

Environmental Science Processes & Impacts

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



rsc.li/process-impacts

Non-contact assessment of COD and turbidity concentrations in water using diffuse reflectance UV-Vis spectroscopy

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Jon Agustsson, Oliver Akermann, D. Andrew Barry, and Luca Rossi*

Water contamination is an important environmental concern underlining the need for reliable real-time information on contaminant concentrations in natural waters. Here, a new non-contact UV-Vis spectroscopic approach for monitoring contaminants in water, and especially wastewater, is proposed. Diffuse reflectance UV-Vis spectroscopy was applied to measure simultaneously the chemical oxygen demand (COD) and turbidity (TUR) concentrations in water. The measurements were carried out in the wavelength range from 200 – 1100 nm. The measured spectrums were analysed using partial-least-squares (PLS) regression. The correlation coefficient between the measured and the reference concentrations of COD and TUR in the water samples were $R^2 = 0.85$ and 0.96, respectively. These results highlight the potential of non-contact UV-Vis spectroscopy for the assessment of water contamination. A system built on the concept would be able to monitor wastewater pollution continuously, without the need for laborious sample collection and subsequent laboratory analysis. Furthermore, since no parts of the system are in contact with the wastewater stream the need for maintenance is minimised.

Introduction

Water contamination is an important issue nowadays, especially in urban areas. Due to the constant and rapid increase of population in cities, water infrastructure is often underdimensioned to manage the subsequent increase in volume and concentration of wastewater. For example, wastewater treatment plants cannot process all wastewater during large rainfall events due to their limited capacity. In this case, the excess wastewater is directly discharged into the natural environment. While this wastewater is diluted by the rain, it has still been shown to be an important source of contamination for the environment.^{1,2} In order to tackle this problem, there is a need for reliable real-time monitoring techniques.

Monitoring the total pollutant mass contained in wastewater during rain events can give valuable information for environmental assessment and sizing of technical solutions. The total pollutant mass can be determined from the water flow rate and the pollutant concentration. Gathering long term, reliable data is, however, a problem. Sewer systems are harsh environments with high humidity, corrosive gases, and wastewater carries large amounts of solid material. Measurement devices that are in contact with the water generally need frequent maintenance due to clogging, something that is avoided using non-contact monitoring methods.

Non-contact techniques, such as radar and ultrasonic devices for example, are able to measure the water velocity and water level, respectively. Recently, an integrated non-contact measurement system, HydroPix, was developed to measure the water flow rate in sewer systems.^{3,4} This system uses cameras placed above the water surface and image processing techniques to measure the water flow rate. The advantage of this system is that it allows remote, real-time monitoring the water flow rate and level over an extended period of time with minimal maintenance needs.

UV-Vis spectroscopy is commonly applied to measurements of wastewater pollution.⁵ In order to measure the pollutant concentrations, different immersed on-line systems were developed. More specifically, a UV-Vis spectroscopic system was recently developed to be installed in sewer systems and measures the concentrations of important contaminant indicators such as chemical oxygen demand (COD) and turbidity (TUR).⁶ The problem of clogging was reduced by blowing pressurized air through the optical path between the light source and the detector. Despite this innovation, frequent maintenance is still needed due to immersion in the wastewater.

The task of extracting pollutant concentrations from optical spectrums can be solved using multivariate correlation methods. A useful method to solve the task is partial-least-square (PLS) regression.⁷ The method looks for orthogonal linear

combinations of wavelengths which maximize the covariance between the measurement and reference values.

In this paper we propose a new system for monitoring the concentration of COD and TUR in wastewater based on diffuse reflectance UV-Vis spectroscopy. The technique has been applied to study lake water quality using sunlight as the light source.⁸ Diffuse reflectance UV-Vis spectroscopy is well known from soil analysis where it is used to identify soil types.⁹ For mixed powders, it has been shown that the absorption and scattering coefficients of the mixture constituents can be summed to obtain the corresponding coefficients for the whole mixture.^{10,11} Since no parts of the measurement equipment are immersed, the need for maintenance is low. The proposed instrument will be integrated in the HydroPix system,^{3,4} allowing a non-contact estimation of pollutant masses and concentrations in water.

The results presented in this paper illustrate the feasibility of the measurement concept and encourage further research on the application of non-contact spectroscopic methods to measure the quality of wastewater released to the environment.

