Ana Teresa Lourenço Santos

Licenciada em Ciências da Engenharia Química e Bioquimica

EVALUATION OF UV SPECTROPHOTOMETRY FOR ESTIMATION OF NITRITE AND NITRATE IN NITRIFIED URINE

Dissertação para obtenção do Grau de Mestre em Engenharia Química e Bioquímica

Orientadores: Kris Villez, Doutor, EAWAG
Alma Mašić, Doutora, EAWAG
Co-orientadora: Maria Ascensão Reis, Professora Doutora, FCT/UNL

Júri:

Presidente: Professora Doutora Isabel Maria de Figueiredo Ligeiro da Fonseca, FCT-UNL
Arguente: Doutor Adrian Michael Oehmen, FCT-UNL
Vogal: Doutor Kris Roger Elie Villez, EAWAG

FCt FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

Setembro 2014

Ana Teresa Lourenço Santos

Licenciada em Ciências da Engenharia Química e Bioquimica

EVALUATION OF UV SPECTROPHOTOMETRY FOR ESTIMATION OF NITRITE AND NITRATE IN NITRIFIED URINE

Dissertação apresentada à Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para obtenção do Grau de Mestre em Engenharia Química e Bioquímica

Setembro 2014

EVALUATION OF UV SPECTROPHOTOMETRY FOR ESTIMATION OF NITRITE AND NITRATE IN NITRIFIED URINE

ABSTRACT

Water is a limited resource for which demand is growing. Contaminated water from inadequate wastewater treatment provides one of the greatest health challenges as it restricts development and increases poverty in emerging and developing countries. Therefore, the connection between wastewater and human health is linked to access to sanitation and to human waste disposal. Adequate sanitation is expected to create a barrier between disposed human excreta and sources of drinking water. Different approaches to wastewater management are required for different geographical regions and different stages of economic governance depending on the capacity to manage wastewater. Effective wastewater management can contribute to overcome the challenges of water scarcity.

Separate collection of human urine at its source is one promising approach that strongly reduces the economic and load demands on wastewater treatment plants (WWTP). Treatment of source-separated urine appears as a sanitation system that is affordable, produces a valuable fertiliser, reduces pollution of water resources and promotes health. However, the technical realisation of urine separation still faces challenges. Biological hydrolysis of urea causes a strong increase of ammonia and pH. Under these conditions ammonia volatilises which can cause odour problems and significant nitrogen losses.

The above problems can be avoided by urine stabilisation. Biological nitrification is a suitable process for stabilisation of urine. Urine is a highly concentrated nutrient solution which can lead to strong inhibition effects during bacterial nitrification. This can further lead to process instabilies. The major cause of instability is accumulation of the inhibitory intermediate compound nitrite, which could lead to process breakdown. Enhanced on-line nitrite monitoring can be applied in biological source-separated urine nitrification reactors as a sustainable and efficient way to improve the reactor performance, avoiding reactor failures and eventual loss of biological activity. Spectrophotometry appears as a promising candidate for the development and application of on-line nitrite monitoring.

Spectroscopic methods together with chemometrics are presented in this work as a powerful tool for estimation of nitrite concentrations. Principal component regression (PCR) is applied for the estimation of nitrite concentrations using an immersible UV sensor and off-line spectra acquisition. The effect of particles and the effect of saturation, respectively, on the UV absorbance spectra are investigated. The analysis allows to conclude that *(i)* saturation has a substantial effect on nitrite estimation; *(ii)* particles appear to have less impact on nitrite estimation. In addition, improper mixing together with instabilies in the urine nitrification process appears to significantly reduce the performance of the estimation model.

Key-words: Biological nitrification, Chemometrics, Nitrite estimation, Principal Component Regression, Source-separated urine, Spectral sensor, UV Spectroscopy.

RESUMO

A água é um recurso limitado para o qual a procura é crescente. A água contaminada proveniente de tratamento inadequado de águas residuais conduz a um dos maiores desafios ao nível da saúde, condicionado o desenvolvimento de países em vias de desenvolvimento. O vínculo entre águas residuais e saúde humana está directamente relacionado com o acesso a saneamento e com a eliminação de resíduos humanos. Espera-se que sistemas de saneamento adequados permitam a separação eficaz de dejectos humanos de correntes de águas limpas. Desta forma, são requeridas diferentes abordagens na gestão de águas residuais de acordo com diferentes áreas geográficas e respectivos níveis de desenvolvimento económico. A gestão eficaz de águas residuais pode contribuir para solucionar a problemática de escassez de água.

A separação diferenciada de urina humana surge como uma promissora abordagem no sentido de reduzir grandemente as exigências económicas e energéticas em estações de tratamento de águas residuais. O tratamento de urina diferenciada apresenta-se como um sistema de saneamento económicamente viável, que possibilita a produção de um fertilizante de grande valor, reduzindo ainda a poluição dos recursos hidricos e promovendo a saúde humana. No entanto, a técnica de separação diferenciada de urina enfrenta ainda alguns desafios. A hidrólise biológica da ureia causa um aumento notável de amónia e pH. Nestas condições ocorre volatilização da amónia o que pode originar problemas de odor e percas significativas de azoto. Os problemas referidos podem ser evitados através da estabilização da urina, sendo a nitrificação biológica um processo apropriado para a sua estabilização. Sendo a urina uma solução de nutrientes extremamente concentrada, podem ocorrer durante a nitrificação significativos efeitos de inibição. A principal causa de falhas no processo é a acumulação de nitrito (intermediário da reacção) que funciona como inibidor, o que pode levar ao colapso do processo. A monitorização *on-line* do nitrito pode ser aplicada em reactores de nitrificação de urina diferenciada como uma forma sustentável e eficiente para melhorar o desempenho do mesmo, evitando falhas e, eventualmente, a perda de atividade biológica. A espectrofotometria é uma técnica promissora para o desenvolvimento e aplicação da monitorização *on-line* de nitrito. Métodos espectroscópicos juntamente com métodos quemiométricos são apresentados neste trabalho como uma ferramenta valiosa para estimar as concentrações de nitrito. Uma regressão de componentes principais (PCR) é desenvolvida para estimar as concentrações de nitrito recorrendo a um sensor de UV imersível e a aquisição espectral *off-line*. Neste trabalho são investigados o efeito de partículas e o efeito de saturação no espectro de absorção UV. A análise permite concluir que *(i)* a saturação apresenta um efeito significativo na quantificação de nitrito; *(ii)* as partículas afiguram-se como tendo uma importância menor na quantificação de nitrito. Adicionalmente, agitação imprópria juntamente com instabilidades no processo de nitrificação de urina reduzem significativamente a eficiência do modelo.

Palavras-chave: Espectroscopia de UV, Nitrificação biológica, Previsão de nitrito; Quemiometria, Regressão de Componentes Principais, Sensor espectral, Urina diferenciada.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

x

INDEX OF ACRONYMS

AOB	Ammonia-Oxidising Bacteria
CV	Cross Validation
EMR	Electromagnetic Radiation
HCO_3^-	Carbonic Acid
HNO_2	Nitrous Acid
LOOCV	Leave-One-Out Cross Validation
MBBR	Moving Bed Biofilm Reactor
NOB	Nitrite-Oxidising Bacteria
NH_3	Ammonia
NH_4^+	Ammonium
NO_2^-	Nitrite
NO_3^-	Nitrate
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
SSR	Sum of Squared Residuals
SVD	Singular Value Decomposition
WWTP	Wastewater Treatment Plant

1 INTRODUCTION

1.1 BACKGROUND

The production of waste from human activities is inevitable. A significant part of this waste ends up as wastewater (Henze & Comeau, 2008). The primary aim of wastewater treatment is the removal of nutrients, organic matter and suspended solids contained in wastewater influent. This is usually carried out by using several unit processes, which include biological, chemical and physical treatment methods (Sedlak, 1991). Current treatment of wastewater presents several shortcomings throughout the treatment process, e.g. consumption of high amounts of resources, loss of valuable nutrients into water bodies and poor elimination of micropollutants (Udert *et al.*, 2006). Furthermore, incomplete wastewater treatment contributes to eutrophication (Lienert & Larsen, 2009), caused by the water enrichment in nutrients, particularly phosphorus and nitrogen, which lead to great development of algae and consequent deterioration of water quality (Fernández *et al.*, 2014). Thus, managing wastewater and the resources it contains is an important step towards ecologically sustainable development (Ganrot, 2005).

The importance of wastewater management is more striking in emerging and developing countries, which have inadequate infrastructure and resources to address wastewater management in an efficient and sustainable way. Therefore, providing adequate sanitation is a major challenge for these countries (Udert *et al.*, 2006). Adapting centralized sanitation systems from industrialized countries to poorer and water scarce countries is nearly impossible as these systems require enormous capital investments and high amounts of water (Lienert & Larsen, 2009). Moreover, the typical wastewater treatment turns valuable resources into pollutants (Ganrot, 2005). Innovative approaches in wastewater management will generate significant returns, as addressing wastewater is a key step in reducing poverty and sustaining ecosystems (Corcoran *et al.*, 2010). New technologies focus on the recovery of resources such as water, nutrients and energy instead of simply preventing pollution (Udert & Wächter, 2012). One very promising approach is the separation of wastewater streams and their specific treatment in decentralized reactors (Udert & Wächter, 2012). The separate collection and treatment of urine has attracted considerable attention in the engineering community in the last few years and has been seen as a viable option for enhancing the sustainability of wastewater management (Maurer *et al.*, 2006).

Urine contains most of the nutrients in domestic wastewater (typically 80% of nitrogen and 50% of phosphorus (Lienert & Larsen, 2009)), however makes up less than one percent of the total wastewater volume and contributes with many problematic micropollutants such as synthetic hormones, pharmaceuticals and their metabolites (Maurer *et al.*, 2006; Udert *et al.*, 2003; Udert *et al.*, 2006). Separation of urine at the source allows for high nutrient recycling from a concentrated nutrient solution and at the same time avoids advanced nutrient removal, including phosphorus elimination and the nitrogen removing processes nitrification and denitrification (Maurer *et al.*, 2006). Urine separation relieves conventional wastewater treatment, facilitates the elimination of organic micropollutants originating from the human metabolism (Udert *et al.*, 2003 a)) and mitigates the negative impact of eutrophication in water bodies (Lienert & Larsen, 2009).

Moreover, urine is a natural resource, which is available in all human societies, even in the poorest ones (Heinonen-Tanski *et al.*, 2007).

Urine contains high amounts of nitrogen, phosphorus and potassium, which makes its use as a fertilizer in agriculture the most appropriate and obvious application (Maurer *et al.*, 2006). Since fertilizers have a market value, separate collection and treatment of urine has the potential to motivate private business initiatives, which help to promote sanitation in emerging and developing countries (Udert & Wächter, 2012). In addition, urine treatment is also essential regarding the prevention of the release of micropollutants into the environment (Udert *et al.*, 2006).

Apart from all the positive aspects of source-separated urine, some process aspects can be enhanced (Larsen & Gujer, 1996). Source-separated urine is a highly concentrated and unstable solution (Udert *et al.*, 2003). Fresh urine contains salts, soluble organic matter and urea. During storage, urine undergoes microbial contamination. Urea, the main component of urine, decomposes quickly (hydrolysis) into ammonia (NH_3) and carbonic acid (HCO_3^-). Ammonia is released into the bulk liquid and the pH increases strongly. At the same time, organic matter is degraded and malodorous compounds are produced (Maurer *et al.*, 2006). After storage, urine contains a large amount of ammonia, which can volatilise and be lost during transport or application as fertilizer (Udert *et al.*, 2003). This can be avoided by stabilisation of urine. Urine stabilisation can be performed by means of biological nitrification (Udert & Wächter, 2012). The urine stabilisation is essential to prevent both environmental pollution and negative effects on human health as well as to retain nitrogen in solution for later recovery (Udert *et al.*, 2006).

Biological nitrification is a microbial process by which reduced nitrogen compounds (primarily ammonia) are sequentially oxidized to nitrite (NO_2^-) and nitrate (NO_3^-), by ammonia oxidising bacteria (AOB) and nitrite oxidising bacteria (NOB), respectively. High concentrations in urine, mainly of salt, ammonia and nitrous acid (HNO_2), can inhibit biological nitrification and make the process sensitive to instabilities. A major concern is accumulation of the inhibitory intermediate NO_2^-, since it can lead to process breakdown (VUNA, 2013). In a well-functioning biological nitrification reactor the NO_2^- concentration is normally very low. However, under special circumstances enrichment of nitrite can be caused by disturbances in the microbiological processes. Inhibition due to toxic substances or to unfavourable conditions for the nitrite oxidizer (NOB) are also reported in conventional wastewater treatment plants (WWTP) (Rieger *et al.*, 2008).

Stable biological nitrate production requires the nitrite oxidation to be as fast as the ammonia oxidation. Thus, it is of major importance to monitor the nitrification process carefully. Monitoring of the NO_2^- concentrations is particularly crucial (Maurer *et al.*, 2006). Among the potential candidates for the development and application of on-line NO_2^--N measurements, spectroscopy is promising and could lead to interesting results (Paulo, 2008). Spectroscopy is the basis for non-invasive and non-destructive measuring systems (Pons *et al.*, 2004). Rieger *et al.*, 2008, demonstrate that for conventional wastewater systems measuring variables on-line with a single sensor will increase the monitoring capability and enable appropriate control of the process. Currently, there are several continuous sensors and analysers capable of operating on-line, among them the UV spectrophotometry sensors. UV spectrophotometry is promising for measuring NO_2^- concentrations in the urine nitrification reactor because nitrite dissolved in water

absorbs UV light mostly at wavelengths below 250 nm. This absorption makes it possible to have a photometric determination without reagents by using a sensor positioned in the media (Drolc & Vrtovsek, 2010).

Although the peak absorbance wavelengths in the UV range are well known for nitrite, the urine nitrification system causes a number of interferences. Turbidity is the most important influence on in-situ absorption, typically occurring due to suspended solids causing light scattering and shading, which influences absorption over the entire spectrum (Langergraber, 2003). Sample filtration has been proposed to reduce turbidity (Pons *et al.*, 2004). Interferences stemming from organic matter and other ions which absorb in the same wavelength range should also be taken into account (Rieger *et al.*, 2008). All these interferences could lead to unreliable measurement results (Drolc & Vrtovsek, 2010).

Spectroscopic methods together with chemometrics are being presented as a powerful tool for process monitoring and control of biotechnological processes. They combine a "hard" part, for sensing, and a "soft" part, for data treatment (Pons *et al.*, 2004). Models for on-line estimation address the problem of reconstructing the relationship existing between the process inputs (easy-to-measure variables) and the process outputs (hard-to-measure variables). Relationships regarding the input and output variables can be modelled by solving a regression problem. The simplest regression techniques assume the existence of a linear input-output relationship. For instance, they fit a linear model to reconstruct it by means of multivariate statistics (Haimi *et al.*, 2013). The term multivariate calibration refers to the process of constructing a mathematical model that relates the absorbances of a set of known reference samples at more than one wavelength to a property of the sample such as a concentration or identity (Drolc & Vrtovsek, 2010).

Multivariate modelling can provide an interpretable description of how the inputs affect the outputs. When large numbers of inputs exist, multivariate statistical methods combine reduction techniques and linear regression to reduce the dimensionality of the modelling problem. This usually makes the model easier to understand and increases the quality of the produced estimates. Commonly used methods include principal component regression (PCR) and partial least square (PLS) regression (Haimi *et al.*, 2013).

Principal components regression (PCR) is used when there is a large number of predictor variables and those predictors are highly correlated or even collinear. PCR constructs new predictor variables, known as components, as linear combinations of the original predictor variables (Maesschalck *et al.*, 1999). This method can be used to make data easier to understand by extracting relevant information and modelling it. PCR is typically used to deal with large amounts of data, such as spectral data. It makes use of data directly collected from the process to build an empirical model, providing graphical tools that are easy to apply and to interpret, making them very useful for real-time control and monitoring (Aguado & Rosen, 2008).

PCR is applied in this work as a method to create a model for the estimation of nitrite concentrations using an immersible UV-sensor and off-line spectra acquisition. The successful implementation of a nitrite estimation model in the nitrification reactor is promising in terms of process monitoring and prevention of process breakdown.

1.2 OBJECTIVES

The main goal of this thesis is to develop a chemometric PCR model for estimation of nitrite concentrations in a nitrification reactor treating stored source-separated urine. Estimation of nitrite concentrations in the urine nitrification reactor is vital for the operation of the same, since it could avoid system failures and, potentially, a total system breakdown. A PCR model for estimation of nitrate concentrations is developed as well.

A number of important factors that could affect the performance of the sensor and the associated PCR model are investigated. These are the effect of particles, which can cause back-scattering, and the effect of saturation, which could make linear models impractical.

Ultimately, the proposed work is expected to contribute to biological urine nitrification process stabilization.

1.3 STRUCTURE OF THE THESIS

The thesis includes the following chapters and contents:

1. Introduction
2. Literature review
3. Materials and methods
4. Results and discussion
5. Conclusions and future developments
6. References

Chapter 1 includes the background, objectives and the structure of the thesis. Chapter 2 provides an introduction to the concepts presented and discussed in this thesis. Additionally, it serves the purpose of reviewing the pertinent literature, and therefore, the main advances put forward in the last years. Thus, it attempts to create the necessary base for a constructive and supported discussion of the results found from the current work. The materials and methods applied are presented in Chapter 3. Chapter 4 reports the results and discussion for all experiments and was divided in two parts. The first part (4.1) includes the effects of saturation in UV absorbance spectra and the second part (4.2) discusses the effects of particles and saturation in UV absorbance spectra. Main conclusions and proposals for future work developments are presented in Chapter 5.

2 LITERATURE REVIEW

The most important concepts discussed in this thesis will be presented in the following paragraphs. In Section 2.1 a brief introduction to urine source separation technology is presented, followed by a description of the main biological processes involved in source-separated urine stabilisation (Section 2.2). Sections 2.3 and 2.4 introduce the UV spectroscopy fundamentals and chemometric principles by means of a review of the pertinent literature.

2.1 URINE SOURCE SEPARATION

Managing wastewater and the resources it contains is tremendously important for an ecologically sustainable development (Ganrot, 2005). Conventional wastewater treatment system functions well, at least in developed countries. However, this system always lags behind with respect to upcoming environmental problems (Corcoran *et al.*, 2010). In addition, adapting the same system, which is extremely wasteful of water and energy, is nearly impossible in poor and arid countries. This impracticality leads to devastating consequences for human health (Lienert & Larsen, 2009). Nutrient losses to the environment can be strongly reduced if nutrients from wastewater are recycled, for instance for agricultural applications (Udert, 2002).

One promising alternative to the current mixing of waste streams is the separate collection and treatment of urine. Separate collection of urine allows to recover valuable nutrients such as nitrogen, phosphorous and potassium. Subsequently, by keeping urine out of the sewage system, wastewater treatment plants would reportedly receive 80 percent less nitrogen and 50 percent less phosphorous. Hence, the recovered nutrients can be used as fertiliser in agriculture (VUNA, 2013). An additional advantage of urine separation could be the elimination of micropollutants, mainly excreted with urine (Udert, 2002).

Urine source separation comes in several variations. The main principle is to always divert urine from faeces in the toilet bowl by a separate outlet. In urine-collecting systems the separated urine is directed to a collection tank with or without flushing water and stored for later transport to a treatment facility (Lienert & Larsen, 2009). Microbial urea hydrolysis occurs in the collection tank. The present work addresses a setting with water-flushed urine diverting NoMix-technology applied at Eawag, Switzerland. According to Lienert & Larsen, (2009), NoMix-technology is a promising innovation aiming at a decentralized approach in urban wastewater. Sweden is a pioneering country regarding the above technology, followed by other Northern and Central European countries and more recently Australia and China, where several urine-diverting toilets have been installed and successfully tested.

During urine storage both ammonia concentration and pH increase due to urea degradation. High pH values cause ammonia volatilisation leading to odour problems and nitrogen losses during handling of stored urine. Therefore, source-separated urine needs to be pre-treated in order to stabilise it. Urine stabilisation can be achieved by means of biological nitrification.

A brief explanation of the main biological processes involved in the collection and treatment of source-separated urine is introduced in Section 2.2.

2.2 BIOLOGICAL PROCESSES

2.2.1 UREA HYDROLYSIS: UREOLYSIS

Microbial urea hydrolysis, ureolysis, corresponds to the degradation process of urea, the main component of urine (Udert, 2002). The enzyme urease, hydrolyses urea to ammonia and carbonate (Maurer *et al.*, 2006). Carbonate decomposes spontaneously to carbonic acid and ammonia (Udert & Wächter, 2012). The overall reaction can be written as

$$CO(NH_2)_2 + 2\,H_2O \xrightarrow{\text{urease}} 2NH_3 + H_2CO_3 \tag{2.1}$$
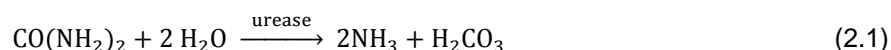
During urea degradation, ammonia (NH_3) is released and the pH increases strongly (Udert *et al.*, 2003). At the same time, anaerobic bacteria produce malodorous compounds, which are responsible for the characteristic strong smell of stored urine (Maurer *et al.*, 2006). In addition, urea degradation is a very fast process. All urea is hydrolysed after little more than one day of retention time. Under these conditions large amounts of phosphate minerals precipitate, which can lead to pipe blockages (Udert, 2002).

Thus, urine stabilisation is essential to prevent both environmental pollution and negative effects on human health. In addition, it retains nitrogen in solution for later recovery (Udert *et al.*, 2006). Biological nitrification is a method for preventing the ammonia volatilisation from stored urine (urine stabilisation) and will be discussed in detail in Section 2.2.2. In addition to the nitrogen stabilisation, the biological treatment of urine has further beneficial effects: 80% of the organic compounds (based on chemical oxygen demand) are degraded and the unpleasant odour is eliminated (Udert, 2002).

2.2.2 BIOLOGICAL NITRIFICATION

Nitrification is the biological oxidation of ammonia to nitrate with nitrite formation as an intermediate (Sedlak, 1991). The nitrification process is primarily accomplished by two groups of autotrophic nitrifying bacteria that can build organic molecules using energy obtained from nitrogen containing inorganic sources, in this case ammonia and nitrite (Bock & Wagner, 2006).

In the first step of nitrification, ammonia-oxidising bacteria (AOB) oxidize ammonia to nitrite according to Equation 2.2.

$$NH_3 + \frac{3}{2}\,O_2 \rightarrow NO_2^- + 3H^+ + H_2O \tag{2.2}$$
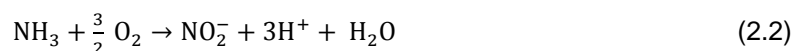
Nitrosomonas is the most frequently identified genus associated with this step, although other genera, including *Nitrosococcus* and *Nitrosospira* can be present (EPA, 2002). In the second step of the process, nitrite-oxidising bacteria (NOB) oxidize nitrite to nitrate according to Equation 2.3.

$$NO_2^- + \frac{1}{2}\,O_2 \rightarrow NO_3^- \tag{2.3}$$

Nitrobacter is the most frequently identified genus associated with this second step, although other genera, including *Nitrospina, Nitrococcus* and *Nitrosospira* can also autotrophically oxidize nitrite (EPA, 2002). Both groups of nitrifying bacteria are obligate aerobic organisms commonly found in terrestrial and aquatic environments. Their growth rates are controlled by substrate (ammonia and nitrite) concentrations, temperature, pH, light, oxygen concentrations and microbiological community composition (EPA, 2002).

Owing to stored urine alkalinity, nitrifying bacteria oxidise just half of the NH_3 to non-volatile nitrate (NO_3^- ion). Ammonia-oxidising bacteria (AOB) convert ammonia to nitrite (NO_2^-), which is the substrate for the second group of bacteria, the nitrite-oxidising bacteria (NOB) that produces nitrate (NO_3^-). The remaining NH_3 is converted to non-volatile ammonium (NH_4^+ ion), due to the release of protons (VUNA, 2013). Nevertheless, almost all ammonium can be oxidized to nitrate, by adding additional alkalinity. Additional alkalinity leads to an equilibrium shift (Udert & Wächter, 2012).

The activities of AOB and NOB are strongly dependent on the pH as the concentration of their substrates, NH_3 (AOB) and HNO_2 (NOB) are in a pH dependent equilibrium with their acid (NH_4^+) or base (NO_2^-) (Udert, 2002). As nitrite is an intermediate product of the two-step process of nitrification and the second step of nitrification is rather fast, the nitrite concentration in a steady biological nitrification is normally very low. Enrichment of nitrite in the system is a major concern and usually suggests that the microbiological process is disturbed/inhibited. Extreme events could lead to an increased influent to the nitrification reactor, increasing NH_3 concentration and consequently pH values. Due to higher availability of NH_3 at higher pH values, AOB activity increases immediately. NOB reacts too slowly to the rising nitrite concentration. As a consequence, nitrite keeps increasing and the HNO_2 concentration approaches values that could cause complete NOB inhibition and lead to process breakdown (Udert & Wächter, 2012). Thus, it is imperative for the well-functioning of the biological nitrification process to monitor the nitrite concentrations.

2.3 ULTRAVIOLET (UV) SPECTROSCOPY

Among the potential candidates for the development and application of on-line measurements, spectroscopy is very promising, since it can be the basis for non-invasive and non-destructive measuring systems (Pons *et al.*, 2004). The use of submersible equipment which can perform a spectra analysis directly in liquid media for determination of several parameters in the effluent of a conventional WWTP such as nitrite and nitrate, has been successfully applied using the UV spectra range 200-400 nm (Rieger *et al.*, 2004).

