critical value for the student's t-distribution, for 95 % of confidence and n-1 degrees of freedom;

n – Number of samples.

3.4.3.1 *In-situ* UV-Visible parameters modelling

3.4.3.1.1 *COD modelling*

The best correlation achieved for COD was obtained for the range between 250 and 500 nm.

A bootstrap object variables selection was performed (

Figure 34). The bootstrap object selected a part of the variables (wavelengths), taking into account only the wavelengths that better correlate with the measured COD.

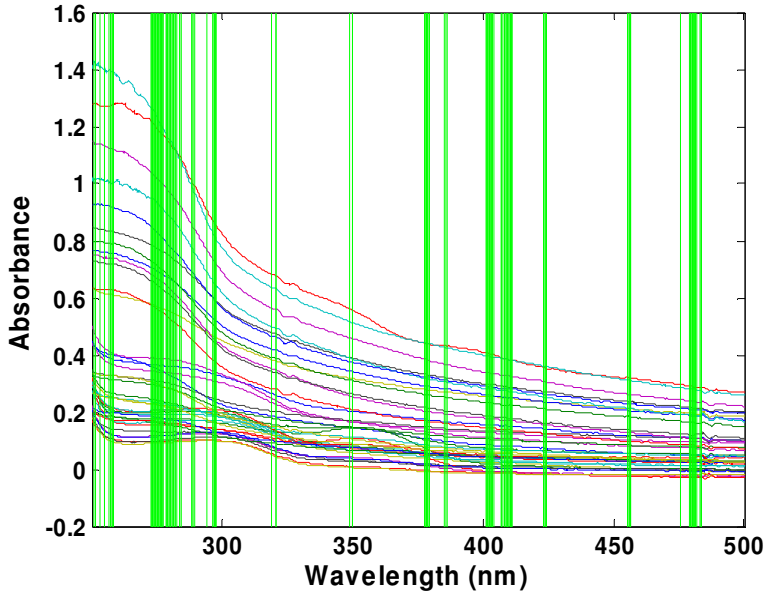


Figure 34. Wavelength selection for COD calibration by performing bootstrap object (X – Wavelength (nm); Y – Absorbance (A.U.)).

The variables selection minimizes the number of wavelengths used for PLS regression, making the calibration process faster and, hence, more suitable for real-time monitoring and control purposes.

This procedure was quite effective for obtaining a significant improvement in the results in terms of RMSECV and correlation coefficient (Table 13). Variables selection enabled the reduction of several interferences, by reducing the necessary wavenumbers for the

correlation. The R^2 value obtained for PLS B indicates that UV-Visible immersion probe can be suitable for COD determinations, what is in agreement with the results from literature.

Table 13. Results obtained for COD calibration with UV-Visible immersible probe, by performing PLS regression without (PLS A) and with variables selection (PLS B)

COD mg O_2/L		
PLS A	PP	SNV
	LV	6
	RMSECV	25.0
	R^2	0.5420
Variables Selection	PP	SNV
	LV	8
PLS B	PP	SNV
	LV	6
	RMSECV	15.4
	R^2	0.8239

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SNV – Standard Normal Variate.

RMSECV value obtained for PLS B can be considered acceptable (15.4 mg O_2/L) if the model is not used for a very precise quantification. Considering the COD Portuguese limit discharge of 150 mg O_2/L (DL n. º 236/98, 1 August) this value of RMSECV is satisfactory. Taking into account the COD concentration range used in this work (20 – 160 mg O_2/L), only the highest values are feasible of being used with better prediction.

The highest errors for the determination of COD by the reference method can be exemplified by the error for one of the lowest obtained concentrations: 30.16 ± 3.07 mg O_2/L. This error is around 10 % of the average concentration, what means that even though lower concentrations should have been measured with more accuracy to obtain a better PLS model, this error is lower than the obtained RMSECV.

The COD measured values were plotted against the COD values predicted by the model (Figure 35). It was possible to achieve a good COD distribution between 20 mg O_2/L and 120 mg O_2/L. However, it is important to refer that lower COD values were subject of a greater dispersion and experimental error.

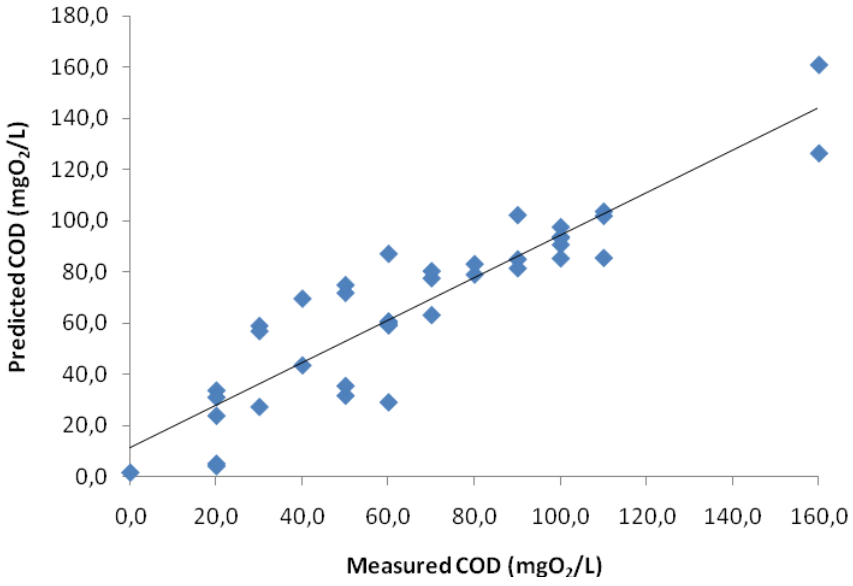


Figure 35. Regression curve for COD with variables selection (bootstrap object).

Higher correlation coefficients were obtained in several *in-situ* UV-Visible spectroscopic studies, however, for a COD range bellow 200 mg O_2/L, more usual in the effluent, calibration coefficients may be lower than the ones achieved for the concentrations found in the influent (Langergraber *et al.*, 2004a). In this study, Langergraber *et al.* (2004) modelled COD for the range between 75 and 175 mg O_2/L, not obtaining very low COD values in the effluent. This may have contributed to the achievement of better results.

An increase in the number of COD analysis and spectra acquisition, with a broader range of concentrations, would help to improve the PLS calibration. Nevertheless, the results obtained can already be very useful for monitoring purposes, by indicating changes in the system in terms of COD in the outlet.

Considering that spectroscopic techniques are not yet prepared to completely substitute reference analytical methods, for more quantitative purposes, the possibility of using a system to monitor and detect major variations, can avoid a number of analysis. In fact, this is already an important achievement. The kind of hazardous residues produced and its management, the sampling and the time spent to perform a COD analysis, are reasons enough to consider the substitution of traditional for these spectroscopic methods that can indicate the periods when is worth to measure COD values decreasing the frequency of measurements.

3.4.3.1.2 Nitrate modelling

The best correlation achieved for nitrate modelling was obtained for the range between 250 and 380 nm, not being expected a great influence of the visible region in its determination.

A bootstrap object variables selection was performed (

Figure 36), improving the results, although not in a greater extend as achieved for COD modelling (Table 14). It is suggested that for nitrate determination, most of the UV region was important, being less affected by interferences.

In fact, a great part of the UV region of the spectra can be used for nitrate calibration, what can originate better correlation results.

Comparing to the COD calibration results, nitrate determination achieved the best correlation coefficient but the RMSECV value must be analyzed in a different perspective. This error will affect more the determination of lower $N\text{-}NO_3^-$ concentration values, regarding that the working range is 0-160 mg/L (Figure 37). However, this value of RMSECV cannot be acceptable when taking into account the Portuguese legislation discharge limit for this parameter – 11.3 mg $N\text{-}NO_3^-$ /L (DL n. º 236/98, 1 August). Though, this model can be more suitable for nitrification process monitoring, since it is expected higher nitrate concentrations in the system outlet.