Experimental

Experimental setup

The system consists mainly of two parts: A light source that illuminates the water surface and a spectrometer that measures the reflected light. The measured spectrum is processed in an attached computer to extract the pollutant concentrations.

Figure 1(a) shows a schematic of the experimental setup. Water samples were placed in 80-ml polypropylene (PP) cups. The cups were coated on the inside with deep black spray paint (Dupli-Color AEROSOL ART RAL 9005). The cup had a rim diameter of 5.5 cm, and the cup height was 4.0 cm. Each water sample tested had a volume of 50 ml resulting in a 2.7 cm water depth in the cup.

The water surface was illuminated using a continuous wave Xenon light source (HPX-2000) supplied by Ocean Optics. A Maya 2000 Pro spectrometer (also Ocean Optics) was used to measure the spectrum of the reflected light in the wavelength range from 200 to 1100 nm with a spectral resolution of 0.47 nm. Two fused silica collimating lenses (Ocean Optics UV-74) were used to emit the light to the water surface and collect the reflected light. The lenses were connected to the light source and spectrometer, respectively, using 2-m optical fibres (Ocean Optics QP600-2-SR).

The light emitter and collector were placed 25 cm above the cup bottom and 20 cm apart, as shown in Fig. 1(a). The light emitter and collector lenses were pointed at the water surface to avoid detecting light reflected from the cup bottom. Furthermore, the position of the light emitter and collector were chosen such that they would not interact with the same area on the bottom of the cup. The illuminated spot on the water surface had a diameter of 1 cm. The angle between the water surface normal, and the light emitter β and collector α , respectively, was not equal.

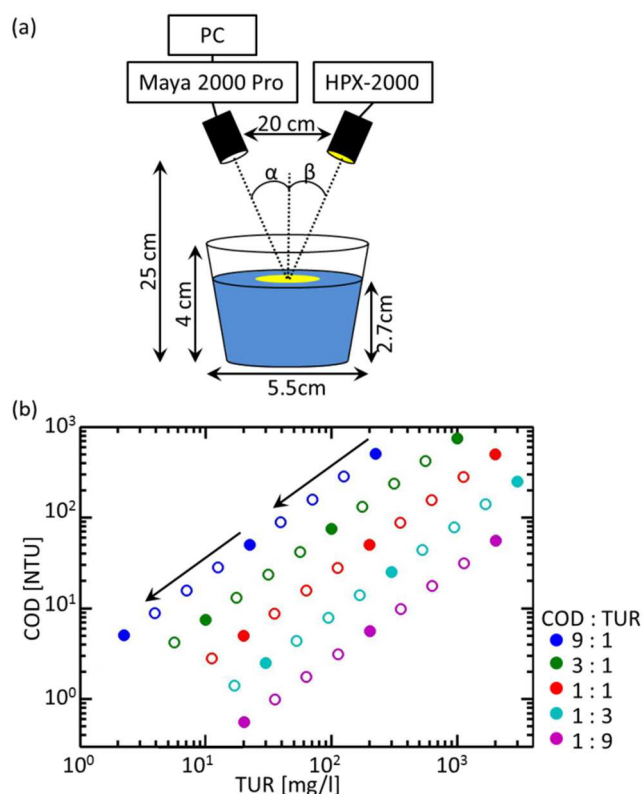


Figure 1: Schematic illustrating the measurement setup (a), and concentrations of COD and TUR in the measured solutions (b). 48 water samples were prepared with different COD and TUR concentrations, as illustrated in (b). The colors represent different ratios of COD and TUR.

The relative positions of the light emitter and the light collector were such that the light reflected specularly from the water surface did not enter the collector.

Water samples

Water samples were prepared using COD and TUR chemical standards. A 1000 mg/l COD standard from Hach and a 4000 NTU Formazine TUR calibration standard from Sigma-Aldrich were used to prepare the water samples. The chemical standards were used as bought. Water samples were prepared by mixing the COD and TUR standards in 5 different ratios. Figure 1(b) shows the ratios of COD and TUR used in the water samples. Dilutions of the water samples were performed using de-ionized water ($R > 18 \text{ M}\Omega\text{-cm}$). The filled circles indicate solution concentrations mixed directly and the open circles indicate solutions for which concentrations were mixed by diluting the prepared solution with di-H₂O. The arrows indicate how the water sample concentrations changed with successive dilutions. Each colour represents a certain ratio of the COD and TUR standard used in the water sample. The concentrations of COD and TUR ranged from 0.56 to 750 mg/l and 2.2 to 3000 NTU, respectively. In total, 50 different combinations of COD and TUR were used.