Spectroscopic processes rely on the fact that electromagnetic radiation (EMR) interacts with atoms and molecules in discrete ways to produce characteristic absorption or emission profiles (Thomas & Theraulaz, 2007). EMR is energy that is propagated through free space or through a material medium in the form of electromagnetic waves. The various types of radiation can be defined in terms of their wave frequency and when EMR is spread out according to its wavelength, the result is a spectrum (Thomas *et al.*, 1996). The types of EMR that make up the electromagnetic spectrum are presented in Figure 2.1.
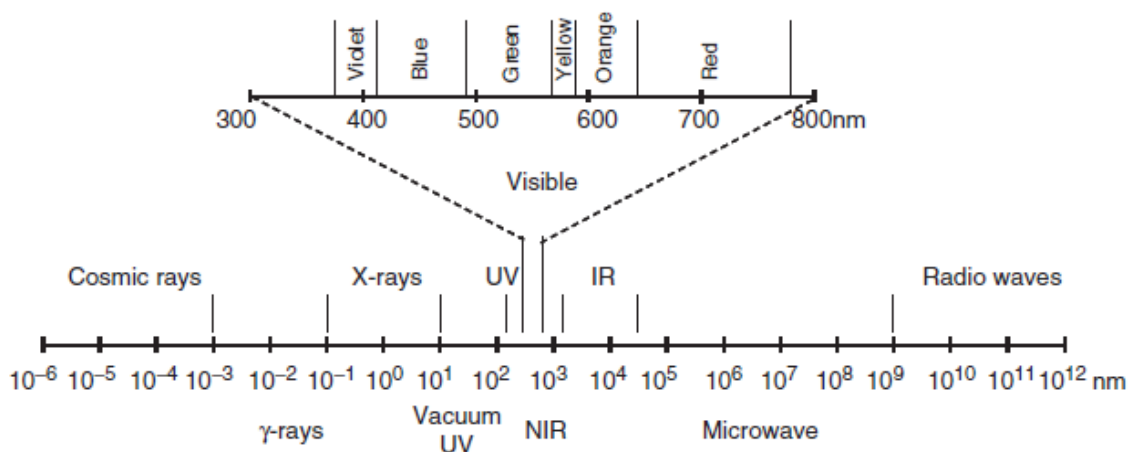
Figure 2.1. Electromagnetic spectrum (Thomas & Theraulaz, 2007)

UV spectroscopy is the study of how a sample responds to ultraviolet light. When a beam of light passes through a substance or a solution, some of the light may be absorbed and the remainder transmitted through the sample. The most important principle in absorption analysis is the Lambert–Beer's law (Upstone, 2000). This law states that, for a given ideal solution, there is a linear relation between absorption and concentration of a single determinant (Langergraber et al., 2003). Thus UV spectroscopic techniques used for quantifying purposes are based on the mentioned law. According to Lambert-Beer's law, for a single wavelength and a single component, absorbance is a linear function of the concentration of the component and the following relation (Equation 2.4) is valid:

$$A = \varepsilon.b.c \tag{2.4}$$

where A is the absorbance, ε is the molar absorptivity ($mol^{-1}.cm^{-1}$), which is constant for each chemical compound and for each wavelength, b is the path length of the cell in which the sample is contained (cm) and c is the concentration of the absorber ($mol.dm^{-3}$).

Provided that ε and b are kept constant for a given set of experiments, a plot of the sample absorbance against the concentration of the absorbing substance should give a linear calibration curve going through the origin. In addition, Lambert–Beer's law allows to change the path length to affect the absorbance. The path length can be increased or reduced when lower or higher limits are required, respectively. Alternatively, it is possible to reduce the absorbance by diluting the sample (Upstone, 2000).

The Lambert-Beer's law describes a linear relationship between sample concentration and absorbance. However, this relationship is based on a number of assumptions, including that radiation is perfectly monochromatic, there are no uncompensated losses due to scattering or reflection, there are no molecular interactions between the absorber and other molecules in solution and the temperature remains constant. In practice, these assumptions are not always met which causes deviations from the ideal Lambert–Beer's law behaviour (Burgess, 2007). Langergraber et al., (2003), and Drolc & Vrtovsek, (2010), report that wastewater monitoring has

to deal with superposition of numerous single substance absorbances, sometimes even with overlapping peaks, due to numerous dissolved and suspended compounds, which can lead to poor performance of the UV sensor. In urine, saturation and turbidity due to high nitrate concentrations and suspended particles, respectively, are additional factors influencing the measurement. The latter effects could be minimized by dilution and filtration, respectively. Previous studies of a saturation effect in a urine nitrification reactor have not been reported. An alternative method to reduce saturation is to decrease the path length of the sensor. However, for the studied case the applied path length (0.5 mm) cannot be minimized further.

2.4 CHEMOMETRICS

The term chemometrics was introduced in 1972 by Svante Wold and Bruce Kowalski and has its foundation as a discipline in chemistry. In chemometrics the main issue is to structure the chemical problem to a form that can be expressed as a mathematical relation, by means of a statistical–mathematical method (Otto, 2007). According to Svante Wold, a reasonable definition of chemometrics is 'How to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data' and the only reasonable way to extract and represent this chemical information is in terms of models (Wold, 1995).

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 background complexity can complicate the direct identification of individual compounds. Therefore, data-reduction techniques such as chemometric tools are essential to rapidly extract the relevant information, presenting the data in a more clear way. Indirect chemometric models are used in wastewater for correlating the concentrations of the required parameters to spectral information.

Multivariate statistical methods of data analysis are applied to find patterns in data and to distinguish those patterns in the samples. Recognition of patterns is feasible with projection methods, such as principal component analysis (PCA) and principal component regression (PCR) (Otto, 2007). These tools are usually used to deal with large amounts of data, such as spectral data. Both methods can be used to make data more understandable by extracting relevant information. These methods make use of data directly collected from the process to build an empirical model which serves as reference of the desired process behaviour and against which new data can be compared. Additionally, they provide graphical tools that are easy to apply and to interpret (Aguado & Rosen, 2008).

Both the PCA and PCR models are linear. The unknown variables are linear functions of the known ones (Hyötyniemi, 2001). Linear models are simple and often provide an adequate and interpretable description of how the inputs affect the outputs. For prediction purposes they can sometimes outperform fancier non-linear models, especially in situations with a small number of training cases (Hastie *et al.*, 2001).

PCA is the multivariate statistical method most frequently used for environmental data analysis. It has already been applied in some cases of wastewater monitoring (Pons *et al.*, 2004; Aguado

& Rosen, 2008). The main purpose of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables to a considerably smaller number, while retaining as much as possible of the variation present in the data set (maximum variance). This reduction is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables (Jolliffe, 2002). PCA is a popular technique to find patterns in data of high dimension and expressing the data in such a way as to highlight their similarities and differences (Smith, 2002). PCA models can be computed efficiently by singular value decomposition (SVD).

Multivariate regression models like principal component regression (PCR) and partial least squares (PLS) regression are very popular in a wide range of fields. The main reason for this is that they have been designed to confront the situation that there are many, possibly related, predictor variables, and relatively few samples (Mevik & Wehrens, 2007). The use of UV/Vis spectroscopy in combination with multivariate methods was reported to be a useful tool for correlating the concentrations of parameters to spectral information. Several studies were developed for nitrite and nitrate prediction. Partial least squares (PLS) regression based on UV/Vis spectra for prediction of nitrate and organic carbon in groundwater was reported as very promising (Dahlen *et al.*, 2000). Rieger *et al.*, 2004, demonstrates that an in-situ UV spectrometer together with a multivariate calibration based on PLS was successfully applied to nitrite and nitrate concentrations prediction in a WWTP. Similar results were obtained by Langergraber *et al.*, 2003, Bouvier *et al.*, 2008, and more recently by Drolc & Vrtovsek, 2010.

3 MATERIALS AND METHODS

3.1 RESEARCH STRATEGY

In order to achieve the central objective of this thesis (defined in Section 1.2), the research strategy developed was divided in 2 main steps.

An evaluation of both particles and saturation effects in the UV absorbance spectra was performed. The evaluation of particles effect aims to clarify how strong the effect of particles is and whether a reasonable estimation of nitrite and nitrate concentrations remains feasible in the presence of particles. The target of the saturation effect experiment is to evaluate how the presence of saturation influences the estimation of nitrite and nitrate concentrations.

Prior to the evaluation of the effects of particles and saturation in the UV absorbance spectra, there is a need to find an **ideal dilution** – a particular dilution to ensure that no saturation is present. This ensures that the best conditions are met for the identification of a linear model. Thus, an experiment to characterize the saturation effect empirically was carried out. On the basis of this experiment, a dilution was selected which guarantees the absence of saturation.

Thus, the research strategy can be described in 2 main experiments:

I. **Primary evaluation of saturation effect in the UV absorbance spectra** on the collected spectra data through **dilution experiments**.
UV spectral measurements of diluted treated urine samples (10 different dilutions) were performed. The collected data were evaluated with linear models in order to determine a reference dilution for the following experiments. The main goal of this step was to find a dilution which guarantees that the saturation effect disappears.

II. Investigate the **effect of particles and the effect of saturation in the UV absorbance spectra**.
UV spectral measurements of treated urine samples, with 3 different types of filtration (0.7µm filter, coffee filter and unfiltered), diluted and non-diluted, with and without nitrite stock-solution addition, were performed. The applied dilution was based on the results of the previous experiment.
The collected data were evaluated with a multivariate model (PCR), in order to evaluate how strong the effect of particles and the effect of saturation are in the referred measurements and to clarify if a reasonable estimation of NO_2^- and NO_3^- concentrations is achievable in the presence of particles and saturation.

3.2 MATERIALS

3.2.1 URINE SAMPLES

The urine samples for the experiments are collected from the nitrification reactor in the Eawag main building (Forum Chriesbach). The stored source-separated urine, collected within the building, is pre-treated in the nitrification reactor in order to stabilise the urine. The reactor is a moving bed biofilm reactor (MBBR) with a volume of 120 L and Kaldnes® K1 biofilm carriers to support the growth of nitrifying bacteria. The Kaldnes® biofilm carriers are polyethylene tubes with a diameter of 9 mm and a length of 7 mm. Air is supplied by a bubble aeration system (Etter *et al.*, 2013; Alexandra Fumasoli, personal communication, July 21, 2014).

3.2.2 UV SPECTROMETER

The spectro::lyser^{TM} UV, a spectrometer probe manufactured by s::can, is a submersible spectrometer of 44 mm diameter and about 0.6 m length, capable of online measurements of UV absorption spectra directly in liquid media (in-situ), without sampling or sample treatment and without reagents. The sensor provides measurements of spectra in the wavelength region between 220 nm and 390 nm (171-dimensional spectra) and displays and/or communicates the result in real time. The measuring path is 0.5 mm which is unusual for UV spectrophotometry in wastewater processes. Figure 3.1 shows a picture of the applied sensor. A single measurement takes about 45 seconds (s::can, 2007; Langergraber *et al.*, 2003).

The probe consists of three main components: the emitter, measuring cell and receiving unit, with a xenon flash lamp as a light source. The measuring section includes (1) optical measuring path, (2) cleaning nozzles and (3) fixtures for the measuring path (Figure 3.1). In the measuring path (1) the light passes through the space between the two measuring windows which is filled with the medium. A second light beam within the probe – compensation beam – is guided across an internal comparison section and performs as a reference.
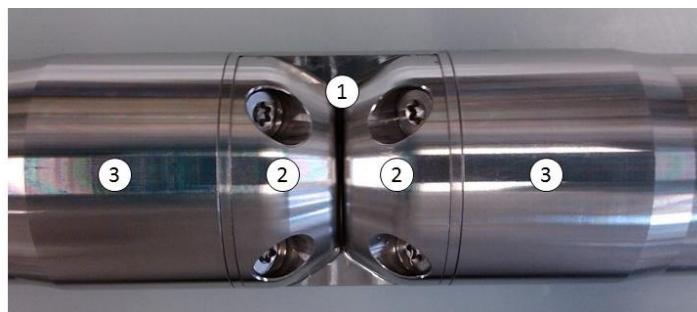
Figure 3.1. s::can spectrometer probe - measuring section

3.2.3 REFERENCE TESTS: Hach-Lange Cuvette Tests

Hach-Lange cuvette tests are colorimetric tests that allow identification of the concentrations of the compounds of a sample, by means of barcoded cuvettes. A Hach DR2800 spectrophotometer is used to automatically identify the cuvette test from its barcode, rotating the cuvette and taking 10 measurements, eliminating outliers, and showing the measurement result in mg/l (HACH, 2012).

The Hach-Lange tests used are LCK 303, LCK 341/342 and LCK 340 for ammonia, nitrite and nitrate concentrations, respectively (Table 3.1).

Table 3.1. Reference tests used for determination of ammonium, nitrite and nitrate concentrations in the collected samples

TESTED COMPOUND	HACH-LANGE TEST	MEASURING RANGE	METHOD
NH_4^+-N	LCK 303	2 - 47 mg/l NH_4^+-N	Indophenol Blue[1]
NO_2^--N	LCK 341 LCK 342	0.015 - 0.6 mg/l NO_2^--N 0.6 – 6.0 mg/l NO_2^--N	Diazotisation[2]
NO_3^--N	LCK 340	5 - 35 mg/l NO_3^--N	2.6-Dimethylphenol[3]

3.2.4 FILTERS

Glass Fiber Filters:

Glass fiber filters are widely applicable in many fields of laboratory use due to the exceptionally good resistance to most organic and inorganic solvents. They allow a fast filtration with simultaneous high particle retention and good loading capacity (Macherey-Nagel, 2014). The general characteristics of the used micro filters are presented in Table 3.2.

Table 3.2. Glass microfiber filters used in experiments

GLASS FIBER FILTER PAPER GRADES	AVERAGE RETENTION CAPACITY (µm)	WEIGHT (g.m^{-2})	DIAMETER (mm)	THICKNESS (mm)	FILTRATION SPEED (s)
MN GF-1	0.7	55	90	0.3	12
MN GF-4	1.4	120	90	0.6	5
MN GF-5	0.4	85	47	0.4	80

---

[1] Ammonium ions react at pH 12.6 with hypochlorite ions and salicylate ions in the presence of sodium nitroprusside as a catalyst to form indophenol blue

[2] Ammonium ions react at pH 12.6 with hypochlorite ions and salicylate ions in the presence of sodium nitroprusside as a catalyst to form indophenol blue

[3] Nitrate ions in solutions containing sulphuric and phosphoric acids react with 2.6-dimethylphenol to form 4-nitro-2.6-dimethylphenol.

Coffee Filters:

Melitta® Original Coffee Filters #4 and #6 made from natural brown paper with multi-layer filtration separated into 3 zones (Figure 3.2), each with a different number of perforations (Melitta®, 2014) are used to remove a majority of the suspended particles from the samples.
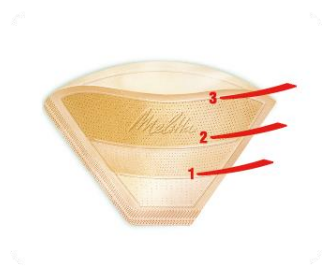
Figure 3.2. Melitta® Coffee filter

3.2.5 REAGENTS

Ethanol is of HPLC gradient grade, with a purity of ≥ 99.8%, according to the supplier (Sigma-Aldrich Chemie GmbH, Buchs, Switzerland). Sodium hydroxide (Sigma-Aldrich Chemie GmbH, Buchs, Switzerland) and Hydrochloric acid (Sigma-Aldrich Chemie GmbH, Buchs, Switzerland) are reagent grade of 50% and 37%, respectively. Sodium nitrite (Merck Millipore, Darmstadt, Germany) is of extra pure Ph Eur, USP. Detailed product information can be found under CAS number 7632-00-0. The laboratory water used in the study is provided by the Thermo Scientific™ Barnstead™ NanoPure™ system (Thermo Fisher Scientific, Basel, USA).

3.3 EXPERIMENTAL PROCEDURE

In order to accomplish the central objective of this thesis (defined in Section 1.2), the research strategy (defined in Section 3.1) is divided in 2 main experiments and explained in detail in Sections 3.3.1 and 3.3.2.

3.3.1 PRIMARY SATURATION EFFECT EXPERIMENT

The goal of the first experiment is to determine a reference dilution for the following experiments and it is schematically represented in Figure 3.3. The experiment was carried out in the Eawag laboratory.
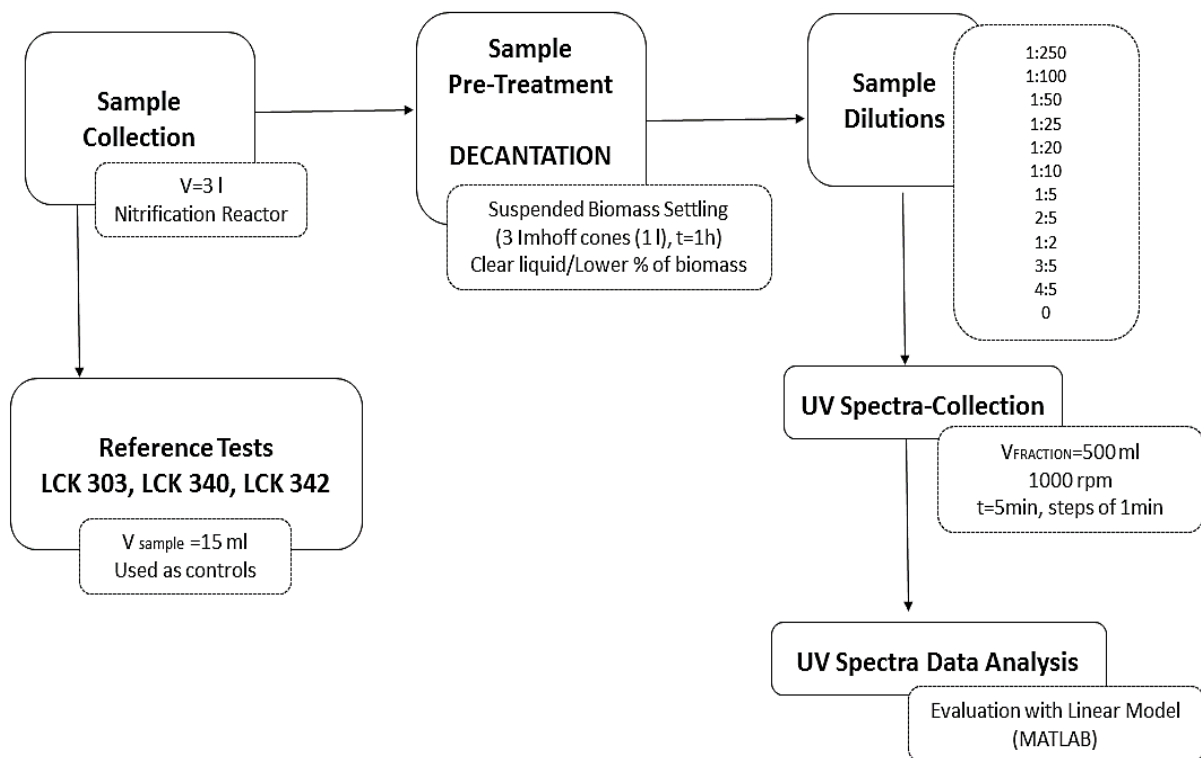
Figure 3.3. Schematic diagram of Saturation Effect Experiment

3.3.1.1 Sample Collection and Pre-treatment (Decantation)

Urine samples measuring at least three liter (3 l) are collected from the nitrification reactor (Figure 3.4 (a) and (b)) in the Eawag main building (Forum Chriesbach) and carried to the Eawag laboratory in a plastic container (Figure 3.4 (c)). The volumetric fraction of the suspended solids, which is approximately 750 ml, settles relatively fast and is removed by sedimentation and decantation. To this end, the urine sample is distributed over 4 Imhoff cones of 1 l (Figure 3.4 (d)). Following this, the supernatant was collected from the top of the Imhoff cones for further analysis.

By following this procedure, effects of improper mixing during spectral measurements can be minimized.

Figure 3.4. Sample Collection and Decantation steps: (a) Nitrification reactor; (b) Urine sampling (c) Urine plastic container; (d) Imhoff cones

3.3.1.2 Sample Dilutions

After the decantation step, several diluted samples are prepared from the supernatant. The dilutions used in each experiment are represented in Table 3.3.

Table 3.3. Sample dilutions applied for UV-spectra measurements

SAMPLE	SAMPLE DATE (dd.mm.yy)	SAMPLE DILUTION	V_{URINE} (ml)	V_{H2O} (ml)	V_{TOTAL} (ml)	URINE FRACTION (%)
01	03.04.14	1:100	5	495	500	1
02	02.04.14	1:50	10	490	500	2
03	03.04.14	1:20	25	475	500	5
04a	02.04.14	1:10	50	450	500	10
04b	03.04.14	1:10	50	450	500	10
05	03.04.14	1:5	100	400	500	20
06	04.04.14	2:5	200	300	500	40
07	02.04.14	1:2	250	250	500	50
08	04.04.14	3:5	300	200	500	60
09	04.04.14	4:5	400	100	500	80
10	02.04.14	0	500	0	500	100
01.1	09.04.14	1:100	5	495	500	1
02.1	09.04.14	1:50	10	490	500	2
03.1	09.04.14	1:25	20	480	500	4
04.1	09.04.14	1:20	25	475	500	5
05.1	09.04.14	1:10	50	450	500	10
06.1	09.04.14	1:5	100	400	500	20
07.1	09.04.14	2.5	200	300	500	40
08.1	09.04.14	3:5	300	200	500	60
09.1	09.04.14	4:5	400	100	500	80
10.1	09.04.14	0	500	0	500	100

3.3.1.3 UV-Spectra Collection

Experiments are carried out in a 1000 ml graduated glass cylinder placed on a magnetic agitator at 1000 rpm. 500 ml of sample and a stirring magnet were added to the graduated cylinder. The UV probe was immersed in the media until the volume of the graduated cylinder read 900 ml. The stirrer was started at 1000 rpm. Absorbance spectra were collected using the s::can UV sensor during 5 minutes, with a measurement interval of one (1) minute. The above procedure is repeated for each sample and each dilution. Between sets of measurements (5 spectra) the UV probe was cleaned with 20% NaOH solution, 15% HCl solution, EtOH and nanopure water, in this order, to avoid deposition of biomass particles and salts in the measuring path.

3.3.1.4 Chemical Analysis: Ammonium, nitrite and nitrate determination

Hach-Lange tests of the collected samples are used for reference in this study. A 15 ml of the collected supernatant sample was filtered through a glass microfiber filter MN GF-5 (Macherey-Nagel AG, Switzerland) to remove suspended particles (Figure 3.5 (a)). Thereafter, an additional dilution (50x) of the sample is performed in order to work in the range of the Hach-Lange test. Stable operation of the reactor allowed the use of the same dilution during all experiments. Ammonium, nitrite and nitrate concentrations were measured by means of Hach-Lange tests LCK 303, LCK 342 and LCK 340, respectively (Figure 3.5 (b) and (c), Table 3.4).
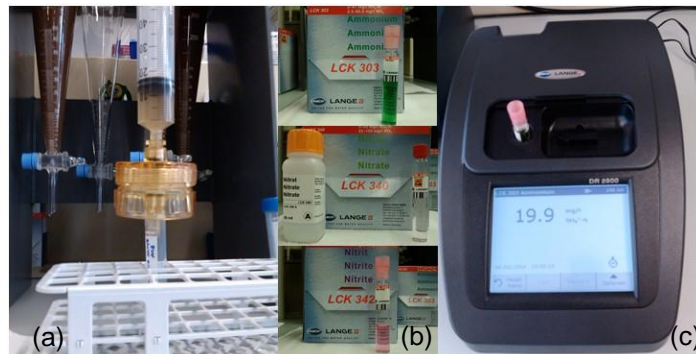
Figure 3.5. Ammonium, nitrite and nitrate determination by means of Hach-Lange reference tests
(a) Sample microfiltration; (b) Hach-Lange reference tests; (c) Hach DR2800 spectrophotometer

Table 3.4. Reference tests used for determination of ammonium, nitrite and nitrate
concentrations in the collected samples of the saturation effect experiments

TESTED COMPOUND	HACH-LANGE TEST	SAMPLE DILUTION	MEASURED CONCENTRATION
NH_4^+-N	LCK 303	1:50	mg/l NH_4^+-N
NO_2^--N	LCK 342	1:1	mg/l NO_2^--N
NO_3^--N	LCK 340	1:50	mg/l NO_3^- -N

3.3.2 PARTICLES AND SATURATION EFFECT EXPERIMENT

The second experiment aims to evaluate the effect of particles and the effect of saturation in the UV-spectra. The experiment was carried out in the Eawag laboratory and is schematically represented in Figure 3.6. Samples for this experiment were collected 3 times a week, during 16 weeks.
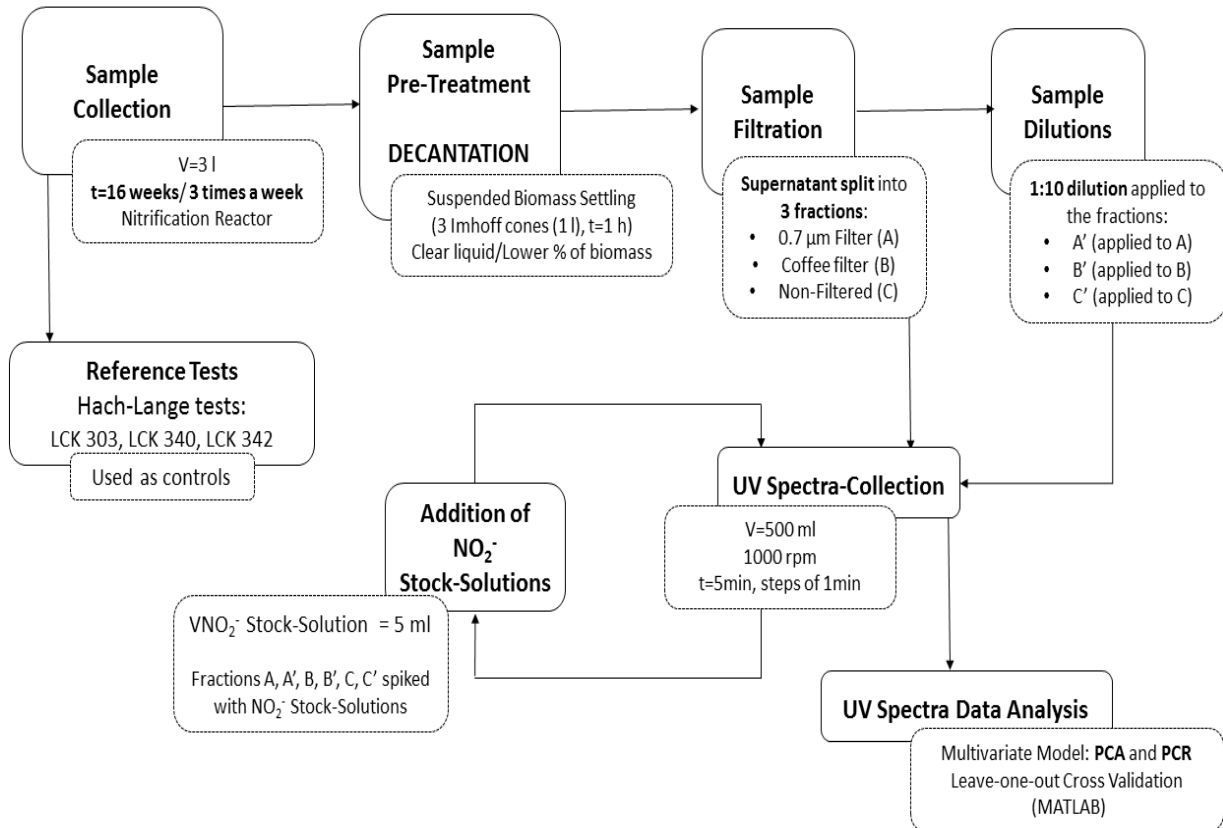
Figure 3.6. Schematic diagram of Particles Effect Experiment

3.3.2.1 Sample Collection and Pre-treatment (Decantation)

The detailed explanation of this step is described above in Section 3.3.1.1.