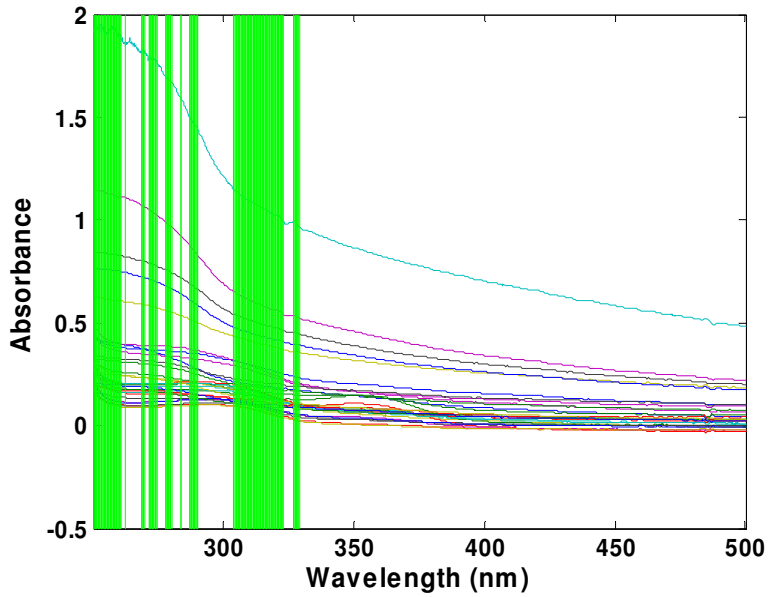


Figure 36. Wavelength selection for $N\text{-}NO_3^-$ calibration by performing bootstrap object (X – Wavelength (nm); Y – Absorbance (A.U.)).

Nitrate concentration values measured during the monitoring periods are not well distributed along the regression curve, and the lack of more middle values can be at the origin of this RMSECV lower value.

The error of determination of nitrate by HLPC can be considered quite low, regarding the greater error found for the concentration of 12.08 ± 0.15 mg $N\text{-}NO_3^-$/L, which corresponds to less than 2 % of the average concentration.

In several multiparametric spectroscopic studies the nitrate calibration achieved better correlation coefficients than COD or TSS (Thomas *et al.*, 1996; El Khorassani *et al.*, 1999; Rieger *et al.*, 2004), what can be explained by the fact that both COD and TSS are aggregate parameters, not having a defined maximum absorbance peak as nitrate has.

Table 14. Results obtained for $N\text{-}NO_3^-$ calibration with UV-Visible immersible probe, by performing PLS regression without (PLS A) and with variables selection (PLS B)

$N\text{-}NO_3^-$ mg/L		
PLS1	PP	SG(31,2,1)
	LV	3
	RMSECV	20.53
	R^2	0.8527
Variables Selection	PP	SG(15,2,1)
	LV	2
PLS B	PP	MNCN
	LV	3
	RMSECV	18.96
	R^2	0.8705

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SG(x,y,z) – Savitzky-Golay (window range, polynomial order, derivative order), MNCN – Mean Centering.

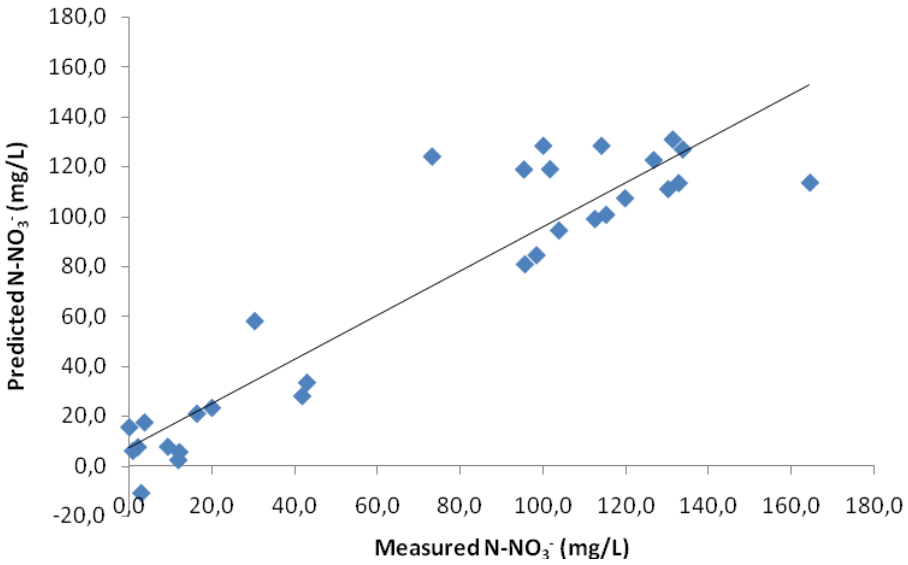


Figure 37. Regression curve for $N\text{-}NO_3^-$ with variables selection (bootstrap object).

3.4.3.1.3 TSS modelling

The best correlation achieved for TSS modelling was obtained for the range between 250 and 380 nm. However only a very small number of wavelengths were selected, by performing bootstrap objects for the PLS regression (Figure 38), showing that TSS had a poorer correlation to the acquired UV-Visible spectra. This is probably in the basis for the unsatisfactory results obtained for this parameter (Table 15).

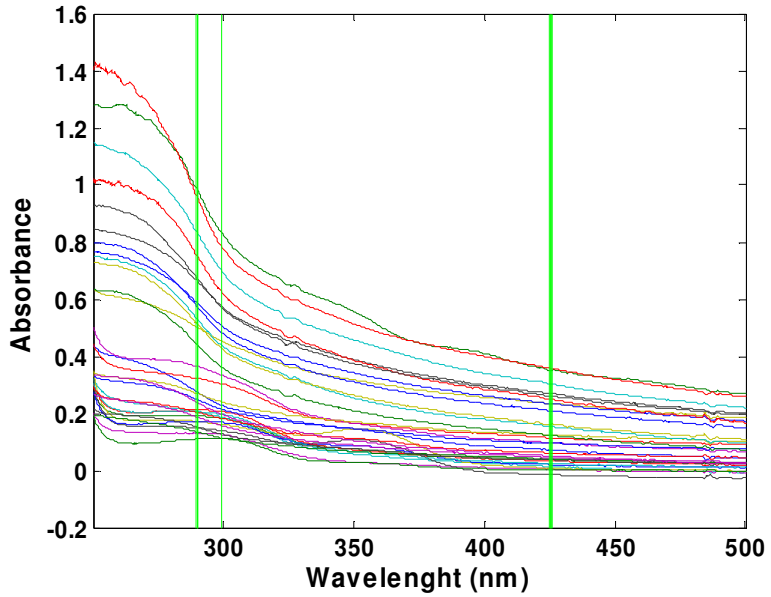


Figure 38. Wavelength selection for TSS calibration by performing bootstrap object (X – Wavelength (nm); Y – Absorbance (A.U.)).

However, since the correlation coefficient only indicates a good correlation between the real and the predicted values, this coefficient alone may not be the best way to evaluate the prediction ability of the model. In fact, Rieger *et al.* (2004), achieved a relatively good R^2 value of 0.845 for TSS modelling with a mean value of 13.5 mg/L and a precision of 5.5 mg/L. This precision compromises the model, once it is a large error when considering the studied TSS range of 0-25 mg/L. The possible explanation for the results obtained in this later work can be related to the spectra range used: 210–400 nm. As a matter of fact, according to Figure 5, TSS are usually detected in the visible spectral range and thus, better correlations could be expected using this spectroscopic range.

Table 15. Results obtained for TSS calibration with UV-Visible immersible probe, by performing PLS regression without (PLS A) and with variables selection (PLS B)

TSS mg/L		
PLS A	PP	SG(31,2,1)
	LV	2
	RMSECV	36.70
	R^2	0.4700
Variables Selection	PP	SG(15,2,1)
	LV	5
PLS B	PP	MNCN
	LV	4
	RMSECV	35.30
	R^2	0.5117

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SG(x,y,z) – Savitzky-Golay (window range, polynomial order, derivative order); MNCN – Mean Centering.

Even though it was possible to achieve good TSS distribution between 20 and 190 mg/L (Figure 39), dispersion for higher concentrations is noticed. In terms of RMSECV, the value of 35.30 mg/L can be acceptable, when considering only the higher concentrations.

