Identification of water pollution sources in wastewater collecting systems using electromagnetic spectra in the UV -VIS range

João Pedro Ferreira de Almeida
joao.pedro.ferreira.de.almeida@ist.utl.pt

Técnico Lisboa, Lisboa, Portugal

July 2016

Abstract

In the context of urban drainage, the protection of surface waters from the effects of wastewater discharges has assumed increasing importance. As a result of human activity, there is sometimes contamination of water resources due not only to the phenomenon of diffuse pollution, but also as a result of specific effluent discharges (such as those from different industries or farming and by some discharges of insufficiently treated urban wastewater.

Currently, wastewater quality evaluation is based on physicochemical and microbiological parameters (contaminants indicators). However, these parameters do not allow the identification of the source of contamination in the effluents. On the other hand, these tests are time-consuming analysis and do not provide real-time results, therefore they cannot be used as a warning system in bathing areas or at the entrance of a wastewater treatment plant (WWTP).

The present study evaluates whether the principal component analysis (PCA) applied to the spectra in the range Ultraviolet-Visible (UV-Vis) allows the real time identification of wastewater quality variations or types, in particular due to the increase of organic load.

The results show that the UV-Vis spectra obtained from wastewater samples includes information that can be used to monitor wastewater quality variations in urban drainage systems. The application of PCA presents the potential to monitor wastewater quality using UV-Vis range spectra, since this approach allows identification of anomalies or changes in effluent quality matrix those results from the discharge of industrial effluents. The suitability of the model for PCA anomaly detection has been demonstrated for two types of industrial effluents. From the results obtained it is believed that, if the spectra of known industrial effluents are included in the PCA model, the scores obtained in this model can be used to identify, in real time, this type of discharges.

Keywords: UV-Vis Spectra, principal component analysis, industrial wastewater, pollution sources.

1 Introduction

The protection of surface waters from the effects of wastewater discharges has assumed increasing importance in the context of urban drainage. In 2009, about 50% of the world population were living in urban areas, the accelerated growth of large urban centers, associated with disordered development has caused profound changes in the natural water cycle, which affected the life quality of the populations and generated impacts on water resources. As a result of those pressures, contamination of water resources is very common, caused by contaminated effluent discharges, some specific (industrial and some urban wastewater discharges insufficiently treated that still exist), but mainly by diffuse pollution (agriculture and, in some cases , agro-livestock), which calls into question the execution of the quality objectives for the environment.

To respond to this increasing pressure, developed countries have specific legislation to try to limit the impact on the receiving means, such as the Water framework directive of the European Union.

Currently, water quality evaluation is based on physicochemical parameters and the presence of fecal bacteria contamination indicators (Escherichia coli (E. coli) and intestinal enterococci (EI)). However, these parameters and bacteria do not allow the identification of effluent contamination source. Moreover, these parameters do not achieve results in real time, so it cannot be used as a warning system at a bathing site or in an input from a WWTP. Sometimes, tributary pollution loads into the processing stations are relevant and have high toxicity, which leads to different issues, such as toxicity shock phenomenon, inhabitation [1] which may result in shutting down the stations.

Dissolved compounds in a wastewater sample have specific characteristics of light absorbance, depending on their molecular structure, the wavelength and the concentration of the compound in the sample [2], this approach is therefore a possible option to monitor effluents.

The main application of spectroscopic is to use the UV-Vis spectra response (eg absorbency) to estimate a parameter [3]. With the support of statistical techniques, the obtained spectral information allows the estimation of water quality parameters, such as BOD (biochemical oxygen demand), TSS (total suspended solids), once known the expected relationship between the spectra shape and the value obtained for the analytical parameter in the laboratory. An in-situ spectrophotometer was used successfully for the determination of various chemical parameters in wastewater, such as COD (chemical oxygen demand), TSS, nitrates and nitrites [4]. Using the PCA multivariate analysis applied to the UV-Vis spectra [5] TSS were monitored in a municipal WWTP.

PCA is a high potential tool in urban drainage, it reduces the information to a limited number of independent factors, studies the most striking features of the data and highlights what relates (or sets) the various magnitudes in question.

In this context, the use of simple, rapid and relevant techniques such as the use of spectrophotometry should be considered when thinking of a method of detection and alert to waterlines pollution sources [3].