Measurements

Each water sample, containing a specific combination of COD and TUR concentrations, was measured 14 times. The cup was placed in two different positions in order to add variability to the measurements and reduce systematic effects caused by the cup edges on the measurement.

The measurements were carried out automatically. Since the turbidity of the water sample affects the amount of diffusively reflected light, the measurements were performed at seven different integration times (Table 1). A long integration time results in a high signal to noise ratio. For highly turbid water samples, the spectrometer may saturate if the integration time used is too long. Therefore, a short integration time is needed to measure highly turbid samples while a long integration time is needed water samples with low turbidity.

To minimize the influence of water movement the total measuring time was chosen to be 60 s, regardless of the integration time. This was done by measuring several spectrums and using their average as the final measured spectrum. Typically, integration times of 0.5 or 1 s were used. The measurements were carried out in a dark environment, as is normally the case in sewer systems.

Table 1: Integration times and number of averages used to measure the spectrums. The total measurement time (integration time \times number of averages) was set to 60 s.

Integration time [ms]	Number of averages
1000	60
500	120
200	300
100	600
50	1200
20	3000
10	6000

The background was determined for each spectrum individually by calculating the mean value of the spectrum intensity at the shortest (202 – 214 nm) and longest (1104 – 1114 nm) wavelengths, where the measured signal did not exceed the noise floor of the spectrometer and represents the base line of the spectrum.

With our spectrometer, the highest signal-to-noise ratio in the spectrums was achieved by using the longest possible integration time. The spectrums measured with the highest integration time for each water sample at each cup position were marked as *best measurements*.

Figure 2 shows a schematic of how the PLS training and testing was performed. Before performing the PLS regression the data were grouped by COD and TUR concentrations. All measurements carried out on a unique combination of COD and TUR concentrations were grouped together. When training the PLS model all groups, but one, were used as training data. The *best measurement* in the group left out was used to validate the model. This accomplished two things: (i) The combination of COD and TUR used for model validation were not present in the training data; and (ii) measurements in the training data carried out at a fixed combination of COD and TUR concentrations using various integration times simulated situations with varying light intensities making the model more robust.

It was important to first determine the TUR concentration of the measurement. To do this, a PLS model was trained using the measured spectrum and the reference TUR values. Using this PLS model the TUR values were calculated for all the measurements (Fig. 2). The measured TUR concentrations were obtained by using this PLS model with the measured spectrum of the *best measurement* in the data group left out of the training.

The calculated TUR values, the COD reference values, and the spectrums were then used to train a second PLS model. The second PLS model was used to calculate the COD levels (Fig. 2). The measured values of COD were found by using the second PLS model with the measured spectrums.

During the PLS training we assumed that the TUR concentration had a linear relation to the spectrum intensities and that the COD had a linear relation to the logarithm of the spectrum intensities. The number of components used in the PLS regression was 20.

Outliers were detected and labelled. If a calculated concentration of COD or TUR of a particular measurement was negative the measurement was labelled as an outlier and excluded from the calculation of the model errors. The error in

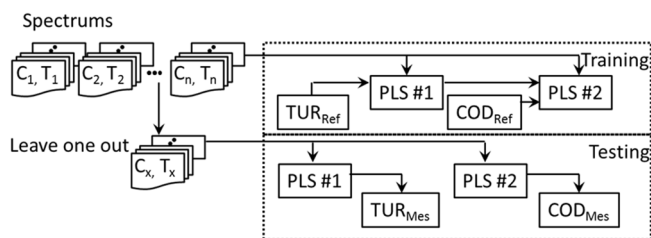


Figure 2: Schematic showing the training and testing of the PLS models. All of the measured spectrums are grouped by COD and TUR concentrations, C_x and T_x respectively, where $1 \leq x \leq n$, and n is the total number of prepared water samples. During training of the PLS models one group of spectrums was left out of the training data. During training the PLS models were constructed using the remaining spectrums and the reference TUR and COD values, TUR_{Ref} and COD_{Ref} respectively. The spectrum with the highest intensity in the group left out was used to validate the PLS models by calculating the TUR and COD concentrations, TUR_{Mes} and COD_{Mes} respectively.

