

# CHEMOMETRIC ESTIMATION OF WASTEWATER COMPOSITION FOR THE ON-LINE CONTROL OF TREATMENT PLANTS

Marie-Noëlle PONS\*, Jing WU<sup>‡</sup>, Olivier POTIER

*Laboratoire des Sciences du Génie Chimique, CNRS-ENSIC-INPL  
1, rue Grandville, BP 451, F-54001 Nancy cedex, France*

*<sup>‡</sup>Present address : Department of Environmental Science and Technology,  
Tsinghua University, Beijing 100084, P. R. China*

*\*Corresponding author*

Abstract : Estimation of wastewater composition (Chemical Organic Demand – COD and nitrogen-related species such as ammonia and organic nitrogen) for on-line control of treatment plants based on Principal Components Regression (PCR) and Partial Least Squares (PLS) applied to spectrometric data (turbidity, UV-visible and synchronous fluorescence spectra). Satisfactory models were obtained based on the total UV-visible spectra. The synchronous fluorescence spectra should be divided into regions of interest that could be related to urine and humic and fulvic acids. *Copyright © IFAC 2005.*

Keywords : PLS, PCA, PCR, spectrometry, wastewater, fluorescence

## 1. INTRODUCTION

Municipal wastewater is a mixture of sewage from residential and office buildings, restaurants, schools, hospitals, factories, etc. Its variations with time reflect the human activity as modified by weather conditions. The variability of wastewater is one of the causes of the difficulties met to control the treatment plants as those by activated sludge, whose performance should be maintained within strict limits. The complexity of the physical and biochemical phenomena and to the large range of time constants (from a few minutes to several days) inherent in the activated sludge process are other sources of problems for the design of efficient control strategies.

Pollution is assessed in terms of global variables such as Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), Total or Dissolved Organic Carbon (TOC or DOC), Total or Kjeldahl Nitrogen (TN or TKN). Ammonium and nitrate are some of the few species that can be

measured directly. Furthermore these measurements are time-consuming (from a couple of hours for COD, TN and TKN to several days for BOD) and may need sophisticated equipment and/or dangerous reagents.

Surrogate measurements based on spectrometry and electrochemistry (pH, conductivity, ion-selective electrodes) have been proposed to monitor in real-time the variations of wastewater characteristics.

For many years now UV-visible spectroscopy has been used as an alternate and rapid way to estimate the level of pollution of wastewater and surface water. The 200-300 nm range has been found to be of particular interest. The 254nm wavelength has been more particularly selected and correlations between the absorbance at 254nm (A<sub>254</sub>) and Chemical Oxygen Demand (Mrkva, 1983) or Total Organic Carbon (Dobbs et al., 1973) have been established. On-line devices are commercially available. By miniaturisation of the optical systems, full UV-visible spectra can now be

collected in-situ (Langergraber et al., 2002). The technique is however sensitive to turbidity and sample filtration or correction using a second wavelength in the visible range have been suggested. Saturated hydrocarbons and sugars cannot be detected by such a method. Figure 1 presents some UV-visible spectra related to municipal wastewater and some of its anthropogenic substances.

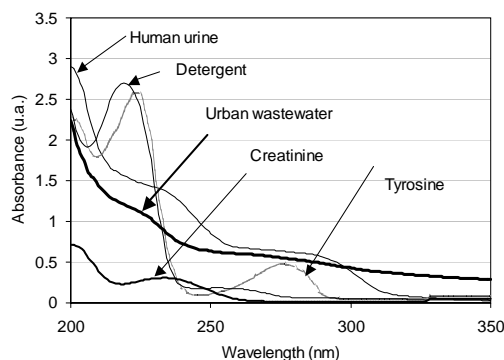


Fig. 1: UV-Visible spectra of urban wastewater and of some of its possible anthropogenic components.