One of the major sources of errors related to the obtained results for this parameter can be due to the used reference method. When considering the lowest concentrations of TSS measured in the settler, the error can be quite significant: 28.33 ± 4.87 mg/L. This error corresponds to 17.2 % of the estimated concentration, being considered high. For increased concentrations the error is still high, corresponding to 12 % of the estimated concentration (165.0 ± 19.5 mg/L).

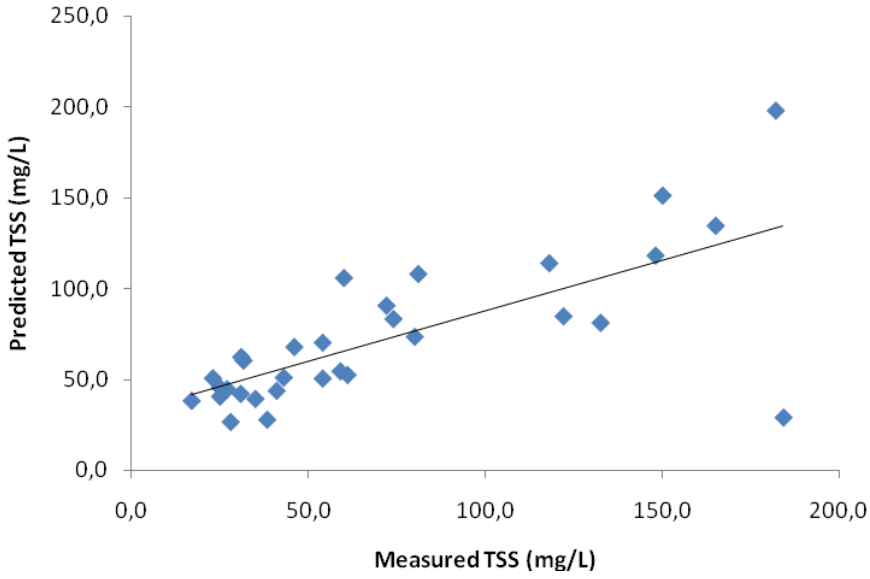


Figure 39. Regression curve for TSS with variables selection (bootstrap object).

Some other suggestions can be made to try to explain the lack of correlation between the visible region of the spectrum and the suspended solids content. Possible sampling errors could be the cause for these results, namely related to the representativeness of the sample taken from the settler and to lower solids concentration. Since very small particles of suspended solids present in the settler supernatant sometimes made the filtration process difficult, maybe the volume of sample, for TSS lower concentrations, was not sufficient, even though it was considered the minimum weight of solid residues required for TSS measurement. However, a system imbalance could happen if larger volume of samples were collected.

Another hypothesis which can be in the basis of the problem related to the UV-Visible determination of TSS, can be due to fouling. In fact, even though, in a regular monitoring day, the probe was not kept in the settler for very long periods, the particles accumulation in the sample window could interfere with the visible radiation transmitted to the detector.

3.4.3.2 Off-line UV-Visible parameters modelling

3.4.3.2.1 COD modelling

The best correlation for COD was achieved for the spectral range of 250-500 nm, being performed a bootstrap residuals variable selection (Figure 40).

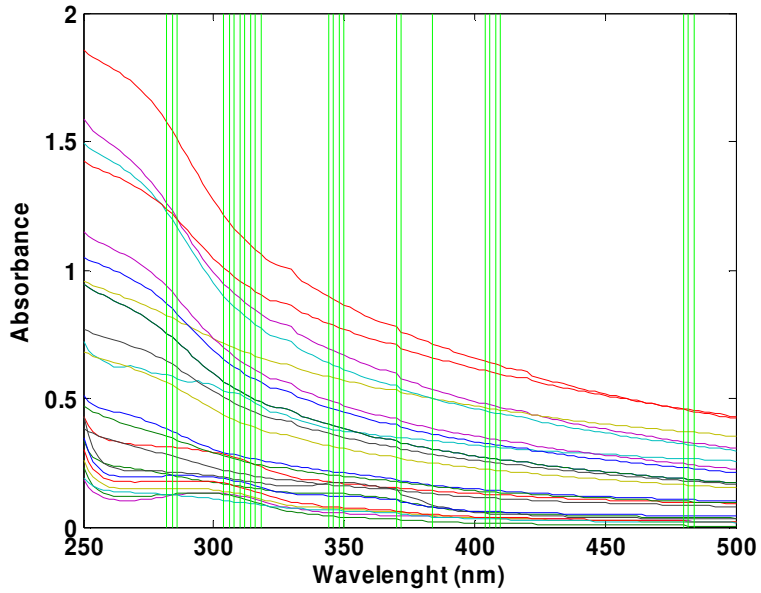


Figure 40. Wavelength selection for COD calibration by performing bootstrap residuals (X – Wavelength (nm); Y – Absorbance (A.U.)).

Although the results improvement after the wavenumber selection were significant in terms of R^2 and RMSECV (Table 16), these were not so good as the results obtained with *in-situ* immersion probe.

In this case, errors caused by sampling, that could modify the sample, being less representative of the system status, and also the possibility of solids settling during the spectra acquisition, could affect the measurements. However, a R^2 of 0.77 and a RMSECV of 18.2 mg O_2/L, cannot be considered as a bad result. In the absence of *in-situ* techniques, an

off-line spectroscopic analysis can also indicate COD concentration values in the sample, without performing any pre-treatment.

A satisfactory COD values distribution was also achieved in a similar way to the *in-situ* technique (Figure 41).

Table 16. Results obtained for COD calibration with UV-Visible off-line spectra acquisition, by performing PLS regression without (PLS A) and with variables selection (PLS B)

COD mgO_2/L		
PLS A	PP	SG(31,2,2)
	LV	3
	RMSECV	24,0
	R^2	0.6084
Variables Selection	PP	SG(15,2,1)
	LV	3
PLS B	PP	MNCN
	LV	9
	RMSECV	18,2
	R^2	0.7719

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SG(x,y,z) – Savitzky-Golay (window range, polynomial order, derivative order); MNCN – Mean Centering.

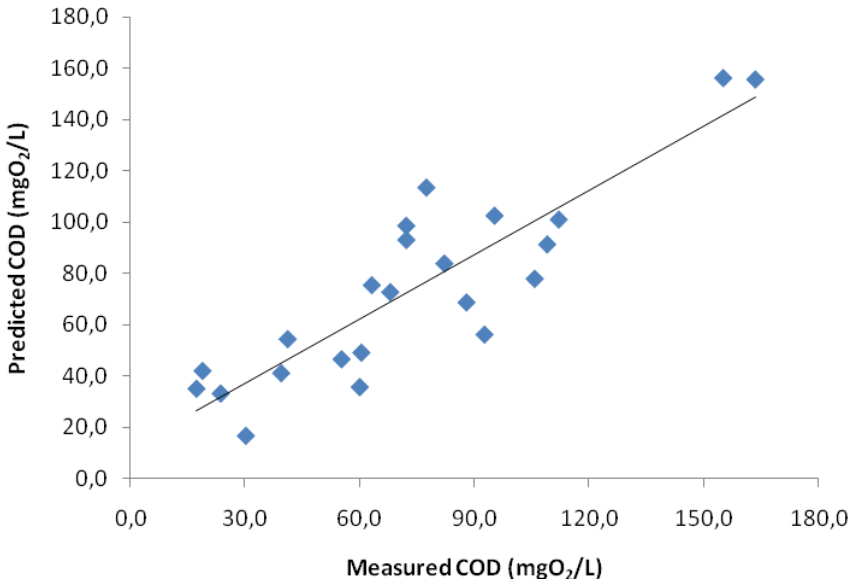


Figure 41. Regression curve for COD with variables selection (bootstrap residuals).

3.4.3.2.2 Nitrate modelling

As for *in-situ* measurements, the best correlation was achieved for the range 250 – 380 nm, without pre-processing the spectra. Wavenumber selection was performed using bootstrap objects (Figure 42). When compared to the variables selection performed for the immersible probe, it is possible to notice that the number of wavelengths selected is lower.

Even though this selection improved the obtained results (Table 17), these are not so good when compared to the *in-situ* technique.