Processing

The COD and TUR of the water samples were calculated by applying PLS regression on the measured spectrums.⁷ Before performing the PLS regression all saturated spectrums were removed from the measurement set. Saturated spectrums were defined as spectrums having intensity above 97% of the spectrometer measurement range.

The spectrums were recorded using different integration times. The integration time used influences the spectrum intensity and the background level. To normalize the spectrums, the measured intensity was divided by the corresponding integration time and a constant, representing the background, was subtracted.

the calculated COD and TUR values was calculated by the distance between the calculated value and the reference concentration in the corresponding water sample. The uncertainty of the measured concentrations was shown to be heteroscedastic. The variance scales linearly with the reference concentration of COD and TUR. Therefore, the logarithms of the reference and measured concentrations were used in the error calculations. An estimate of the measurement error was calculated using a standard error of prediction (SEP) defined by Varmuza⁷.

$$\log(SEP) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i - bias)^2} \quad (1)$$

$$bias = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (2)$$

$$y = \log(C_r), \hat{y} = \log(C_m) \quad (3)$$

where C_r and C_m are the reference and measured concentrations of the solutions, respectively.

The correlation coefficient of the measured concentrations with regards to the reference concentrations was calculated using

$$R(y, \hat{y}) = \frac{COV(y, \hat{y})}{\sqrt{COV(y, \hat{y})COV(y, \hat{y})}} \quad (4)$$

where $COV(y, \hat{y})$ is the covariance of the variables y and \hat{y} from Equation 3. Since the concentrations were spread over three orders of magnitude and the data was heteroscedastic we chose to use the logarithms of the measured and reference concentrations when calculating the correlation coefficient. Using the logarithms gives more weight to the low concentration values resulting in less optimistic correlation coefficients.

Results

Figure 3 shows three typical measured spectrums after normalization and background subtraction. The spectrums were measured in the wavelength range from 200 – 1100 nm. Overall, the spectrums show the profile of the light source used to illuminate the water sample. The spectrum intensity scales with the TUR concentration in the water sample. In total 672 spectrums having 48 different combinations of COD and TUR concentrations were measured. After removing saturated measurements 545 spectrums were left. Of those, 167 outliers were identified. The measurements marked as outliers were used in the PLS training and omitted in the error calculations. In the following calculations, 545 measurements were used during the PLS training, and 378 measurements were used to calculate the SEP.

Using 20 components, more than 97% of the measurement variance in COD and 99% of that in TUR was explained (Fig. 4). Increasing the number of components did not significantly increase the predictive power of the model. The risk of over fitting using 20 parameters is minimal since the PLS training was performed on over 540 measured spectrums over 2068 wavelengths.

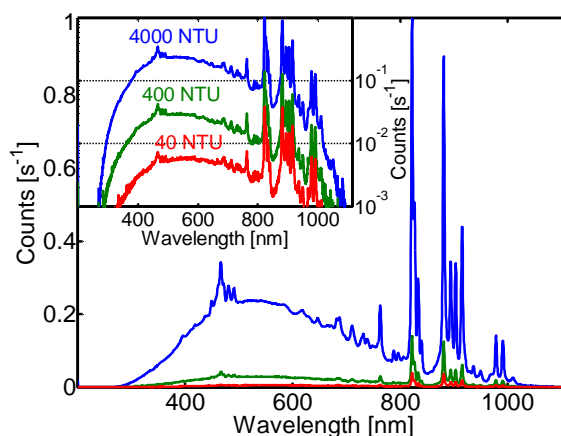


Figure 3: Measured spectrums of three water samples with different turbidity levels. The intensity of the measured signal decreases with decreasing turbidity. The inset shows the same spectrums plotted on a logarithmic scale.

