Intelligent multispectral vision system for non-contact water quality monitoring for wastewater

Karen Preitner*[a], Sébastien Blanc[a], David Honzatko[a], Clément Kündig[a], Pedram Pad[a], Sareh Saeedi[a], Salvador Peña-Haro[b], Pierre Lechevallier[c,d], Jörg Rieckermann[c], L. Andrea Dunbar[a]
[a]Edge AI and Vision Group, CSEM SA, Rue Jaquet-Droz 1, 2002 Neuchâtel, Switzerland; [b]Photrack AG, Ankerstrasse 16A, 8004 Zürich, Switzerland; [c]EAWAG, Überlandstrasse 133, 8600 Dübendorf, Switzerland, [d]Institute of Civil, Environmental and Geomatic Engineering, ETH Zürich, 8093, Zurich, Switzerland.

ABSTRACT

Water quality monitoring in sewer networks remains a technical challenge even though water pollution and control are high priorities since decades. Current water quality monitoring usually analyzes samples in laboratories, allowing only sporadic measurements, or uses immersed sensors in the wastewater, leading to clogging and sensor fouling resulting in expense due to intensive maintenance. Both techniques thus have serious limitations.

Previous research showed that UV-Vis reflectance spectrometry can be used for non-contact monitoring of turbidity (TUR) and Chemical Oxygen Demand (COD), which are two key water quality indicators. Although spectrometer achieve high spectral resolution their limited spatial field of view is problematic for highly inhomogeneous surfaces as is the case wastewater

In this study, we obtain beyond state-of-art measurement accuracies by combining machine learning techniques with increased spatial field-of-view Multi-Spectral Imaging (MSI) whilst substantially reducing the spectral resolution. We designed and built a dedicated setup with a monochromatic camera and an active illumination of thirteen LEDs covering the spectrum range of 200-700 nm. We acquired and calibrated data on 27 samples with different concentrations of TUR and COD. Machine learning regression models were trained and evaluated with the extracted spectra. We tested the Partial Least Square (PLS), Support Vector Machine (SVM) and Random Forest (RF). PLS regression performed best with excellent correlation coefficients (R^2) of the 0.99 for TUR and 0.93 for COD. We obtained similar results with the SVM algorithm (R^2 = 0.99 and 0.92), whilst RF had lower scores (R^2 = 0.96 and 0.71).

Keywords: Multi-spectral imaging, Water quality control, Machine learning, Wastewater

1. INTRODUCTION

Water pollution is an increasingly serious problem in our society[1]. Water quality monitoring in sewer networks remains a technical challenge due to the high solid content and aggressive matrix of raw wastewater.

Currently, the assessment of the water quality is performed by taking samples and analyzing them in a remote laboratory or by using immersive sensors in direct contact with the polluted water. The first solution allows accurate but sporadic and costly analysis. Such that it cannot capture the high variability of sewer pollution, especially during rain events. The immersive solution has the advantage of a much higher measurement rate. However, sensors in contact with raw wastewater are quickly degraded by grease, fat, and clogging through particulates, which leads to inaccurate measurements and sensor failure. Weekly maintenance is required for accurate measurements which often means this technique prohibitively expensive.

An autonomous and contact-less system would benefit from real time analysis with lower maintenance compared to these techniques. Previous research by J. Agustsson et al.[2] demonstrated the feasibility of using reflectance spectrometry methods in the ultraviolet to visible (UV-Vis) spectrum range for remote and continuous water quality assessment. However, their solution is not easily scalable as it relies on a single point measurement and expensive and complex equipment, e.g. halogen illumination.

Our approach uses the fact that UV-Vis water quality monitoring is a correlation-based approach. Making it possible to reduce the spectral resolution for this application whilst retaining accuracy. In this work we designed and built just such a Multi-Spectra Imaging (MSI), tested its performance, and showed beyond state-of-art results.

As water quality variables, we followed Agustsson et al.[2] and also focused on the chemical oxygen demand and turbidity. The chemical oxygen demand (COD) is a measure of the amount of oxygen required to fully oxidize the organic matter contained in a sample. COD is therefore an indirect measurement of organic pollution, expressed in mass of oxygen consumed over the volume of solution, which in SI units is mgO_2/L. Organic content in wastewater can be measured at 254 nm because many organic compounds absorb UV light at that wavelength and the amount of light absorbed can be used to calculate the concentration of the organic compounds in the wastewater[3]. Here, we used a 1000 mgO_2/l COD stock solution. Turbidity is a measure of the relative clearness of a liquid[3]. Depending on the amount of solids in suspension in the sample, light rays are reflected and attenuated. Therefore, the higher the number of suspended particles, the higher the turbidity. The most common method to measure turbidity, is called nephelometry, which measures the backscattered light with spectrometers. Thus, turbidity is usually expressed in Nephelometric Turbidity Units (NTU).

This paper is set out as follows. First, we describe the experimental setup, then the data preparation, acquisition protocol and the image processing. Finally, we will evaluate the models: Partial Least Square (PLS), Random Forest (RF) and Support Vector Machine (SVM).

2. EXPERIMENTAL SETUP

Our setup was a redesign of that used by Agustsson *et al.*[2].Here, we can only give a brief description, see Preitner's Thesis[4] for more details. The camera and illumination are placed above the samples which are held in a black non-reflective cup (dimensions/volume), see Figure 1. The setup is placed in a dark environment to avoid parasitic noise and mimicing real conditions in sewers. The camera is fixed 40 cm above the samples. At this distance the sample fills the field of view of the camera, a 4Mpxl camera was chosen for sufficient spatial resolution. The distance between the LEDs and the solution is fixed and at the as distance to the camera. An image of the sample is taken thirteen times, one for each spectral band. The data obtained is a data cube with the dimensions 13 x 2048 x 2046.