3.3.2.2 Sample Filtration

After the decantation step, the remaining liquid is split in 3 fractions (A, B and C, V=800 ml each). Fraction A is used as is without filtration. Fractions B and C are subjected to a filtration step. Fraction B is filtered through a coffee filter to remove a majority of suspended particles and collected in an 800 ml beaker. Fraction C is initially filtered through a coffee filter. Subsequently, the sample is filtered through a glass microfiber filter MN GF-4 (pore size 1.4µm), followed by

filtration through a glass fiber filter MN GF-1 (pore size 0.7μm) to remove suspended particles and collected in an 800 ml beaker. The apparatus used for filtrations is presented in Figure 3.7.

(a) (b)

Figure 3.7. Sample Filtration: (a) Coffee filtration and (b) Microfiltration

3.3.2.3 Sample Preparation/Dilution

Fractions A, B and C are each collected undiluted by means of a 500 ml volumetric flask. The remaining liquid of each fraction is used to prepare diluted fractions (A', B', C'). The dilutions performed are based on the results of the previous experiment. Fraction A', B', C' are collected by means of a 500 ml volumetric flask as well (Figure 3.8).

Figure 3.8. Non-diluted and 10x diluted samples

3.3.2.4 Preparation of Nitrite Stock-Solutions

Three different nitrite (NO_2^-) stock solutions, V=500 ml, are prepared by dissolving sodium nitrite ($NaNO_2$) in water. The target concentrations in the sample fractions analysed were 50 mg NO_2^--N/l, 100 mg NO_2^--N/l and 150 mg NO_2^--N/l.

3.3.2.5 UV-Spectra Collection and Analysis

Experiments are carried out in 1000 ml graduated glass cylinder placed on a magnetic agitator at 1000 rpm. Sample fractions (A, A', B, B', C, C', Table 3.5) are added to the graduated cylinder together with a stirring magnet. The UV probe is submerged in the media until the volume of the graduated cylinder reads 900 ml. The stirrer is started. Measurements are performed during 5 minutes, in steps of one (1) minute. Between sets of measurements, the UV probe is cleaned with 20% NaOH solution, 15% HCl solution, EtOH and nanopure water. All samples are also spiked with 5 ml of NO_2^- stock solution to ensure measurements of a wide range of NO_2^- concentrations. The concentrations of the added NO_2^- stock solution are 50 mg NO_2^--N/l, 100 mg NO_2^--N/l and 150 mg NO_2^--N/l, for each of the 3 days of the week, respectively. To save time, no cleaning is executed before NO_2^- stock solution is added. The apparatus used is presented in Figure 3.9.
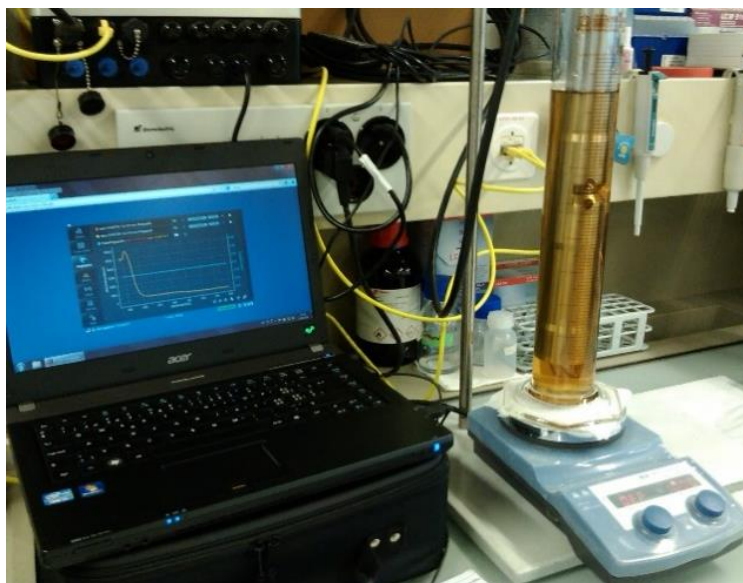
Figure 3.9. Data collection

Table 3.5. Sample fractions used for UV-spectra measurements

SAMPLE FRACTION		TYPE OF FILTRATION	SAMPLE DILUTION (times)	V_{URINE} (ml)	VH_2O (ml)	$V_{NO_2^- \ STOCK \ SOLUTION}$ (ml)	V_{TOTAL} (ml)
A	01	0.7 µm Filter	0	500	0	0	500
	02	0.7 µm Filter	0	500	0	5	505
A'	03	0.7 µm Filter	10	50	450	0	500
	04	0.7 µm Filter	10	50	450	5	505
B	05	Coffee filter	0	500	0	0	500
	06	Coffee filter	0	500	0	5	505
B'	07	Coffee filter	10	50	450	0	500
	08	Coffee filter	10	50	450	5	505
C	09	Non-filtered	0	500	0	0	500
	10	Non-filtered	0	500	0	5	505
C'	11	Non-filtered	10	50	450	0	500
	12	Non-filtered	10	50	450	5	505

3.3.2.6 Chemical Analysis: Ammonium, nitrite and nitrate determination

The detailed explanation of this step is described above in Section 3.3.1.4.

3.4 MODELLING

In the current Section the methods used for data analysis will be introduced. In Section 3.4.1 the required steps to apply a linear regression on the acquired data of the first experiment are explained in detail. Section 3.4.2 presents the methods used to build a PCR model for estimation of nitrite concentrations, based on the acquired data of the second experiment.

3.4.1 PRIMARY SATURATION EFFECT EXPERIMENT

The obtained UV-spectra are analysed using Matlab®, by means of a linear regression (Matlab code is presented in Appendix I).

Consider a vector of inputs x and the desired estimation of an output y, where x correspond to the urine fractions and y to the UV absorbances. The linear regression model has the form:

$$f(x) = \beta_0 + \sum_{j=1}^{p} x_j \beta_j \, , \tag{3.1}$$

where β_0 and β_j's are unknown parameters to be estimated - intercept and slope.

The most popular estimation method is least squares, in which we pick the coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ to minimize the sum of squares of residuals (SSR).

$$SSR(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 . \tag{3.2}$$

The solution is easiest to characterize in matrix notation, given by:

$$SSR(\beta) = (y - X\beta)^T (y - X\beta), \qquad X = [1 \ \ X]$$

where X is a *N x p* matrix with each row an input and y an N-vector of the outputs of the calibration set. A column of ones is added to the X matrix for determination of β_0 . Differentiating the above equation with respect to β, gives

$$\frac{\partial SSR}{\partial \beta} = -2X^T (y - X\beta). \tag{3.3}$$

Setting the first derivative to zero

$$X^T (y - X\beta) = 0 \tag{3.4}$$

the estimated values for the intercept and the slope are obtained as follows:

$$\widehat{\beta} = (X^T X)^{-1} X^T y. \tag{3.5}$$

26

The fitted values for the training inputs (predicted UV absorbances) are given by:

$$\hat{y} = X\hat{\beta} = X\,(X^T X)^{-1} X^T y. \tag{3.6}$$

After fitting the data with a linear model, there is a need to evaluate the accuracy of the model. In the present work residuals allow to measure the total deviation of the predicted absorbances from the measured absorbances. The prediction error is computed as:

$$e_i = y_i - \hat{y}_i\,, \tag{3.7}$$

where, y_i is the measured absorbance, $\hat{y}_i$ is the predicted absorbance, e_i is the residual error.

3.4.2 PARTICLES AND SATURATION EFFECT EXPERIMENT

The collected data are evaluated firstly with principal component analysis (PCA) followed by multivariate regression (PCR) in order to build an estimation model for nitrite concentrations. A leave-one-out cross validation (LOOCV) is performed as well, to choose the best number of principal components (PCs) for the PCR model (Matlab code is presented in Appendix II).

3.4.2.1 Data Pre-Processing

Consider that the collected data are represented as an *m x n* matrix, X, where the n columns are the samples (observations) and the m rows are the variables (wavelengths).

$$X = \begin{bmatrix} x_{1.1} & \cdots & x_{1.n} \\ \cdots & \cdots & \cdots \\ x_{m.1} & \cdots & x_{m.n} \end{bmatrix} \tag{3.8}$$

Prior to data analysis, the absorbance measurements are centred. Mean centring is commonly applied for any multivariate calibration model. This involves calculating the average for each column in the training data set and then subtracting the result from each element of that column. This operation ensures that results are interpretable in terms of variation around the mean (Maesschalck *et al.*, 1999). The mean-centring operation is given by:

$$\mathrm{X_{i,cent}} = \mathrm{X_i} - \overline{\mathrm{X}}, \tag{3.9}$$

where, $X_{i,cent}$ is the centred value of X_i and $\bar{X}$ the mean value of X.

After mean centring, the matrix is scaled as follows:

$$Y = \frac{X_{i,cent}}{\sqrt{(n-1)}}$$

(3.10)

3.4.2.2 PCA Calibration: Singular Value Decomposition (SVD)

The n-dimensional spectra are reduced to a lower number of principal components (PCs), by means of SVD. Y can be rewritten as follows:

$$Y = U * S * V^T$$

(3.11)

where U and V are (m x r) and (n x r) orthogonal matrices, S is a (r x r) diagonal matrix which contains the variance described by each singular value and r is the rank of Y (Jolliffe, 2002). The diagonal entries $d_1 \geq d_2 \geq \ldots \geq d_m \geq 0$, are called the singular values of Y (Hastie *et al.*, 2001). The eigenvalues of the covariance matrix of Y are equal to the singular values of the matrix Y and are found on the diagonal of S.

3.4.2.3 PCA Application

The main purpose of PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables to a much smaller number, while retaining as much as possible of the variation present in the data set (maximum variance). The PCs are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables. In doing so, the measured variables (absorbances at different wavelengths) are converted into new variables called scores.

The score matrix T (n x r) containing the PCs can be calculated as

$$T = U * S = Y * V$$

(3.12)

3.4.2.4 PCR Calibration and Application

In the present work PCA is used as a precursor for a predictive model (PCR). The PCs with the higher variance are the ones chosen to perform the linear regression. A PCR is then made to estimate nitrite concentrations by means of the selected PCs in the PCA model. The estimated values for the intercept and the slope are obtained as follows:

$$\widehat{\beta} = (T^T T)^{-1} T^T y.$$

(3.13)

The fitted values of the training inputs (predicted UV absorbances) are given by:

$$\hat{y} = T\hat{\beta} = T\ (T^TT)^{-1}T^Ty. \tag{3.14}$$

3.4.2.5 Cross-validation

Finally, the optimal value for the number of PCs is defined by means of cross validation (CV). In this study, leave-one-out cross validation (LOOCV) is chosen. Model performance evaluation is based on the sum of squared residuals (SSR). A set of data is removed as a validation set (1 sample corresponding to 5 measurements) from the calibration set once and a model is fitted with the remaining data. Then the model is used to estimate the target variables in the validation data set and the sum of the squared residuals over all removed data samples is calculated. The number of significant principal components is obtained from the SSR minimum. SSR can be rewritten as follows:

$$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_{ij} - \hat{y}_{ij}(r)\right)^2, \tag{3.15}$$

where $\hat{y}_{ij}(r)$ are the predicted elements of the validation data with r PCs used to reconstruct the data matrix, and y_{ij} are the original calibration data values.

4 RESULTS AND DISCUSSION

4.1 AMMONIUM, NITRITE AND NITRATE DETERMINATION

The measured concentrations for each type of samples and compounds are represented in Figure 4.1. The data set consists of reference measurements of treated urine samples from 02/04/2014 to 11/08/2014. The first four (4) samples correspond to samples of the primary evaluation on saturation effect in the UV spectra experiment (from 02/04/2014 to 09/04/2014) and the remaining values from 5 to 49 correspond to samples of the evaluation on particles and saturation effects experiment (from 14/04/2014 to 11/08/2014).
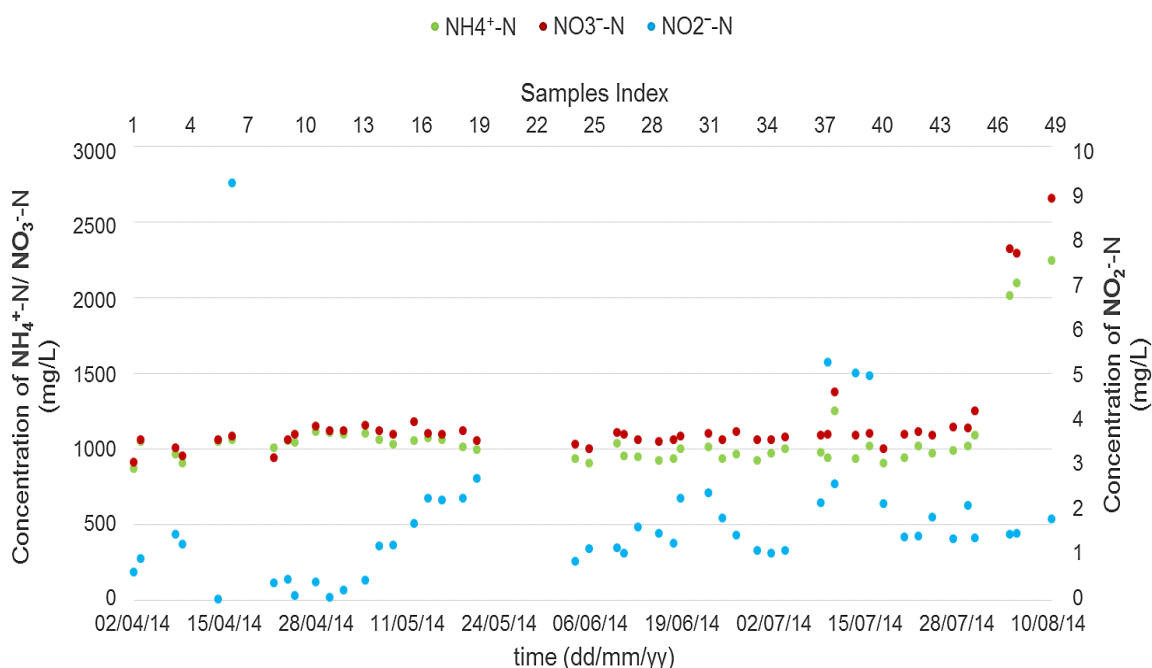
Figure 4.1. Measured ammonium, nitrite and nitrate concentrations in the collected samples, by means of reference tests

Analysing the figure, it is possible to infer that NH_4^+-N and NO_3^--N concentrations hover closely around 1000 mg/l. Exceptions from this behaviour are visible for sample 37 (11/07) and for the last three (3) samples 47, 48 and 49 (05/08, 06/08 and 11/08). Performance of sample 37 could be attributed to kinetic variability in the nitrification reactor. The atypical behaviour of the last 3 samples is due to a change in the reactor feed source. The source feed was changed from female urine to male urine, which is more concentrated. Ammonia concentrations in the influent underwent a huge increase, which leads to a large increase of NH_4^+-N and NO_3^--N concentrations in the nitrification reactor. Regarding NO_2^--N concentrations some variation during the experimental period is visible. Samples 7 to 13 (between 22/04 and 07/05) show NO_2^--N concentrations values close to zero. This behaviour could be a consequence from the reset of the nitrification reactor after being disabled for a period of 7 days (between 17/04 and 22/04). During

the restarting period some disturbances in the nitrification reactor were reported. Sample 6 (16/04) was collected from the nitrification reactor effluent tank during the reactor deactivation period. Samples 36, 38 and 39 (10/07, 14/07 and 16/07) also exhibit an increase of the NO_2^--N concentration, which could be attributed to kinetic variability in the nitrification reactor that probably led to NOB partial inhibition and subsequent nitrite accumulation.

UV absorbance measurements could be influenced by the above referred deviations from the normal behaviour. In addition, during the period between 14/04 and 12/05 (samples 1 to 15) a smaller magnet agitator was used for mixing. Imperfect mixing could lead to interferences in the sensor reading. For these reasons, data screening and selection was performed as follows. Data collected during the second experiment (particles and saturation effect experiment) are divided in three (3) different groups. Data set 1, corresponding to guaranteed normal operation and female urine feeding from 12/05 to 31/07 (samples 16 to 46). Data set 2, corresponding to abnormal operation plus normal operation and female urine feeding from 14/04 to 31/07 (samples 5 to 46). Finally, data set 3, which includes all samples from 12/05 to 11/08 (samples 16 to 49). This third data set includes normal data corresponding to female urine and male urine feeding. This data set excludes data for which mixing conditions are considered suboptimal.

4.2 PRIMARY EVALUATION ON SATURATION EFFECT

The analysed data set consists of UV spectral measurements of treated urine samples, acquired between 02/04/2014 to 09/04/2014 (4 days). 21 samples with different dilutions are analysed (Table 3.3). The spectral measurements are grouped in groups of 5, each corresponding to a certain sample dilution and nitrite (NO_2^-) and nitrate (NO_3^-) concentration. This results in a total of 105 absorbance spectra. As discussed above, UV absorbance measurements are recorded in the range between 220 and 390 nm (171-dimensional spectra). The evaluation of the acquired data is accessed by means of a linear regression, in order to find out which dilution is necessary to ensure that no saturation is present. The methods used for data analysis are explained in detail in Chapter 3, Section 3.4.1.

UV spectra, based on the collected data results from the UV sensor, are displayed in the following figure. In Figure 2.1, it is possible to see that the nitrite and nitrate compounds result in peaks in the UV spectra in a range between 220 nm and 240 nm. Saturation of absorbance measurements take place in the same range. For this reason the chosen wavelengths to pursuit the analysis are from 220 nm up to 240 nm, in steps of 5 nm.
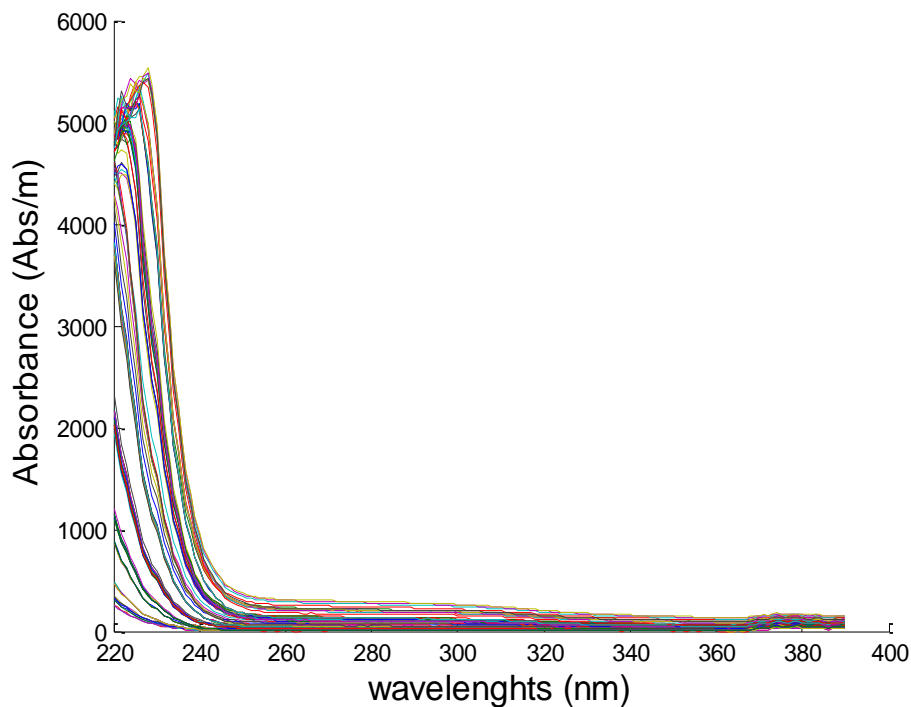
Figure 4.2. UV spectra for all the measured samples and for all applied dilutions

Prior to data analysis, the original data is split in two sets – data set 1 from 02/04 to 04/04, including sample dilutions 01 to 10, and data set 2 from 09/04, including sample dilutions 01.1 to

10.1 (Table 3.3, Section 3.3.1.2). In Figure 4.3, the absorbance measurements for the chosen wavelengths (220, 225, 230, 235, 240 nm) are plotted as a function of the urine fraction.
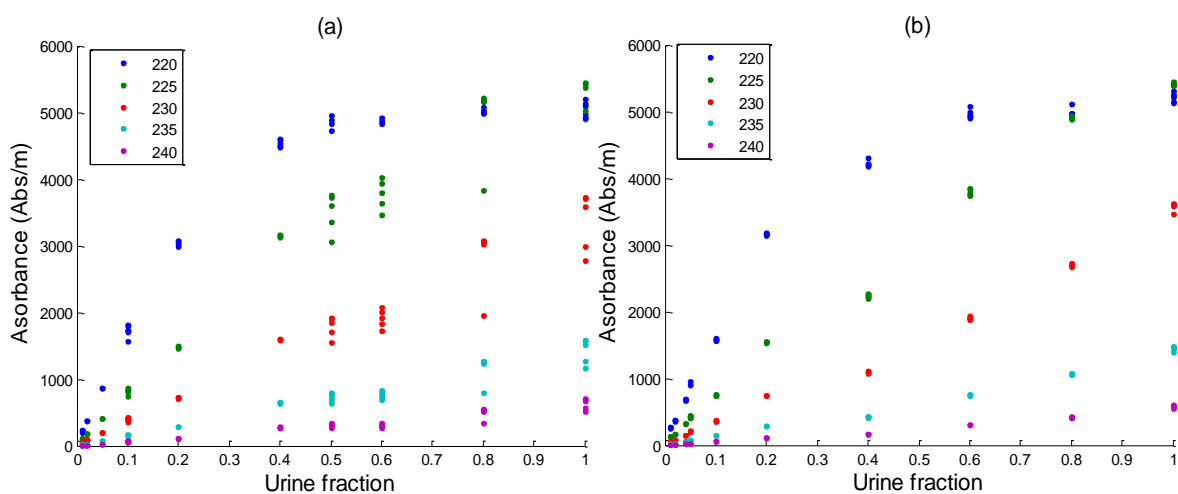
Figure 4.3. Absorbance for the chosen wavelengths (220, 225, 230, 235, 240 nm) as a function of Urine fraction for (a) data set 1 and (b) data set 2