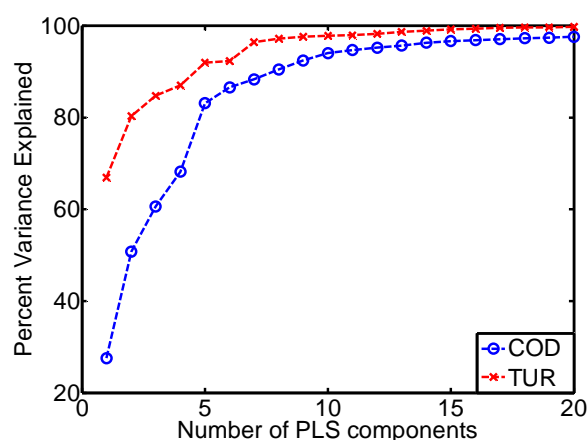


Figure 4: The percent of the measurement variance explained by the PLS as a function of the number of components used in the PLS mode. The improvement in the model levels off after about 15 components are included in the model.

Figures 5(a) and (b) show the measured concentrations of COD and TUR, respectively. The data are plotted as a function of the reference concentration of the constituents in the solution on a double logarithmic scale. The marks show the measured values, the solid line is the optimal one-to-one correlation. The dotted lines show the 95% confidence interval. The 95% confidence interval is defined as the interval within a distance of two SEP from the optimal one to one correlation. Using the whole measurement range the correlation coefficients for the measurement were $R^2 = 0.69$ and 0.95 for the COD and TUR respectively. The SEP of the measurements is found to be 35% for COD and 21% for TUR. The figures show that the constant error in the measurement sets the detection limit at around 30 mg/l for COD. At higher concentrations the relative error becomes more important.

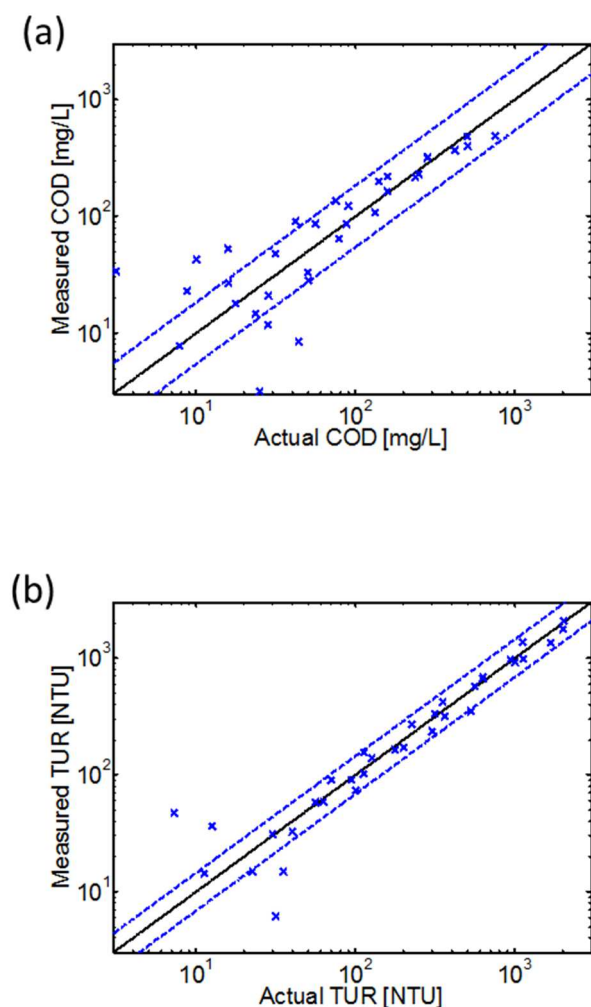


Figure 5: Measured values of COD (a) and TUR (b) as a function of the actual concentration of the parameters in the water samples. The x marks are the measured points, the solid line shows the optimal placement of the measured values, and the dashed lines indicate the 95% confidence interval. The correlation coefficients from COD and TUR are $R^2 = 0.69$ and 0.95 , respectively. The SEP for COD is 35% and TUR is 21%.

Figure 6 shows a measure of the relative error for all the available combinations of COD and TUR concentrations. The figure shows the combinations of COD and TUR used in the measurements on a double logarithmic scale. The colour indicates the error in the measurement of COD (Fig. 6(a)) and TUR (Fig. 6(b)) for each combination in units of SEP. As seen in the figure the highest errors in COD occur at low COD concentrations, lower than about 30 mg/l. The same is true for TUR values lower than about 10 NTU. These values determine the detection limits of the device. Calculating the correlation coefficients of the data measured at reference concentrations above the detection limit of the COD and the TUR gives correlation coefficients of $R^2 = 0.85$ and 0.96 , respectively.