The aim of this work is to evaluate the potential to identify improper discharges in waterlines, using UV-Vis spectra in real time, to provide a warning system against industrial discharges. The spectra would be analyzed using mathematical procedures that characterize and distinguish the usual effluent, allowing the identification of a discharge (improper).

To sum up, the aim is to propose a methodology based on principal components analysis, and once set the UV-Vis spectra, comparing them with the previously collected samples and analyze the identification potential in terms of pollution load.

2 Methods

2.1 Case study

In order to test the applicability and the possibility to monitor the water quality of wastewater through UV-Vis spectra, the basin that drains into the subsystem Frielas was identified. By the time of this study it was operated by SIMTEJO – Saneamento Integrado dos Municípios do Tejo e Trancão, S.A. (do Grupo Águas de Portugal) currently LVT. SIMTEJO collects and treats wastewater from the Municipalities of Amadora, Loures, Lisboa, Vila Franca de Xira and Sintra. The subsystem consists of the WWTP located in Frielas, four pumping stations and about 82 km of interceptors and outfalls. The WWTP of Frielas receives domestic effluents, rain and effluents from industrial origin, being the later quite relevant.

In one of the tributaries to the WWTP, the emissary of Rio da Costa, were already made earlier studies under spectrophotometry in wastewater systems [6] [7] so we tried to focus on the study of industrial units that contribute to this basin emissary.

It was decided to collect wastewater in a cheese factory, a slaughterhouse, a dairy and a poultry industry. These industrial effluents are characterized by high values of COD and TSS, values much higher and different than typical domestic wastewater (DWW).

2.2 Sampling strategy and spectra acquisition

The applied methodology scheme is summed up in Figure 1 . Data from previous spectral studies in Rio da costa was gathered [6] [7]. Specifically for this study were carried out two experimental campaigns to collect wastewater samples. The industrial effluents were collected in manholes immediately downstream of the industrial units. Taking into account the possibility that these industrial units may have a unit to carry out a pre-treatment to the effluent, the collections were made outside the premises, since it is important to know what debit goes to the network and not the debiting process within the facility. Was collected approximately 1L of effluent (for each sample) considered representative of the effluent produced in the industry. Several samples were collected with a time interval of 15 minutes between each one. DWW was collected at the entrance of the WWTP.

To simulate different contribution scenarios of the industrial effluent to the WWTP, laboratory composite samples were created. Composite samples (MT) were made of industrial wastewater and domestic wastewater, 75%, 50% and 25% of industrial effluent concentration. For the acquisition of the spectra was used a submersible spectrophotometer. The equipment used was a probe :: Spectro Lyser Scan Messtechnik GmbH (Vienna, Austria) having a capacity to record the absorbance for wavelengths between 200 and 750 nm (in increments of 2.5 nm).

Whenever the spectra showed noise due to signal saturation, dilutions of industrial effluent with domestic wastewater (Dil) 1:2 were made, in order to overcome this situation. The spectra were acquired again, multiplying by two the results of absorbance based on the Lambert-Beer law. These samples that were diluted were considered pure in this study, to overcome the signal saturation. At least two spectra were acquired for each sample, in order to look for outliers in measures. For this study were collected 220 spectra of domestic wastewater and 140 industrial wastewater spectra.