Many biological compounds, including proteins, humic and fulvic substances, steroids, phenols, oils, surfactants, vitamins, etc., exhibit a characteristic fluorescent emission after excitation by visible or near-UV light (Figure 1). Fluorescence methods are much more sensitive than absorbance methods and they are much less affected by the medium turbidity. In wastewater treatment proteins should be hydrolysed prior to the biodegradation of the organic nitrogen they contain. Tartakovsky et al. (1996) have monitored the dynamics of anoxic and aerobic sequential batch reactors fed with synthetic substrates based on albumin and casein. Good correlations (correlation coefficient > 0.9) were established between wastewater BOD<sub>5</sub> and fluorescence ( $\lambda_{exc} = 248 \text{ nm}$  or  $280 \text{ nm}$ ,  $\lambda_{em} = 340 \text{ nm}$ ) by Reynolds and Ahmad (1997) and Ahmad and Reynolds (1999). This range corresponds to the tryptophan-like fluorescence.

To monitor the effect of effluent discharge to receiving water bodies, the fluorescence of Natural Organic Mater (NOM) has to be taken into account. Two regions are generally recognized, one related to fulvic-like acids ( $\lambda_{exc} \approx 330 \text{ nm}$ ,  $\lambda_{em} \approx 420 \text{ nm}$ ) and the other to humic-like acids ( $\lambda_{exc} \approx 380 \text{ nm}$ ,  $\lambda_{em} \approx 470 \text{ nm}$ ). Various studies have shown the potentialities of the method to evaluate the quality of reservoir water (Goslan et al., 2004), the contamination by untreated domestic sewage (Baker et al., 2003), farm wastes (Baker, 2002) or industrial works (Santos et al., 2000), or simply the effect of urbanization (Westerhoff and Anning, 2000). But, even with modern fluorometers, the

collection of a two-dimensional spectrum might be too time-consuming. As seen in Figure 2, most substances of interest line up in the  $\lambda_{exc}$ - $\lambda_{em}$  plane. Synchronous fluorescence has been suggested for faster analysis with  $\Delta\lambda$  values ( $\lambda_{em} - \lambda_{exc}$ ) between 20 and 60 nm (Galapate et al., 1998; Ahmad and Reynolds, 1995). For well chosen conditions the synchronous fluorograms exhibit more peaks than simple emission or excitation fluorograms.

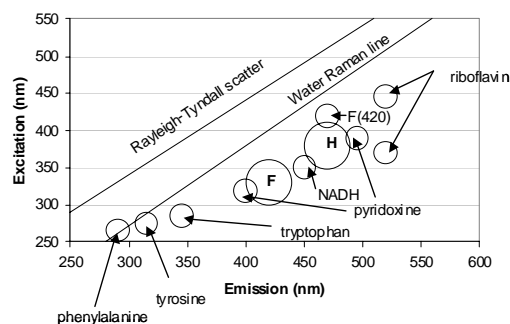


Fig. 2: Localisation of excitation-emission zones for some biological substances

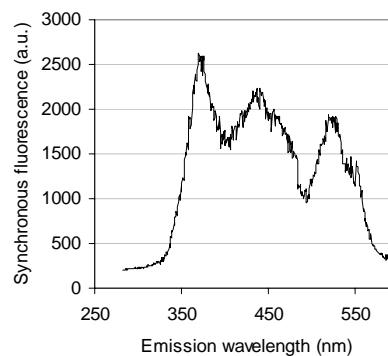


Fig. 3: Synchronous fluorescence ( $\Delta\lambda=30 \text{ nm}$ ) spectra of a wastewater sample grabbed in the primary settler

Different types of data-mining methods of these spectrometric spectra have been proposed. By spectral deconvolution of UV-visible spectra, Thomas et al. (1995) monitored the fate of anionic detergents, suspended and colloidal matter, DOC and nitrates. Similar deconvolution have been applied to estimate the temporal variability of DOC profiles along a wastewater treatment plant (Escalas et al., 2003). A Partial Least Squares (PLS)-based calibration procedure combining off-line lab analyses and in-situ measurements can be set up to improve the quantification of organic matter and nitrate in wastewater by taking into account the long-range variations of the matrix (Langergraber et al., 2003). PLS have also been proposed for fluorometric spectra from bioprocesses. Recently Morel et al. (2004) have compared the relative merits of PLS, non linear principal components regression (nPCR) and step-wise regression models (SWR) for the on-line

monitoring of an anaerobic digestion process for wastewater treatment.