It is important to refer that for off-line spectral acquisition, a lower number of spectra were acquired. In order to have an accurate PLS model, it could be important to increase the amount of spectra measured.

The nitrate modelling results are the best obtained for off-line spectral acquisition, but not so different from COD results. Once more, regarding effluent discharge limits, the RMSECV value is not satisfactory, but for the nitrification process monitoring, with higher nitrate concentrations, it can be acceptable.

Figure 43 presents the obtained regression curve for nitrate.

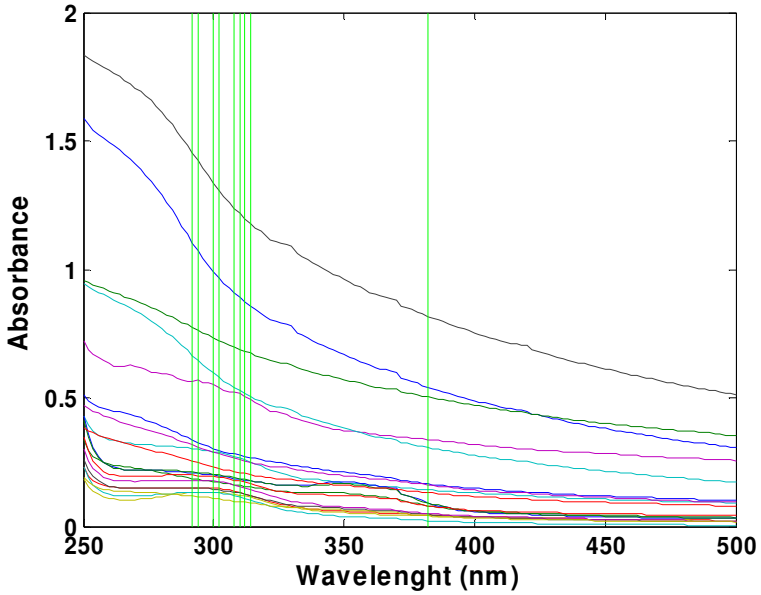


Figure 42. Wavelength selection for N-NO_3^- calibration by performing bootstrap objects (X – Wavelength (nm); Y – Absorbance (A.U.)).

Table 17. Results obtained for $N\text{-}NO_3^-$ calibration with UV-Visible off-line spectra acquisition, by performing PLS regression without (PLS A) and with variables selection (PLS B)

$N\text{-}NO_3^-$ mg/L		
PLS A	PP	MNCN
	LV	5
	RMSECV	25,82
	R^2	0.7578
Variables Selection	PP	SG(15,2,2)
	LV	3
PLS B	PP	MNCN
	LV	3
	RMSECV	23,77
	R^2	0.7935

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SG(x,y,z) – Savitzky-Golay (window range, polynomial order, derivative order), MNCN – Mean Centering.

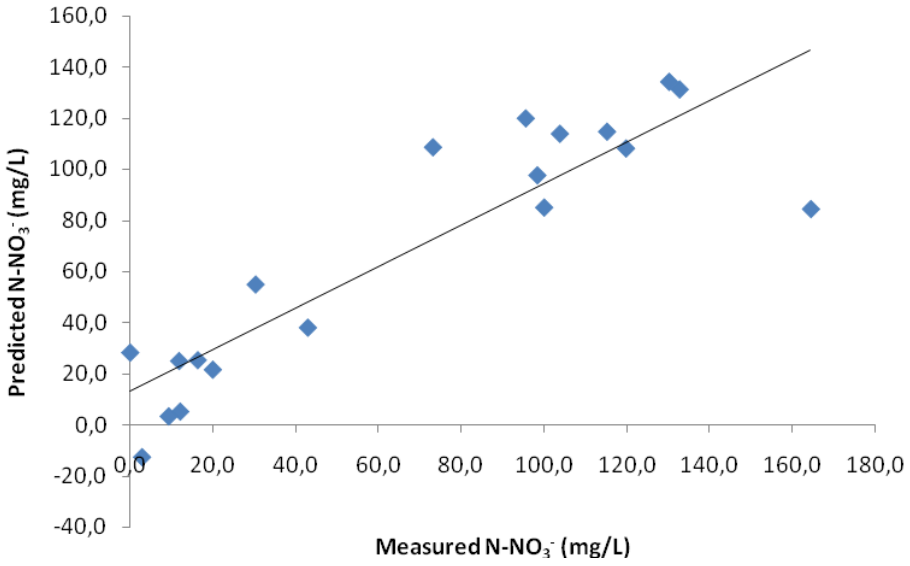


Figure 43. Regression curve for $N\text{-}NO_3^-$ with variables selection (bootstrap object).

3.4.3.2.3 TSS modelling

The best correlation for TSS was obtained for the spectral range of 250-380 nm. This was the only case were the variables selection did not improve the results. Table 18 presents the obtained final results.

Table 18. Results obtained for TSS calibration with UV-Visible off-line spectra acquisition, by performing PLS regression without variables selection (PLS A)

TSS mg/L		
PLS A	PP	SG(5,2,2)
	LV	5
	RMSECV	33.60
	R^2	0.6308

Legend: PP – Pre-processing method; LV – Latent variables; RMSECV – Root Mean Square Error of Cross Validation; R^2 – Correlation coefficient; SG(x,y,z) – Savitzky-Golay (window range, polynomial order, derivative order).

Even though the data distribution can be satisfactory (Figure 44), the results are still far from what was expected. However, compared to TSS calibration results obtained for the *in-situ* technique, these results are improved, mainly in terms of correlation coefficient.

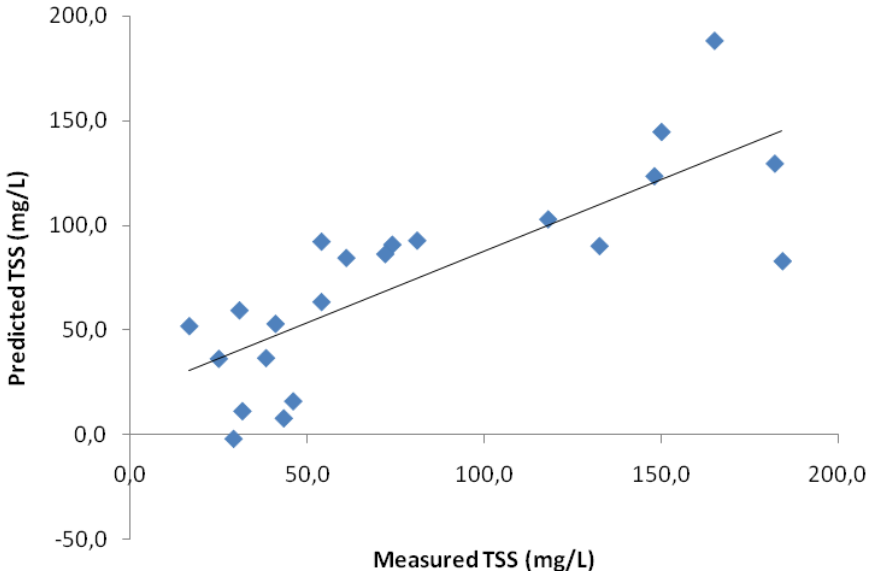


Figure 44. Regression curve for TSS without variables selection.

4 General Conclusions

The results obtained in this work permitted to conclude that the investigated monitoring system allowed correctly following of the status of an activated sludge process. Different tests and system imbalances were performed in order to obtain the maximum information about the system response and ability of the probes to detect them.

It was possible to conclude that UV-Visible and NIR probe can distinguish between solutions with different composition, being able to detect variations in composition at the inlet of a wastewater treatment system. Regarding the discrimination of solutions with different COD concentrations, UV-Visible probe was more effective than NIR probe in its identification. PCA results showed that the different concentration and variability among replicates were easily detected by the UV-Visible probe, in all tested solutions. NIR probe could differentiate among the different concentrations, although this was not as clear as for the UV-Visible probe, probably due to physical interferences.

During *in-situ* monitoring of the activated sludge process, UV-Visible achieved better performances in both monitoring periods. However, several experimental problems affected the NIR probe and because of that a fair comparison would only be possible by performing more measurements with both probes in similar conditions.