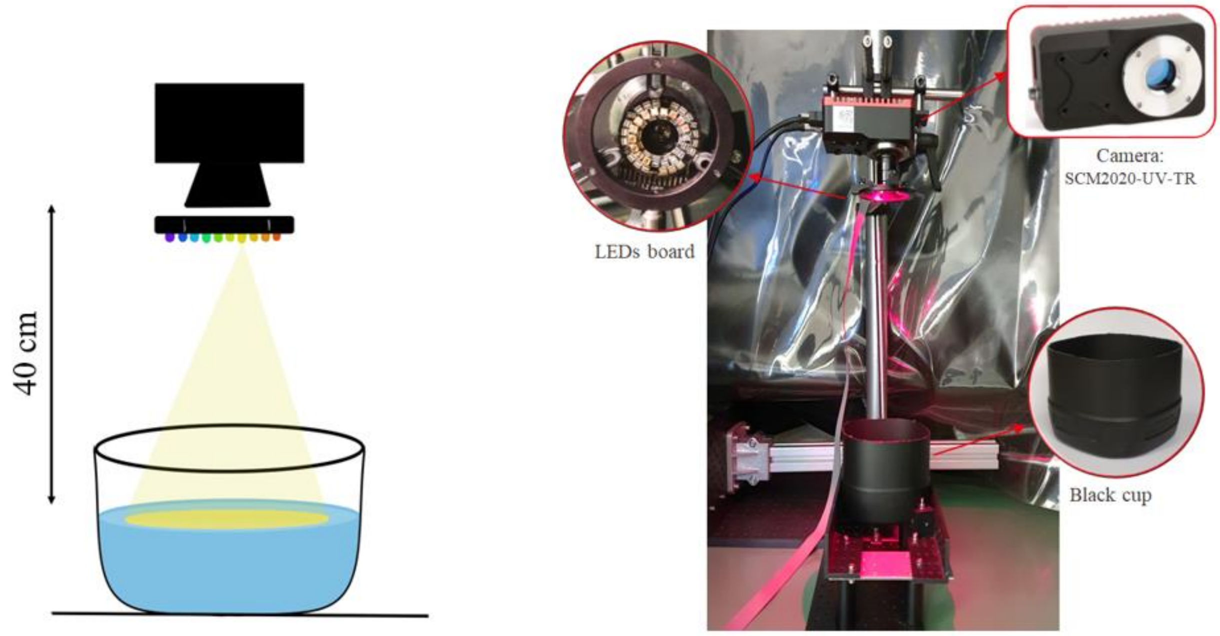


Figure 1. Setup used for the acquisition (left) the schema (right) the setup.

Material

The setup is composed of three main elements: the illumination, the camera, and the black cup which holds the samples. The illumination consists of thirteen LEDs mounted on a circular board. The monochromatic camera (SCM2020-UV-TR) is sensitive in the ultra-violet to visible (UV-VIS) spectrum range, which is the necessary spectrum range to capture the turbidity (TUR) and Chemical Oxygen Demand (COD). This camera has a rolling shutter and an angle of view of 22 degrees. The Figure 2 shows the camera and its quantum efficiency.

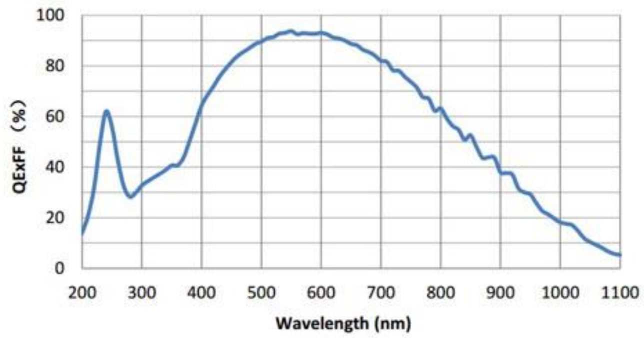


Figure 2. SCM2020-UV-TR camera (left) and its quantum efficiency (right).

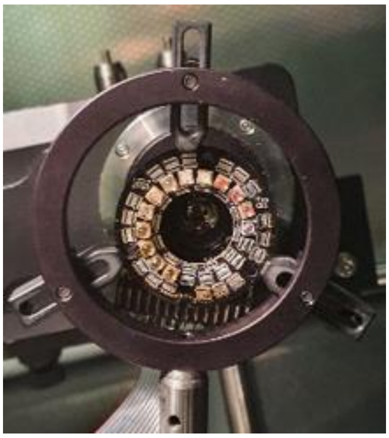
The acquisition system uses a custom-made active illumination, composed of a prototype with thirteen LEDs mounted on a circular board. Figure 3 shows the LED spectra with intensity normalized by the exposure time. Intensities were measured with a spectrometer see Table 1. The theoretical peaks show the values described in the datasheets of the LEDs.. Each LEDs has a different optical power, and the quantum efficiency of the camera is not linear; therefore, the LEDs are modulated to balance the intensity throughout the different wavelengths. Furthermore, the spatial homogeneity is calibrated using a white reference to obtain homogenous illumination for all the bands.

Table 1. LEDs peak wavelengths in nm.

Theoretical peak	250	270	290	310	365	380	400	455	520	590	620	660	700
Measured peak	256	277	293	308	368	386	403	458	530	591	632	656	696
FWHM*	11	11	10	11	10	9	15	14	25	13	12	12	15

* FWHM: full width half max, i.e. the spectral width of the signal at 50% of the peak intensity

The position of the illumination source is critical to obtain the minimum specular reflection, maximum light homogeneity and highest diffuse reflectance. For this reason, the illumination source was placed directly under the lens of the camera. During the experiments, the specular reflections were not problematic because they typically only cover a small area of the total part of the image and can be removed later on using machine vision algorithms.



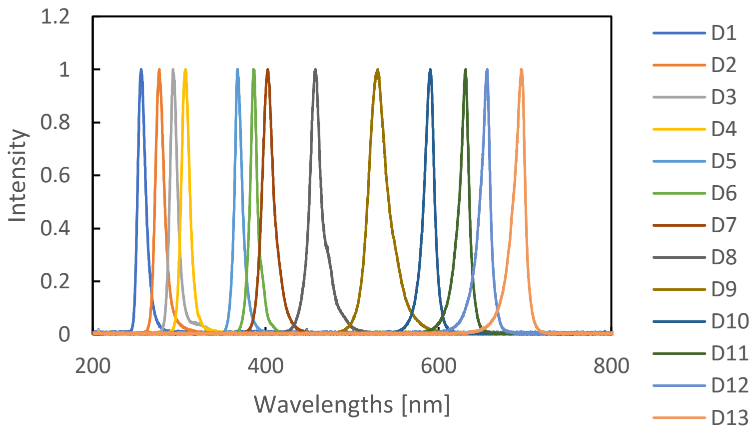


Figure 3. Illumination of the system. (left) Circular board of the LEDs. (right) Spectra of the LEDs, normalized to 1.

The properties of the cup should have no impact on the reflectance of the samples and should not interact with the components of the solutions. We used a black high-density polyethylene (HDPE) flask whose top has been removed.

The samples are produced using standard solutions of TUR and COD, different concentrations are obtained by changing the solutions ratio and dilution adding deionized water mainly for cost reasons: standard solution of COD costs around 60CHF/200mL. Different units are used to quantify the TUR concentration[5], depending on the methods used to measure the TUR. We used the Nephelometric Turbidity Unit (NTU). An NTU Aluminum oxide (Al_2O_3) solution, with a concentration of approx. 0.3 vol.% is used as TUR standard solution.

To have a better insight about the TUR, an example of different concentrations of TUR can be found at the Figure 4.