In the present contribution PLS and PCR will be compared for the estimation of the characteristics of urban wastewater based on UV-visible and synchronous fluorescence spectra.

## 2. MATERIALS AND METHODS

### 2.1. Wastewater analysis

Wastewater grabbed samples have been collected on the primary settler of the Nancy-Maxéville wastewater treatment plant (350 000 PE) in the North-East part of France with a auto-sampler (Sigma 990P, Hach Company, Loveland, Colorado). This treatment unit combines sand and grease removal with primary settling.

After filtration of the samples (pore diameter  $\gg 7 \mu\text{m}$ ) dissolved COD (DCOD), ammonia and total nitrogen were determined by spectrophotometry (Hach method). Total COD (TCOD) was also measured on unfiltered samples. Turbidity (TU) was determined on a Hach DR2400 spectrophotometer and is given in FTU. The UV-visible spectra of filtrated samples were collected on an Anthelie Light (Secomam, Domont, France) spectrophotometer (wavelength stepsize = 1 nm). The synchronous fluorescence spectra were obtained either on a Jobin-Yvon JY3 (Longjumeau, France) with a 150W xenon lamp fluorometer (wavelength stepsize = 1 nm) or a Perkin-Elmer L50B with a  $\Delta\lambda=50 \text{ nm}$ .

Xlstat-Pro (Addinsoft, Paris, France) and Statbox (Grimmerlogiciels, Paris, France) toolboxes for Microsoft® Excel were used for PCA, PCR and PLS.

## 3. RESULTS

### 3.1. UV-visible spectra

Linear correlations have been obtained between the absorbance at 254 nm (A254) and wastewater characteristics (Figure 4), especially DCOD. These results confirm results from literature on the use of UV-visible spectrometry as a surrogate method for pollution estimation. However the apparent correlation between A254 and N-NH4 is due to the simultaneity of the variations of N-NH4 (mostly related to human urine in municipal wastewater) and DCOD and not to a real detection of N-NH4 by UV-visible spectrometry.

PCA was applied to sub-sets of the original UV-visible spectra. 20 spectra (collected over 24 hours on Aprils 12<sup>th</sup>- April 13<sup>th</sup>, 2004 with a sampling period of 1 hr) were considered for training. Subset 1 contains spectra with a wavelength step size of 5

nm in the range 200 to 400 nm (i.e. 41 wavelengths). Subset 2 contains spectra with a wavelength step size of 10 nm in the range 205 to 395 nm (i.e. 20 wavelengths). Table 1 summarizes the percentage of variance captured for both subsets in function of the number of Principal Components. It was concluded that two PCs were sufficient to represent the spectra. The subsequent results were obtained with the UV-visible spectral subsets of type 2.

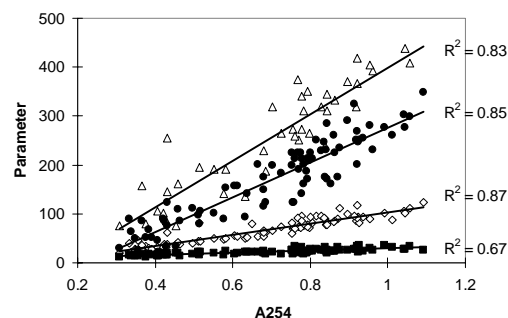


Fig. 4: Simple linear correlations between A254 and turbidity (FTU) ( $\diamond$ ), DCOD (mg O<sub>2</sub>/l) ( $\bullet$ ), TCOD (mg O<sub>2</sub>/l) ( $\Delta$ ) and N-NH<sub>4</sub> (mg N/l) ( $\blacksquare$ ).