The UV-visible probe could detect changes in the spectra due to variations in the composition of the effluent. These changes could be visually identified, helping in the differentiation of diverse effluent characteristics in terms of composition. Nitrate and COD concentration variations were clearly detected by UV-Visible probe. Regarding the nitrification process, UV-Visible probe can be quite suitable for the identification of the system status before performing off-line analysis.

Relations between nitrate concentration, MLVSS and pH were confirmed using the UV-Visible spectra, showing that a spectra acquisition can provide information identical to the use of different monitoring parameters. Other associations were possible related to the COD degradation process, but not as evident as for nitrification, mainly due to the lack of broader range of outlet concentrations.

COD largest variation in the process (disturbance I – monitoring period II), was easily detected by the UV-Visible probe in real-time, showing that UV-Visible probe can be a fast

monitoring technique and very suitable for control purposes. COD variations in the outlet were not so easily obtained as for nitrate, since the COD degradation by the activated sludge system, was always very effective, recovering very fast after disturbances.

In-situ NIR probe results were not as satisfactory as for UV-Visible probe, mainly due to an incident (monitoring period I) that limited the amount of data obtained in the same conditions. PCA showed the differences between the spectra acquired before and after the incident, which could not be detected by visual spectra comparison. Even though, NIR probe was able to detect the disturbances occurred during monitoring period I (inflow adjustment and MLVSS decrease), showing that this spectroscopic range has potentialities for *in-situ* monitoring. The disturbance applied during monitoring period II (HRT sudden decrease) was not able to be detected by NIR probe, mostly due to a variation in the spectra acquisition in the moment before the disturbance. Further spectra processing will be needed to conclude about this event. The limitations and possible interferences, due to physical variations found in the biological process, limited a better insight of the potentialities of this range for wastewater monitoring.

During *in-situ* monitoring, fouling was detected as an obstacle as it could mask the identification of biochemical variations in the process that could occur at the same time. Spectra displacement in PCA scores-plot were easily identified as indicators of solids accumulation in the sample window. Complementary tests showed that the spectra variation due to exposure time while immersed in the settler was more obvious in NIR range, although in reality the effect was similar to both probes.

The results obtained from residuals statistics (Q) for UV-Visible probe, in the detection of the disturbance applied in monitoring period II, needed a careful interpretation, since fouling occurred more than once during this period. Spectra variation due to fouling can lead to a misleading since these events are identified as outliers by residuals statistics (Q). However, during bulking events fouling can be an important indicator. An automated cleaning system can be a good option to reduce problems due to fouling, for longer immersion periods.

COD, nitrate and TSS modelling using PLS regression was promissory using UV-Visible probe, and better than for off-line spectra acquisition. For COD, a RMSECV of 15.4 mg O_2/L was obtained for a concentration range between 20.0 and 160.0 mg O_2/L. For nitrate, a RMSECV of 19.0 mg N-NO_3^-/L was obtained for a concentration range between zero and 170.0 mg/L. For TSS, a RMSECV of 35.3 mg/L was obtained for a concentration range between 20.0 and 190.0 mg/L. Considering the lower concentration values, for each parameter, the errors obtained were not acceptable. Even though, these results can be considered satisfactory since they indicate that the *in-situ* probe is able of establishing differences between lower and higher values. The errors obtained by the reference methods for COD and TSS, regarding the lowest concentrations, could have limited the achievement of better results. For nitrate concentration prediction, the lack of a broader range of concentrations can be one possible explanation for the high RMSECV obtained.

The better results obtained for *in-situ* acquisition when compared to off-line spectra acquisition confirm that UV-Visible probe is the most suitable technique for wastewater monitoring.

The conclusions obtained help in the understanding of the limitations and benefits of *in-situ* UV-Visible wastewater monitoring. NIR application in this work showed promissory results, even though it was affected by problems that limited the achievement of better results.

This work allowed to recognize that all steps of *in-situ* spectroscopic applications can be determinant for the final results, existing all sort of factors and variables that must be considered during the spectra acquisition, system monitoring and also during spectra processing. Nevertheless, very good perspectives for future use of *in-situ* UV-Visible technique have been shown through these results, namely for the application in real time control of wastewater treatment systems.

5 Perspectives for Future Research

Regarding the importance of this type of studies for the development of more suitable techniques for monitoring and control of biological processes, the following ideas are suggested as future research:

- Improvement and validation of the obtained models by continuing to use the *in-situ* UV-Visible probe in the lab scale activated sludge system;
- To perform more studies where UV-Visible and NIR ranges may be more comprehensively compared;
- The extent of the actual work to other biological wastewater treatment systems;
- Selection of other parameters for calibration purposes;
- The application of immersible probes to real wastewater monitoring.

6 REFERENCES

Aguado D., Rosen C., (2008). Multivariate statistical monitoring of continuous wastewater treatment plants. Engineering Applications of Artificial Intelligence 21 (7): 1080-1091.

APHA, (1999). Standard Methods for the Examination of Water and Wastewater – 20th Edition. American Public Health Association, Washington, DC.

Arnold S. A. , Gaensakoo R., Harvey L.M., McNeil B., (2002). Use of At-Line and *In-Situ* Near-Infrared Spectroscopy to Monitor Biomass in an Industrial Fed-Batch *Escherichia coli* Process. Biotechnology and Bioengineering 80 (4): 405-413.

Aske N., Kallevik H., Sjöblom J., (2001). Determination of Saturate, Aromatic, Resin, and Asphaltenic (SARA) Components in Crude Oils by Means of Infrared and Near-Infrared Spectroscopy. Energy & Fuels 15: 1304-1312.

Balabin R. M., Safieva R. Z., Lomakina E. I., (2007). Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. Chemometrics and Intelligent Laboratory Systems 88: 183–188.

Barnes R., Dhanoa M., Lister S. (1989). Standard normal variate transformation and detrending of near infrared diffuse reflectance. Applied Spectroscopy 43: 772-777.

Blanco M., Villarroya I., (2002). NIR spectroscopy: a rapid-response analytical tool. Trends in Analytical Chemistry 21 (4): 240-250.

Bourgeois W., Burgess J. E., Stuetz R. M., (2001). Review On-line monitoring of wastewater quality: a review. Journal of Chemical Technology and Biotechnology 76: 337-348.

Büning-Pfaue H., (2003). Analysis of water in food by near infrared spectroscopy. Food Chemistry 82: 107–115.

Burgess C., (2007). The Basics of Spectrophotometric Measurement. *In* Thomas O., Burgess C., UV-Visible Spectrophotometry of Water and Wastewater. Techniques and Instrumentation in Analytical Chemistry, Volume 27, Elsevier.

Carvalho A. R., Sánchez M. N., Wattoom J., Brereton R. G., (2006). Comparison of PLS and kinetic models for a second-order reaction as monitored using ultraviolet visible and mid-infrared spectroscopy. Talanta 68: 1190–1200.

Chen D., Hu B., Shao X., Su Q., (2004a). Removal of major interference sources in aqueous near-infrared spectroscopy techniques. Analytical and Bioanalytical Chemistry 379: 143–148.

Chen J., Jönsson P., Tamura M., Gu Z., Matsushita B., Eklundh L., (2004b). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sensing of Environment 91: 332–344.

Christensen J., Hansen A. B., Tomasi G., Mortensen J., Andersen O., (2004). Integrated Methodology for Forensic Oil Spill Identification. Environmental Science and Technology 38 : 2912-2918.

Čurda L., Kukačková O., (2004). NIR spectroscopy: a useful tool for rapid monitoring of processed cheeses manufacture. Journal of Food Engineering 61: 557–560.

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

Dobbs R. A., Wise R. H., Dean R. B. (1972). The use of ultraviolet absorbance for monitoring the total organic carbon content of water and wastewater. Water Research 6: 110–1173.

Einax J.W., Zwanziger H.W., Geiß S., (1997). Chemometris in Environmental Analysis. VCH, Germany.

El Khorassani H., Trebuchon P., Bitar H., Thomas O., (1999). A simple UV spectrophotometric procedure for the survey of industrial sewage system. Water Science and Technology 39 (10): 77-82.