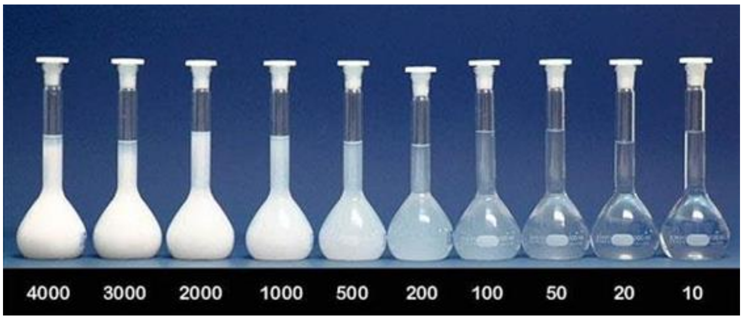


Figure 4. Examples of TUR for a concentration of 10 to 4000 NTU. Source: DEAL Guyane, 2018 [6].

The concentrations of TUR and COD vary widely in the wastewater systems. The ranges of concentrations that can be found in the sewers and wastewater are 150-2500 NTU and 42-1000 mg/L respectively[7].

Twenty-seven samples of 400ml with different TUR and COD concentrations were prepared mostly within the previously defined range, to represent at best the conditions inside the wastewater systems. Turbidity and COD levels were determined by the mixing proportions of standard solutions used in each samples. In addition, turbidity was measured with. standard laboratory portable turbidity meter. However, further COD analysis couldn't be performed. (Table 2).

We generated 27 samples, initially ten samples were obtained by mixing the COD and turbidity standards under different proportions. In total, seven different COD-turbidity ratio were studied. Then, eight of those ten samples were diluted between one and four times to generate more samples. Figure 5 gives an overview of the different sample's concentrations.

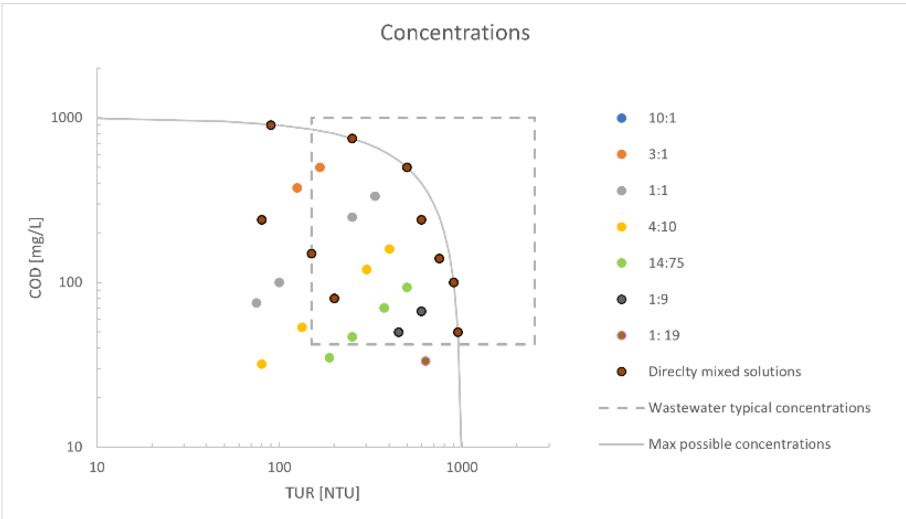


Figure 5. Samples concentrations. The range of the concentrations that could be measured in wastewater is represented by the rectangle. Each color represents a different ration of COD:TUR.

Calibration

The system was calibrated to compensate the variation in the intensities of the LEDs. The intensities of the LEDs are modified with a pulsed-width modulation (PWM) such that the intensity of the light reflected by the white target is similar for all the LEDs with a fixed exposure time of 80ms. We used a Zenith Lite™ square target of 500x500 mm with a diffuse reflectance of 95%. The dark reference is measured by taking an image in a dark environment with no light source. The acquisition of the references is done for each LED. Moreover, the images of the dark reference are normalized with the exposure time. For example, if the image has been acquired with an exposure of 250ms and the white reference with 80ms, the image would be normalized by dividing the intensity with $\frac{80}{250}$. The final reflectance is computed using the equation (1).

$$I_{calib} = \frac{\left(I_{image,t1} - I_{dark,t1}\right) * (t2/t1)}{I_{white,t2} - I_{dark,t2}} \quad (1)$$

Where I_{calib} is the image obtained after calibration, $I_{image,t}$ the original image acquired with an exposure time t, $I_{white,t}$ the white reference acquired with an exposure time t, $I_{dark,t}$ the dark reference acquired with an exposure time t. t_1 and t_2 are the two exposure times used for the acquisition.

3. DATA ACQUISITION

Protocol and final dataset

Data acquisition was the same for the twenty-seven samples, each acquisition was performed in under 30 minutes, to prevent any degradation of the sample. The samples are according to Table 2. TUR is measured with the turbidimeter, the concentration of COD is assumed to correspond to the theoretical values of the mix. 400ml of the solution is poured into the black cup and placed under the camera. Between five and twelve acquisitions were performed with different exposure times each set of acquisitions is composed of 13 images, one for each LED. The cup and all the tools used were rinsed between all the acquisitions to avoid any contamination of the previous sample. The result of an acquisition is a data cube (3D array) with the dimension 13 x 2048 x 2046. Finally, the experimental test led to a dataset of 225 data cubes, i.e. multi-spectral images, and corresponding ground truth concentrations for TUR and COD. Figure 6 shows the images for a solution with high TUR(917 NTU) and low COD (50 mgO2/l), before calibration. One can see the specular reflection of the LED at the center, and at the rim. Little diffusion is observed in the deep UV range (250nm – 270nm) and in the near infrared (700nm).

Table 2. Sample concentrations.