Table 1: Cumulated percentage of variance explained in function of the number of Principal Components (PC) for the UV-visible spectra

	1 PC	2 PC	3 PC
Subset 1	98.3	99.4	99.8
Subset 2	98.5	99.6	99.9

PCR model were built for DCOD and TCOD, using only the PC1 and PC2 for DCOD and adding turbidity for TCOD. The internal validation results are shown in Table 2 and the corresponding recovery functions in Figure 5.  $R^2$  is the coefficient of determination. The adjusted coefficient of determination is:

$$\hat{R}^2 = \frac{(n-1)R^2 - p}{n - p - 1} \quad (1)$$

where  $n$  is the number of observations and  $p$  the number of explanatory variables. RMSE is the root mean square error.

Table 2: Internal validation results for PCR models based on UV-visible spectra

	$n$	$p$	$R^2$	$\hat{R}^2$	RMSE (mg O <sub>2</sub> /l)
DCOD	20	2	0.86	0.84	24
TCOD	10	3	0.96	0.94	21

An external validation was then conducted using three sets of samples collected between April 14<sup>th</sup> and April 16<sup>th</sup>, 2004 (Table 3). RMSE increases with time for DCOD and TCOD. It indicates that there is a slow change with time of the wastewater characteristics and of the PCR models. This change

is visualised on the recovery functions in Fig. 6 and is related to a change in the urban activity between weekend and week days.

**Table 3: External validation results for PCR models based on UV-visible spectra**

	<i>n</i> (DCOD)	RMSE (DCOD) (mg O <sub>2</sub> /l)	<i>n</i> (TCOD)	RMSE (TCOD) (mg O <sub>2</sub> /l)
Set 1	24	35	11	47
Set 2	24	60	11	68
Set 3	6	79	3	71

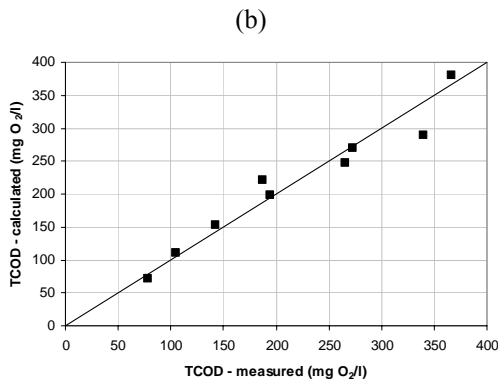
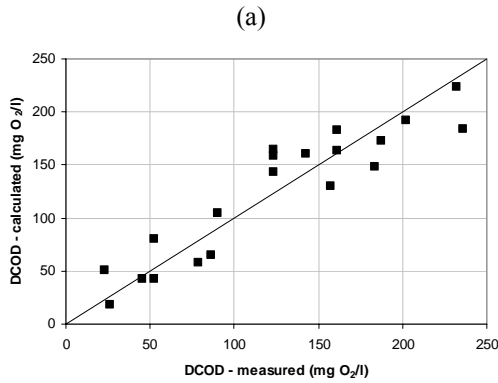


Fig. 5: Recovery functions of DCOD (a) and TCOD (b) for internal validation

PLS models were then built for DCOD and TCOD based on UV-visible spectral subsets of type 2 and taking into account turbidity for TCOD (data from April 12<sup>th</sup>, 2004). Two latent variables were considered sufficient in both cases (Table 4). The RMSEs obtained for PLS are similar to those for PCR and they exhibit the same increase with time (Table 5).