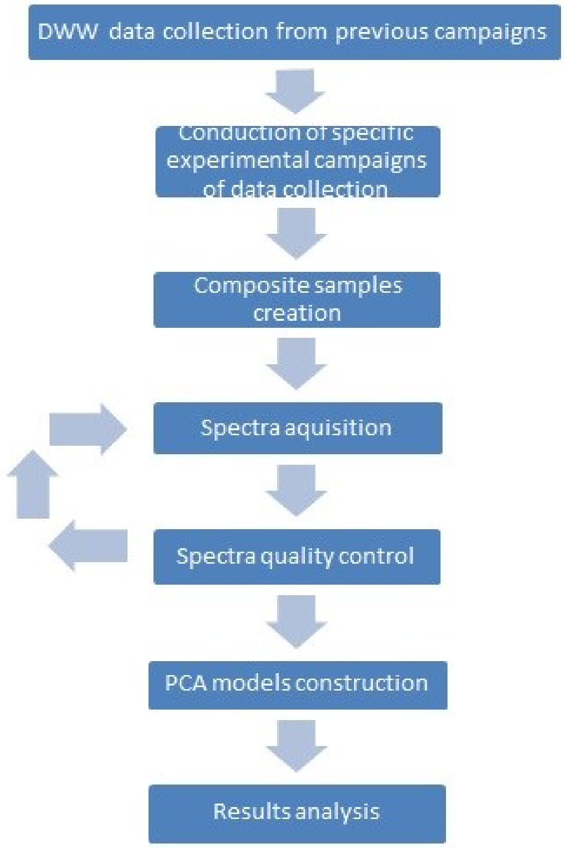

Figure 1 – Methodology applied to the study

2.3 Principal components analysis

This technique aims to detect the most striking features of data in order to try to understand what relates or differentiate the various elements of the samples in data analysis. Thus, the technique is intended to describe and represent the numerous variables starting from a smaller number of hypothetical variables called principal components (PC), orthogonal to each other, without significant loss of information contained in the set. Under these circumstances, an extensive data set can be arranged in a two-dimensional data matrix where the rows are the columns and samples are different variables (absorbances, for example). Therefore, this technique will not explain the correlations between variables, but will find linear combinations of the initial variables that explain much of the existing variation in the data, reducing the variables and describing them otherwise [8] [9]. A PCA model can be represented by the following [10] equation:

$$X(n \times p) = T(n \times d).L^T(d \times p) + E(n \times p)$$

where n represents the number of elements in the sample, p the number of variables (wavelengths in this case), d the number of PC's, T represents the scores matrix, L the loadings matrix, L^T the transposed loadings matrix and E the residuals matrix.

3 Results and discussion

3.1 PCA application

Two PCA models were built, one where we tested the spectra collected in the campaign 1, and a second model where was tested the spectra collected in the campaign 2.

The PCA models were created with DWW spectra, seeking to adapt the model to the typical matrix of DWW effluent and then to undergo this model a predictive test with industrial wastewater spectra, in order to try to understand whether the industrial effluent spectra would stand from the array of typical DWW. The difference between the two created models rely on the fact that in the first model were used not only the samples from previous studies but also samples collected in campaign 1, and in model two were used samples of DWW from previous studies and samples of DWW collected in campaign 2.

Since the data to be processed is only absorbances at different wavelengths, a single type of quantity, was selected mean centering preprocessing. As a calibration method was used contiguous blocks, the eingenvalue criteria was used to select the number of PC, whereas only the PC over eingenvalue one are relevant [11]. For this reason, and since in both models created the first two principal components capture a higher variance than 90% (Table 1) it was decided that only two PC would be considered in the models.

Table 1 – PCA Models

Model	Model Construction	Model Test	*PC*	*Eigenvalue*	Variance
	(nº of domestic wastewater spectra)	(nº of industrial wastewater spectra	(-)	(-)	(%)
1	211	39	2	1,45	99,80
2	213	56	2	1,23	99,85

3.2 Score Plots

3.2.1 Model 1

The Figure 2 shows samples scores results of cheese factory and poultry. It appears that the pure effluent samples from cheese factory are in the confidence interval limit of 95% of DWW, and its variance is almost totally explained by PC1. Regarding the composite samples, is clearly observed three clusters, corresponding to concentrations of 75%, 50% and 25% of industrial effluent. As the concentration of industrial waste decreases, the results in graph scores go towards the direction of DWW sample by decreasing scores on PC1. It can be concluded that up to a concentration of 75% of industrial effluent, it can be distinguished the dairy effluent scores compared to DWW, since the scores are characterized by a high score in PC1 and almost no score in PC2. In relation to poultry industry is notorious that it's scores location are quite different from DWW samples, although the difference is only verifiable in PC2, which explains only 0.67% of the variability of the model. It can be considered that, despite the effluent quality is similar to DWW, since the values in the PC1 scores are similar, the effluent has a component in its constitution which is clearly different from typical DWW, explained by PC2. Looking at the scores plot of composite samples, surprisingly, samples with concentrations of 75% and 50% are located virtually in the same spot of pure samples, standing out clearly from domestic wastewater. It was expected that by decreasing the concentration of the samples they walked towards the DWW. Samples at 25% concentration of industrial effluent show a more expectable behavior, reducing the score for CP2, and it is difficult to distinguish them from DWW. Thus, can be concluded that it is possible to clearly distinguish the samples of poultry industry effluent up to a concentration of 50% in industrial effluent from DWW. The fact that the scores results of pure and composite samples up to a concentration of 50% are similar may be due to the fact that the characteristic that makes them having similar scores remains until the moment that the concentration of industrial effluent is equal to or higher than the DWW.