Solution	Ratio	COD [mg/L]	TUR [NTU]	CODStand [ml]	TURStand [ml]	Water [ml]
1	10:1	900	90	360	36	4
2	3:1	750	250	300	100	0
3		500	166.67	200	66.67	133.33
4		375	125	150	50	200
5		240	80	96	32	272
6	1:1	500	500	200	200	0
7		333.33	333.33	133.33	133.33	133.33
8		250	250	100	100	200
9		150	150	60	60	280
10		100	100	40	40	320
11		75	75	30	30	340
12	4:10	240	600	96	240	64
13		160	400	64	160	176
14		120	300	48	120	232
15		80	200	32	80	288
16		53.33	133.33	21.33	53.33	325.33
17		32	80	12.8	32	355.2
18	14:75	140	750	56	300	44
19		93.33	500	37.33	200	162.67
20		70	375	28	150	222
21		46.67	250	18.67	100	281.33
22		35	187.5	14	75	311
23	1:9	100	900	40	360	0
24		66.67	600	26.67	240	133.33
25		50	450	20	180	200
26	1:19	50	950	20	380	0
27		33.33	633.33	13.33	253.33	133.33

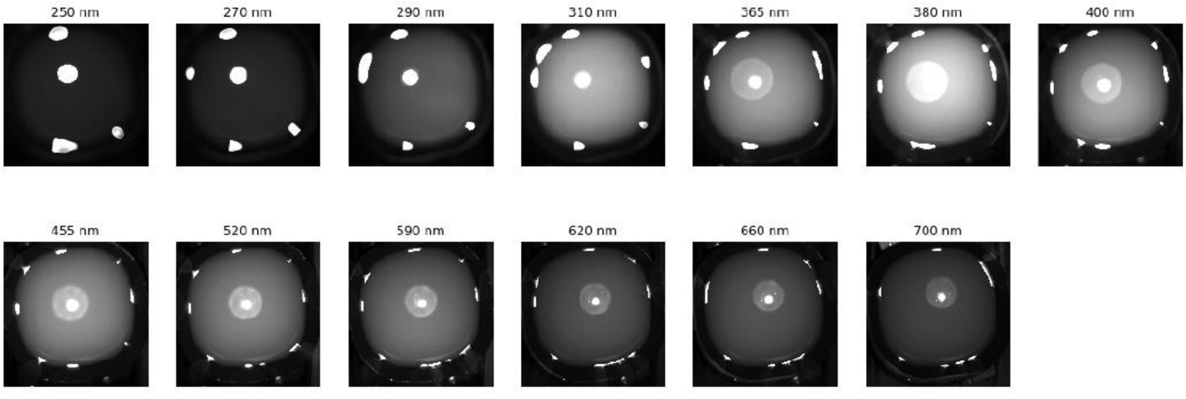


Figure 6. Example of images obtained during the acquisition, for a high TUR solution.

Image pre-processing

After the calibration the image is smoothed with a Gaussian filter, sigma set at 0.7. Figure 7 shows the intensity at the yellow line, before the calibration (Figure 7a and 7c), and after the calibration (Figure 7b and 7d) for the concentrations 100 mg/l and 854 NTU, and for the illumination at 620 nm. We set at zero the regions outside the region of interest (ROI) to improve the visualization of the calibration result on the reflectance

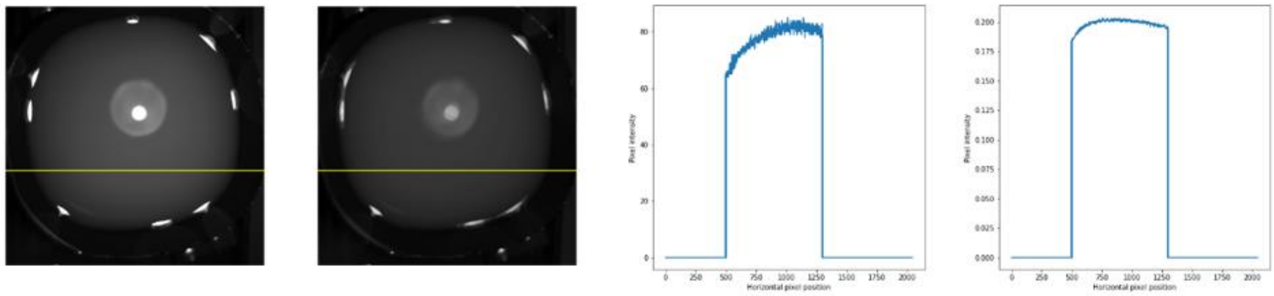


Figure 7. Intensity of the pixels before and after the calibration for a concentration of 100 mg/l and 854 NTU for an illumination at 620 nm. Left from right: (a) Image of the solution before the calibration. (b)Image of the solution after the calibration. (c) Intensity of the pixels before the calibration. (d) Intensity of the pixels after the calibration

A mask is generated to define the ROI corresponding to light reflection over the sample surface. In particular, the mask was designed to exclude the background and areas with high specular reflections (e.g. rim of the cup, reflections of the surface, etc.).

Although the bottom of the cup reflects only little light, this effect is considered as noise and has not been compensated for during the analysis. Also, the cup has a small stabilizing diagonal bulge at its bottom, which presents a non-homogenous background reflection of the bottom surface for low TUR (see Figure 8). Therefore, this region is also removed from the ROI (see Mask in Figure 8). Each individual specular reflection is segmented using the flood fill algorithm with manually selected seeds. The union of the segmented specularities is then removed from the ROI. Figure 8 shows the final ROI and the images of a single set with the mask applied to them, no specular reflection can be observed in the ROI.

Finally, the agglomeration of particles of alumina from the TUR solution created highly reflective floating particles; therefore, they were removed from the ROI. They were detected with blob detection on each image.

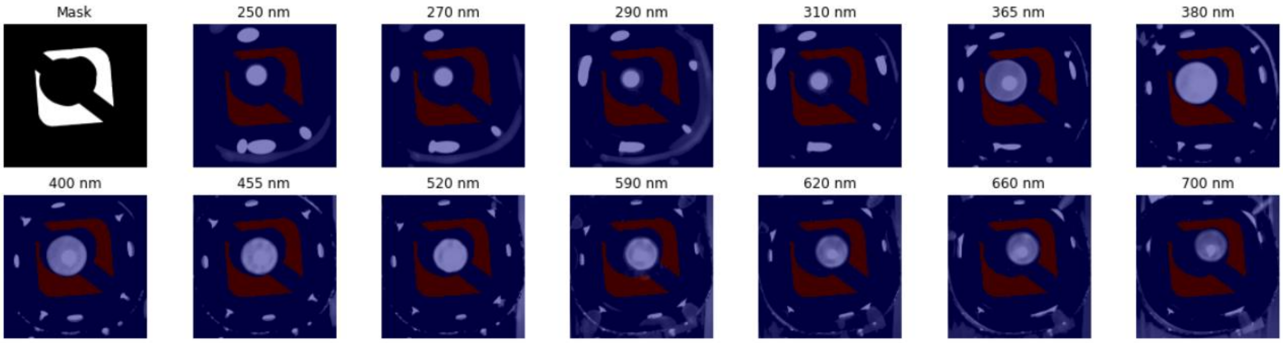


Figure 8. Mask that defines the region on interest, i.e., the region of the solution. It is shared for all the sets of images. The mask removes all the specular reflections of the images, as well of the surroundings, such as the rim of the cup and the background.