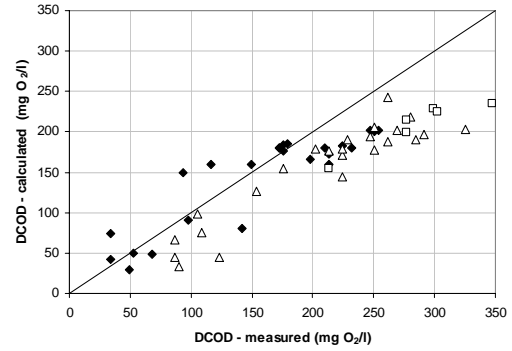
**Table 4: Cumulated percentage of variance explained in function of the number of latent variables (LV) for the UV-visible spectra**

	1 LV	2 LV	3 LV	4 LV	5 LV
DCOD	78.6	85.7	86.7	88.8	90.7
TCOD	94.7	95.6	96.2	97.2	97.9

**Table 5: RMSE for internal (April 12<sup>th</sup>) and external (April 14<sup>th</sup> to 16<sup>th</sup>) validation of PLS-models for DCOD and TCOD with two LVs.**

	RMSE (DCOD) (mg O <sub>2</sub> /l)	RMSE (TCOD) (mg O <sub>2</sub> /l)
Training set	24	16
Set 1	36	31
Set 2	59	76
Set 3	78	57

(a)



(b)

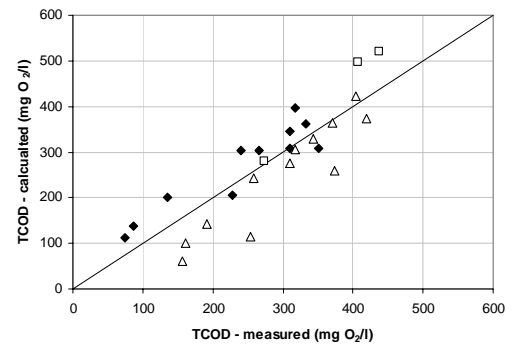


Fig. 6: Recovery functions of DCOD – External validation (♦) Set 1, (Δ) Set 2 and (□) Set 3.

### 3.2. Synchronous fluorescence spectra

The spectra were sub-sampled with a wavelength step size of 5 nm between 300 and 570 nm (i.e. 55 wavelengths per spectrum). In view of Table 6, three PCs were considered to built PCR models for N-NH<sub>4</sub>, TN and organic nitrogen (Norg). Norg is defined as  $Norg = TN - N-NH_4 - N-NO_3$  where N-NO<sub>3</sub> is the nitrogen due to nitrates and nitrites, which is negligible in wastewater. Very low coefficients of determination were obtained for either type of nitrogen species, as shown by the large data scatter in Fig. 7. PLS models were built tentatively but similar results were obtained. These results are not as good as those obtained by Vassel and Praet (2002) with emission fluorescence spectra (at  $\lambda_{exc} = 280$  nm) of urban wastewater. On the other hand Baker et al. (2003) could not find any correlation between ammonia and tryptophan-

like fluorescence ( $\lambda_{em} = 340 \text{ nm} / \lambda_{exc} = 280 \text{ nm}$ ) in sewage-polluted river water.

**Table 6: Percentage of variance captured par PCA in function of the number of PC for synchronous fluorescence spectra**

Nb. PC	1	2	3	4	5
Cum. Var. (%)	80.9	87.1	91.7	94.6	95.9

**Table 7: Internal validation results for PCR on synchronous fluorescence spectra**

	$R^2$	$\hat{R}^2$	RMSE (mg N/l)
N-NH4	0.31	0.20	21
TN	0.22	0.09	53
Norg	0.13	-0.015	41