Escalas A., Droguet M., Guadayol J.M., Caixach J., (2003). Estimating DOC regime in a wastewater treatment plant by UV deconvolution. Water Research 37: 2627–2635.

Faber N. M., (2002). Uncertainty estimation for multivariate regression coefficients. Chemometrics and Intelligent Laboratory Systems 64:169–179.

Ferree M.A., Shannon R.D., (2001). Evaluation of a second derivative UV/Visible spectroscopy technique for nitrate and total nitrogen analysis of wastewater samples. Water Research 35 (1): 327-332.

Fogelman S., Zhao H., Blumenstein M., (2006). A rapid analytical method for predicting the oxygen demand of wastewater. Analytical and Bioanalytical Chemistry 386: 1773–1779.

Franco D., Nuñez M. J., M. P., Sineiro J., (2006). Applicability of NIR spectroscopy to determine oil and other physicochemical parameters in Rosa mosqueta and Chilean hazelnut. European Food Research and Technology 222: 443–450.

Gerardi M. H., (2002). Nitrification and Denitrification in The Activated Sludge Process. Wastewater Microbiology Series, John Wiley and Sons, Inc., New York.

Giavasis I., Robertson I., McNeil B., Harvey L. M., (2003). Simultaneous and rapid monitoring of biomass and biopolymer production by *Sphingomonas paucimobilis* using Fourier transform-near infrared spectroscopy. Biotechnology Letters 25: 975–979.

Geladi P., Hadjiiski L., Hopke P., (1999). Multiple regression for environmental data: nonlinearities and prediction bias. Chemometrics and Intelligent Laboratory Systems 47: 165–173.

Hansson M., Nordberg A., Mathisen B., (2003). On-line NIR monitoring during anaerobic treatment of municipal solid waste. Water Science and Technology 48 (4): 9-13.

Heikka R. A., Immonen K.T., Minkkinen P. 0., Paatero E. Y.O., Salmi T. 0., (1997). Determination of acid value, hydroxyl value and water content in reactions between dicarboxylic acids and diols using near-infrared spectroscopy and nonlinear partial least squares regression. Analytica Chimica Acta 349: 287-294.

Holm-Nielsen J. B., Dahl C. K., Esbensen K. H., (2006). Representative sampling for process analytical characterization of heterogeneous bioslurry systems—a reference study of sampling issues in PAT. Chemometrics and Intelligent Laboratory Systems 83: 114–126.

Karlsson M., Karlberg B., Olsson R.J.O., (1995). Determination of nitrate in municipal waste water by UV spectroscopy. Analytica Chimica Acta 312: 117-113.

Langergraber G., Fleischmann N., Hofstaedter F., Weingartner A., (2004a). Monitoring of a paper mill wastewater treatment plant using UV/VIS spectroscopy. Water Science and Technology 49 (1): 9–14.

Langergraber G. , Gupta J.K. , Pressl A. , Hofstaedter F. , Lettl W. , Weingartner A., Fleischmann N., (2004b). On-line monitoring for control of a pilot-scale sequencing batch

reactor using a submersible UV/VIS spectrometer. Water Science and Technology 50 (10): 73–80.

Laporte M. F., Paquin P., (1999). Near-Infrared Analysis of Fat, Protein, and Casein in Cow's Milk. Journal of Agriculture and Food Chemistry 47: 2600-2605.

Li X., He Y., Wu C., Sun D., (2007). Nondestructive measurement and fingerprint analysis of soluble solid content of tea soft drink based on Vis/NIR spectroscopy. Journal of Food Engineering 82: 316–323.

Lopes J.A., Costa P. F., Alves T. P., Menezes J. C., (2004). Chemometrics in bioprocess engineering: process analytical technology (PAT) applications. Chemometrics and Intelligent Laboratory Systems 74: 269–275.

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

Maribas A., da Silva M. C. L., Laurent N., Loison B., Battaglia P., Pons M.N., (2008). Monitoring of rain events with a submersible UV/VIS spectrophotometer. Water Science and Technology 57 (10): 1587-1593.

Mark H., (2001). Part B: Applications Based on Near-Infrared Absorbance – Fundamentals of Near-Infrared Spectroscopy. *In* Raghavachari R., Near-Infrared Applications in Biotechnology: Practical Spectroscopy Series. Volume 25. Marcel Dekker, Inc., New York.

Márquez M. C., Costa C., and Jul M., (2004). Kinetics and Effects of Hydraulic Residence Time and Biomass Concentration on Removal of Organic Pollutants in a Continuous Unsteady State Activated Sludge Process. Chemical and Biochemical Engineering Quarterly 18(4): 423-428.

Matsché N., Stumvöhrer K., (1996). UV absorption as control parameter for biological treatment plants. Water Science and Technology 33 (12): 211–218.

Metcalf and Eddy (2003). Wastewater Engineering – Treatment and Reuse. Fourth Edition, McGraw-Hill International Editions, New York.

Miettinen T., Hurse T. J., Connor M. A., Reinikainen S., Minkkinen P., (2004). Multivariate monitoring of a biological wastewater treatment process: a case study at Melbourne

Water's Western Treatment Plant. Chemometrics and Intelligent Laboratory Systems 73: 131– 138.

Miller J. N., Miller J.C., (2000). Statistics and Chemometrics for Analytical Chemistry. Fourth Edition, Pearson – Prentice Hall, England.

Mrkva, M. (1975). Automatic UV-control system for relative evaluation of organic water pollution. Water Research 9: 587–589.

Naes T., Isaksson T., Fearn T., Davies T., (2002). A User-Friendly Guide to Multivariate Calibration and Classification. NIR Publications, United Kingdom.

Nake A., Dubreuil B., Raynaud C., Talou T., (2005). Outdoor in situ monitoring of volatile emissions from wastewater treatment plants with two portable technologies of electronic noses. Sensors and Actuators B 106: 36–39.

Nicolaï B. M., Beullens K., Bobelyn E., Peirs A., W. Saeys, Theron K. I., Lammertyn J., (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology 46: 99–118.

Nordberg A., Hansson M., Sundh I., Nordkvist E., Carlsson H., Mathisen B., (2000). Monitoring of a biogas process using electronic gas sensors and near-infrared spectroscopy (NIR). Water Science and Technology 41 (3): 1-8.

Ozen B. F., Mauer L.J., (2002). Detection of Hazelnut Oil Adulteration Using FT-IR Spectroscopy. Journal of Agricultural and Food Chemistry 50: 3898-3901.

Páscoa R., (2006). Monitorização *in-situ* de processos de tratamento de águas residuais com espectroscopia de infra-vermelho próximo. Tese de Mestrado. Faculdade de Farmácia da Universidade do Porto.

Páscoa N.M., Lopes J. A., Lima J.L.F.C., (2008). *In situ* near infrared monitoring of activated dairy sludge wastewater treatment processes. Journal of Near Infrared Spectroscopy 16: 409-419.

Pasquini C., (2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. Journal of the Brazilian Chemical Society 14 (2): 198-219.

Peré-Trepat E., Petrovic M., Barceló D., Tauler R., (2004). Application of chemometric methods to the investigation of main microcontaminant sources of endocrine disruptors in coastal and harbour waters and sediments. Analytical and Bioanalytical Chemistry 378: 642–654.

Pons M. N., Le Bonté S., Potier O., (2004). Spectral analysis and fingerprinting for biomedia characterization. Journal of Biotechnology 113: 211–230.

Qu N., Zhu M., Mi H., Dou Y., Ren Y., (2008). Nondestructive determination of compound amoxicillin powder by NIR spectroscopy with the aid of chemometrics. Spectrochimica Acta Part A 70: 1146–1151.

Raghavachari R., (2001). Part A: Applications Based on Near-Infrared Fluorescence – Introduction. *In* Raghavachari R., Near-Infrared Applications in Biotechnology: Practical Spectroscopy Series. Volume 25. Marcel Dekker, Inc., New York.

Reich G., (2005). Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. Advanced Drug Delivery Reviews 57: 1109–1143.

Rieger L., Langergraber G., Thomann M., Fleischmann N., Siegrist H., (2004). Spectral *in-situ* analysis of NO_2, NO_3, COD, DOC and TSS in the effluent of a WWTP. Water Science and Technology 50 (11):143–152.