4. ML MODEL

Spectra creation

The reflectance of the sample is constant throughout the whole ROI, meaning that there is few spatial information within the image. Therefore, there is no reason to directly use the images for the model to predict the concentrations. The pseudo-spectrum of the reflectance is created by using the intensity of the pixels for a given band. The peak wavelength of the LED is used as the wavelength value for the pseudo-spectra.

Seven points are arbitrary defined within the ROI to reduce the amount of data fed to the algorithm. This eventually increase speed of computation. The distribution of the seven points is described in Figure 9 (up left). To reduce the chance of selecting an outlier pixel, a region of 7x7 is selected around the point and the median of the intensity is kept. The spectrum of the point is created from those intensities computed for the thirteen images of the set; as a result, seven spectra are obtained for each set. The Figure 9 (down) show the reflectance spectra of the twenty-seven samples, obtained by taking the mean of all the spectra for each sample.

An outlier detection is also performed on all the spectra for a same concentration using the median and median absolute deviation (MAD), see equation (2). The spectrum is designated as outlier if the difference between its intensity and the median of the one of all spectra is higher than 10*MAD for at least one wavelength. Usually, the threshold would be set at 3*MAD but setting it at 10*MAD allows a higher variance in the values of the spectra and should prevent to a certain extent the models to overfit. The Figure 9 (up right) shows the remaining spectra (full line) and the removed ones (dotted lines) for the 100 mgO_2/l COD and turbidity of 94 NTU.

$$mad = median(|spectrum - median(spectra)|) \tag{2}$$

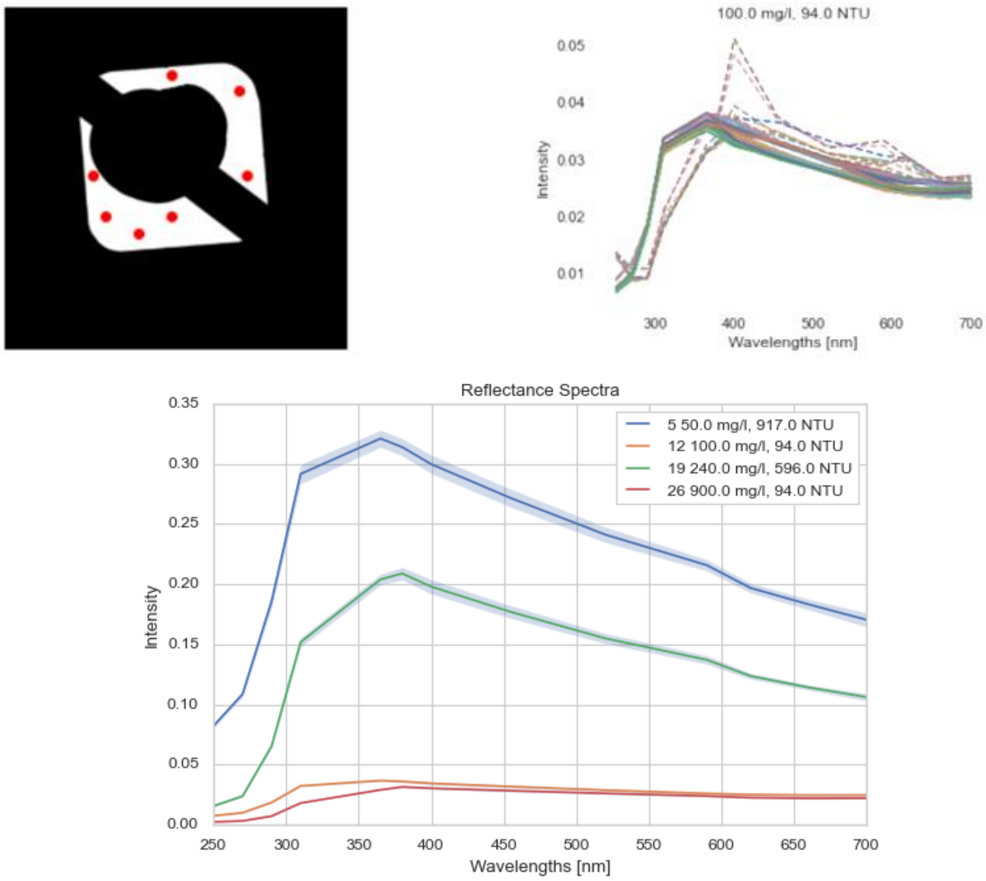


Figure 9. Distribution of the seven locations of the spectra extraction for one sample (up left), the outlier detection for one concentration (up right) and the reflectance spectra of 4 of the 27 concentrations (down)

Data

The dataset is made up of 27 groups of concentrations, each group has between 30 and 84 elements after the outliers were removed. This makes a total of 1414 spectra out of the 1491 original dataset, with concentrations between 32 and 900 mgO_2/l for COD and 67 - 917 NTU for TUR. Figure 10 shows the effect of variation on individual elements, COD and TUR. The effect of TUR seems to involve mainly the global intensity of the reflectance (see right), while COD (see left) has an effect in the lower wavelengths, 250 - 310 nm, mainly in the slope between the intensity at 365 and 310 nm.

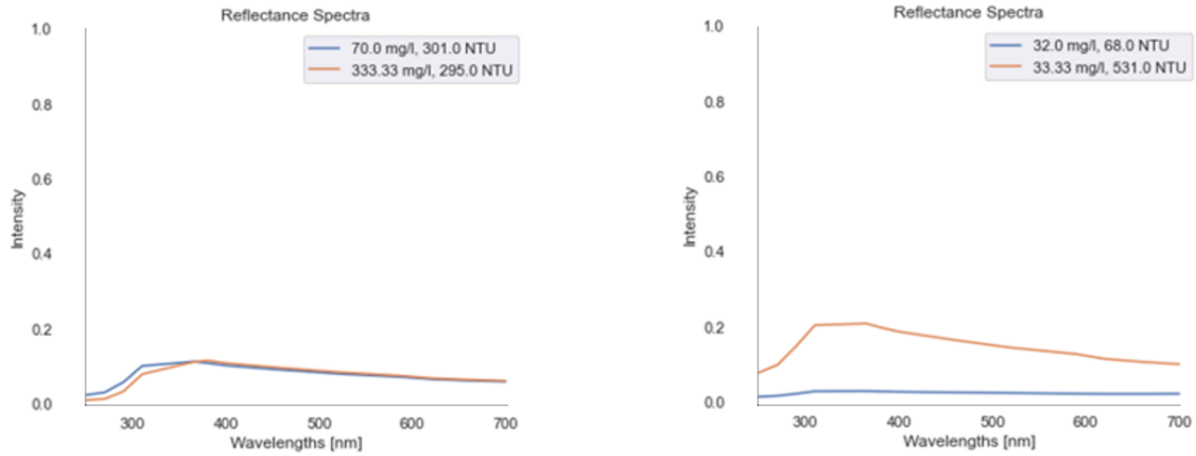


Figure 10. Samples with only one component that strongly varies, used to assess the effect of individual elements on the spectrum. (left) The spectrum of samples with similar COD concentration but different COD. (right) The spectrum of sample with similar COD but different level of COD.