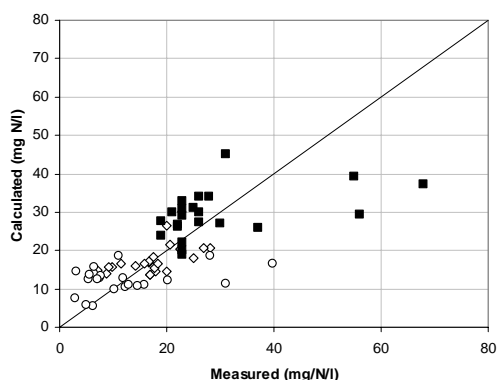


Fig. 7. Recovery functions of TN (■), N-NH4 (◇) and Norg (○)

Instead of looking for global correlations between nitrogen substances and synchronous fluorescence spectra taken at a whole, it was finally decided to focus on certain regions of the spectra. As seen on Figure 3, three peaks can be detected on the synchronous spectra. Their coordinates for  $\Delta\lambda = 50 \text{ nm}$  are: P1 ( $\lambda_{em}=366 \text{ nm} / \lambda_{exc} = 316 \text{ nm}$ ), P2 ( $\lambda_{em} = 430 \text{ nm} / \lambda_{exc} = 380 \text{ nm}$ ), and P3 ( $\lambda_{em} = 520 \text{ nm} / \lambda_{exc} = 470 \text{ nm}$ ). P1 can be related to some indole-type urine components (Dubayová et al., 2003) and contain some tryptophan fluorescence. P2 can be linked to humic-like and fulvic-like fluorescence. P3 is partially due to riboflavin (also contained in urine) but some lipidic substances can also contribute. Better coefficients of determination are obtained between P1 and N-NH4 than by considering the whole synchronous fluorescence spectra (Table 8). These correlations are confirmed by examination of the variations of the surrogate variables P1, P2 and A254 and wastewater characteristics such as N-NH4 (Figure 8). The coefficient of determination found for N-NH4 is similar to the one obtained by Vasel and

Praet (2002) for fluorescence in the region corresponding roughly to P1.

**Table 8: Coefficients of determination for simple linear regressions**

Correlation		$R^2$
between	and	
P1	DCOD	0.82
P2	DCOD	0.85
P1	N-NH4	0.61
P2	N-NH4	0.55
P1	A254	0.96
P2	A254	0.89

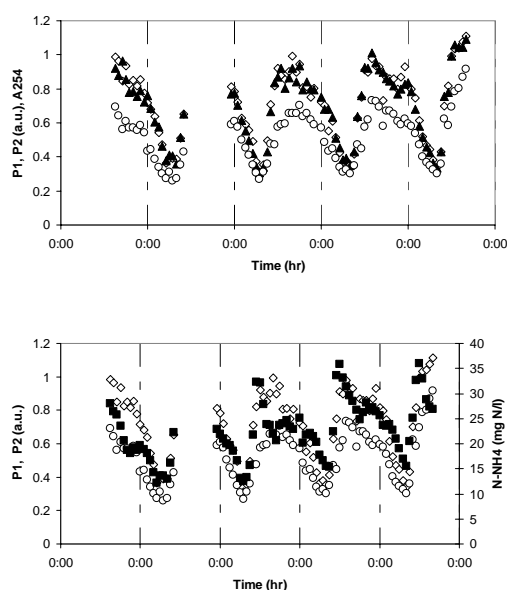


Fig. 8: Variations of the intensities of P1 (◇) and P2 (○), of A254 (■) and N-NH4 (▲) with respect to time.

#### 4. CONCLUSIONS

PCR and PLS models have been built between spectrometric data (turbidity, UV-visible and synchronous fluorescence spectra) and wastewater characteristics such as Chemical Organic Demand and nitrogen-related species. No difference was found between both types of models when they were applied to UV-visible spectrometric data and turbidity. It was however noticed that the models need to be adaptive to take into account the variations of human activity with respect to time (weekday / weekend for example). For synchronous fluorescence spectra, it is better to consider specific regions which can be related to urine and humic and fulvic acids. These surrogate measurements could help to improve control of wastewater treatment plants by providing information needed for feed-forward algorithms.