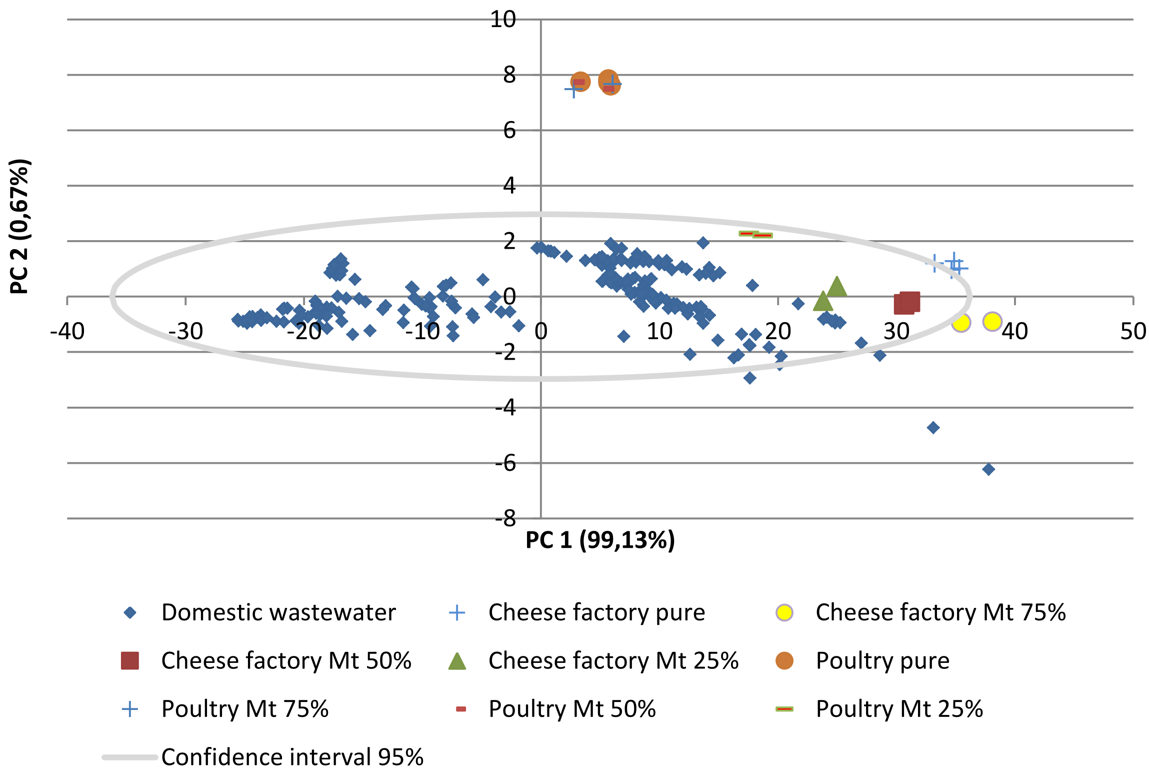

Figure 2 – Results of model 1: score-plot with cheese factory and poultry samples

The pure slaughterhouse effluent samples (diluted but considered pure in the study, Figure 3, are the ones that in the created model stand out more from the DWW matrix, having a high score in PC1, component with greater variance, standing out clearly in the model. These results suggest that the typical array of this type of industrial effluent is quite different in its composition from typical DWW. Observing the composite samples of this type of effluent there are three groups, corresponding to three different concentrations, with these three groups clearly outside the confidence interval of 95%. As the concentration decreases, the scores in either PC1 or PC2 decrease, approaching the DWW samples that were used in the model construction. Even the samples with 25% of concentration in industrial effluent can be distinguished in the model, standing outside of the confidence interval at 95%. Thus it is seen that all the samples, pure or composite differ in the scores plot in relation to DWW.

By looking the scores of dairy effluent spectra samples, it is observed that in the model created these industrial effluent samples spectra are indistinguishable from typical DWW. One of the explanations might be that the effluent collected in the campaign is not representative of typical dairy effluent, and it is only an effluent from washes and not an effluent from the existing transformation processes in this type of activity. As pure samples do not stand out, it is expected that the composite samples at different concentrations did not stand out in the scores plot, which in reality is clearly viewable in the same figure. Due to these poor results, it was decided to abandon the study of this type of industrial effluent given their poor distinction from DWW in the PCA model.