As expected, the reflectance and the concentrations of COD and TUR are highly correlated with one another (Figure 11). The TUR is best correlated with the intensity of the reflectance in the higher wavelengths (365 - 700 nm) with a correlation value above 0.99. On the other hand, the COD concentration doesn't seem to have any correlation with the reflectance. As Agustsson et al.[2], we observed a linear relationship between the COD and the logarithm of the reflectance intensities, Figure 11 (middle) confirm this idea, it shows that the COD concentration has a strong negative correlation with the intensities in lower wavelength, 250-290, with correlation coefficients from -0.52 to -0.68. Furthermore, the COD seems to be defined by the drop between the intensity at 365 nm and the one at 310 nm. The correlation map of this feature (Figure 11 right) shows that the COD and the subtraction of the logarithm of the intensity at 365 nm with the one at 310 nm are indeed correlated with a coefficient of 0.72.

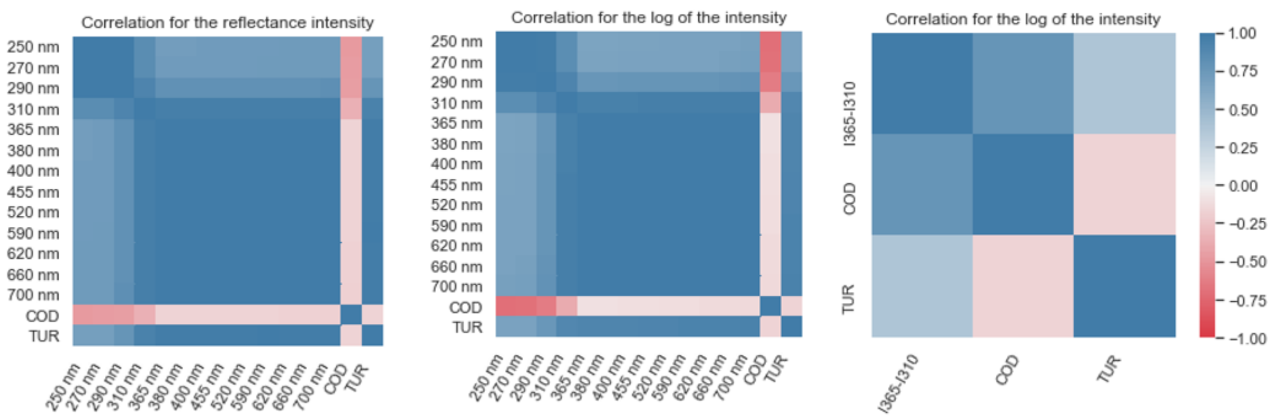


Figure 11. Correlations map of the reflectance intensities (left), the logarithm of the intensities (middle) and the subtraction between the log of the intensities at 365 and 310 nm (right).

Regression models

Independently of the regression algorithm used, we tested three main types of models: a multi-output model, two independent single-output models, and two separated single output models using the TUR prediction as a feature to train the COD model. The Figure 12 shows the training and testing methods used for those three types.

We trained the models with different types of inputs extracted from the TUR and COD data. TUR models are tested with the following inputs: the intensity of the whole spectrum, the intensity within the near-visible and visible range (365 - 700 nm), the mean of the intensity at 250 - 700 nm, and the mean of the intensity at 365 to 700 nm. The chemical demand oxygen model is tested with the intensity of the spectrum, the intensity in the UV range, the logarithmic of the two previous inputs, the mean of the logarithmic inputs, and the subtraction of the intensity at 365 nm with the one at 310 nm. The TUR prediction can be added as a feature for each of the inputs cited previously. The comparison is done always using the same TUR model, the one with the highest R^2 score.

If the model were randomly separated within a train and test set, the results would be biased because another spectrum of the same concentration would have probably been used for the training of the model, which would lead to overly good results. Therefore, each group of samples must be considered as one data, the training and testing are performed using the leave-one-out cross-validation. The dataset is separated into twenty-seven groups corresponding to the concentrations. The following method is done for each group. The model, either the multi-output, the TUR, or the COD one is trained with the twenty-six other groups, then we predict the concentrations of all the elements in the remaining group. Those predictions are kept to evaluate the model when all the concentrations are predicted.

Three classical machine learning algorithms are trained and evaluated for the measurement of the TUR and COD: the Partial Least Square Regression, a popular method in Chemometrics, the Support Vector Machine Regression, and the Random Forest Regression.

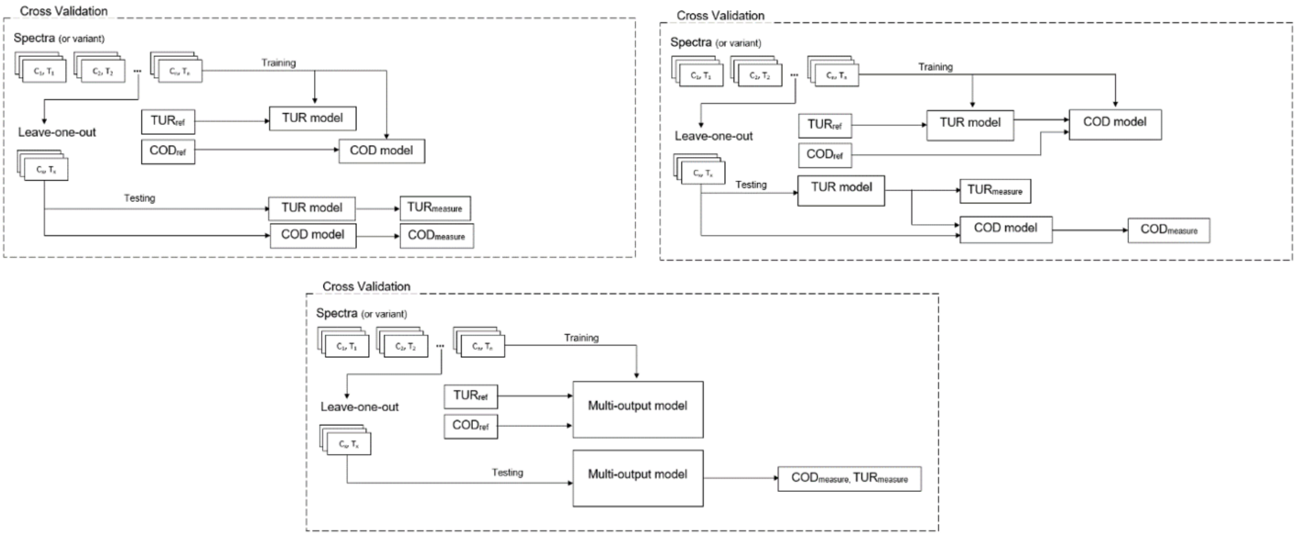


Figure 12. Training of (up left) two single outputs (up right) two single-output models using the TUR prediction to train the COD model (down) the multi-output model, and testing with the cross-validation leave-one-out technique.