A linear regression model is fitted to the presented data. This results in an estimate for the intercept and slope. Linear regression is executed for different subsets of the available data. Firstly, a linear regression model is fitted by including only the data corresponding to fractions of urine lower or equal than 10% (Figure 4.4 (a) and (b)). Then, the same linear regression was executed by also including the data obtained corresponding to a urine fraction of 20% (Figure 4.4 (c) and (d)). This iterative addition of data to the calibration data set is repeated for all fractions until all available data is included (Figure 4.4 (e)-(h)). By doing so, one can experimentally evaluate at which concentration the linear law of Beer-Lambert (Equation 2.4) becomes invalid due to saturation. Analysing the figures below it is possible to infer that for urine fractions up to 20% (see Figure 4.4 (a)-(d)) the linear regression fits rather well. However, when increasing the highest urine fraction to 40%, the linear regression does not fit properly anymore (see (Figure 4.4 (e)-(h)). This results suggest that for urine fractions below 20% no saturation is present and a linear model can be applied.
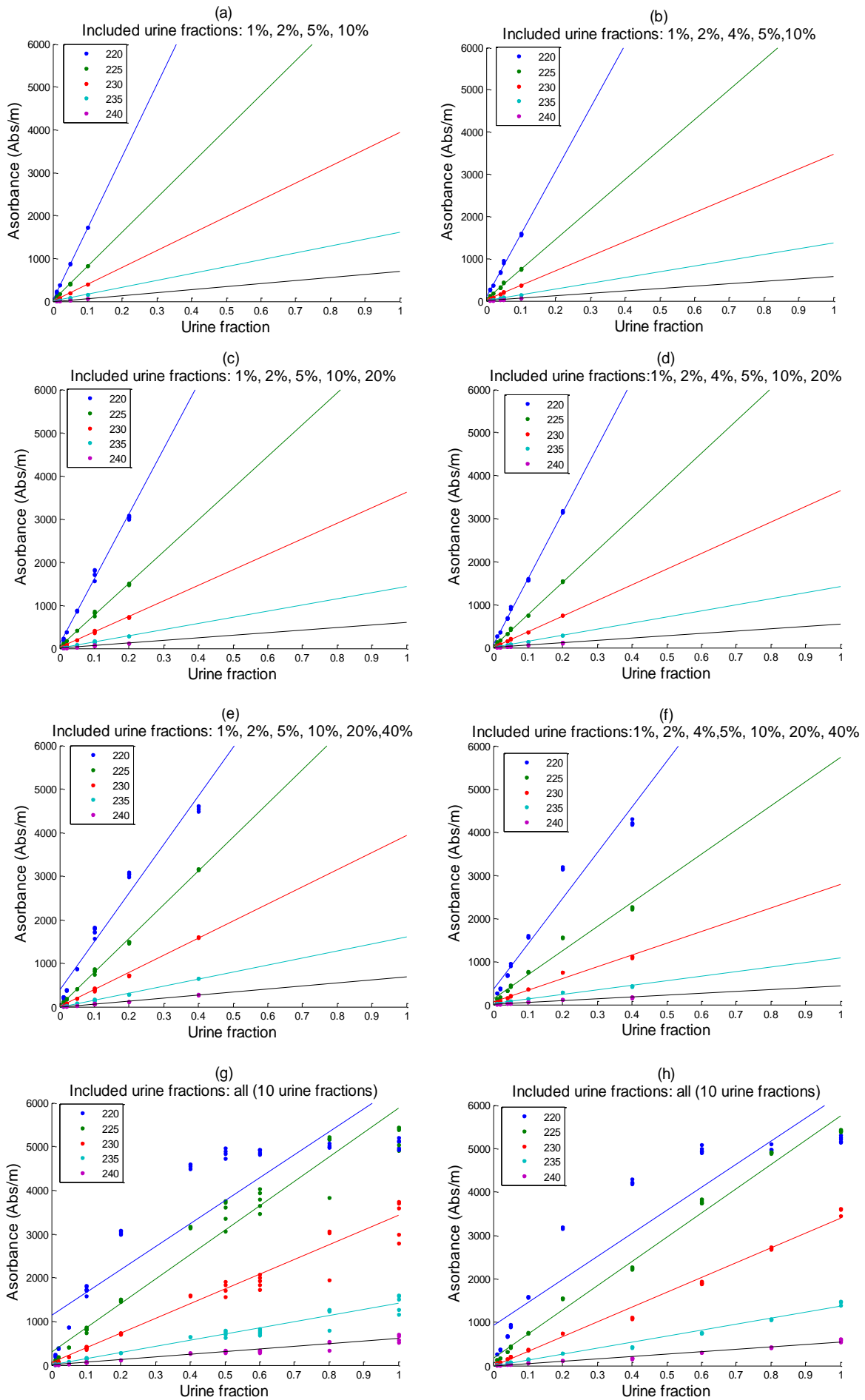
Figure 4.4. Linear regression of urine fraction as a function of Absorbance for the chosen wavelengths (220, 225, 230, 235, 240 nm) for (a), (c), (e), (g) data set 1 and (b), (d), (f), (h) data set 2

The above estimated parameters (intercept and slope) together with the input variables (urine fractions) allow to predict absorbance measurements, for the chosen wavelengths. The absorbance measurements predicted by the linear models are plotted as a function of the measured absorbances in Figure 4.5 (a)-(d). The prediction error is also displayed as function of the same measurements in Figure 4.6 (e)-(h). The prediction error is computed based on the difference between the measured absorbances and the predicted absorbances. The figures below are based on data set 2. Results of data set 1 are in accordance with the presented ones and are shown in Appendix III.
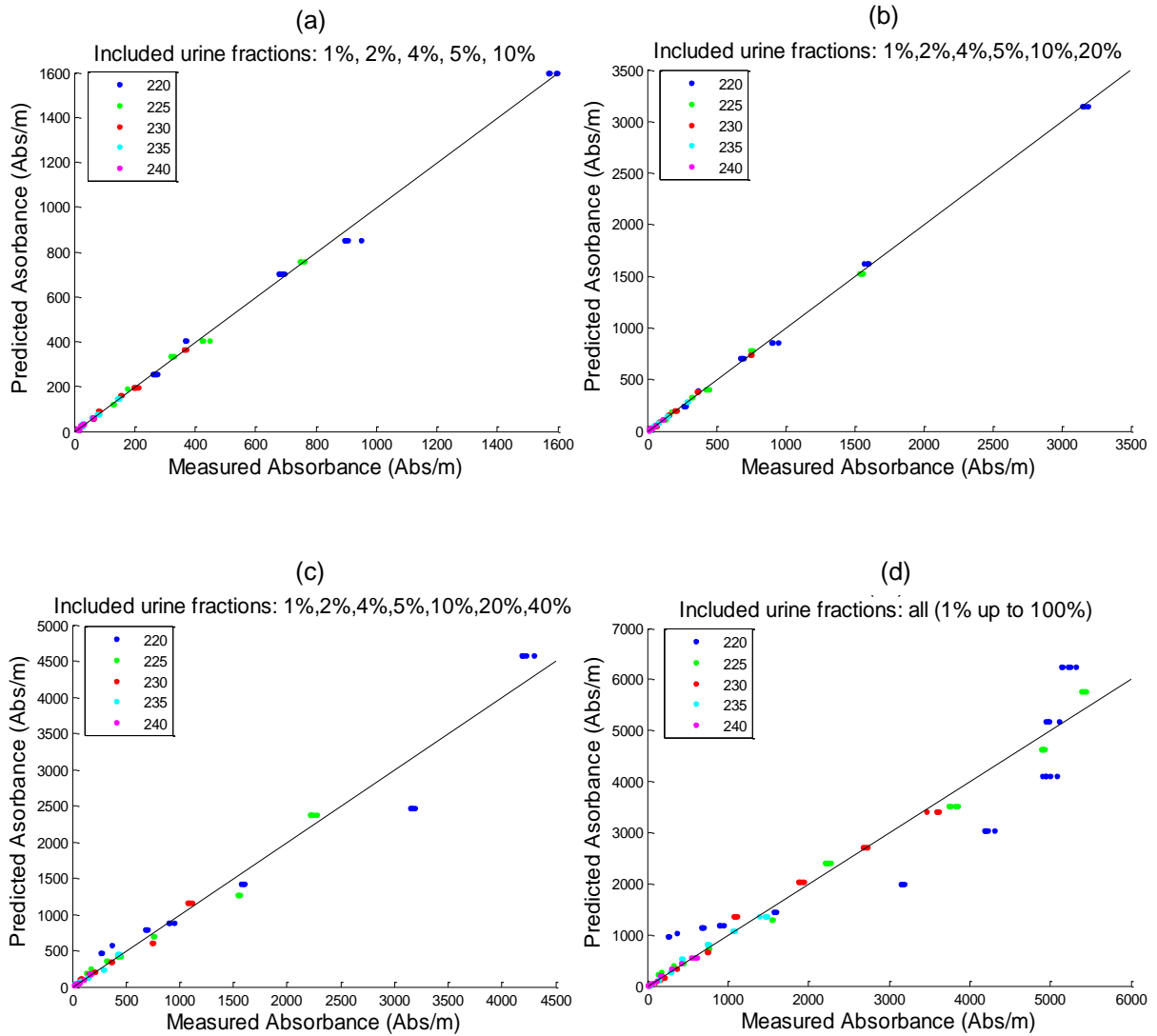
Figure 4.5. (a)-(d) Predicted absorbances as function of measured absorbance, for the chosen wavelengths (220, 225, 230, 235, 240 nm) of data set 2
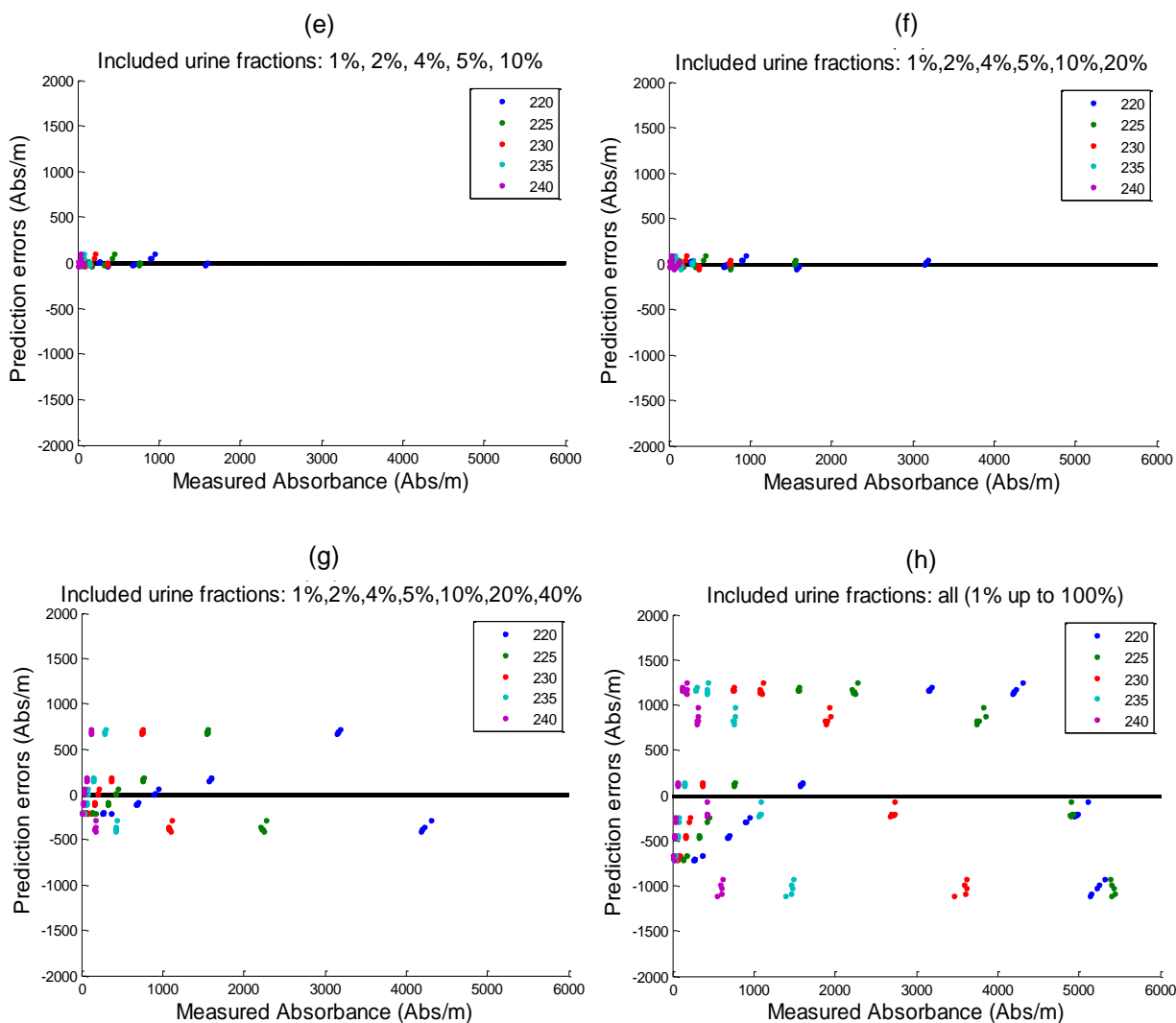
Figure 4.6. (e)-(h) Prediction errors as function of measured absorbance, for the chosen wavelengths (220, 225, 230, 235, 240 nm) of data set 2

Once more, it is possible to infer that for urine fractions up to 20% the linear regression model fits well. By increasing the urine fraction further than 20%, the linear regression does not fit anymore. With increasing values for the highest included urine fraction, the prediction errors increase drastically. Therefore, it is obvious that the saturation effect challenges the assumption of linearity quite dramatically. The above results suggest that for urine fractions below 20% no saturation is present and Lambert-Beer's law is valid. In order to work in a safe range a 10% urine fraction corresponding to a 10x dilution is considered a proper experimental setting in which the saturation effect is expected to be removed entirely. The following experiment in Section 4.3 therefore uses this dilution as a means of reference for evaluation of the quality of nitrite estimates.

4.3 PARTICLES EFFECT AND SATURATION EFFECT EXPERIMENT

The data set consists of UV spectral measurements of treated urine samples, from 14/04/2014 to 11/08/2014 (16 weeks, 3 times a week) and corresponding nitrite (NO_2^-) and nitrate (NO_3^-) concentration measurements. The spectral measurements are grouped in groups of 5 (corresponding to one particular sample), each corresponding to a certain nitrite (NO_2^--N) and nitrate (NO_3^--N) concentration. This results in a total of 2720 absorbance spectra. UV spectra are recorded in the range between 220 and 390 nm (171-dimensional spectra). The number of absorbance spectra for each type of sample fractions are listed in Table 4.1.

Table 4.1. Type of samples acquired and corresponding number of measurements in the particles effect and saturation effect experiment

FRACTION		TYPE OF FILTRATION	TYPE OF DILUTION	NO^- STOCK –SOLUTION ADDITION	NUMBER OF SAMPLES ACQUIRED	NUMBER OF ABSORBANCE SPECTRA
A	01	0.7 µm Filter	Non-diluted	No addition	46	230
	02	0.7 µm Filter	Non-diluted	5 ml addition	46	230
A'	03	0.7 µm Filter	Diluted (1:10)	No addition	46	230
	04	**0.7 µm Filter**	**Diluted (1:10)**	**5 ml addition**	46	230
B	05	Coffee filter	Non-diluted	No addition	45	225
	06	Coffee filter	Non-diluted	5 ml addition	45	225
B'	07	Coffee filter	Diluted (1:10)	No addition	45	225
	08	Coffee filter	Diluted (1:10)	5 ml addition	45	225
C	09	Non-filtered	Non-diluted	No addition	45	225
	10	**Non-filtered**	**Non-diluted**	**5 ml addition**	45	225
C'	11	Non-filtered	Diluted (1:10)	No addition	45	225
	12	Non-filtered	Diluted (1:10)	5 ml addition	45	225
				TOTAL	**544**	**2720**

The spectral data are analysed according to the procedure described in detail in Chapter 3, Section 3.4.2.

4.3.1 ESTIMATION OF NITRITE (NO2⁻-N) CONCENTRATIONS

PCR modelling is initiated with evaluation of data set 1 (described in Section 4.1), corresponding to normal operation and female urine feeding. The spectral data are firstly evaluated for the ideal case corresponding to diluted and filtered samples (case 1). In order to gradually challenge the calibration/validation procedure for PCR, the spectral data are evaluated for unfiltered and diluted samples (case 2), followed by the analysis for filtered and non-diluted samples (case 3). These cases enable the evaluation of the effect of particles and the effect of saturation in the nitrite estimation model, respectively. Hereafter, data set 2 (described in Section 4.1) including urine samples corresponding to irregular operation is analysed. The same three cases are evaluated for data set 2 as well and are referred to as case 1.1, case 2.1 and case 3.1.

4.3.1.1 Case (1) - Micro Filtration (0.7 μm), Dilution (1:10) and NO₂⁻ stock-solution addition (5 ml)

The evaluation of the results are started for what is considered as a most ideal case – a sample subjected to filtration, dilution and NO₂⁻ stock-solution addition – since in this case the saturation effect and the particles effect are minimized due to the dilution (1:10) step and to the filtration (0.7 μm) step, respectively. Data set 1 only includes data of guaranteed normal operation of the reactor and excellent experimental conditions.

Prior to data analysis, the absorbance measurements are centred. In Figure 4.7 UV absorbance spectra and UV centred absorbance spectra of the sample fractions above described are presented. It is clear that nitrite absorbs mostly in UV range between 220 nm and 240 nm.
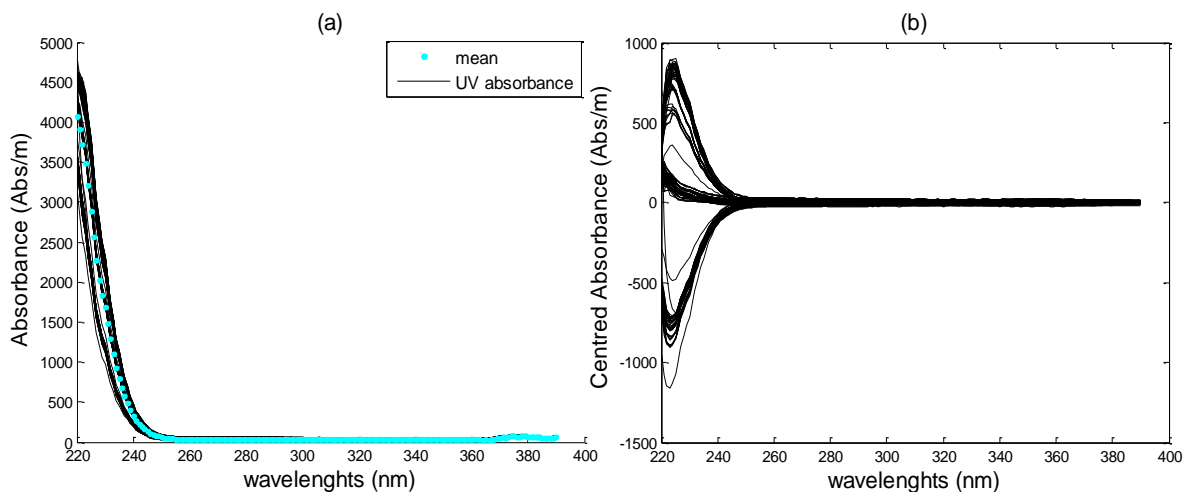
Figure 4.7. Case 1. (a) UV absorbance spectra and (b) UV absorbance spectra centred, acquired for all the measured samples

Hereafter, the remaining spectral data are first analysed by PCA, to reduce the dimensionality of the data. The 171-dimensional spectra are reduced to a lower number of principal components (PCs), by means of singular value decomposition (SVD). The singular values of the covariance matrix are indicative of the importance of each of the orthogonal components and are displayed in Figure 4.8 (a). The eigenspectrum suggests that a single principal component explains most of the information in the spectral data. The variance of the first PC correspond to a value of $3.906(*10^6)$ and the variance of the second PC decrease radically to a value of $0.021(*10^6)$. PC1 correspond to 99% of the total variance on the eigenspectrum. This effect was expected, since the first PC is assumed to correspond to the nitrite concentration and no other major effects are expected with this data set. In the present case the saturation effect and the particles effect are minimized due to the dilution (1:10) step and to the filtration (0.7 μm) step, respectively, and the only effect expected is variation on nitrite concentration. The first component is displayed in Figure 4.8 (b), showing that the spectra vary in a band between 220 nm and 240 nm, which corresponds to nitrite range of absorbance.
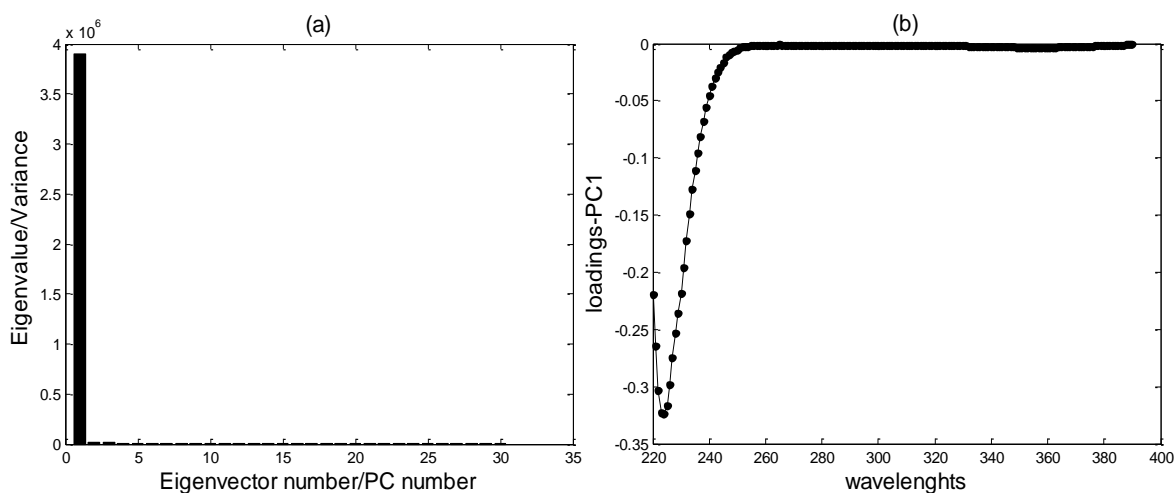
Figure 4.8. Case 1. (a) Eigenvalues corresponding to the first 30 Principal Components; (b) Effect of original variables (loadings) upon 1st PC.

The above PCA model is used as a precursor for a predictive model – PCR. A regression is made to predict nitrite concentrations by means of a single PC in the PCA model. Figure 4.9 (a) and (b) shows the score for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations, respectively. The spectral measurements are grouped in sets of 5, each group corresponding to a certain original urine sample and nitrite concentration.
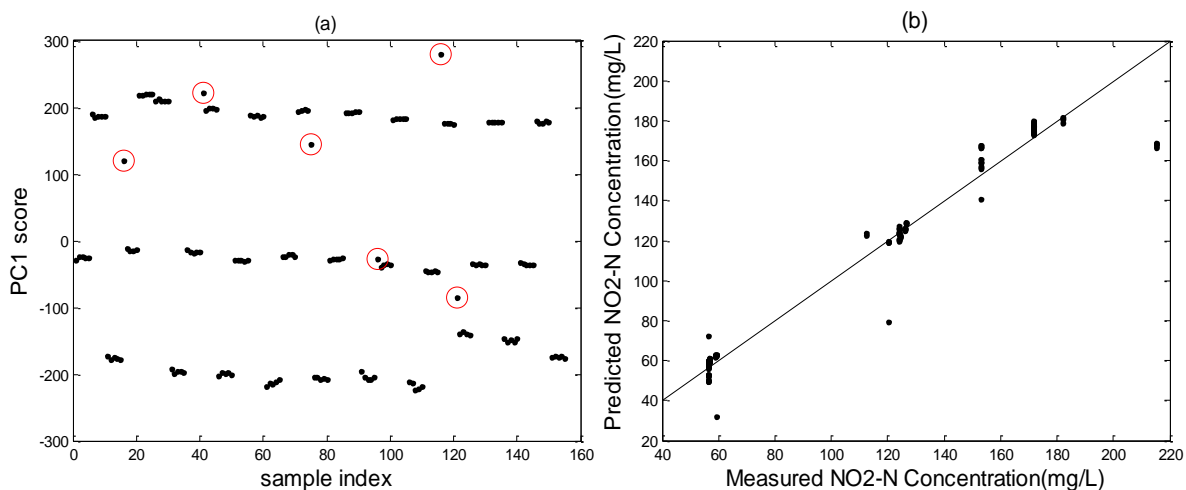
Figure 4.9. Case 1. (a) PC1 score as a function of sample index and (b) Nitrite
Concentrations - predicted vs. measured with 1-PC model

In Figure 4.9 (a) some outliers can easily be distinguished (indicated by red circles in Figure 4.9 (a)). During addition of nitrite stock-solution, proper mixing is needed to mix it properly in the sample. Outliers could occur however, typically in the first spectral measurement of a group of 5, due to incomplete mixing. Six (6) outliers are removed and PCR model is repeated for the remaining samples. The results are displayed in Figure 4.10 (a) and (b).
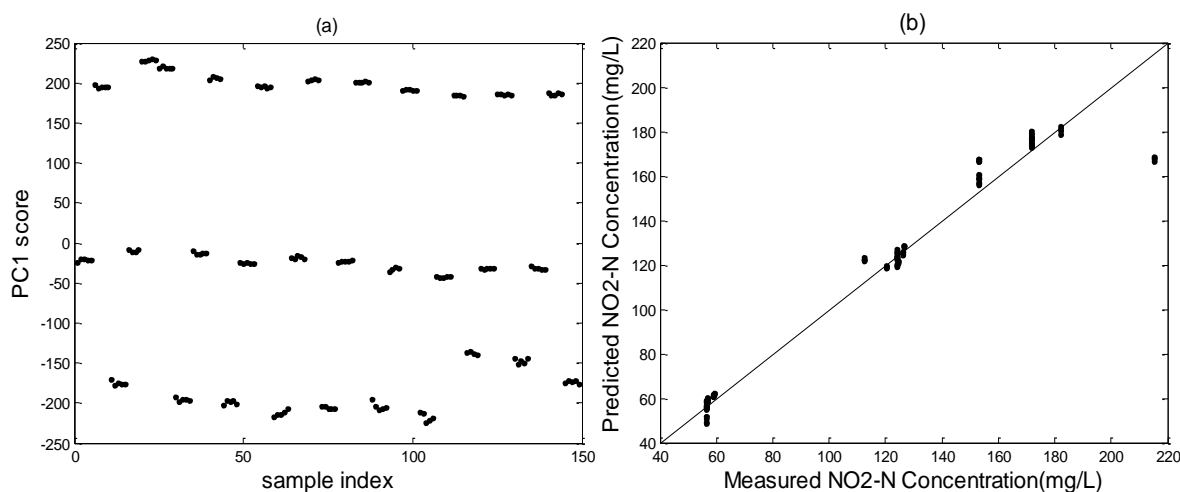
Figure 4.10. Case1. (a) PC1 score as a function of sample index and (b) Nitrite
Concentrations - predicted vs. measured with 1-PC model, after outliers removal

Figure 4.10 (b) shows the predicted nitrite concentrations as a function of measured nitrite concentration for the calibration data set for one dimension. The PCR model seems to fit accurately to the nitrite concentrations. The results above are performed without cross-validation. A leave-one-out cross validation (LOOCV) is then performed to determine the best number of PCs for the above PCR model. The number of significant principal components (PCs) is obtained from the minimum residual error (minimum SSR). The number of PCs is given on the x-axis and the SRR value on the y-axis and the plot is represented in Figure 4.11 below.
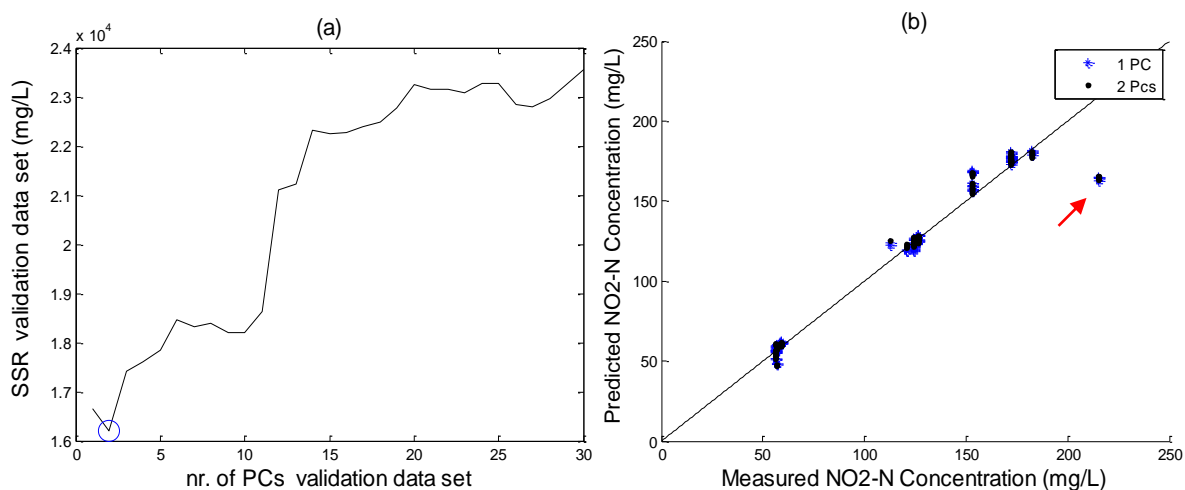
Figure 4.11. Case1. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set.