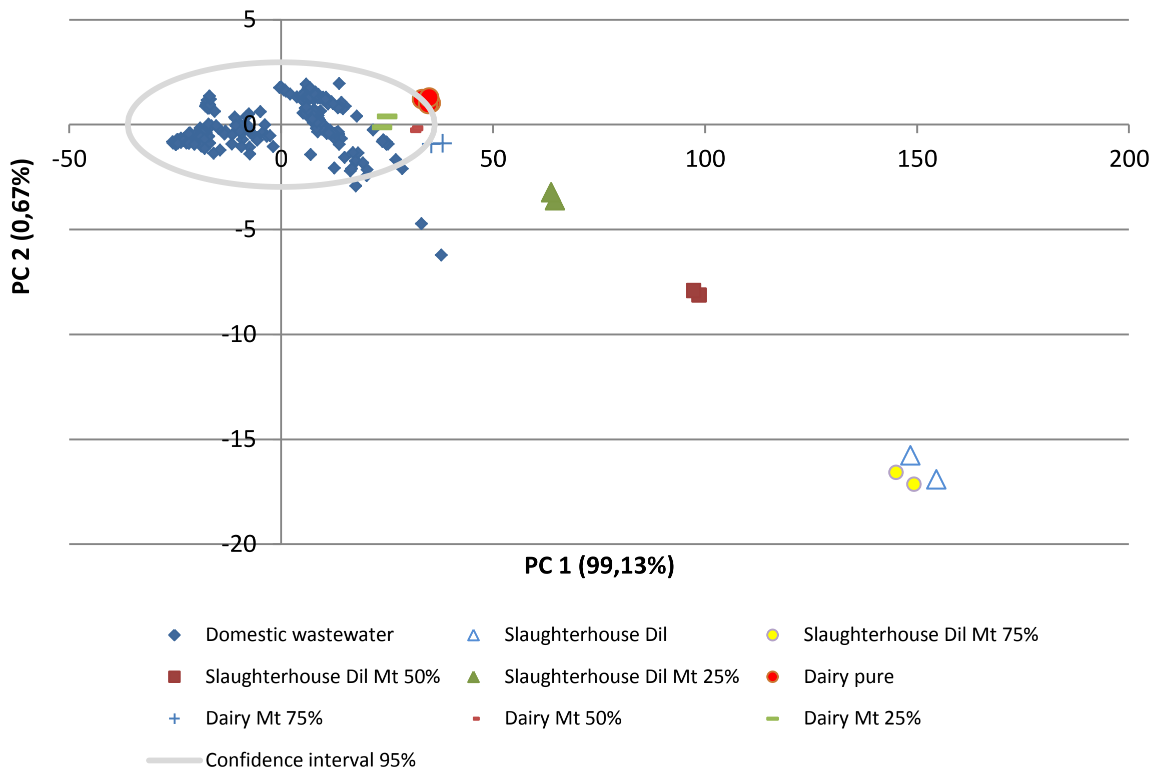

Figure 3 – Results of model 1: score-plot with slaughterhouse and dairy samples

3.2.2 Model 2

In Figure 4 it is shown the scores graph from a forecasting model built with industrial samples from the dairy industry. Starting by analyzing the diluted samples (considered to be pure in the study), the samples in model 1 were near the confidence interval of 95%, in this new model new samples are much more distant from that limit. The scores in PC2 are similar to the samples from campaign 1, on the other hand the scores at PC1 are much higher and distant from the typical DWW matrix. Looking at the scores of composite samples there is a decrease in scores in both PC, in other words, there is an approximation to the established model that is centered on the average. This approach is expected due to the reduction of industrial effluent concentration; however, the samples still have a rather different behavior from typical DWW even at 25% of concentration. Comparing to the data of the first campaign it is observed that pure samples of the first campaign show the same results as the composite samples of the second campaign, it may show that the samples taken in the second campaign, although having the same composition matrix were more concentrated than the first.

The second campaign reinforces the results of the first, that is, it shows again that pure and composite effluent from dairy industry up to 50% concentration of industrial effluent stand out in the scores plot. And that in the second campaign even concentrations of 25% exhibit different behavior of typical DWW.