Rieger L., Langergraber G., Siegrist H., (2006). Uncertainties of spectral in situ measurements in wastewater using different calibration approaches. Water Science and Technology 53 (12): 187–197.

Rittman B. E., McCarty P.L., (2001). Environmental Biotechnology: Principles and Applications. McGraw-Hill International Editions, New York.

Rodionova O. Y., Houmøller L. P., Pomerantsev A. L., Geladi P., Burger J., Dorofeyev V. L.., Arzamastsev A. P., (2005). NIR spectrometry for counterfeit drug detection: A feasibility study. Analytica Chimica Acta 549: 151–158.

Rosen C., (2001). A Chemometric Approach to Process Monitoring and Control with Applications to Wastewater Treatment Operation. Ph.D. Thesis, Lund Institute of Technology, Sweden.

Rosen C., Lennox J.A., (2001). Multivariate and multiscale monitoring of wastewater treatment operation. Water Research 35 (14): 3402–3410.

Sousa A. C., Lucio M. M. L. M., Neto O.F. B., Marcone G.P.S. , Pereira A. F.C. , Dantas E. O., Fragoso W.D., Araujo M. C. U., Galvão R. K.H., (2007). A method for determination of COD in a domestic wastewater treatment plant by using near-infrared reflectance spectrometry of seston. Analytica Chimica Acta 588: 231–236.

Šašić S., Ozaki Y., (2001). Short-Wave Near-Infrared Spectroscopy of Biological Fluids. 1. Quantitative Analysis of Fat, Protein, and Lactose in Raw Milk by Partial Least-Squares Regression and Band Assignment. Analytical Chemistry 73: 64-71.

Savitzky A., Golay M. J. E., (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36 (8): 1627-1639.

Sporea D. G., Sporea R. A., (2005). Setup for the in situ monitoring of the irradiation-induced effects in optical fibers in the ultraviolet-visible optical range. Review of Scientific Instruments 76: 113110.

Stephens A.B., Walker P.N., (2002). Near-infrared spectroscopy as a tool for real-time determination of BOD5 for single-source samples. Transactions of the ASAE 45 (2): 451-458.

Thomas M.J.K., (1996). Analytical Chemistry by Open Learning – Ultraviolet and Visible Spectroscopy. Second Edition. University of Greenwich UK, John Wiley and Sons, Ltd., England.

Thomas O., Burgess C., (2007). UV-Visible Spectrophotometry of Water and Wastewater. Techniques and Instrumentation in Analytical Chemistry, Volume 27, Elsevier.

Thomas O., El Khorassani H., Touraud E., Bitar H., (1999). TOC versus UV spectrophotometry for wastewater quality monitoring. Talanta 50: 743–749.

Thomas O., Gallot S., Mazas, N. (1990). Ultraviolet multiwavelength absorptiometry (UVMA) for the examination of natural waters and wastewaters. Part II: determination of nitrate. Fresenius Journal of Analytical Chemistry 338: 238–240.

Thomas O., Theraulaz F., Agnel C., Suryani S., (1996). Advanced UV examination of wastewater. Environmental Technology 17: 251-261.

Thomas O., Theraulaz F., Cerdà V., Constant D., Quevauviller P., (1997). Wastewater quality monitoring. Trends in Analytical Chemistry 16 (7): 419-424.

Thomas O., Theraulaz F., Domeizel M., Massiani C., (1993). UV spectral deconvolution: a valuable tool for waste water quality determination. Environmental Technology 14: 1187-1192.

Uddin M., Okazaki E., Ahmad M. U., Fukuda Y., Tanaka M., (2006). NIR spectroscopy: A non-destructive fast technique to verify heat treatment of fish-meat gel. Food Control 17: 660–664.

Vaidyanathan S., Arnold S. A., Matheson L., Mohan P., McNeil B., Harvey L. M., (2001). Assessment of Near-Infrared Spectral Information for Rapid Monitoring of Bioprocess Quality. Biotechnology and Bioengineering 74 (5): 376-388.

Vaidyanathan S., McNeil B. (1998). Near-infrared spectroscopy—A panacea in pharmaceutical bioprocessing? European Pharmaceutical Review 3 (2): 43–48.

Vaidyanathan S., White S., Harvey L. M., B. McNeil, (2003). Influence of Morphology on the Near-Infrared Spectra of Mycelial Biomass and Its Implications in Bioprocess Monitoring. Biotechnology and Bioengineering 82 (6): 715-724.

Vaillant S., Pouet M.F., Thomas O., (2002). Basic handling of UV spectra for urban water quality monitoring. Urban Water 4: 273–281.

Vargas A., and Buitrón G., (2006). On-line concentration measurements in wastewater using nonlinear deconvolution and partial least squares of spectrophotometric data. Water Science and Technology 53: 457–463.

Wang Y., Veltkamp D. J., Kowalski B. R., (1991). Multivariate Instrument Standardization. Analytical Chemistry 63 (23): 2750-2756.

Wehrens R., Van der Linden W. E., (1997). Bootstrapping principal component regression models. Journal of Chemometrics 11: 157-171.

Winiarski T., Thomas O., Charrier C., (1995). Analysis of the spatial and temporal variations in the water quality of a karstic aquifer using UV spectrophotometry. Journal of Contaminant Hydrology 19: 307-320.

Workman J., Springsteen A.W., (1998). Applied Spectroscopy – A compact reference for practioners. Academic Press, New York.

Wu J., Pons M.N., Potier O., (2006). Wastewater fingerprinting by UV-visible and synchronous fluorescence spectroscopy. Water Science and Technology 53: 449-456.

APPENDIX A

Immersible probes *in-situ* monitoring procedure

1 – The equipments (light sources and detectors) were turned about 10 minutes before *in-situ* monitoring for the light sources to stabilize.

2 – Probes tips and sample windows were rinsed with distilled water and the sample window was cleaned with smooth paper sheet. NIR probe sample window was dry before starting spectra acquisition.

3 – NIR detector's temperature was calibrated to – 4 ºC. Only after this temperature was reached the reference spectrum was acquired.

4 – UV-Visible probe was immersed in tap water and NIR probe was positioned in contact with air in a stable position for acquisition of the reference spectrum. During this procedure locations were always the same for both probes.

5 – After starting the Ocean Optics software (OOIBase32/Ocean Optics) integration time was adjusted till the peak for UV-Visible between 550 and 600 was close to 60 000 intensity units, as recommended by the manufacturer. The same procedure was done for NIR probe where integration time was increased till the peak between 1100 and 1200 nm was around 60 000 units of intensity (Figure 45). For this work it was chosen a value of 10 scans to make a spectral average, a boxcar width of 5 and electric dark correction for both probes. A reference and dark spectra were saved. The reference spectrum was saved with the amount of light intensity that was selected. A dark spectrum was acquired when no light passed through the probe (light source was disconnected). New reference and dark spectra were saved each time it was decided to perform a new calibration during probes operation, by defining new integration time values.