Analysing Figure 4.11 (a), it can be clearly seen that the minimum SSR corresponds to 2 dimensions (represented by a blue circle in Figure 4.11 (a)), thus the PCR model should include 2 PCs. The model dimension obtained by means of SSR minimum (2 PCs) is very close to the dimension achieved by studying the eigenspectrum (1 PC). Figure 4.11 (b) shows the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set for one and two dimensions. The PCR model seems to fit well to the nitrite concentrations in both cases. The figure demonstrates small deviations from the linear model, which could be explained by the non-linear behaviour of the samples or by drastic changes in the concentrations range that could not be accurately predicted by the sensor. Sample from 16/05 ((indicated by a red arrow in Figure 4.11 (b)) presents a concentration range that hinders the model fit. The high nitrite concentration in the sample led to saturation. The saturation effect in the sample appears to influence the model fitting.

The first two components are displayed in Figure 4.12 (a) showing that the spectra vary similarly in a band between 220 nm and 240 nm for both eigenvectors, as expected. The second eigenvector starts to decrease slowly at around 260 nm, which could indicate the effect of organic compounds on the spectra. Figure 4.12 (b) shows the scores for the second PC as a function of sample. Substantial variation is present.
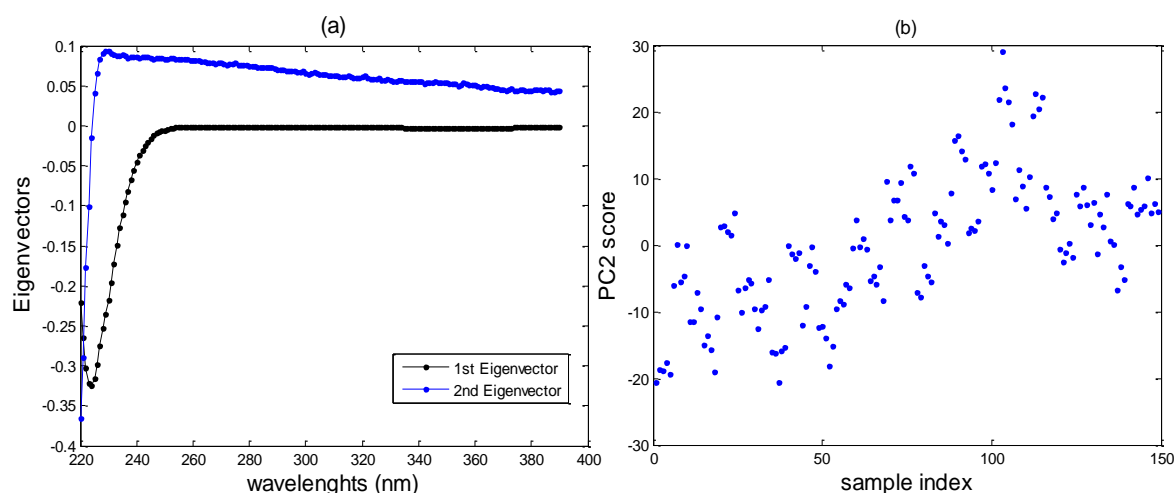
Figure 4.12. Effect of original variables (loadings) upon 1st and 2nd PC for the calibration data set; (b) PC2 score as a function of sample index

4.3.1.2 CASE (2) –Non-Filtration, Dilution (1:10) and NO_2^- stock-solution addition (5 ml)

The evaluation of the results proceed with samples subjected to dilution and NO_2^- stock-solution addition, however without the filtration step. The goal of this is to evaluate the effect of particles in the UV spectra. The saturation effect remains minimized due to the dilution (1:10) step.

Prior to the evaluation of the data, the absorbance is centred as in the previous analysis. Once more, the spectral data are then primarily analysed by PCA to reduce the dimensionality of the data. The singular values of the covariance matrix are displayed in Figure 4.13 (a). The eigenspectrum suggests again that a single principal component dominates all the rest, which could mean that the effect of particles in the UV spectra appears to be of less importance. The variance of the first PC correspond to a value of 3.876($*10^6$) and the variance of the second PC decrease considerably to a value of 0.019($*10^6$). PC1 represents 99% of the total variance on the eigenspectrum. The first component is displayed in Figure 4.13 (b), showing that the spectra vary in a band between 220 nm and 240 nm, which corresponds to the known range of absorbance for nitrite.

Figure 4.13. Case 2. (a) Eigenvalues corresponding to the first 30 Principal Components;
(b) Effect of original variables (loadings) upon 1st PC.

The PCA model is used as a precursor for a PCR model. A regression is made to predict nitrite concentrations by means of a single PC in the PCA model. Figure 4.14 (a) and Figure 4.14 (b) display the score for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations, respectively.
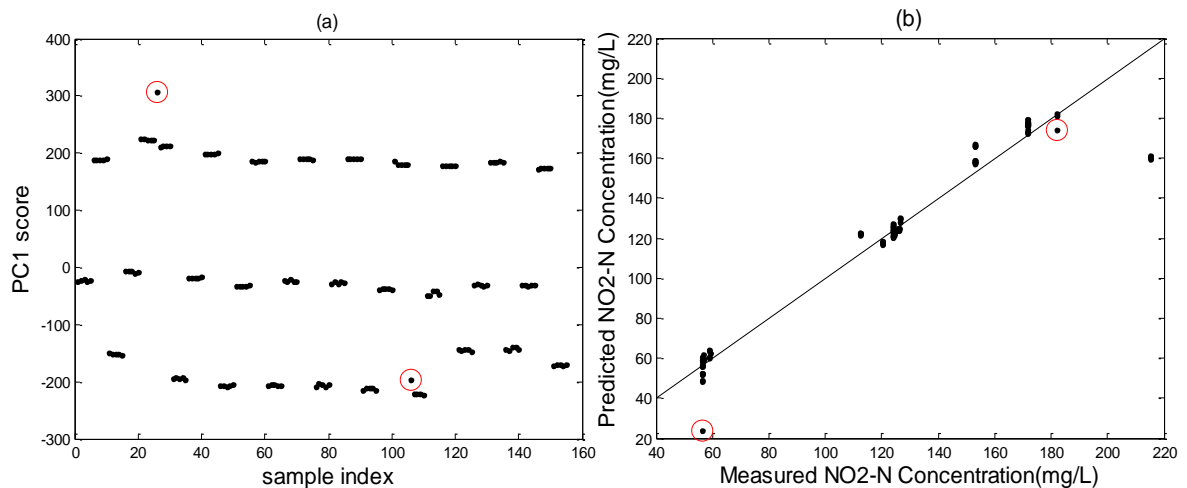
Figure 4.14. Case 2. (a) PC1 score as a function of sample index: (b) Nitrite
Concentrations - predicted vs. measured with 1-PC model

Similarly to the case above, two (2) outliers are thereafter removed and PCR model is repeated for the remaining samples. The results are shown in Figure 4.15 (a) and (b).
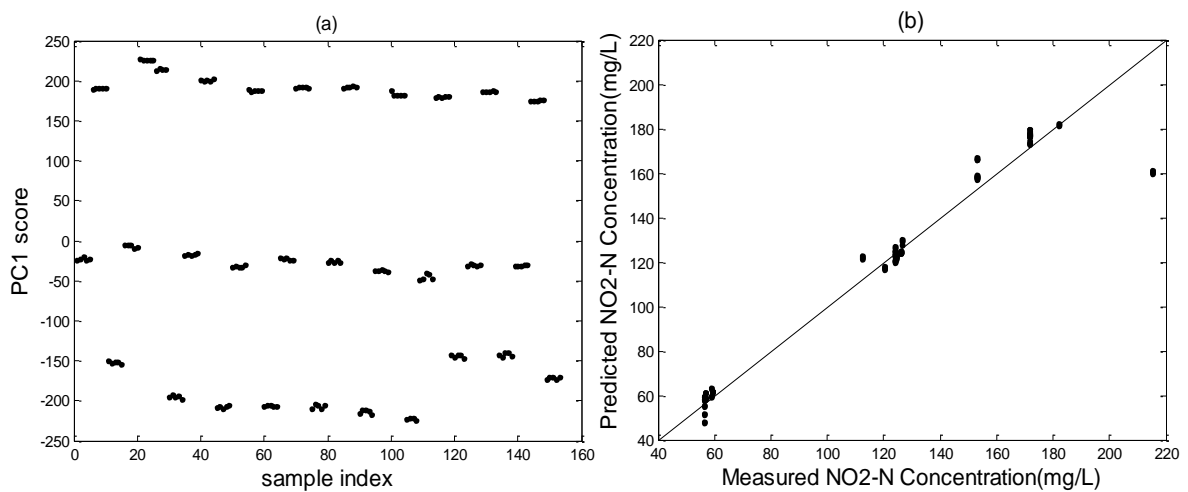
Figure 4.15. Case 2. (a) PC1 score as a function of sample index and (b) Nitrite
Concentrations - predicted vs. measured with 1-PC model, after outliers removal

Analysing Figure 4.15 (b) it appears that the one dimensional PCR model fits with reasonable accuracy to the data. However, the results above are performed without cross-validation. A cross validation is then performed to determine the ideal number of PCs for the above PCR model. The number of significant PCs is obtained from the minimum residual error (minimum SSR). The number of PCs is displayed in Figure 4.16 (a).

Figure 4.16. Case 2. (a) SSR as a function of model dimensions for the validation data set; (b)
Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing Figure 4.16 (a), it can be clearly seen that the minimum SSR corresponds to 2 dimensions, thus the PCR model should include 2 PCs as well. The model dimension obtained by means of SSR minimum (2 PCs) approaches the dimension achieved by the eigenspectrum (1 PC) and is according to expectations. Figure 4.16 (b) displays the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set for one and two dimensions. The figure demonstrates some deviations from the linear model, which could be explained by the non-linear behaviour of the samples or by drastic changes in the concentrations range that could not be accurately predicted by the sensor. These deviations

45

appear to decrease when the dimensionality of the PCR model is increased. Sample from 16/05 (indicated by a red arrow in Figure 4.16 (b)) is not predicted well. The high nitrite concentration in the sample led to saturation. The saturation effect in the sample appears to influence the model fitting.

4.3.1.3 CASE (3) – Micro Filtration (0.7 µm), non-Dilution and NO_2^- stock-solution addition (5 ml)

The third evaluation is based on samples subjected to filtration and NO_2^- stock-solution addition, however without the dilution step. The goal is to evaluate the effect of saturation on the UV spectra. The particles effect remains minimized due to the filtration (0.7 µm) step.

Spectral data are analysed by PCA, after mean centring. The singular values of the covariance matrix are indicative of the importance of each of the orthogonal components and are shown in Figure 4.17 (a). The eigenspectrum suggests a gradual decrease of importance of each orthogonal component and that two principal component explain most of the information in the spectral data. This is different from cases (1) and (2). The variance of the first PC correspond to a value of $5.379(*10^5)$ and the variance of the second PC decrease considerably to a value of $2.125(*10^5)$, followed by the third that presents a value of $0.298(*10^5)$. PC1, PC2 and PC3 correspond to 65%, 26% and 4%, respectively, of the total variance. PC1 and PC2 together correspond to 91% of the variance on the eigenspectrum.
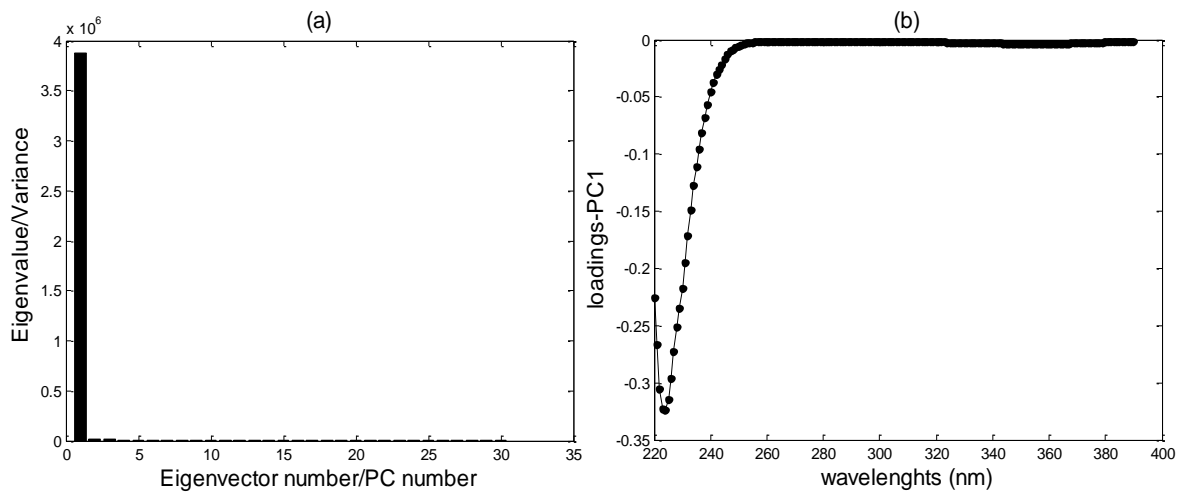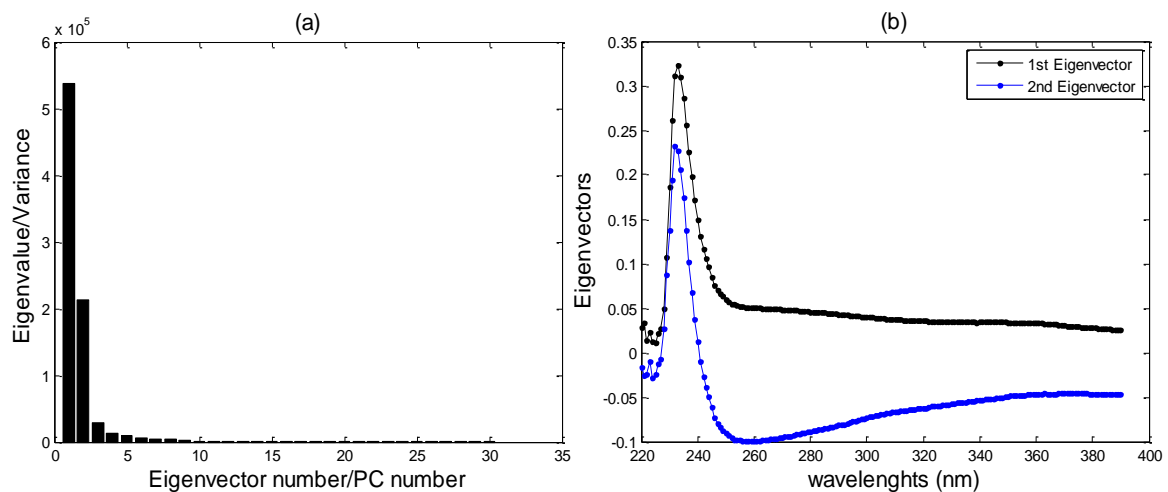

Figure 4.17. Case 3. (a) Eigenvalues corresponding to the first 30 Principal Components;
(b) Effect of original variables (loadings) upon 1st and 2nd PC

The first two components are displayed in Figure 4.17 (b), respectively, showing that the spectra vary in a band between 220 nm and 240 nm for the first eigenvector and for the second eigenvector vary simultaneously in a band between 220 nm and 240 nm and a broader band around 260 nm. Nitrite absorbances in the range between 220 nm and 240 nm are expected. However, absorbance at 260 nm is unexpected and suggests that the presence of saturation is possibly leading to poor sensor reading.

A PCR is made to predict nitrite concentrations by means of two PCs in the PCA model. Figure 4.18 and Figure 4.19 (a) shows the scores for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations for a two dimensions (2 PCs) model, respectively.
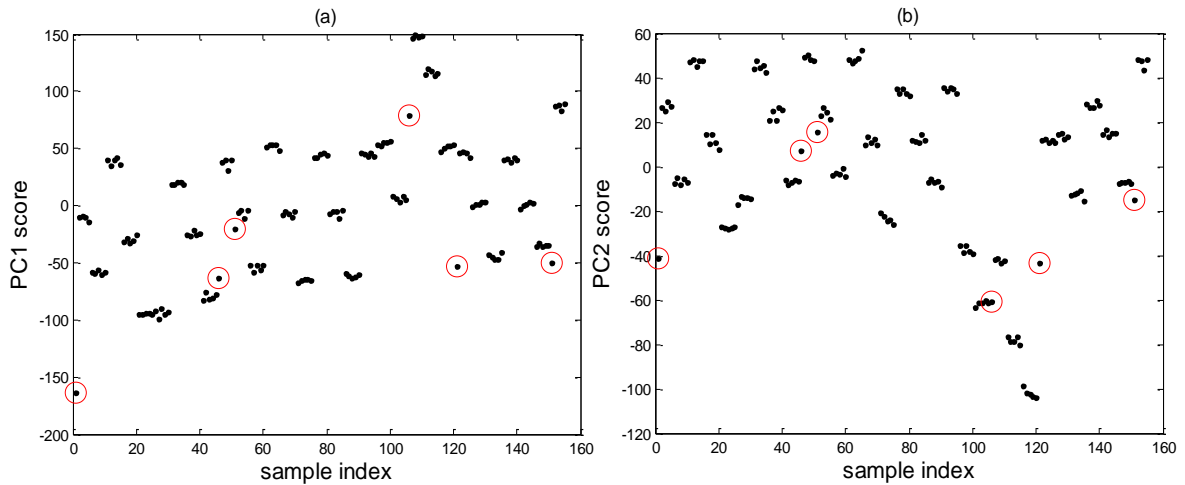

Figure 4.18. Case 3. (a) PC1 and (b) PC2 scores as a function of sample index

Six (6) outliers are removed, after analysing the above figures, and PCR model was repeated for the remaining samples.
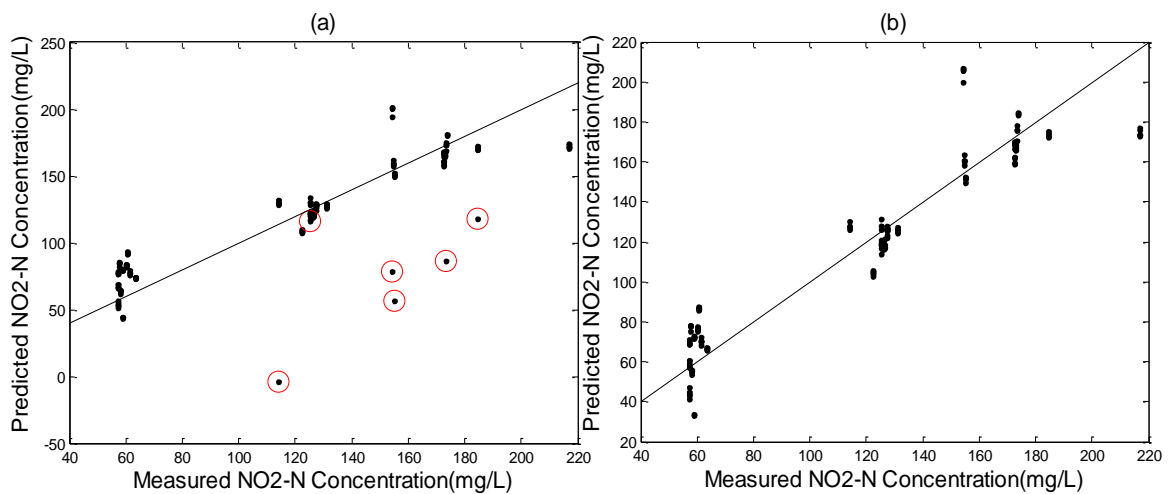
Figure 4.19. Case 3. Nitrite Concentrations - predicted vs. measured with 2-PCs model
(a) before outlier removal and (b) after outlier removal

The predicted nitrite concentrations as a function of the measured nitrite concentrations for a two dimensions (2 PCs) model are shown in Figure 4.19. Analysing the figure appears that the two dimensional PCR model fit with acceptable accuracy to the results which is coherent with the presence of saturation effect due to non-dilution. Nevertheless, the results above are performed without cross-validation. A cross validation is then performed to determine the best number of

PCs for the PCR model. The number of significant principal components (PCs) was obtained from the SSR minimum, as indicated in Figure 4.20 (a).
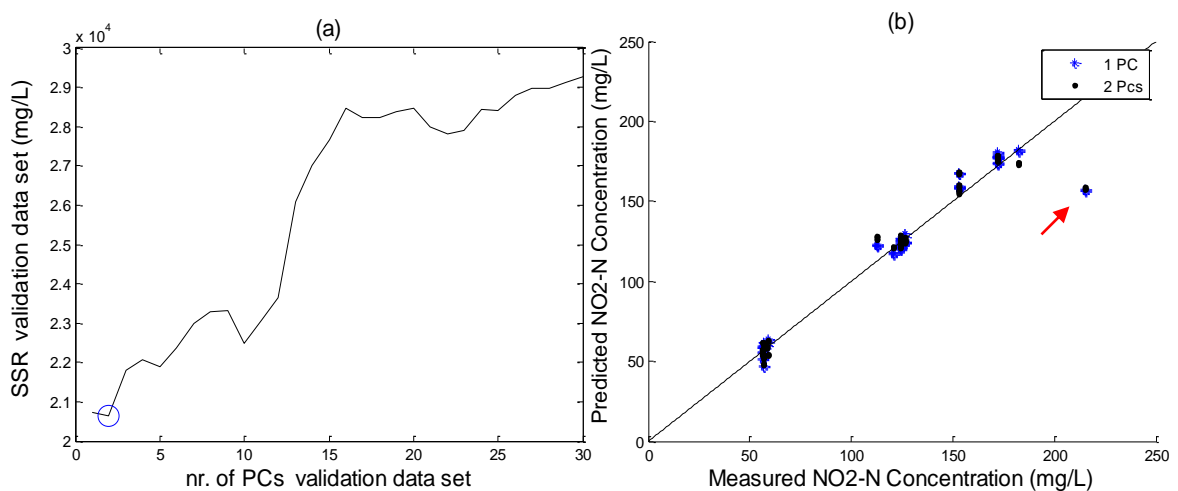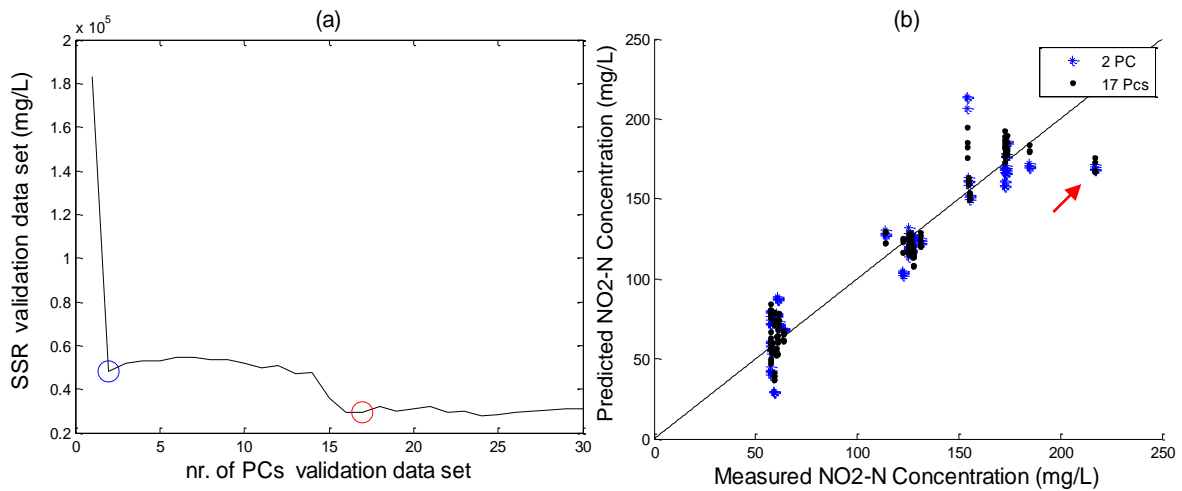

Figure 4.20. Case 3. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing the figure, it can be seen that the SSR minimum corresponds to 17 dimensions (represented by a red circle in Figure 4.20 (a)), thus the PCR model should include 17 PCs as well. The model dimension obtained by means of SSR minimum (17 PCs) differs drastically from the dimension given by the eigenspectrum (2 PCs). Even so, for 2 PCs the SSR presents a pronounced minimum as well ((represented by a blue circle in Figure 4.20 (a)). This discrepancy is explained as a result of saturation. To evaluate the impact of this discrepancy on the nitrite estimations, PCR models are inspected in more detail for 2 PCs and 17 PCs. Figure 4.20 (b) displays the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set for these models. The figure demonstrates some deviations from the linear model. However, increasing the model dimension from 2 PCs to 17 PCs minimizes these deviations. The sample from 16/05 (indicated by a red arrow in Figure 4.20 (b)) presents a concentration range that negatively influences the model fit. The high nitrite concentration in the sample led to saturation. The saturation effect in the sample appears to influence the model fitting.