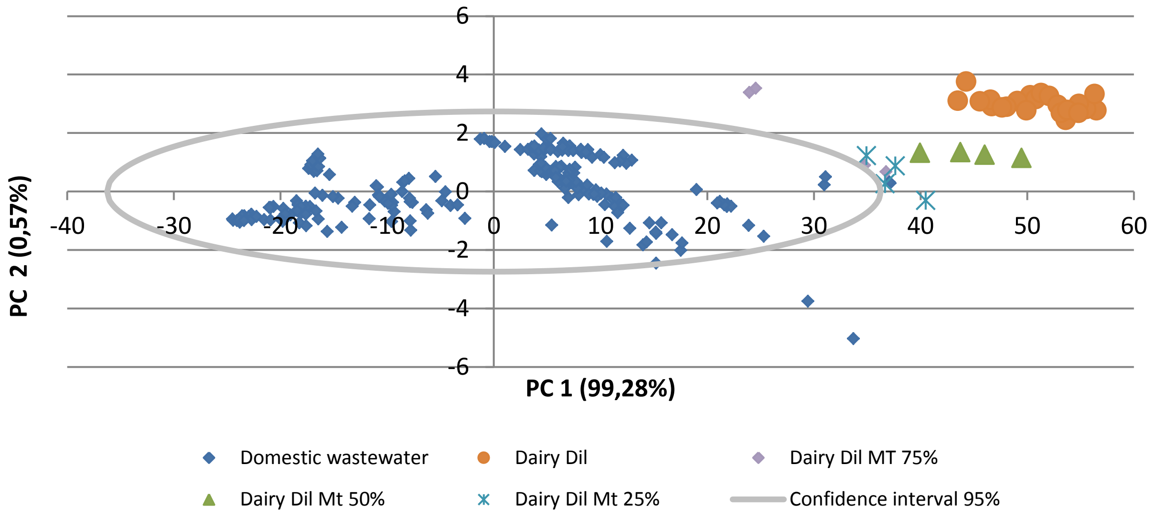

Figure 4 – Results of model 2: score-plot with dairy samples

Regarding the predictive model of industrial wasterwater from the poultry industry, Figure 4, starting by analyzing the diluted samples, looking at the results it is clearly observed they stand out in the PCA model. In terms of PC1 scores, they have similar ranges of values as DWW, however in terms of PC2 the results are on average three to four times higher than the 95% confidence interval. When analyzing composite samples of this kind of industrial wastewater with DWW, the results are similar as the pure samples scores up to concentrations of 75% and 50%, but with an approximation to the DWW samples used while building the model. Samples with concentration of 25% in industrial effluent while still stand out in the model are close to the confidence interval of 95%. Therefore, the results and behavior are very similar to those shown in the model 1, Figure 2. In short, both models show that the effluents from poultry industry are detected in scores plot up to a concentration of 50%.