Results

The best PLS model for the TUR is obtained with the intensity of the spectra in the range 365 - 700 nm as inputs, with 1 component as parameter. The model has a score of R^2= 0.99 with a mean absolute error of 15.48 and a mean square error of 490.48. The best model for the COD model using the TUR prediction is the one using the logarithmic of the intensity of the whole spectra with 14 components. The model has a score of R^2 = 0.93 with a mean absolute error of 40.10 and a mean square error of 3349.91. The best model for the COD that doesn't use the TUR prediction as a feature is the same one as the TUR prediction. It has a score of R^2 = 0.93 with a mean absolute error of 39.44 and a mean square error of 3243.65. The difference between the two models is negligible. Figure 13 shows the prediction obtained with the leave-one-out method, COD model tends to underestimate the concentrations for high COD levels.

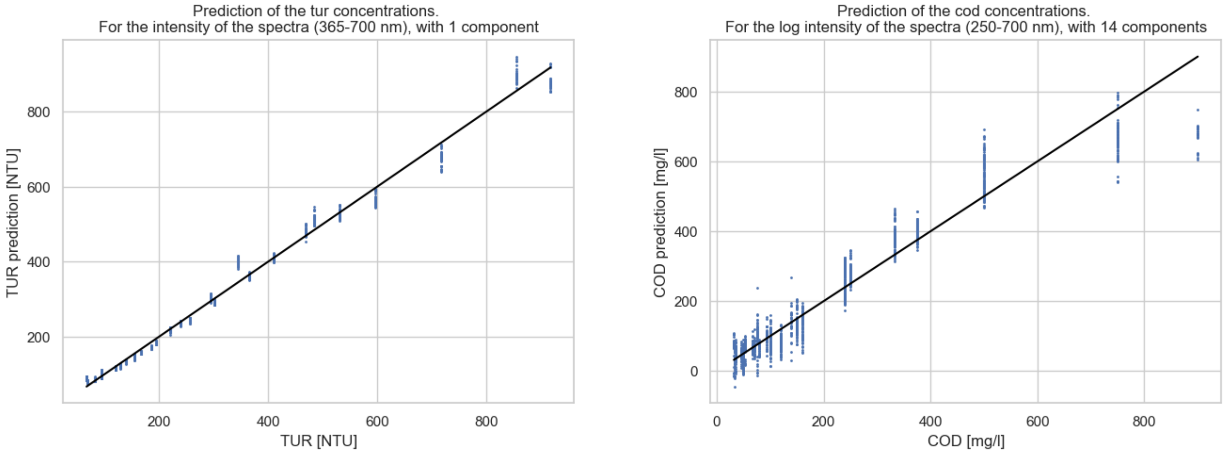


Figure 13. Predictions of the PLS models. (left) The TUR model with the intensity in the range 365 - 700 nm as input and 1 component. (right) The COD model, with the logarithmic intensity of the whole spectra without using the tur prediction as feature and with 14 components

The best SVM model for the turbidity, shown in Figure 14, is the model using the intensities at the wavelengths 365 to 700 nm with a penalty parameter of 0.14. The model has a score of R^2 = 0.99 with a mean absolute error of 15.31 and a mean square error of 518.27. The best COD model is the one with the logarithmic intensities in the band 250 to 700 nm not using the predictions of the turbidity, and a penalty set at 10,000. The model has a score of R^2 = 0.92 with a mean absolute error of 35.75 and a mean square error of 3540.35.

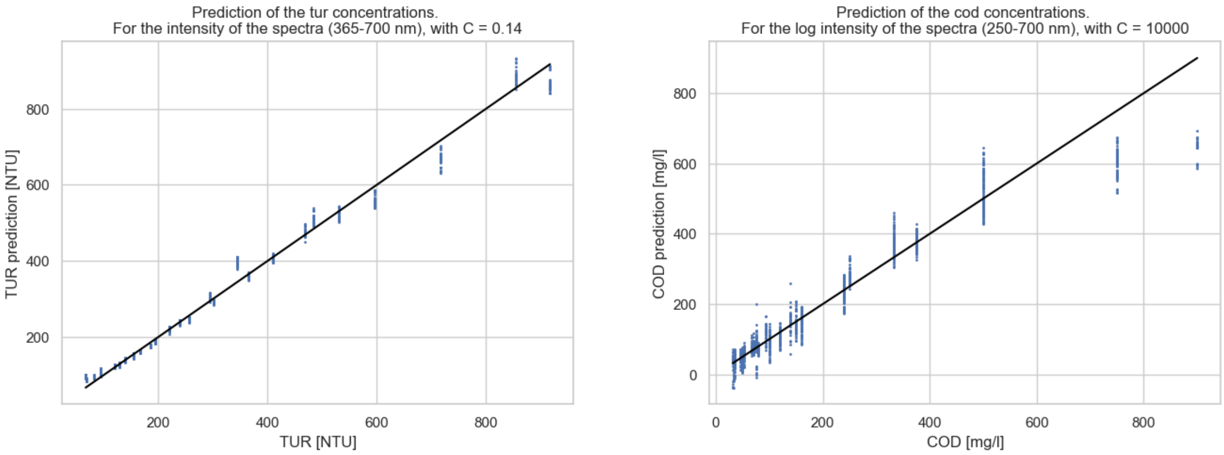


Figure 14. Prediction of the best SVM models. (left) The turbidity model with the intensity in the range 365-700 nm as input and C=0.14 (right) The COD model, with the logarithm of the intensities at the range 250-700 nm with the TUR prediction, with the parameter set at 10,000.

Random Forest (RF) is popular for the prediction of regression problems, it is an ensemble algorithm as it includes multiple decision trees. When using the RF for regression, there are two important parameters to optimize: i) the number of features to consider when looking for the best split and ii) the criterion to measure the quality of a split. The optimization is performed with a grid search using the leave-one-out cross validation to compute the score. The parameters leading to the highest R^2 score is kept as the optimal parameters. The values of the parameters are the following:

- *max_features* = ['sqrt', 'log2', 0.001, 0.1, 0.33, 0.7, 1, 'auto']
- *criterion* = ['squared_error', 'absolute_error', 'poisson']

Where sqrt = 3.6, log2 = 1.1 and auto is the number of features. The number of trees is set at 10 because the dataset is quite small.