4.3.1.4 Summary of Results for cases (1), (2) and (3)

Comparing Figure 4.16 (case 2 - unfiltered and diluted samples) with Figure 4.11 (case 1- filtered and diluted samples) it is possible to infer that the effect of particles appears to have a minimal impact on nitrite concentration estimation. The dimensionality of both models is 2 PCs, based on SSR minimum. The quality of the obtained estimation for a two dimension model presents a suitable accuracy.

Comparing Figure 4.20 (case 3 – filtered and non-diluted) with Figure 4.11 (case 1 – filtered and diluted) it is possible to infer that the effect of saturation has a drastic impact on nitrite concentration estimation. The main difference from case 1 to case 3 is the need to increase the dimension of the model from 2 PCs to 17 PCs, based on SSR minimum. Nevertheless, the quality of the obtained estimation for a 17 dimension model might be acceptable for practical application

In summary, it is possible to infer that the UV absorbance measurements and, consequently, nitrite concentration estimation are severely affected by saturation while the particle effect is of limited importance.

4.3.1.5 CASE (1.1) - Micro Filtration (0.7 μm), Dilution (1:10) and NO_2^- stock-solution addition (5 ml)

The evaluation of the acquired data is now calculated for data set 2, in order to evaluate in what way unstable operation of the nitrification reactor could influence the nitrite estimation. Data set 2 includes both data from abnormal operation and of guaranteed normal operation.

The new evaluation of the results is started once more for what is considered as a most ideal case. The samples are subjected to filtration, dilution and NO_2^- stock-solution addition. Indeed, in this case the saturation effect and the particles effect are minimized due to the dilution (1:10) step and to the filtration (0.7 μm) step, respectively.

Outliers identified for case 1 are removed. Thereafter, PCA is performed. The singular values of the covariance matrix are shown in Figure 4.21 (a). The variance of the first PC correspond to a value of 4.190(*10^6) and the variance of the second PC decrease considerably to a value of 0.021(*10^6). PC1 represents 99% of the total variance on the eigenspectrum. The eigenspectrum suggests once more that a single principal component explains most of the information in the spectral data. This effect was expected, since in this case the saturation effect and the particles effect are minimized due to the dilution (1:10) step and to the filtration (0.7 μm) step, respectively. The first component is displayed in Figure 4.21 (b), showing that the spectra vary in a band between 220 nm and 240 nm, which corresponds to the range of absorbance for nitrite.

Figure 4.21. Case 1.1. (a) Eigenvalues corresponding to the first 30 Principal Components; (b) Effect of original variables (loadings) upon 1st PC.

The above PCA model is used as a precursor for a predictive model – PCR. A regression is made to predict nitrite concentrations by means of a single PC in the PCA model. The following Figure 4.22 (a) and (b) shows the score for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations, respectively.

Figure 4.22. Case 1.1. (a) PC1 score as a function of sample index and (b) Nitrite Concentrations - predicted vs. measured with 1-PC model, after outliers removal

Observing the figures above, the abnormal behaviour is very obvious for the first 60 samples which correspond to the reactor restart period. This behaviour could be caused by imperfect mixing during the experiment (a smaller magnet agitator was used during the referred period) or to the presence of higher percentage of organic compounds, related to the reactor restart. Both hypotheses potentially explain the poorer fit. Inappropriate agitation appears to have a great impact on the PCR model which leads to poor accuracy in the model adjustment. Nevertheless, the results above are performed without cross-validation. A LOOCV is then performed to evaluate the best number of PCs for the PCR model. The number of significant PCs is obtained from the minimum residual error (minimum SSR). The number of PCs is given in the Figure 4.23 below.

Figure 4.23. Case 1.1. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing Figure 4.23 (a), it can be clearly seen that the minimum SSR corresponds to 12 dimensions (represented by a blue circle in Figure 4.23 (a)), thus the PCR model should include 12 PCs. The model dimension obtained by means of SSR minimum (12 PCs) differs drastically from the dimension given by the eigenspectrum (1 PC). To evaluate the impact of this discrepancy

51

on the nitrite estimations, PCR models are performed including 1 PC and 12 PCs, respectively. Figure 4.23 (b) shows the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set. The figure demonstrates some deviations from the linear model. Samples affected by improper mixing or abnormal operation appear to present poorest fitting.

4.3.1.6 CASE (2.1) –Non-Filtration, Dilution (1:10) and NO_2^- stock-solution addition (5 ml)

The new evaluation of the results is pursued with samples subjected to dilution and NO_2^- stock-solution addition, however without the filtration step. The goal is to evaluate the effect of particles in the UV spectra. The saturation effect remains minimized due to the dilution (1:10) step.

Outliers referred to in previous Section for case 2 are removed. The spectral data are then primarily analysed by PCA. The singular values of the covariance matrix are displayed in Figure 4.24 (a). The eigenspectrum suggests again that a single principal component dominates all the rest, which could mean that the effect of particles in the UV absorbance spectra appears to be of less importance. The variance of the first PC correspond to a value of 3.827($*10^6$) and the variance of the second PC decrease considerably to a value of 0.023($*10^6$). PC1 corresponds to 99% of the total variance on the eigenspectrum. The first component is displayed in Figure 4.24 (b), showing that the spectra vary in a band between 220 nm and 240 nm, which, once more, corresponds to the range of absorbance for nitrite.
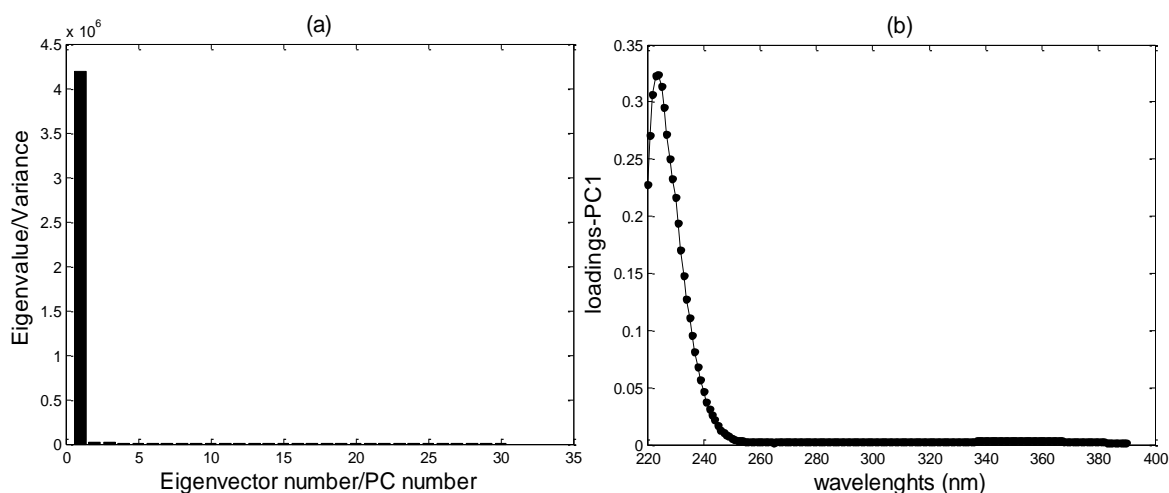

Figure 4.24. Case 2.1. (a) Eigenvalues corresponding to the first 30 Principal Components; (b) Effect of original variables (loadings) upon 1st PC.

A regression is made to predict nitrite concentrations by means of a single PC in the PCA model. Figure 4.25 (a) and (b) shows the score for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations, respectively.
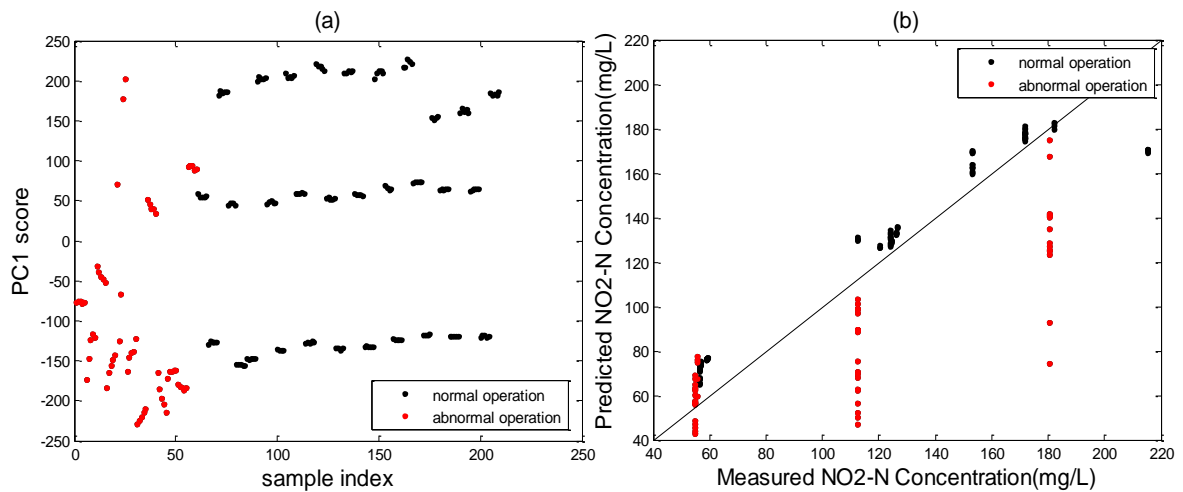
Figure 4.25. Case 2.1. (a) PC1 score as a function of sample index after outliers removing; (b) Nitrite Concentrations - predicted vs. measured with 1-PC model after outliers removal

A LOOCV is performed to determine the correct number of PCs for the PCR model. The number of significant PCs are displayed in the following Figure 4.26 (a) and is obtained from the SSR minimum.
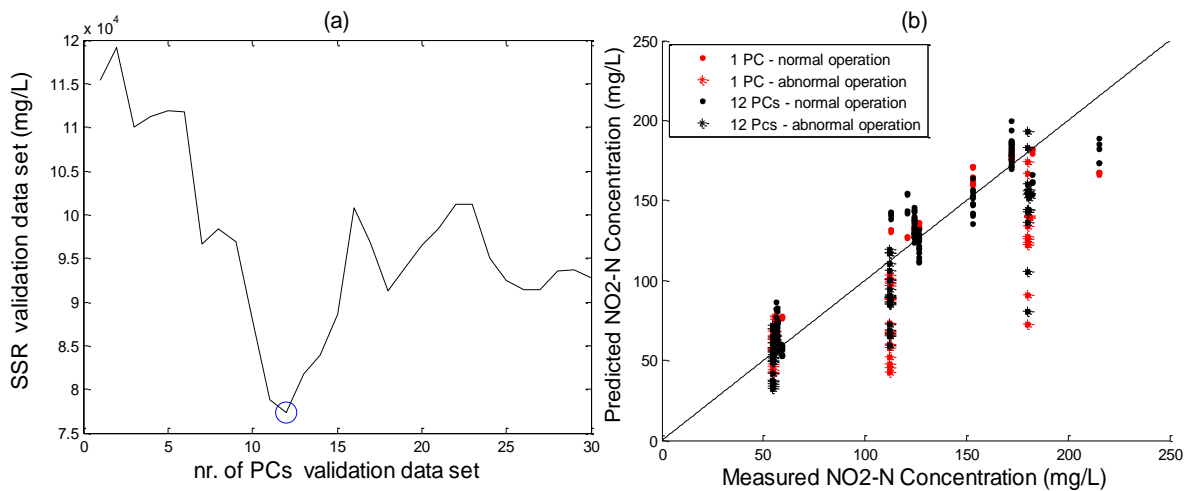

Figure 4.26. Case 2.1. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing Figure 4.26 (a), it can be seen that the minimum SSR corresponds to 19 dimensions (represented by a red circle in Figure 4.26 (a)), thus the PCR model should include 19 PCs as well. However, for 2 PCs the SSR presents a pronounced minimum as well (represented by a blue circle in Figure 4.26 (a)). The model dimension obtained by means of SSR minimum (19 PCs) differs drastically from the dimension given by the eigenspectrum (1 PC). This discrepancy is not entirely expected since particles appear to have a lower impact in the UV absorbance spectra, based on the previous evaluation. To evaluate the impact of this discrepancy on the nitrite predictions, PCR models are performed including 1 PC and 19 PCs. Figure 4.26 (b)

displays the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set for the referred dimensions. The figure demonstrates several deviations from the linear model. Samples affected by improper mixing and abnormal operation appear to present more deviations from the model. However, by increasing the model dimension up to 19 PCs the referred deviations appears to be minimized.

4.3.1.7 CASE (3.1) – Micro Filtration (0.7 μm), non-Dilution and NO$_2^-$ stock-solution addition (5 ml)

Lastly, the third evaluation is based on samples subjected to filtration and NO$_2^-$ stock-solution addition, however without the dilution step. The goal is to evaluate the effect of saturation in the UV spectra. The particles effect remains minimized due to the filtration (0.7 μm) step.

Outliers indicated in the previous Section for case 3 are removed again. The spectral data are then primarily analysed by PCA. The singular values of the covariance matrix are displayed in Figure 4.27 (a). The variance of the first PC correspond to a value of 7.104(*10^5). In the second PC the value of the variance decreases considerably to 1.717(*10^5), followed by the third PC that presents a value of 0.355(*10^5). PC1, PC2 and PC3 correspond to 74%, 18% and 4%, respectively, of the total variance. PC1 and PC2 together correspond to 92% of the total variance on the eigenspectrum. Thus, the eigenspectrum suggests a gradual decrease of importance of each orthogonal component and that two principal components dominate all the rest, which suggests that saturation has a larger effect in the UV absorbance spectra, comparing to the effect of particles.
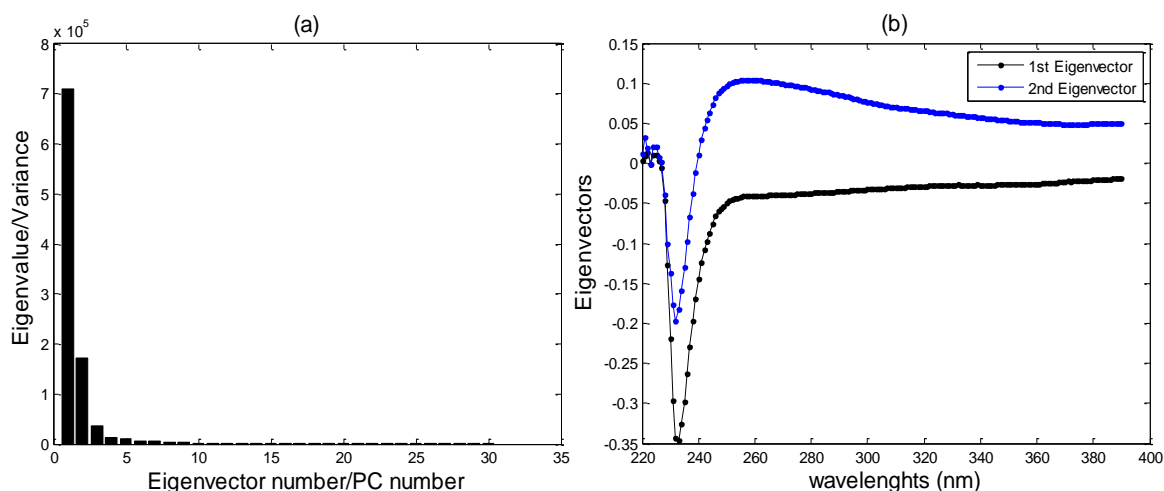
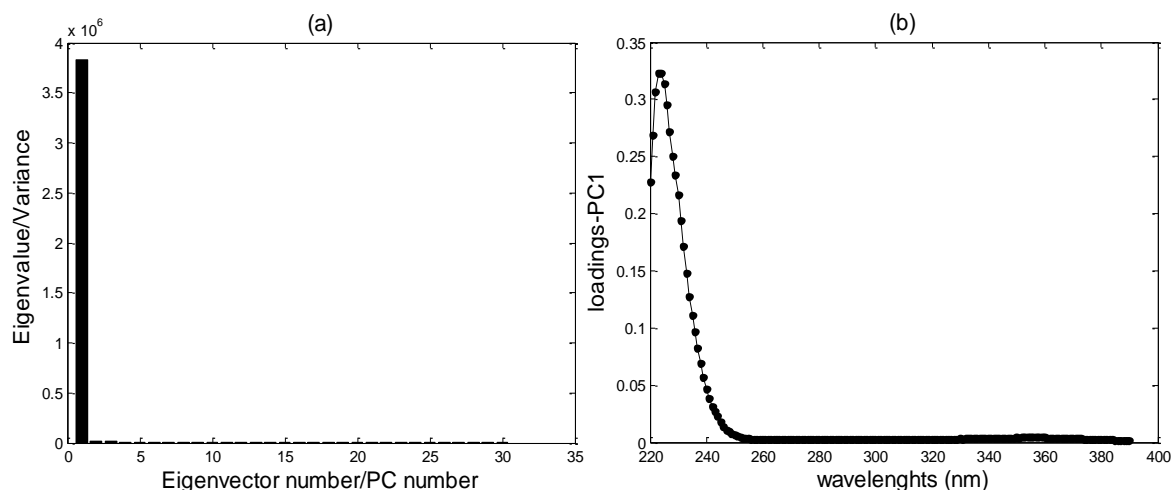
Figure 4.27. Case 3.1. (a) Eigenvalues corresponding to the first 30 Principal Components; (b) Effect of original variables (loadings) upon 1st and 2nd PC

The first two components are displayed in Figure 4.27 (b), respectively, showing that the spectra vary in a band between 220 nm and 240 nm for the first eigenvector and for the second

eigenvector vary simultaneously in a band between 220 nm and 240 nm and a band around 260 nm. Nitrite absorbances in the range between 220 nm and 240 nm are expected. However, absorbance at 260 nm are unexpected and could indicate the effect of organic compounds on the spectra

A PCR is made to predict nitrite concentrations by means of two PCs in the PCA model. Figure 4.28 and Figure 4.29 show the scores for each spectrum as a function of sample index and predictions of nitrite vs. the measured concentrations for a two dimensions (2 PCs) model, respectively.
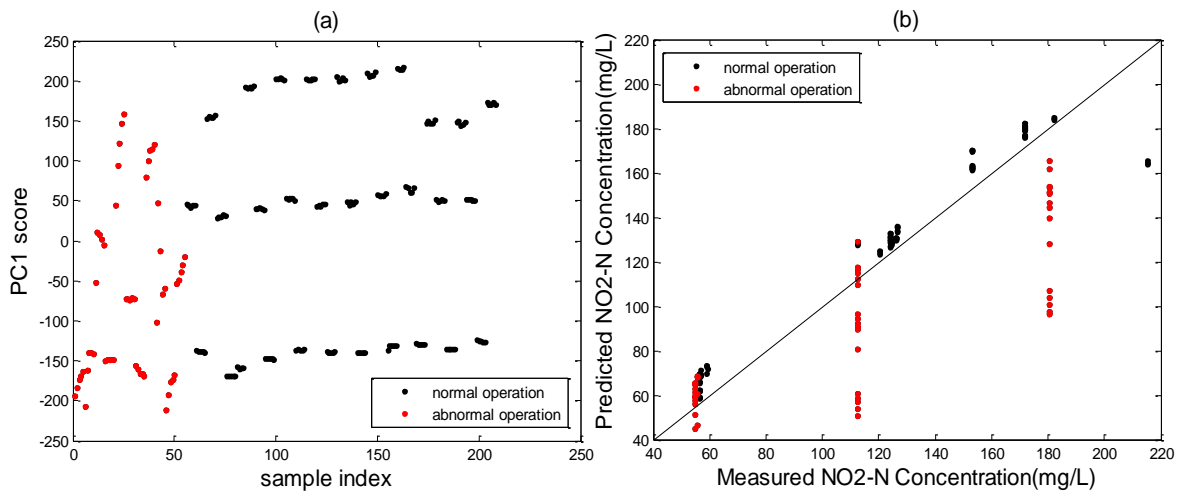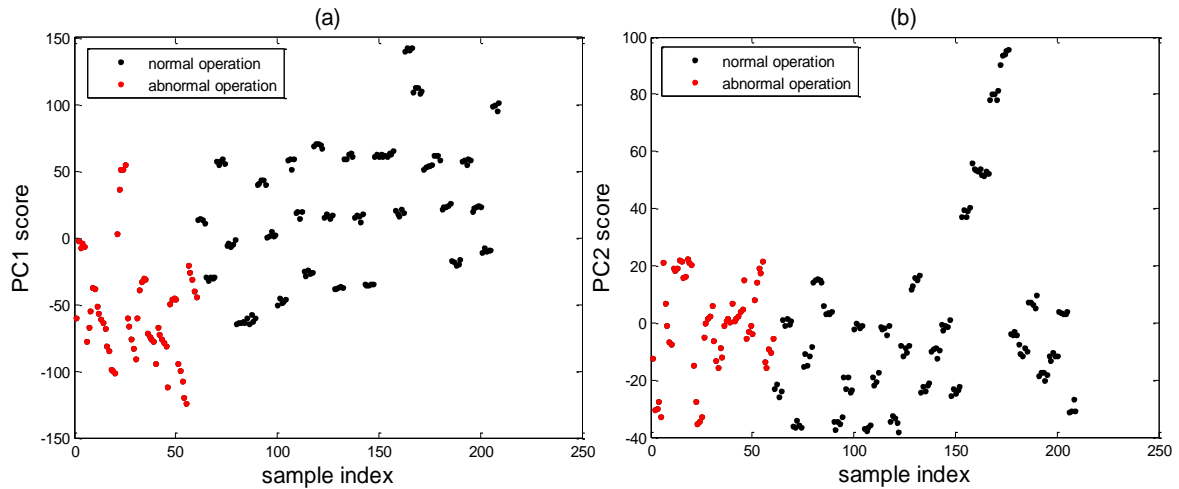

Figure 4.28. Case 3.1. (a) PC1 and (b) PC2 scores as a function of sample index after outliers removal
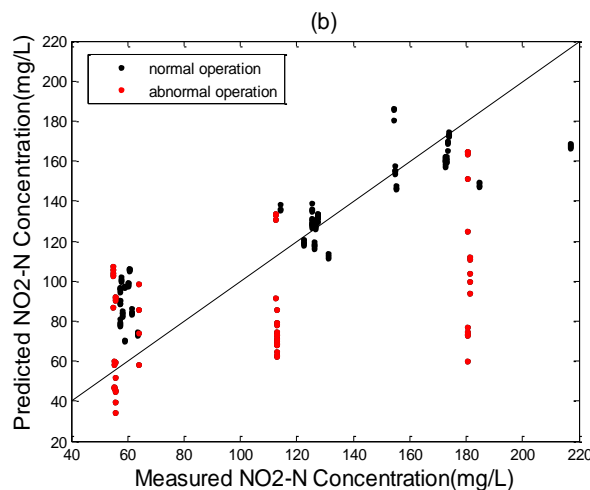
Figure 4.29. Case 3.1. Nitrite Concentrations - predicted vs. measured with 2 PCs model after outliers removal

Analysing the figure above it seems that the two dimensional PCR model fit with poor accuracy to the results which is coherent with the presence of saturation effect due to non-dilution and samples acquired from unstable nitrification process with poor agitation. Nevertheless, the results above are performed without cross-validation. Subsequently, a LOOCV is performed to evaluate

the best number of PCs for the PCR model. The number of significant principal components (PCs) was obtained from the minimum residual error (minimum SSR), as displayed in Figure 4.30 (a).
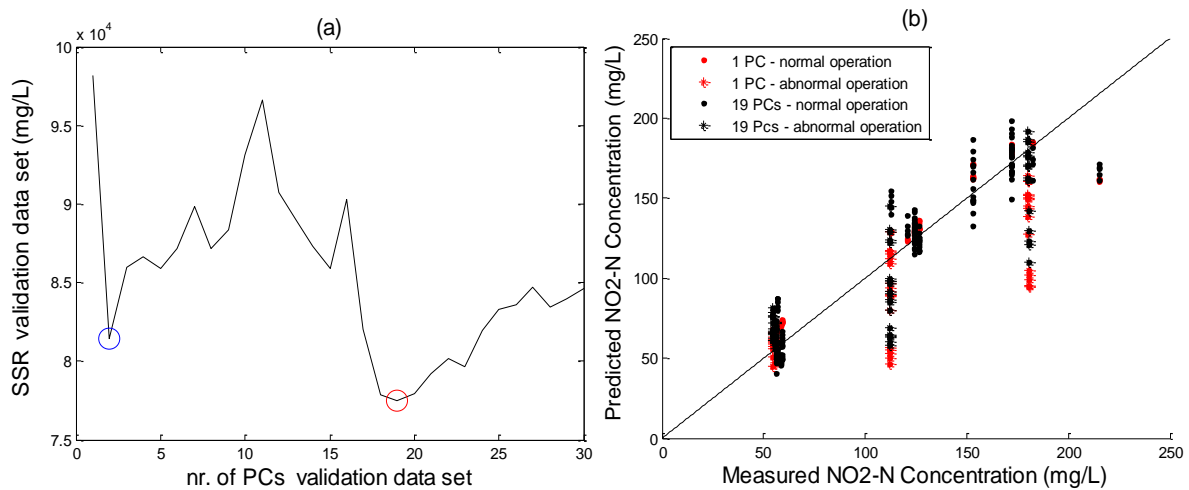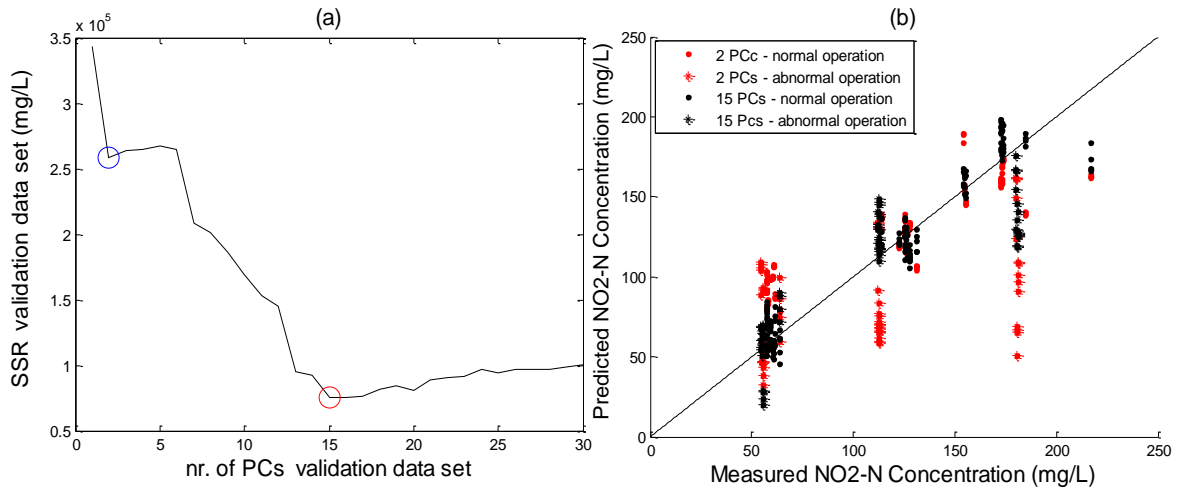

Figure 4.30. Case 3.1. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing Figure 4.30 (a), it can be seen that the minimum SSR corresponds to 15 dimensions (represented by a red circle in Figure 4.30 (a)), thus the PCR model for the cross validation should include 15 PCs. The model dimension obtained by means of SSR minimum (15 PCs) differs drastically from the dimension given by the eigenspectrum (2 PCs). Even so, for 2 PCs the SSR presents a pronounced minimum as well. This discrepancy is explained as an effect of saturation. Figure 4.30 (b) shows the predicted nitrite concentrations as a function of measured nitrite concentration for the validation data set. To evaluate the impact of the referred discrepancy on the nitrite predictions, PCR models are performed including 2 PCs and 15 PCs. The figure demonstrates some deviations from the linear model, which could be explained by the non-linear saturation effect. Samples affected by improper mixing or abnormal operation appear to lead to the poorest fit.