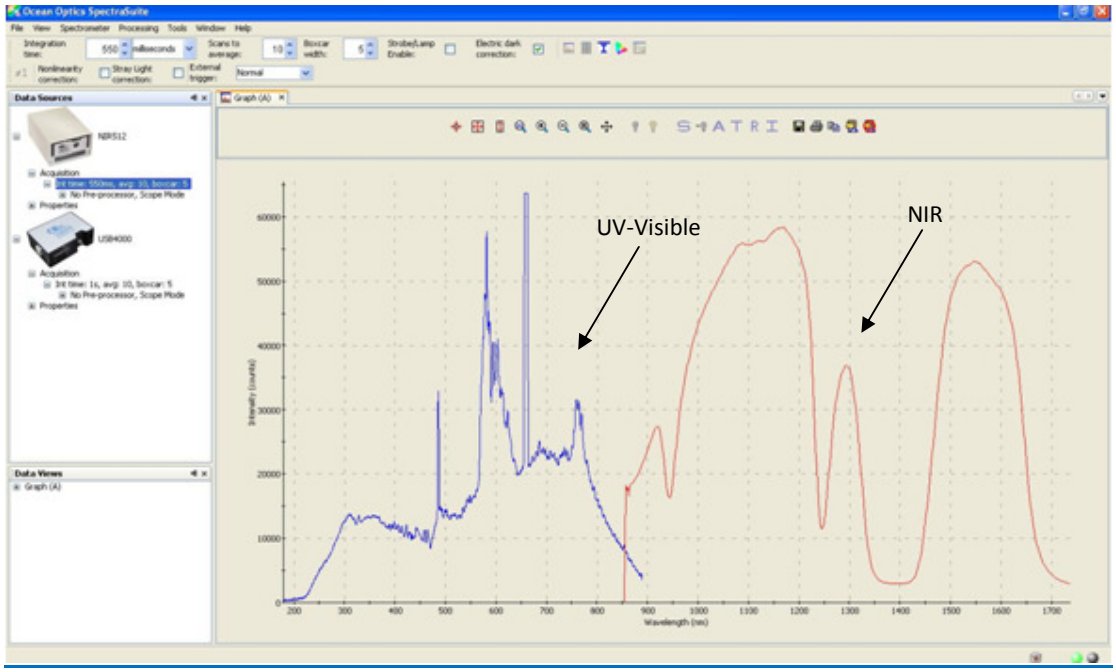


Figure 45. Intensity spectra for both probes during parameters adjustment, using Ocean Optics SpectraSuite software.

6 – Before initializing *in-situ* or sample spectra acquisition for both reference spectra (UV-Visible and NIR) were close to zero (within ± 0.05) absorbance units, throughout the entire wavelength range. Every time baseline was not close to zero absorbance units, new reference and dark spectra were performed.

7 – Probes were immersed in the settler and acquisition started after defining the period of time for automatic spectra acquisition (45 minutes). When necessary, spectra were saved manually. An example of simultaneous spectra acquisition is presented in Figure 46.

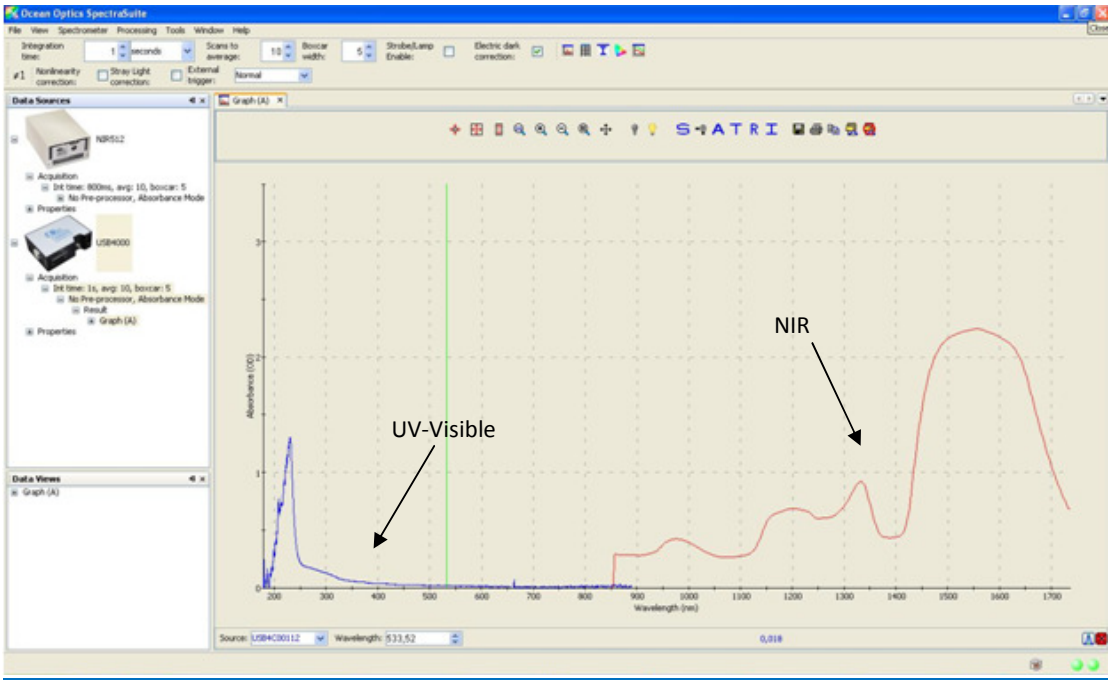


Figure 46. Instantaneous visualization of both spectra acquired in the settler.

COD determination – Reagents

- 1 L Digestion solution (500 mL distilled water + 10.216 g $K_2Cr_2O_7$ (previously dried at 103ºC during 2 hours) + 167 mL H_2SO_4 conc. + 33,3 g $HgSO_4$);
- 1 L Sulfuric Acid Reagent (9.715 g Ag_2SO_4 + 1 L H_2SO_4 concentrated);
- 1 L Potassium hydrogen phthalate (KHP) stock solution – 1000 mg CQO/L (425 mg KHP (previously dried at 105 ºC during one day) + distilled water).

N-Kjeldahl determination – Reagents and detailed procedure description

- NaOH solution (400 g/L);
- NaOH solution (40 g/L);
- NaOH solution (4 g/L);
- Bromocresol green indicator solution (1 g/L : 100 mg of bromocresol green in 100 mL of 96 % ethanol);
- Metil red indicator solution (1 g/L: 100 mg of metal red in 100 mL of 96 % ethanol);
- Boric acid solution (40 g/L: 40 g of boric acid were added to 600 mL of boiled ultrapure water and well dissolved; 300 mL of the same water were added. After the solution has reached room temperature 10 mL bromocresol green and 7 mL of metil red were added and the solution was diluted to 1000 mL. 25 mL of this solution were pipeted to a beaker and 100 mL of ultrapure water were added. Since this solution remained red, it was titrated with NaOH (4 g/L) to obtain a grey color, by adding 10 µL each time. A volume of NaOH (40 g/L) solution four times the volume of NaOH (4 g/L) spent for titration, was added to the solution. This solution lasted only one month);
- Concentrated H_2SO_4;
- H_2SO_4 (200 mL/L);
- H_2SO_4 (0.025 mol/L);
- Kjeltabs with selenium.

All the solutions were prepared with ultrapure water.

The digestion tubes were previously washed with H_2SO_4 (200 mL/L), followed by washing with ultrapure water and then were completely dried. In each digestion tube 10 mL of

concentrated H_2SO_4 were added to 10 mL of sample or ultrapure water (blank), followed by the addition of one Kjeltab with selenium. The tubes were carefully mixed and digested at 400 ºC connected to a system for vapors aspiration. The digestion was usually prolonged for 1 hour till the moment when white vapors appeared and the samples were limpid. After the digestion was considered finished, the digestion equipment was turned off, the tubes were cooled down outside the digester with vapors aspiration on for some minutes. After reaching room temperature 50 mL of ultrapure water were added to each tube.

An automatic distillation system (Tecator Kjeltec 1026) was used to rapidly distillate all the samples and blanks after digestion. For the distillation system it was necessary to supply ultrapure water and a very concentrated NaOH solution (400 g/L). A beaker with 25 mL of acid boric solution receives the distilled solution which results from each digestion tube. After distillation each solution in a beaker is titrated with a solution of H_2SO_4 (0.025 mol/L) until the color changes from green to grey.

Nitrate determination – HPLC operational conditions

HPLC Jasco

- Column: Varian Metacarb 87H;
- Column temperature: 60 ºC;
- Eluent composition: H_2SO_4 (0.005 mol/L);
- Eluent inflow : 0.70 mL/min;
- Eluent inflow pressure: 70-80 kg/cm^2;
- Detector: UV at 210 nm;
- Analysis time: 12 min.

Nitrite determination – Reagents

- Sulfanilamide reagent (in a hood 50 mL of concentrated HCl were added slowly in 300 mL of distilled water, in a volumetric flask; 5 g of sulfanilamide were dissolved in this solution which was completed with distilled water to 1000 mL);
- N(1-naftil) etilenodiamine-dihydrochloride reagent (NEDD) (500 mg of N(1-naftil) etilenodiamine dihydrocloride were dissolved in 500 mL of distilled water; this solution was kept in a dark place lasting only 1 month);