## ACKNOWLEDGEMENTS

The authors are thankful to the Ministry of Higher Education in France for the post-doctoral sponsorship of Jing WU, to GEMCEA and to the staff of the Nancy-Maxéville WWTP.

## REFERENCES

- Baker, A., R. Inverarity, M. Charlton and S. Richmond (2003). Detecting river pollution using fluorescence spectrophotometry: case studies from the Ouseburn, NE England. *Environm. Poll.*, **124** (1) 57-70.
- Baker, A. (2002). Fluorescence of some farm wastes: implication for water quality monitoring. *Wat. Res.* **36**, 189-195.
- Ahmad, S.R., Reynolds, D.M. (1995) Synchronous fluorescence spectroscopy of wastewater and some potential constituents. *Wat. Res.* **29** (6) 1599-1602.
- Ahmad, S.R. and D.M. Reynolds (1999). Monitoring of water quality using fluorescence technique: prospect of on-line process control. *Wat. Res.* **33** (9) 2069-2074.
- Dobbs, R.A., R.H. Wise and R.B. Dean (1973) The use of ultraviolet absorbance for monitoring the total organic carbon content of water and wastewater. *Wat. Res.* **6**, 1173-1180.
- Dubayová, K., J. Kušnir and L. Podracká (2003) Diagnostic monitoring of urine by means of synchronous fluorescence spectrum, *J. Biochem. Biophys. Methods*, **55**, 111-119.
- Escalas, A., M. Droguet, J.M. Guadayol and J. Caixach (2003). Estimating DOC regime in a wastewater treatment plant by UV deconvolution. *Wat. Res.* **37**, 2627-2635.
- Galapate, R.P., A.U. Baes, K. Ito, T. Mukai, E. Shoto, and M. Okada (1998) Detection of domestic wastes in Kurose river using synchronous fluorescence spectroscopy. *Wat. Res.* **32** (7) 2232-2239.
- Goslan E.H., S. Voros, J. Banks, D. Wilson, P. Hillis, A.T. Campbell, S.A. Parsons (2003) A model for predicting dissolved organic carbon distribution, *Wat. Res.*, **38**, 783-791.
- Langergraber, G., N. Fleischmann and F. Hofstädter, F. (2003). A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater. *Wat. Sci. Technol.* **47** (2) 63-71.
- Morel E., K. Santamaria, M. Perrier, S.R. Guiot and B. Tartakovsky (2004). Application of multi-wavelength fluorometry for on-line monitoring of an anaerobic digestion process. *Wat. Res.* **38**, 3287-3296
- Mrkva, M. (1983). Evaluation of correlations between absorbance at 254 nm and COD of river waters. *Wat. Res.* **17** (2) 231-235.
- Reynolds, D.M. and S.R. Ahmad (1997). Rapid and direct determination of wastewater BOD values using a fluorescence technique. *Wat. Res.* **31** (8) 2012-2018.
- Tartakovsky, B., L.A. Lishman and R.L. Legge, (1996). Application of multi-wavelength fluorimetry for monitoring of wastewater treatment process dynamics. *Wat. Res.* **30** (1) 2941-2948.
- Thomas, O., F. Theraulaz, C. Agnel and S. Suryani (1996). Advances in UV examination of wastewater, *Envir. Technol.*, **17**, 251-261.
- Vasel, J.L. and E. Praet (2002). On the use of fluorescence measurements to characterize wastewater. *Wat. Sci. Tech.* **45** (4-5)109-116.
- Santos, E.B.H., O.M.S. Filipe, R.M.B.O. Duarte, H. Pinto and A.C. Duarte (2000). Fluorescence as a tool for tracing the organic contamination from pulp mill effluents in surface waters. *Acta Hydrochim. Hydrobiol.* **28** (7) 364-371.
- Westerhoff, P. and D. Anning, (2000). Concentrations and characteristics of organic carbon in surface water in Arizona: influence of urbanization. *J. Hydrol.* **236**, 202-222.