4.3.1.8 Summary of Results for cases (1.1), (2.1) and (3.1)

Comparing Figure 4.26 (case 2.1 - unfiltered and diluted samples) with Figure 4.23 (case 1.1-filtered and diluted samples) it is possible to infer that the effect of particles show a pronounced impact on nitrite concentration estimation. The main difference from case 1.1 to case 1.2 is the need to increase the dimensionality of the model from 12 PCs to 19 PCs, based on the SSR minimum.

Comparing Figure 4.30 (case 3.1- filtered and non-diluted) with Figure 4.23 (case 1.1- filtered and diluted) it is possible to infer that saturation appears to have a great effect on nitrite concentration estimation. Based on SSR minimum, the dimensionality of the model increases drastically from 12 PCs (case 1.1) to 15 PCs (case 3.1).

The described results for case (1.1), (2.1) and (3.1) are different from those obtained in case (1), (2) and (3). These results appear to be a consequence of the inclusion of data with an atypical behaviour corresponding to samples with improper mixing and/or corresponding to the restart of the nitrification reactor. Improper mixing could lead to sensor malfunction, since sedimentation of suspended biomass could occur in the sensor measuring path. In addition, during the start-up period there appear to exist some disturbances/failures in the nitrification reactor. Thus, the UV absorbance measurements could also be influenced due to the possible presence of suspended biomass in higher concentrations than for the stable nitrification process, which could lead to back-scattering. Unfortunately, the available data cannot be used to test these hypotheses.

Under the above conditions the UV absorbance measurements appear to be limited. Incomplete mixing together, or not, with abnormal operation of the nitrification reactor appears to have a great impact on the nitrite estimation model which leads to poor accuracy in the model fitting.

4.3.2 ESTIMATION OF NITRATE (NO$_3^-$-N) CONCENTRATIONS

PCR modelling for nitrate estimation is started with evaluation of data set 1, corresponding to normal operation and female urine. Thereafter, data set 3 including female and male urine is analysed. Data set 1 and 3 are described in detail in Section 4.1.

4.3.2.1 CASE (1.1.1) - Micro Filtration (0.7 µm), Dilution (1:10) and NO$_2^-$ stock-solution addition (5 ml)

As in the previous evaluation for nitrite concentrations, the evaluation for nitrate is started with samples subjected to filtration, dilution and NO$_2^-$ stock-solution addition. The saturation effect and the particles effect are minimized due to the dilution (1:10) step and to the filtration (0.7 µm) step, respectively.

The spectral data are then primarily analysed by PCA. The singular values of the covariance matrix are displayed in Figure 4.31 (a). The eigenspectrum suggests that a single principal component explains most of the information in the spectral data. The variance of the first PC corresponds to a value of 3.906(*10^6) and the variance of the second PC decreases considerably to a value of 0.021(*10^6). PC1 represents 99% of the total variance on the eigenspectrum. The first component is displayed in Figure 4.31 (b), showing that the spectra vary in a band between 220 nm and 240 nm, which corresponds to the range of absorbance for nitrate.
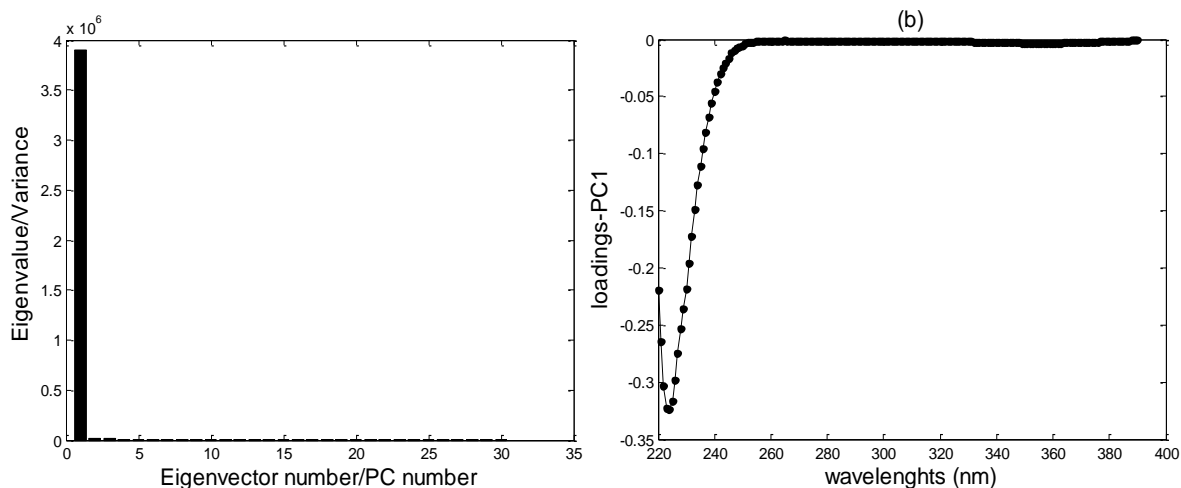
Figure 4.31. Case 1.1.1. (a) Eigenvalues corresponding to the first 30 Principal Components; (b) Effect of original variables (loadings) upon 1st PC.

A PCR is made to predict nitrate concentrations by means of a single PC in the PCA model. The spectral measurements are grouped in sets of 5, each group corresponding to a certain nitrate concentration. Figure 4.32 (a) and (b) shows the score for each spectrum as a function of sample index and predictions of nitrate vs. the measured concentrations, respectively.
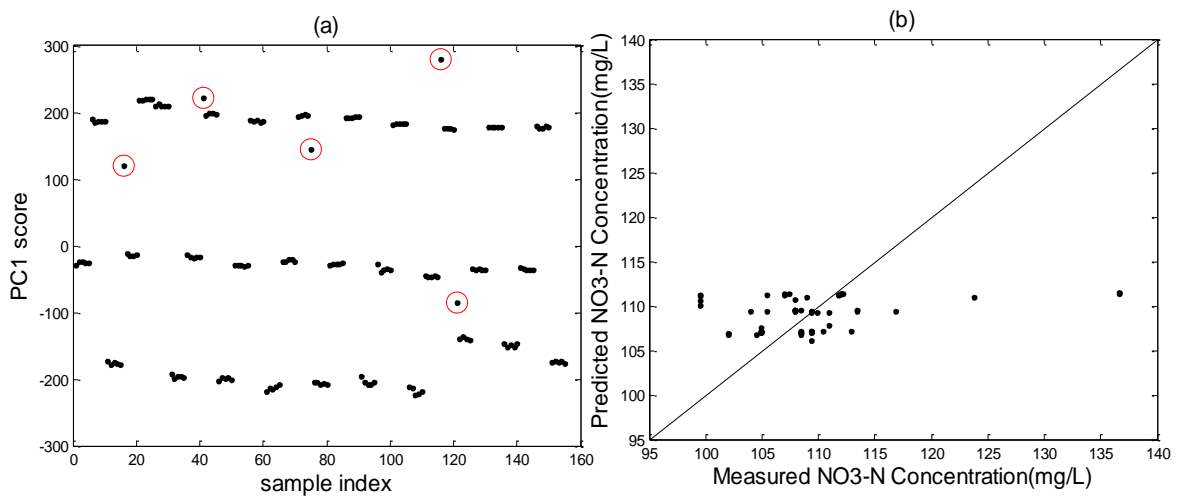
Figure 4.32. Case1. (a) PC1 score as a function of sample index; (b) Nitrate Concentrations - predicted vs. measured with 1-PC model

Five (5) outliers were thereafter removed and PCR model is repeated for the remaining samples. The results are shown in Figure 4.33.
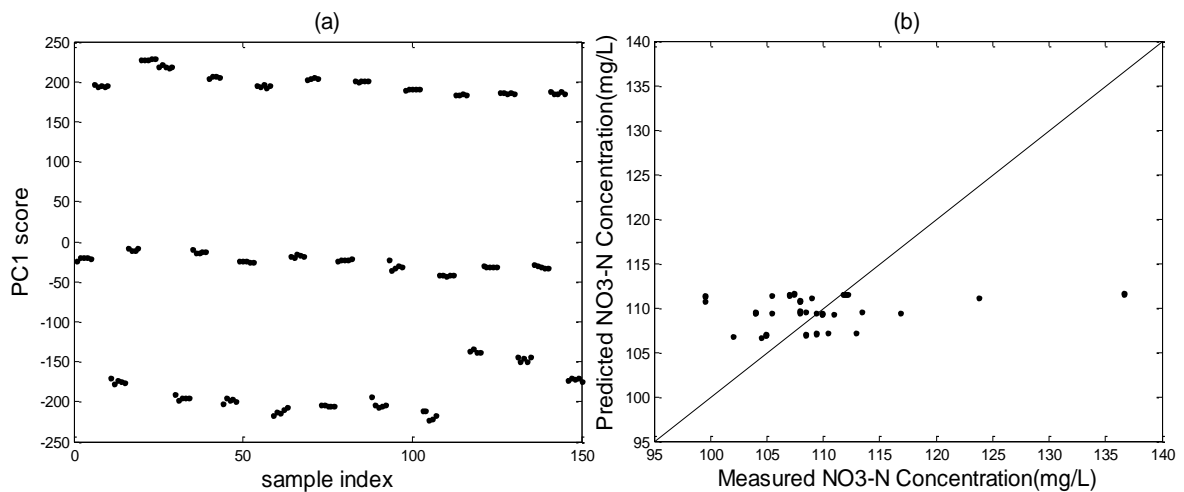
Figure 4.33. Case 1.1.1. (a) PC1 score as a function of sample index after outliers removing; (b) Nitrate Concentrations - predicted vs. measured with 1-PC model after outliers removing

The above figure demonstrates several deviations from the linear model, which could be explained by the limited range of nitrate concentration in the evaluated samples. However, the results above are performed without cross-validation. A LOOCV is then performed to choose the best number of PCs for the PCR model. The number of significant principal components (PCs) is obtained from the minimum residual error (minimum SSR), as displayed in Figure 4.34 (a).
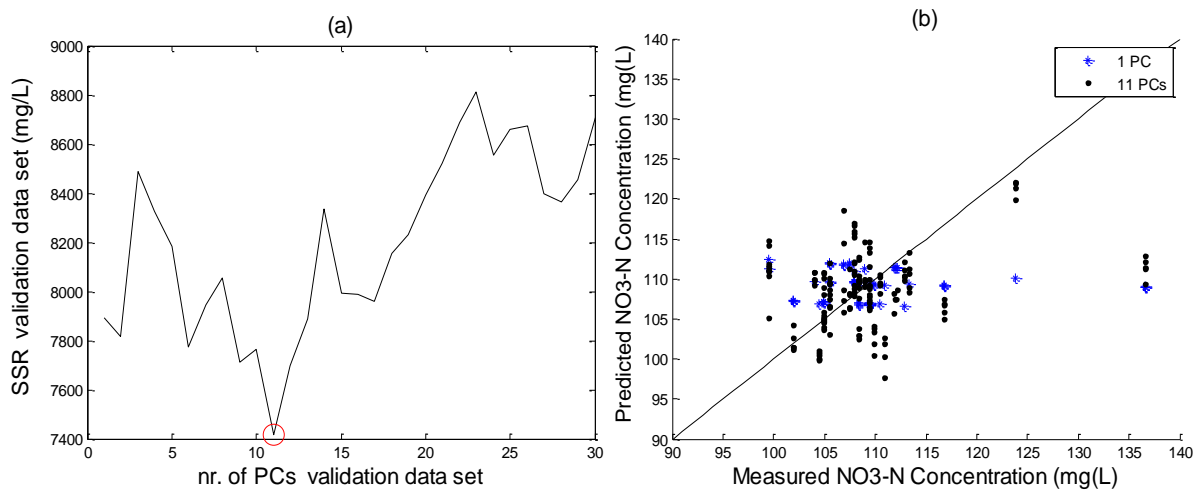
Figure 4.34. Case1. (a) SSR as a function of model dimensions for the validation data set; (b) Predicted Concentration as a function of Measured Concentration for the validation data set

Analysing Figure 4.34 (a), it can be seen that the minimum SSR corresponds to 11 dimensions, thus the PCR model should include 11 PCs as well. The model dimension obtained by means of SSR minimum (11 PCs) differs drastically from the dimension given by the eigenspectrum (1 PC). This discrepancy is expected due to limited range of nitrate concentrations, which appear to have a large effect in the UV absorbance spectra. Figure 4.34 (b) displays the predicted nitrate concentrations as a function of measured nitrate concentration for the validation data set. To evaluate the impact of the referred discrepancy on the nitrate estimations, PCR models are performed including 1 PC and 11 PCs. The figure demonstrates several deviations from the linear model, which could be explained by the limited range of nitrate concentration in the evaluated samples. The same behaviour could be observed analysing Figure 4.33 (b).

To demonstrate the above assumptions, data set 3 is now used as it contains a wider range of nitrate concentrations. The PCR modelling step and LOOCV are repeated. The results are shown in Figure 4.35.
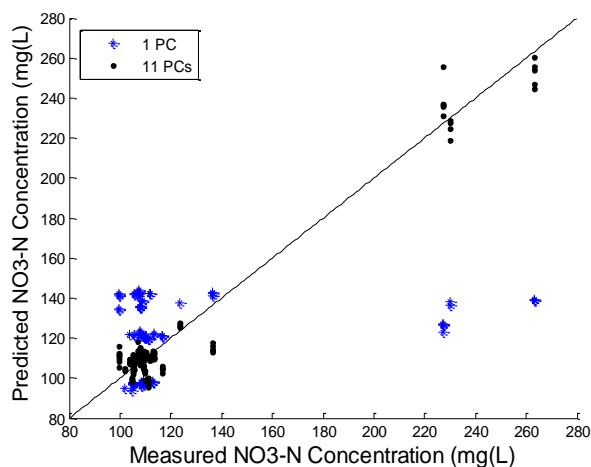
Figure 4.35. Case1.1.1. Predicted Nitrate Concentration as a function of Measured Nitrate Concentration for the validation data set with 1-PC and 11-PCs

Analysing Figure 4.35 it is possible to infer that by increasing the range of nitrate concentrations in the model, the deviations from the linear model decrease. In addition, increasing the dimensionality of the model, from 1 PC up to 11 PCs, appears to have a positive impact on nitrate concentrations estimation.

The focus of the present work is nitrite concentrations estimation. As a result, the obtained data contain limited information regarding the nitrate concentrations and their effect on the spectra. Thus, the creation of a linear regression model that links the acquired spectra to nitrate concentrations is challenged. Experimentally induced variations in the latter would allow the development of a better estimation model for nitrate and could be achieved, for instance, by adding nitrate stock-solutions with different concentrations to the samples, identically to what was performed for nitrite. This is considered as an idea for further study.

5 CONCLUSIONS AND FUTURE DEVELOPMENTS

The detection of contaminants in water sources is raising global concerns regarding their potential risks to the environment and human health, limiting development and increasing poverty in emerging and developing countries. Different approaches to wastewater management are required for different geographical regions and different stages of economic governance depending on the capacity to manage wastewater. One promising approach is the separate collection and treatment of urine. Source-separation of urine represents a treatment opportunity for removal of these compounds from water sources and preventing significant environmental exposure. However, realisation of urine separation still faces challenges. Ammonia volatilization is one that presents bigger concerns, since significant nitrogen losses occur. In addition, inconvenient odour problems arise. Thus, ammonia stabilization is imperative to prevent environmental pollution and negative effects on human health. Importantly, it is essential to retain nitrogen in solution for a later recovery.

Biological nitrification is a suitable pre-treatment to stabilise urine. The major concern regarding biological nitrification is the accumulation of the inhibitory intermediate nitrite, which can lead to process breakdown. Currently, there is no on-line measurement available which could help to prevent nitrite accumulation events. The application of a chemometric model, such as principal component regression (PCR), is a promising approach for estimation of nitrite in urine nitrification processes. UV spectrophotometry together with PCR were used in this work in order to develop a model for estimation of nitrite concentrations. Monitoring of nitrite concentrations in the urine nitrification reactor could avoid system failures and, possibly, a total system breakdown.

An evaluation of particles and saturation effects on the UV absorbance spectra was performed. In the first part of this work the saturation effect on the UV absorbance spectra was characterized empirically. The goal of the experiment was to determine an ideal dilution where no effect of saturation was present and Lambert-Beer's law was applicable. The latter was investigated by means of a linear regression. An ideal dilution, where saturation is absent and Lambert-Beer's law conditions are present, was achieved and determined to correspond to a 10% urine fraction. In the second part of this work, both the effect of particles and the effect of saturation in the UV absorbance spectra were investigated. The aim was *(i)* to clarify how strong the effect of particles is, and *(ii)* to evaluate how saturation influences the nitrite estimation. Nitrite concentrations were estimated through the use of a PCR model.

The model developed in this study estimates the nitrite concentrations accurately under stable operation conditions. A PCR model performs well when the saturation effect and the particles effect are minimized through dilution and filtration (ideal case). The accuracy of the PCR model remains good when particles are present. This allows to conclude that particles have a minimal impact on nitrite concentration estimation. Furthermore, the PCR model remains acceptable when saturation effect is present. However, a large increase in model dimension is required. This allows to conclude that effect of saturation shows a large impact on nitrite concentration estimation.

Despite the above positive result, the model remains sensitive to operational conditions and to considerable changes in the samples composition, revealing lower performance in these circumstances. The quality of mixing appears particularly crucial in the sensor's ability to sense nitrite.

The work presented in this thesis represents a step forward in the biological urine nitrification process stabilization. However there are still some aspects that can be studied in further research. A similar experimental approach could be carried out to evaluate the effect of changing nitrate concentrations. This would lead to a full understanding of the sensor's estimation abilities under different operational conditions. Furthermore, it is necessary to evaluate whether good mixing conditions can be achieved under realistic conditions when using the sensor as an on-line instrument.

6 REFERENCES

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

Bock, E. & Wagner, M., 2006. Oxidation of Inorganic Nitrogen Compounds as an Energy Source. In: *The Prokaryotes. A Handbook on the Biology of Bacteria. Ecophysiology and Biochemistry.* 3rd ed. Singapore: Springer, p.1107.

Bouvier, J.-C. *et al.*, 2008. On-Line Monitoring of Nitrate and Nitrite by UV Spectrophotometry in a SBR Process Used for the Treatment of Industrial Wastewaters. *International Journal of Chemical Reactor Engineering,* 6, pp. 1-19.

Burgess, C., 2007. The Basis of Spectrophotometric Measurements. In: *UV-Visible Spectrophotometry of Water and Wastewater. Techniques and Instrumentation in Analytical Chemistry.* Startforth, England: Elsevier, p.372.

Corcoran, E. *et al.*, 2010. *Sick Water? The Central Role of Wastewater Management in Sustainable Development. A Rapid Response Assessment,* United Nations Environment Programme: UN-HABITAT, GRID-Arendal.

Dahlen, J. *et al.*, 2000. Determination of nitrate and other water quality parameters in groundwater from UV/Vis spectra employing partial least squares regression. *Chemosphere,* 40, pp. 71-77.

Drolc, A. & Vrtovsek, J., 2010. Nitrate and nitrite nitrogen determination in waste water using on-line UV spectrometric method. *Bioresource Technology,* 101, pp. 4228-4233.

Etter, B., Hug, A. & Udert, K. M., 2013. *Total Nutrient Recovery from Urine – Operation of a Pilot-Scale Nitrification Reactor.* WEF/IWA International Conference on Nutrient Removal and Recovery, 2013, 28-31 July, Vancouver, p. 4.

EPA, 2002. *Nitrification,* USA: United States Environmental Protection Agency. [online] Available at:http://water.epa.gov/lawsregs/rulesregs/sdwa/tcr/upload/nitrification.pdf [Accessed 06.2014].

Fernández, J., Nieto, P., Muñiz, C. & Antón, J., 2014. Modelling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecological Engineering,* 68, pp. 80-89.

Ganrot, Z., 2005. *Urine processing for efficient nutrient recovery and reuse in agriculture,* Ph.D. thesis, Göteborg University, Göteborg, Sweden, p.170.

Haimi, H., Mulas, M., Corona, F. & Vahala, R., 2013. Data-derived soft-sensors for biological wastewater treatment plants: An overview. *Environmental Modelling & Software,* 47, pp. 88-107.

Hastie, T., Tibshirani, R. & Friedman, J., 2001. Linear Methods for Regression. In: *The Elements of Statistical Learning; Data Mining, Inference and Prediction.* 4th ed. Stanford University, USA: Springer, p.739.

Heinonen-Tanski, H., Sjöblom, A., Fabritius, H. & Karinen, P., 2007. Pure human urine is a good fertiliser for cucumbers. *Bioresource Technology,* 98, pp. 214-217.

Henze, M. & Comeau, Y., 2008. Wastewater Characterization. In: *Biological Wastewater Treatment: Principles, Modelling and Design.* IWA Publishing, p.511.

Hyötyniemi, H., 2001. *Multivariate Regression. Techniques and tools,* Helsinki University of Technology, Control Engineering Laboratory, Helsinki, report 125, p.207.

Jolliffe, I., 2002. *Principal Component Analysis.* 2nd ed. University of Aberdeen, UK: Springer, p.487.

Juan, A. & Tauler, R., 2003. Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta,* 500, pp. 195–210.

Langergraber, G., Fleischmann, N. & Hofstädter, F., 2003. A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater. *Water Science and Technology,* 47 (2), pp. 63-71.

Larsen, T. A. & Gujer, W., 1996. Separate Management of Anthropogenic Nutrient Solutions (Human Urine). *Water Science and Technology,* 34(3-4), pp. 87-94.

Lienert, J. & Larsen, T. A., 2009. High Acceptance of Urine Source Separation in Seven European Countries: A Review. *Environmental Science & Technology,* 44, pp. 556-566.

Macherey-Nagel, 2014. Glasfaserfilter. [online] Available at: http://www.mn-net.com/StartpageFiltration/Filterpapers/Glasfaserfilter/tabid/10497/language/en-US/Default.aspx [Accessed 07. 2014].

Maesschalck, R. D. *et al.,* 1999. The development of calibration models for spectroscopic data using Principal Component Regression. *Internet Journal of Chemistry,* 2, pp. 19-36.

HACH, 2012. Manual DR 2800™ Portable Spectrophotometer, p. 4.

S::can, 2007. Manual s::can spectrometer probe, p.62.

Mark, H. & Workman, J., 2007. Calculating the Solution for Regression Techniques: Part 4 - Singular Value Decomposition. In: *Chemometrics in Spectroscopy.* 1st ed. New York: Academic Press, pp. 127-129.

Mathworks, 2014. MathWorks - MATLAB and Simulink for Technical Computing. [online] Available at: http://www.mathworks.com [Accessed 06. 2014].

Maurer, M., Pronk, W. & Larsen, T., 2006. Treatment processes for source-separated urine. *Water Research,* 40, pp. 3151-3166.

Melitta, 2014. Melitta® Original Coffee Filters - Melitta. [online] *International.melitta.de. Available at: https://international.melitta.de/en/Melitta-Original-Coffee-Filters-618.html [Accessed 07. 2014].*

Mevik, B. & Wehrens, R., 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software,* 18, pp.1-24.

Otto, M., 2007. What is Chemometrics? In: *Chemometrics. Statistics and Computer Application in Analytical Chemistry.* 2nd ed. Weinheim: Wiley-VCH, pp. 1-11.

Paulo, A. M., 2008. *Monitoring of Biological Wastewater Treatment Processes using Indirect Spectroscopic Techniques,* Ph.D. thesis, Universidade do Minho, Minho, Portugal, p.135.

Pons, M., Bonté, S. & Potier, O., 2004. Spectral analysis and fingerprinting for biomedia characterisation. *Journal of Biotechnology,* 113, pp. 211–230.

Rieger, L. *et al.,* 2008. Long-term evaluation of a spectral sensor for nitrite and nitrate. *Water Science & Technology,* 57(10), pp. 1563-1569.