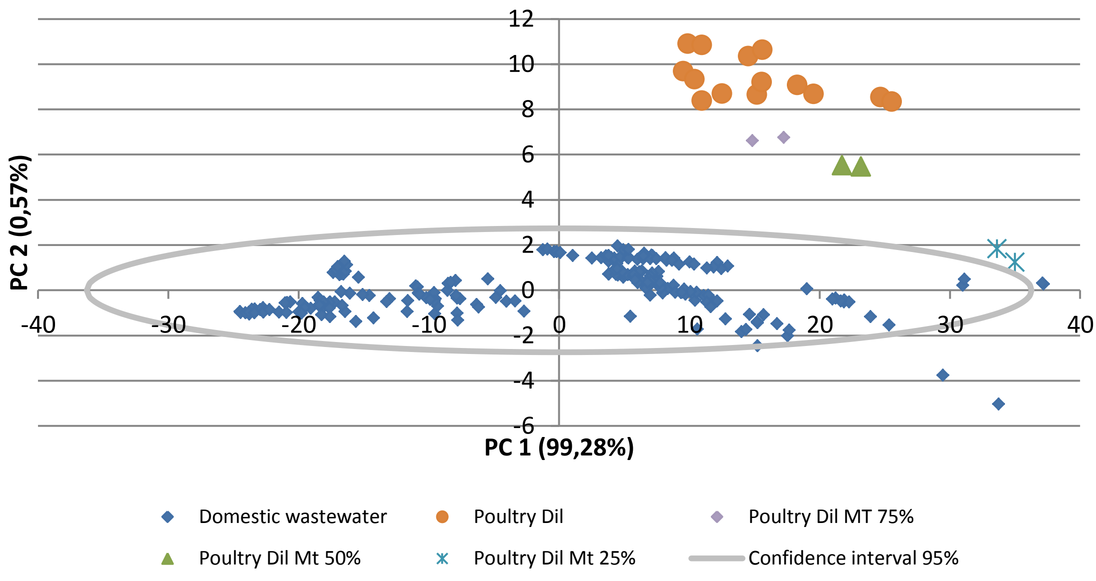

Figure 5 – Results of model 2: score-plot with poultry samples

4 Conclusions

The results presented show that the UV-Vis spectra obtained from wastewater samples include information that can be used to monitor wastewater quality variation in an urban drainage system. The application of PCA holds the potential to monitor water quality by using spectra in UV-Vis range, since it allows the identification of anomalies or changes in effluent quality matrix, like inflow flow rates with industrial pollutants, qualitatively. The power of the PCA model in detecting anomalies was proven to two PC. Through the observed results it seems that if spectra of known industrial effluents are included in a PCA model, the scores obtained in this model can be used for real-time identification of improper and abnormal discharge. This study supports the applicability of online residual water monitoring, allowing such monitoring in real time, providing a more rapid and effective way to identify the pollutant source in a waterline and the optimization of treatment processes in a WWTP; or if the pollutant load is very high and can cause some of the most common failures in treatment plant, like toxic shock, this form of monitoring may be used as a warning to perform a bypass of that industrial wastewater inflow. Although a bypass does not involve treating wastewater, in certain situations it is less onerous to do not treat this effluent than taking the risk of being forced to inactivate a WWTP for days due to this industrial pollutant. For this it is necessary to place a sensor that online collects real-time spectra effluent, and by internet connection sends the results of the spectra to a computer that compares the scores with the PCA model calibrated with samples from the typical DWW. If the real-time samples correspond in the calibrated model to an outlier an alert in real time will be given in order to take steps to minimize the problems that can be caused by the pollution load.

In conclusion, this study shows that a PCA model can provide a way to detect the inflow of a effluent with a differentiated quality matrix and can be applied to an online system that enables obtaining an indication of the effluent quality in real time.

Acronyms

BOD- Biochemical oxygen demand

COD-chemical oxygen demand

DWW- Domestic wastewater

PCA- Principal components analysis

PC- Principal component

UV-Vis- Ultraviolet-visible

TSS- Total suspended solids

WWTP- Wastewater treatment plant

5 References

[1] Garcia-Alvarez, D., Fuente, M., Vega, P., Sainz, G. (2009). Fault Detection and Diagnosis using Multivariate Statistical Techniques in a Wastewater Treatment Plant. Valladolid: National research agency of Spain.

[2] Vanrolleghem, P., Lee, D. (2003). On-line monitoring equipment for wastewater treatment. Water Science and Technology, 47, pp. 1-34.

[3] Thomas, M. (1996). Analytical Chemistry by Open Learning – Ultraviolet and Visible Spectroscopy (2 ed.). Inglaterra: John wiley and Sons, Ltd.

[4] Rieger, L., Langergraber, G., Thomann, M., Fleischmann, M. (2004). Spectral in-situ analysis of NO2, NO3, COD. Proceedings 2nd International IWA C onference on "Automation in Water Quality Monitoring C AutMoNet". Viena, Austria.

[5] Lourenço, N., Paixao, F., Pinheiro, H., Sousa, A. (2010). Use of spectra in the visible and near-mid-ultraviolet range with Principal Component Analysis and Partial Least Squares processing for monitoring of suspended solids in municipal wastewater treatment plants. Applied Spectrocopy, 64(9), 1061-1067.

[6] Brito, R. (2012). Monitoring of Hydraulic Variables and Water Quality in Urban Drainage. Phd thesis , Instituto Superior Técnico, Lisboa. (in portuguese)

[7] Pinheiro, A. (2015). Intercalibration of UV-VIS spectrophotometry in Drainage Network Monitoring. (in portuguese)

[8] Jolliffe, I. (2003). Principal Component Analysis. 2ª. Springer-Verlag Series on Statistics.

[9] Davies, T., Fearn, T. (2005). Back to basics: the principles of principal component analysis. Tom Davies Column, pp. 20-23.

[10] Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. Krishnaiaah, Multivariate Analysis (pp. 391 -420). New York: Academic Press.

[11] Matteau, M., Assani, A., Mesfioui, M. (2009). Application of multivariate statistical analysis methods to the dam hydrologic impact studies. Journal of Hydrology, 371(1), 120-128.