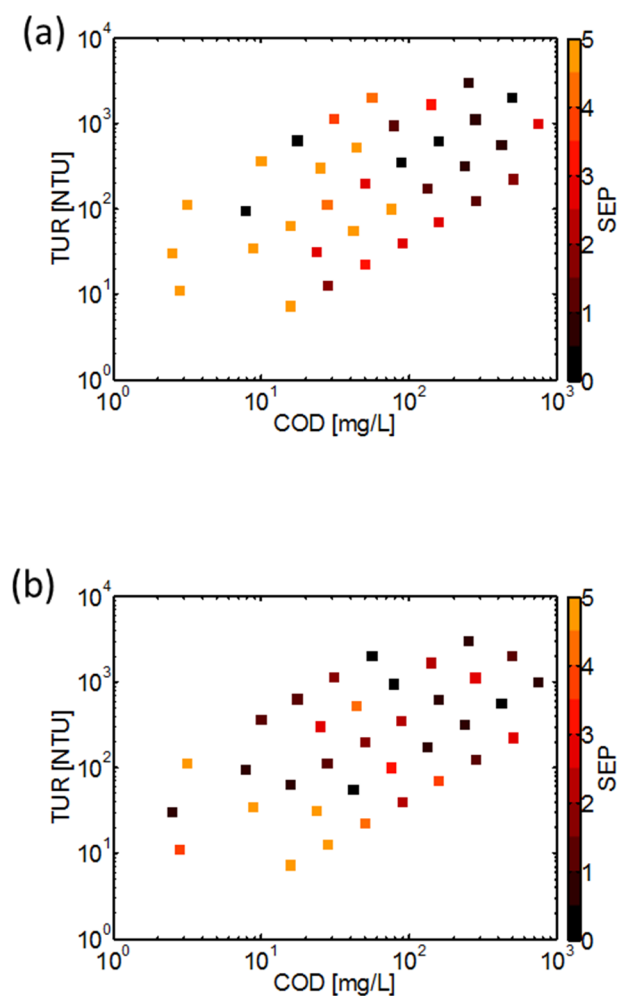


Figure 6: The relative error in the measurements of COD (a) and TUR (b) as a function of COD and TUR concentrations. The marker positions indicate the actual concentration of the parameters used in the measurements. The colour indicates the mean error in the concentration measurements for each unique concentration combination. The error is measured in units of SEP where dark colours indicate mean errors within one standard deviation from the actual values. Red and yellow colours indicate that the mean error at a certain COD and TUR concentration is more than one standard deviation away from the actual concentrations. Figure (a) shows that in the case of COD the highest errors appear around low COD concentrations. This indicates that the detection limit of COD is 30 mg/l. Figure (b) shows that the detection limit for TUR is 10 NTU.

Conclusions

We have shown the feasibility of using diffuse reflectance UV-Vis spectroscopy to monitor the concentrations of COD and TUR in water samples. The concentrations of the constituents used in the measurements ranged over three orders of magnitude from 0.56 - 750 mg/l for COD and 2.2 - 3000 NTU for TUR. We found a detection limit of 30 mg/l for COD. Below the detection limit the uncertainty in the measurement exceeds the 95% confidence interval marked by a distance of two SEP from the expected value. In the case of TUR we are able to predict the concentration down to about 10 NTU. The method measures the diffusely reflected light, from the water sample. When the

turbidity of the water is low no light is reflected to the collector. Therefore, no measurements of pollutants can be performed.