Rieger, L. *et al.*, 2004. Spectral in-situ analysis of NO2, NO3, COD, DOC and TSS in the effluent of a WWTP. *Water Science & Technology,* 50(11), pp. 143-152.

Sedlak, R., 1991. *Phosphorus and nitrogen removal from municipal wastewater: principles and practice.* 2nd ed. New York: The Soap and Detergent Association, p. 88.

Smith, L., 2002. *A tutorial on Principal Components Analysis.* USA: Cornell University, pp.1-26.

Thomas, O. & Theraulaz, F., 2007. Aggregate Organic Constituents. In: *UV-Visible Spectrophotometry of Water and Wastewater. Techniques and Instrumentation in Analytical Chemistry.* Startforth, England: Elsevier, pp. 89-114.

Thomas, O., Theraulaz, F., Agnel, C. & Suryani, S., 1996. Advanced UV Examination of Wastewater. *Environmental Technology,* 17, pp. 251-261.

Udert, K., 2002. *The Fate of Nitrogen and Phosphorus in Source Separated Urine,* Ph.D. thesis, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland, p. 117.

Udert, K., Fux, C., Münster, M., Larsen, T. A., Siegrist, H., Gujer, W., 2003 b). Nitrification and autotrophic denitrification of source-separated urine. *Water Science and Technology,* 48(1), pp. 119-130.

Udert, K., Larsen, T. & Gujer, W., 2006. Fate of major compounds in source-separated urine. *Water Science and Technology,* 54(11), pp. 413-420.

Udert, K. M., Larsen, T. A., Biebow, M. & Gujer, W., 2003 a). Urea hydrolysis and precipitation dynamics in a urine-collecting system. *Water Research,* 37, pp. 2571-2582.

Udert, K. & Wächter, M., 2012. Complete nutrient recovery from source-separated urine by nitrification and distillation. *Water Research,* 46, pp. 453-464.

Upstone, S. L., 2000. Ultraviolet/Visible Light Absorption Spectrophotometry in Clinical Chemistry. In: P. Ltd., ed. *Encyclopedia of Analytical Chemistry.* Beaconsfield, UK: John Wiley & Sons Ltd, pp. 1699–1714.

VUNA, 2013. *Eawag: VUNA - Nutrient harvesting in South Africa: VUNA.* [online] Available at: http://www.vuna.ch [Accessed 07.2014].

Wold, S., 1995. Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems,* 30, pp. 109-115.

APPENDIXES

APPENDIX I. Matlab Code: Primary Saturation Effect Experiment

```matlab
clc; clear all; close all

% load DilutionData1
% templegend = {'10x','2x','1x','20x','5x','100x','50x','10x',
'(5/3)x','2.5x','1.25'};
% dilutions = [10 2 1 20 5 100 50 10 5/3 2.5 1.25]; %c=1:11

load DilutionData2
templegend = {'5x',
'25x','2.5x','(5/3)x','10x','100x','50x','1.25x','20x','1'};
dilutions = [5 25 2.5 5/3 10 100 50 1.25 20  1]; %c=1:10

x = 1./dilutions; %Compute urine fractions
x = repmat(x,5,1);
x = x(:) ;
x_sort=sort(x); %Sort urine fractions
[x_sort,index]=sort(x);
index;

y = spectra_out(:,[23:5:43]);%Choose the wavelenght (ex:23=220nm)
y_sort=y(index,:);

c=10;%Change the value of c according data set 1 or 2 and desired
urine fractions

%Plot for chosen dilutions (urine fractions)vs Absorbance
figure
hold on
for j=1:c
  plot(x_sort((j-1)*5+1),y_sort((j-1)*5+1,:),'.')
  axis([0 1 0 6000])
  xlabel('Urine fraction','Fontsize',14, 'Fontname', 'Arial');
  ylabel('Asorbance (Abs/m)','Fontsize',14, 'Fontname', 'Arial')
  title('(g) Included urine fractions: ','Fontsize',14, 'Fontname','Arial')
  newlegend={'220','225','230','235','240'};
  legend(newlegend)
end
for j=1:c
plot(x_sort ((j-1)*5+(2:5))',y_sort((j-1)*5+(2:5),:),'.')
end

%Equation to obtain the matrix X=[ones x]
X=[ones(c.*5,1) x_sort(1:c.*5)];

%Equation for the matrix [b; m], b=intercept, m=slope
p=((X'*X)^-1)*X'*y_sort(1:c.*5,:)

%Definition of b(intercept) and m(slope) to the regression line
b1= p(1,1) ; b2=p(1,2); b3= p(1,3); b4= p(1,4); b5= p(1,5);
m1= p(2,1) ; m2= p(2,2) ;m3= p(2,3) ; m4= p(2,4) ; m5= p(2,5);
v=0:1;
z1=v.*m1+b1; z2=v.*m2+b2; z3=v.*m3+b3; z4=v.*m4+b4; z5=v.*m5+b5;
hold on
plot(v,z1,v,z2,v,z3,v,z4,v,z5,'k') %Plot regression line

yp=X*p; %Equation that defines the predicted values of absorbance
```

```
%Plot for Measured Absorbance vs Predicted Absorbance (for chosen
urine fractions)
figure, hold on
str={'b.','g.','r.','c.','m.'};
for k=1:5
plot(y_sort(1:5*c,k),yp(1:5*c,k),str{k})
xlabel('Measured Absorbance (Abs/m)','Fontsize',14,'Fontname', 'Arial');
ylabel('Predicted Asorbance (Abs/m)','Fontsize',14, 'Fontname', 'Arial')
title('(a) Included urine fractions: ','Fontsize',14, 'Fontname',
'Arial')
newlegend={'220','225','230','235','240'};
legend(newlegend)
end
R =get(gca,'Xlim');
plot(R,R,'k-');


e=y_sort(1:c.*5,:)-yp; %%Compute the prediction error

% Measured Absorbance vs Prediction Errors(for chosen urine fractions)
figure
hold on
for j=1:c
  axis([0 6000 -2000 2000])
  plot(y_sort ((j-1)*5+1,:)',e ((j-1)*5+1)' ,'.')
  xlabel('Measured Absorbance (Abs/m)','Fontsize',14, 'Fontname', 'Arial')
  ylabel('Prediction errors (Abs/m)','Fontsize',14, 'Fontname', 'Arial')
  hold on
  plot(0:6000,0,'k-');
  newlegend={'220','225','230','235','240'};
  legend(newlegend)
end
for j=1:c
  plot(y_sort ((j-1)*5+(2:5),:)',e ((j-1)*5+(2:5))' ,'.')
end
```

APPENDIX II. Matlab Code: Particles and Saturation Effect Experiment

```matlab
close all, clear all, clc

% Pre Processing

load SamplesData4

SelectSamples = [1:240] ;%SamplesData2 and SamplesData4
% SelectSamples = [1:235] ;%SamplesData12
lambda = [220:390];

SelectSamples = setdiff(SelectSamples,[1:60 76 86:95 111 145 166 186
191 226:240]); %setdiff remove "undesired" samples %SamplesData4
% SelectSamples = setdiff(SelectSamples,[1:55 81:90 91 171 221:235]);
%SamplesData12
% SelectSamples = setdiff(SelectSamples,[1:60 61 86:95 176 191 221
226:240]); %SamplesData2


% outliers=[11 36 70 91 111 116] %outliers (SamplesData 4) %original
outliers: 76 111 145 166 186 191
% outliers=[21 101] %outliers (SamplesData 12) %original outliers: 91
171
% outliers=[1 41 46 101 116 146] %outliers (SamplesData 2) %original
outliers: 61 176 191 221

D=spectra_out(SelectSamples,21:191)';

X=D; %Define X data matrix [m x n] (n=rows and m=columns)
    %m=Variables(waveleghts from 220nm(21) up to 390nm(191) = 171
    spectras)
    %n=observations(samples)

[m n]=size(X) %Define m and n

mn=mean(X,2); %Compute the mean

figure,
 hold on,
    plot(lambda,mn,'c.','markersize',10);legend('mean') %plot mean of
    the original data
    plot(lambda,X,'k-') %plot original data
    plot(lambda,mn,'c.','markersize',10);legend('mean')
    xlabel('wavelenghts (nm)','Fontsize',14,'FontName','Arial');
    ylabel('Absorbance (Abs/m)','Fontsize',14,'FontName','Arial');
    title('(a)','Fontsize',14,'FontName','Arial')

X=X-repmat(mn, 1, n); %Compute centred absorbance

figure, plot(lambda, X, 'k-')%plot centred absorbance
xlabel('wavelenghts (nm)', 'Fontsize',14,'FontName','Arial');
ylabel('Centred Absorbance (Abs/m)','Fontsize',14,'FontName','Arial');
title('(b)','Fontsize',14,'FontName','Arial')

Y=X'/sqrt(n-1); %Create matrix Y


%Singular Value Decomposition (SVD): Y=USV'
[U,S,V]=svd(Y);
```

```
var=diag(S).*diag(S); %Compute covariance matrix
figure, bar(var(1:30), 'k') %Plot Eigenspectrum including 30PCs
xlabel('Eigenvector number/PC number','Fontsize',14,'FontName','Arial');
ylabel('Eigenvalue/Variance','Fontsize',14,'FontName','Arial');
title('(a)','Fontsize',14,'FontName','Arial')

var_p=var/sum(var)*100; % %Compute covariance matrix in %
% figure, bar(var_p(1:30), 'b')%Plot Eigenspectrum including 30PCs
% xlabel('Eigenvector number/PC number','Fontsize',12);
ylabel('Eigenvalue/Variance(%)','Fontsize',12);
% title('Eigenspectrum')

% Principal Component Analysis

PC=V'*Y;%Compute scores (PC=T)
figure,plot(PC(1,:),'k.')%plot PC1
% hold on, plot(outliers,PC(1,outliers),'ro','Markersize',14) %plot
outliers
xlabel('sample index','Fontsize',14,'FontName','Arial');
ylabel('PC1 score','Fontsize',14,'FontName','Arial');
title('(a)','Fontsize',14,'FontName','Arial');

figure,plot(PC(2,:),'b.')%plot PC2
% hold on, plot(outliers,PC(2,outliers),'ro','Markersize',14) %plot
outliers
xlabel('sample index','Fontsize',14,'FontName','Arial');
ylabel('PC2 score','Fontsize',14,'FontName','Arial');
title('(b)', 'Fontsize',14,'FontName','Arial');

figure, plot(PC(1,1:end),'k.') %plot PC1 with different colours for
data set 1 (black)
hold on, plot(PC(1,1:60),'r.') %plot PC1 with different colours for
data set 2 (red)
xlabel('sample index','Fontsize',14,'FontName','Arial'); ylabel('PC1
score','Fontsize',14,'FontName','Arial');
title('(a)','Fontsize',14,'FontName','Arial');
legend('normal operation', 'abnormal operation')

figure, plot(PC(2,1:end),'k.')
hold on, plot(PC(2,1:60),'r.')
xlabel('sample index','Fontsize',14,'FontName','Arial');
ylabel('PC2 score','Fontsize',14,'FontName','Arial');
title('(b)','Fontsize',14,'FontName','Arial');
legend('normal operation', 'abnormal operation')

% figure,plot(PC(1,:),0,'.','markersize',15) %plot PC1
% xlabel('PC1 score'); title('Projection onto 1st Principal Component
(score plot)')
% figure,plot(PC(2,:),0,'.','markersize',15)%plot PC2
% xlabel('PC2 score'); title('Projection onto 2nd Principal Component
(score plot)')

figure,plot(PC(1,:),PC(2,:),'k.','markersize',12)%plot PC1 vs PC2
xlabel('PC1'), ylabel('PC2'),
title('Projections onto first 2 Principal Components')

figure, plot(lambda, V(:,1),'k.-','markersize',15)
xlabel('wavelenghts (nm)','Fontsize',14,'FontName','Arial');
ylabel('loadings-PC1','Fontsize',14,'FontName','Arial');
title('(b)','Fontsize',14,'FontName','Arial')
% title('effect of original variables (loadings) upon PC1');
```

```matlab
% figure, plot(lambda,V(:,2),'r.-','markersize',15),
% xlabel('wavelenghts','Fontsize',14,'FontName','Arial');
% ylabel('loading-PC2','Fontsize',14,'FontName','Arial');
% title('(b)','Fontsize',14,'FontName','Arial')
% % title('effect of original variables (loadings) upon PC2');

figure, plot(lambda, V(:,1),'k.-','markersize',12), hold on,
plot(lambda,V(:,2),'b.-','markersize',12),
legend('1st Eigenvector','2nd Eigenvector')
xlabel('wavelenghts (nm)','Fontsize',14,'FontName','Arial');
ylabel('Eigenvectors','Fontsize',14,'FontName','Arial');title('(b)','F
ontsize',14,'FontName','Arial')
% %
% figure, plot(V(:,1),V(:,2),'.')%plot influence of each original
variables upon PC1 and PC2 % good to see a trend
% xlabel('effect (PC1)','Fontsize',14,'FontName','Arial'),
% ylabel('effect (PC2)','Fontsize',14,'FontName','Arial')
% % title('Load plot')

% Principal Component Regression

T=PC'; %=U*S %Create matrix T

nPCs=1
for d=1:nPCs
Ts=T(:,1:d); %define T for d scores
Vs=V(:,1:d); %define V for d scores

[r c]=size(Ts);
Z=[ones(r,1) Ts]; %Equation to obtain the matrix Z=[ones Ts]

Ym=[NO2_conc_out(SelectSamples)']; %matrix Ym=[NO2_measured]

p=((Z'*Z)^-1)*Z'*Ym;%Equation for the matrix [b; m]

Yp=Z*p; %Equation for the predicted NO2 concentrations

figure, plot(Ym(1:end,:)',Yp(1:end,:)','k.'),
% hold on, plot(Ym(outliers,:),Yp(outliers,:),'ro','Markersize',14)
%plot outliers
% %
% for i=1:n
%      text(Ym(i)+1/2,Yp(i)+1/2,num2str(i));
% end
hold on, R =get(gca,'Xlim');plot(R,R,'k-'); %Compute y=x
% plot(0:250, 0:250,'k-')
xlabel('Measured NO2 Concentration(mg NO2-N/l)','Fontsize',14,'FontName',
'Arial'),
ylabel('Predicted NO2 Concentration(mg NO2-N/l)','Fontsize',14,'FontName',
'Arial');
title('(b)','Fontsize',14,'FontName','Arial')
% title('Predicted Concentration as a function of Measured Concentration')

figure, plot(Ym(1:end,:)',Yp(1:end,:)','k.') %plot PC1 with different
colour for data set 1
hold on,plot(Ym(1:60,:)',Yp(1:60,:)','r.') %plot PC1 with different
colour for data set 2
hold on, R =get(gca,'Xlim');plot(R,R,'k-');
xlabel('Measured NO2 Concentration(mg NO2-N/l)','Fontsize',14,'FontName',
'Arial'),
ylabel('Predicted NO2 Concentration(mg NO2-N/l)','Fontsize',14,'FontName',
```

```matlab
'Arial');
legend('normal operation', 'abnormal operation')
title('(b)','Fontsize',14,'FontName','Arial')
end


e=Ym-Yp;%Compute the prediction error
figure, plot(Ym(:,:),e(:,:),'.')

for i=1:n
    text(Ym(i)+1/2,e(i)+1/2,num2str(i));
end

hold on, R=2*max(Ym); plot(0:R,0:0,'k-'); %Compute y=0
xlabel('Measured NO2 Concentration','Fontsize',12),
ylabel('Prediction error','Fontsize',12)



% Leave-One-Out Cross Validation

clear all, close all, clc

load SamplesData4

SelectSamples = [1:240] ;%SamplesData2 and SamplesData4
% SelectSamples = [1:235] ;%SamplesData12

% SelectSamples = setdiff(SelectSamples,[1:60 86:95 226:240]);
%SamplesData2 and SamplesData4
% SelectSamples = setdiff(SelectSamples,[1:55 81:90 221:235]);
%SamplesData12

N=length(SelectSamples);

nPCs = 30; % Define number of PCs

SampleIndex = nan(N,1) ;
for j=1:N/5
    SampleIndex((1:5)+(j-1)* 5) = j;
end

j_index=1:N/5;

% outliers = [];
outliers = [76 111 145 166 186 191]; %SamplesData4
% outliers = [91 171]; %SamplesData12
% outliers = [61 176 191 221]; %SamplesData2




outliers_index=[];
for k = 1:length(outliers)
    outliers_index = [outliers_index
find(SelectSamples==outliers(k))];
end

SelectSamples = setdiff(SelectSamples,outliers);

SampleIndex = SampleIndex(setdiff(1:N,outliers_index),:);
```

```
A=spectra_out(SelectSamples,21:191)'; %matrix for the selected spectra

C=NO2_conc_out(SelectSamples)'; %matrix for the selected NO2 concentrations

X=A;

Ymval_matrix =[];% Ymval_matrix = zeros(length(SampleIndex),nPCs);  %
predefine matrix for Measured NO2 concentrations storage, for the
validation data set
Ypval_matrix = [];% Ypval_matrix = zeros(length(SampleIndex),nPCs);  %
predefine matrix for Predicted NO2 concentrations storage, for the
validation data set
error_matrix = [];% error_matrix = zeros(length(SampleIndex),nPCs);  %
predefine matrix for error storage

for jj = 1:length(j_index)

    j = j_index(jj);

    cal_temp = SampleIndex~=j ; cal = cal_temp~=0;
    val = SampleIndex==j ;

    Xcal = X(:,cal);
    Xval = X(:,val);
    val_length = size(Xval,2);

    Ccal = C(cal);
    Cval = C(val);


% 1st. Centering/scaling of calibration and validation data

[m n]=size(Xcal); %Define m and n
mn=mean(Xcal,2); %Calculate the mean
Xcaln=Xcal-repmat(mn, 1, n);
[o p]=size(Xval);
Xvaln=Xval-repmat(mn, 1, p);


% 2nd. PCA modelling (SVD)

Ycal=Xcaln'/sqrt(n-1); %Create matrix Ycal

[Ucal,Scal,Vcal]=svd(Ycal); %Singular Value Decomposition (SVD): Y=USV'

var=diag(Scal).*diag(Scal); %Compute covariance matrix


PCcal=Vcal'*Ycal';%Calculate scores (PC it's equal to T on PCR)

Yval=Xvaln'/sqrt(n-1); %Create matrix Yval

PCval=Vcal'*Yval';%Calculate scores (PC it's equal to T on PCR)

    % 3. Regression for calibration and validation data (PCR)

Tcal=PCcal'; %=Ucal*Scal; %Create matrix T
Tval=PCval'; %Create matrix T

Ymval_matrix_small = zeros(val_length,nPCs);
```

```
Ypval_matrix_small = zeros(val_length,nPCs);
error_matrix_small = zeros(val_length,nPCs);

for d=1:nPCs

    Tscal=Tcal(:,1:d); %define Tcal for d scores
    Vscal=Vcal(:,1:d); %define Vcal for d scores

    Tsval=Tval(:,1:d); %define Tval for d scores
    Vsval=Vval(:,1:d); %define Vval for d scores

    [r c]=size(Tscal);
    Zcal=[ones(r,1) Tscal]; %Equation to obtain the matrix Z=[ones
    Tscal]

    [s f]=size(Tsval);
    Zval=[ones(s,1) Tsval]; %Equation to obtain the matrix Z=[ones
    Tsval]

    Ymcal=Ccal; %matrix Ymcal=[NO2_measured]for the calibration set;
    Ymval=Cval; %matrix Ymcal=[NO2_measured]for the validation set
    Ymval_matrix_small(:,d) = Ymval;

    pcal=((Zcal'*Zcal)^-1)*Zcal'*Ymcal; % Equation for the matrix [b; m]

    Ypcal=Zcal*pcal; %Equation for the predicted NO2 concentrations
    for the calibration set
    Ypval=Zval*pcal;%Equation for the predicted NO2 concentration for
    the validation set
    Ypval_matrix_small(:,d) = Ypval;

    eval=Ymval-Ypval;%Equation for the prediction error
    error_matrix_small(:,d) = eval;

    end

    Ymval_matrix = [Ymval_matrix; Ymval_matrix_small];
    Ypval_matrix = [Ypval_matrix; Ypval_matrix_small];
    error_matrix = [error_matrix; error_matrix_small];

end

SSR=sum(error_matrix(:,1:nPCs).^2)%compute SSR(Sum of Squared Residuals)
figure, plot(1:nPCs,SSR,'k-') %choose SSR minimum to know the optimal
number of PCs
hold on, plot(2,SSR(2),'bo','Markersize',14)
hold on, plot(13,SSR(13),'ro','Markersize',14)
xlabel('nr. of PCs _ validation data set','Fontsize',14,'FontName','Arial'),
ylabel('SSR _ validation data set (mg/L)','Fontsize',14,'FontName','Arial'),
title('(a)','Fontsize',14,'FontName','Arial')

%plot Predicted Concentrations vs Measured Concentrations based on SSR
model
figure, hold on,
plot(Ymval_matrix(:,2),Ypval_matrix(:,2),'r.','Markersize',6) %plot
for the 2nd PC
% plot(Ymval_matrix(1:60,2),Ypval_matrix(1:60,2),'r*','Markersize',6)
%plot for the 2nd PC
%  plot(Ymval_matrix(:,2),Ypval_matrix(:,2),'ro','Markersize',6) %plot
for the 2nd PC
 plot(Ymval_matrix(:,nPCs),Ypval_matrix(:,nPCs),'k.')%plot for the 1st
PC
```

```
%     plot(Ymval_matrix(1:60,nPCs),Ypval_matrix(1:60,nPCs),'k*')%plot
for the 1st PC
hold on, plot(0:250, 0:250,'k-')
%  legend('2 PCc - normal operation','2 PCs - abnormal operation','15
PCs - normal operation','15 Pcs - abnormal operation'),
legend('2 PCs','13 PCs'),
%  legend('nr PCs'),
xlabel('Measured NO2-N Concentration (mg/L)','Fontsize',14,'FontName','Arial'),
ylabel('Predicted NO2-N Concentration (mg/L)','Fontsize',13,'FontName','Arial'),
title('(b)','Fontsize',14,'FontName','Arial')
```

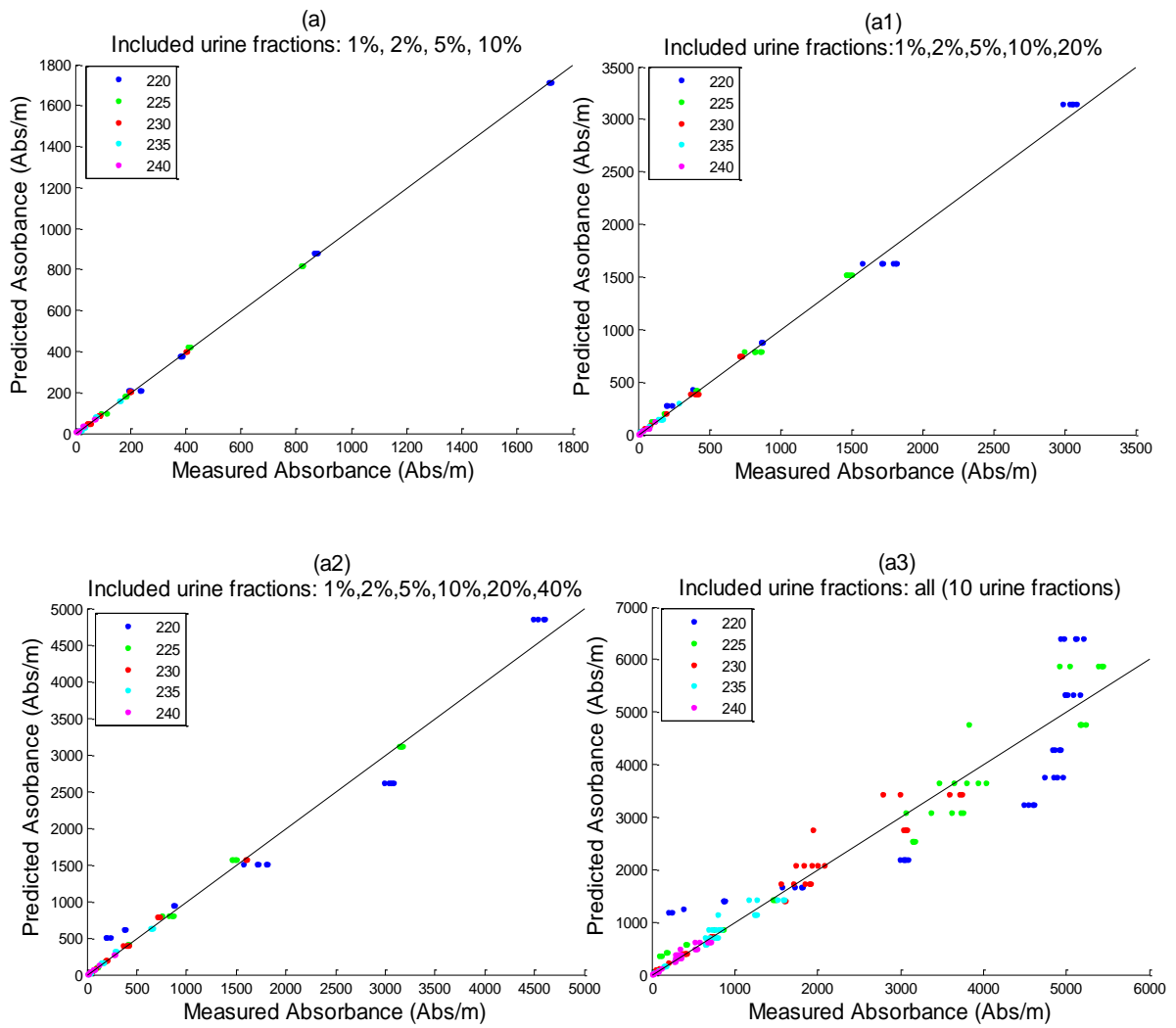**APPENDIX III. Primary Saturation Effect Experiment: Results of data set 1**

Figure A.0.1. Predicted absorbances as function of measured absorbance for the chosen wavelengths (220, 225, 230, 235, 240 nm) of data set 1