The best model for the turbidity, shown in Figure15, is the model using the intensities at the wavelengths 365 to 700 nm, with the parameters: *max_features* = sqrt and *criterion* = poisson. The model has a score of $R^2 = 0.97$ with a mean absolute error of 23.44 and a mean square error of 1171.70. The best COD model is the one with the intensities in the band 250 to 310 nm not using the predictions of the turbidity. The parameters are *criterion* = poisson, *max_features* = auto = 4. The model has a score of $R^2 = 0.71$ with a mean absolute error of 82.94 and a mean square error of 13107.67.

Random forest algorithm support multi-output regression, however, the results with either the intensity or the logarithm of the intensity as inputs are lower than the independent models. We obtain a score for the first model of 0.86 and 0.34 for the turbidity and COD, and a score of 0.86 and 0.43 for the second model.

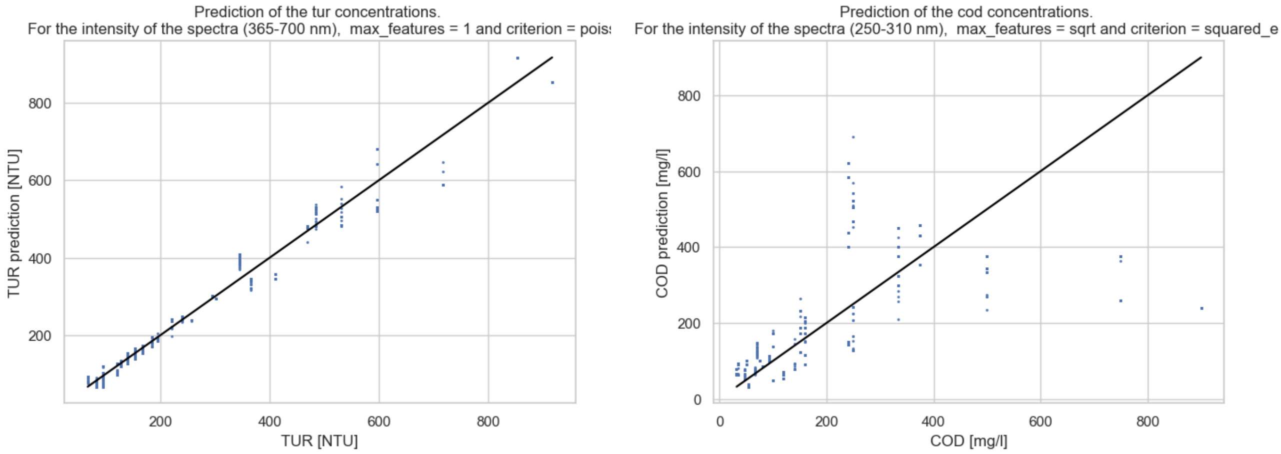


Figure15. Prediction of the best RF models. (left) The turbidity model with the intensity in the range 365-700 nm as input and max features = 1 and criterion = poisson. (right) The COD model, with the intensities at the range 250-310 nm.

Table 3 summarizes the best results of the three algorithms and the previous result obtained in the Agustsson et al., article. This demonstrates that we were able to obtain better results, that those obtained with a more classical spectrometry sensor.

Table 3. Results of the different models.

Concentrations	Model	R^2	MAE	MSE
TUR	Agustsson	0.95		
	PLS	0.99	15.48	490.48
	SVM	0.99	15.31	518.27
	RF	0.97	24.08	1218.13
COD	Agustsson	0.69		
	PLS	0.93	39.44	3243.65
	SVM	0.92	35.75	3540.35
	RF	0.71	85.94	13141.6

5. CONCLUSION

Wastewater quality monitoring in raw wastewater, and especially sewers, remains a challenging topic. Current solutions involve complex infrastructures, and/or costly devices that requires significant maintenance. Our approach improves accuracies in the recent non-contact water quality measurements using multispectral imaging (MSI) in the UV-Vis range and active illumination based on LED technology - at lower cost, and virtually maintenance-free.

In this article, we present the detailed experimental methods and data analysis pipeline we applied to successfully predict turbidity (TUR) and Chemical Oxygen Demand (COD) of synthetic wastewater sample in the range of 75-950 NTU and 32-900 mgO_2/l. This technology shows promising results with a mean absolute error of less than 10% and high R^2 of 0.93 for TUR and 0.99 for COD. Regarding data analysis, our results suggest that the support vector machine algorithm delivers similar results, whilst random forest regression is the least efficient. These results are extremely encouraging as they represent a first step into non-contact water monitoring. However, we used samples synthetic wastewater in a controlled environment and with comparably few samples. Therefore, the next step is to collect a comprehensive dataset of real wastewater, from different sites, which should better capture the real-world variability. Challenges from different monitoring geometries, i.e. changing water levels, and lighting conditions will also need to be addressed. Finally, it will be required to adapt and test the algorithms as well as data transmission and power management for long-term operation in remote monitoring locations to allow an accurate IoT real time wastewater measurement to improve management of these systems.

ACKNOWLEDGEMENT

The results of this research were funded by Innosuisse, CSEM SA and Photrack AG, with the collaboration of the EAWAG.

REFERENCES

[1] J. Rose, «Cleaning the world's water: 'We are now more polluted than we have ever been,» The Guardian, Aug 2016. [En ligne]. Available: https://www.theguardian.com/environment/2016/aug/31/cleaning-the-worlds-water-we-are-now-more-polluted-than-we-have-ever-been.

[2] J. Agustsson, O. Akermann, D. Barry et L. Rossi, «Non-contact assessment of COD and turbidity concentrations in water using diffuse reflectance UV-Vis spectroscopy,» Environmental science. Processes & impacts, vol. 16, 2014.

[3] APHA, AWWA, WEF, A.E. Greenberg, L.S. Clesceri, A.D. Eaton (Eds.), Standard Methods for the Examination of Water and Wastewater (22th edition), NW, Washington (2012)

[4] K. Preitner, «Multi-spectral imaging for contact-less assessment of chemical oxygen demand and turbidity concentrations in wastewater», Master Thesis, EPFL 2022

[5] C. Anderson, «Chapter A6. Section 6.7. Turbidity,» chez U.S. Geological Survey Techniques of Water-Resources Investigations, 2005.

[6] D. Guyane, Le guide technique pour le suivi de la turbidite en Guyane, DEAL Guyane, 2018

[7] O. Thomas et M.-F. Thomas, «Chapter 9 - Urban Wastewater,» chez UV-Visible Spectrophotometry of Water and Wastewater (Second Edition), Elsevier, 2017.