The measurement results show that it is feasible to use diffuse reflectance UV-Vis spectroscopy to measure the levels of pollution in water. In this paper we used a simple and inexpensive experimental setup, but there are several things that could be optimized to obtain even better results. The diameter of the cups used could be enlarged to reduce the influence of the cup edges on the measurements. We suggest that polarizing filters used on the light emitter and collector would reduce the influence of specular reflections on the measurements, and more diffusively reflected light could be collected by optimizing the optics. Even though the tested system can be improved, we were able to predict the COD and TUR concentrations in the water samples with a SEP of 31% and 21%, respectively. This measurement uncertainty is comparative to what has been reported elsewhere for more conventional transmission UV-Vis spectroscopic measurements.¹²

The application of the methodology under real conditions requires the consideration of heterogeneous surfaces with waves, splashes, foam, flowing debris, etc. To cope with these perturbations the optical system needs to be optimized and the measurement strategy developed. A possible solution to this problem is to measure a large area of water, to use long integration times, and to use averages of several measurements as the final value. Moreover, COD can be contributed under real conditions by a lot of organic contaminants which have inconsistent absorption of UV light due to various factors. Therefore local calibration cannot be avoided in system deployment.

The need to monitor the pollution levels in wastewater released into nature will continue to grow. It is therefore important to develop reliable and adaptable instruments to measure the pollutant mass contained in the wastewater. A non-contact method offers flexibility in device installation and reduces the need for maintenance. Incorporating diffuse reflectance UV-Vis spectroscopic measurements into a system that can also perform non-contact measurements of the water flow rate would provide a convenient and complete system to follow the pollutant mass released into nature with wastewater.

Acknowledgements

The thank Michael Matthews at Ocean Optics for help with choosing equipment and fruitful discussions, as well as Prof. Sabine Süssstrunk and Dominic Rüfenacht at the Images and Visual Representation Group at EPFL for helpful discussions.

Notes and references

Ecological Engineering Laboratory, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. Tel: +41 (0) 21 693 57 80; E-mail: luca.rossi@epfl.ch

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

- 1 G.A. Burton and R.E. Pitt, *Stormwater Effects Handbook: A Toolbox for Watershed Managers, Scientists, and Engineers*, Lewis Publishers, Boca Raton, 1st ed., 2002.
- 2 L. Rossi, N. Chèvre, R. Fankhauser, and V. Krejci, *Urban Water Journal*, 2009, **6**, 355 – 367.
- 3 L.S. Nguyen, B. Schaeli, D.Sage, S. Kayal, D. Jeanbourquin, D. A. Barry and L. Rossi, *Water Sci. Technol.*, 2009, **60**, 2281-2289.
- 4 D. Jeanbourqui, D. Sage, L. Nguyen, B. Schaeli, S. Kayal, D. A. Barry and L. Rossi, *Water Sci. Technol.*, 2011, **64**, 1108-1114.
- 5 O. Thomas and C. Burgess, *UV-Visible Spectrophotometry of Water and Wastewater*, Elsevier, Amsterdam, 1st ed., 2007.
- 6 G. Langergraber, N. Fleischmann and F. Hofstädter, *Water Sci. Technol.*, 2003, **47**(2), 63-71.
- 7 K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Boca Raton, 2009, Chapter 4.
- 8 A. Hommersom, S. Kratzer, M. Laanen, I. Ansko, M. Ligi, M. Bresciani, C. Giardino, J.M. Beltrán-Abauza, G. Moore, M. Wernand, and S. Peters, *J. Appl. Remote Sens.*, 2012, **6**, 063615-1 - 063615-21
- 9 L.R. Drees and A.L. Ulery, *Methods of Soil Analysis – Part 5: Mineralogical Methods*, American Society of Agronomy, Madison, 2008, Chapter 13.
- 10 D.R. Duncan, *Proc. Phys. Soc.*, 1940, **52**, 390-401.
- 11 D.R. Duncan, *J. Oil Colour Chem. Assoc.*, 1949, **32**, 269-391.
- 12 L. Rieger, G. Langergraber, and H. Siegrist, *Water Sci. Technol.*, 2006, **53**(12), 187-197.

Environmental impact statement

Water contamination is an important environmental concern underlining the need for reliable real-time information on contaminant concentrations in natural waters. Here, a new non-contact UV-Vis spectroscopic approach for monitoring contaminants in water, and especially wastewater, is proposed. By using non-contact measurements to measure the wastewater pollution the need for maintenance is greatly reduced. This allows the instruments to be mounted at sites that cannot be reached easily and encourages stakeholders to continuously monitor wastewater pollution more carefully. By continuously monitoring the wastewater pollution measures can be taken if unforeseen or unexpected events occur and the treatment process at a receiving wastewater treatment plant can be adjusted to receive unexpected pollution loads.