**In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems**

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Urban Water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>NURW-2012-0116.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Full Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Brito, Rita; Escola Superior de Tecnologia do Barreiro, Instituto Politécnico de Setúbal, Construction and Environment Department Pinheiro, Helena; Instituto Superior Técnico, Universidade Técnica de Lisboa, Centre for Biological and Chemical Engineering, Institute for Biotechnology and Bioengineering (IBB), Department of Bioengineering Ferreira, Filipa; Instituto Superior Técnico, Universidade Técnica de Lisboa, Department of Civil Engineering, Architecture and Georesources Matos, José; Instituto Superior Técnico, Universidade Técnica de Lisboa, Department of Civil Engineering, Architecture and Georesources Lourenço, Nídia; Instituto Superior Técnico, Universidade Técnica de Lisboa, Centre for Biological and Chemical Engineering, Institute for Biotechnology and Bioengineering (IBB), Department of Bioengineering</td>
</tr>
<tr>
<td>Keywords:</td>
<td>UV-Vis, in-line, COD, TSS, wastewater systems</td>
</tr>
</tbody>
</table>
In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems

R. Salgado Brito\textsuperscript{a}, H.M. Pinheiro\textsuperscript{b}, F. Ferreira\textsuperscript{c}, J. Saldanha Matos\textsuperscript{c}, N.D. Lourenço\textsuperscript{b}

\textsuperscript{a}ESTBarreiro/IPS, Rua Américo da Silva Marinho, 2839-001 Lavradio, Portugal; Tel: (351) 21 206 46 60, rita.brito@estbarreiro.ips.pt

\textsuperscript{b}Centre for Biological and Chemical Engineering, Institute for Biotechnology and Bioengineering (IBB), Department of Bioengineering, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001, Lisboa, Portugal; Tel: + 351 21 841 91 25; helena.pinheiro@ist.utl.pt; nidia.lourenco@ist.utl.pt

\textsuperscript{c}Department of Civil Engineering, Architecture and Georesources, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049 - 001 Lisboa, Portugal; Tel: (351) 21 841 83 71; jsm@civil.ist.utl.pt; filipaf@civil.ist.utl.pt

Corresponding Author:

Rita Salgado Brito

ESTBarreiro/IPS, Rua Américo da Silva Marinho

2839-001 Lavradio, Portugal

Tel: (351) 21 206 46 60

Fax: (351) 21 207 50 02

E-mail: rita.brito@estbarreiro.ips.pt

ABSTRACT
Ultraviolet-visible (UV-Vis) spectroscopy is a promising tool for fast and simple evaluation of wastewater quality, as it delivers spectra that may be correlated to various aggregate wastewater quality parameters, such as Total Suspended Solids (TSS) and Chemical Oxygen Demand (COD). Its application in drainage systems has scarcely been reported and is a step forward in water quality monitoring.

Partial Least Squares (PLS) calibration models were developed for TSS and COD prediction, based on UV-Vis spectra. Spectra were acquired in 2 sites, in-line and off-line. Three approaches for model development were studied. Model evaluation statistics and guidelines were assessed, which allowed grading the models from Unsatisfactory to Very Good. Most of the developed models were rated Very Good according to those guidelines. The overall results strongly indicate UV-Vis spectra to be reliable for TSS and COD estimation in sewer systems, despite the rapid variations in hydraulic conditions and water quality.

Keywords: UV-Vis, in-line, COD, TSS, wastewater systems.

1 INTRODUCTION

In Portugal, many drainage systems are either combined or separate with unwanted rainwater inflows and pollutant discharges are frequent (Matos, 2003). Combined sewer overflows may jeopardize public safety and health and damage receiving water quality. However, monitoring of discharges in Portugal is practically inexistent. Traditionally, pollutant load data is acquired from campaigns involving composite or grab sampling, sample conservation, transport and analysis using standard methods. This procedure does not allow for real time decisions. On-line sensors measuring temperature, oxidation-reduction potential, pH, dissolved oxygen, conductivity and turbidity have proven their reliability and
robustness in drainage systems. However, the information provided by each of these sensors is very limited for control actions (Vanrolleghem and Lee, 2003). Multi-wavelength spectroscopy in the ultraviolet-visible range (UV-Vis) has been previously used for wastewater and stormwater quality monitoring (Van der Broeke et al., 2006; Lourenço et al., 2008; Torres and Bertrand-Krajewski, 2008), since most organic compounds and some soluble minerals (such as nitrate) absorb in the UV-Vis region (Vaillant, et al., 2002). On-line UV-Vis sensors have already been used for wastewater treatment plant operation (Vanrolleghem and Lee, 2003; Schuetze, et al., 2004) and robust submersible commercial devices are also becoming available. However, their application in wastewater drainage systems has scarcely been reported (Gruber et al., 2006; Rieger et al., 2006; Torres and Bertrand-Krajewski, 2008). In drainage systems the hydraulic conditions are extremely variable, fouling may occur and there is limited access for maintenance. These circumstances may contribute to the collection of inaccurate data. Additionally, wastewater in drainage sewers may not present a homogenous composition along its cross section. This creates a challenge for spectroscopy, which can be minimized by choosing a location with turbulent flow. In drainage systems with stormwater connections, flow may change rapidly, modifying the water quality matrix (Maribas, 2008), increasing fouling probability and worsening access conditions, which adds pressure to in-line monitoring.

Statistical models can enhance the scope and significance of acquired spectral data (Aguado and Rosen, 2008). Partial Least Squares (PLS) regression is a multivariate statistical method that has been used in this context to predict values for standard aggregate wastewater quality parameters, such as total organic carbon (TOC), chemical oxygen demand (COD) and total suspended solids (TSS) (Rieger, 2006, Lourenço, et al., 2008).

The application of regression algorithms has to follow certain premises, that is, data ought to be independent and have a normal distribution. Using traditional PLS in water quality estimation has
already been criticized because quality processes may not meet the basic algorithm assumptions, due to the lack of a normal distribution (Rosen and Lennox, 2001). Further details on the PLS algorithm can be found in the literature (Wold, 1966; Geladi and Kowalsky, 1986; Tobias, 1997).

A submersible UV-Vis spectrophotometer was used to monitor a paper mill wastewater treatment plant through *in situ* spectra acquisition in the range of 200-750 nm (Langergraber et al., 2004a). Simultaneous prediction of COD and filtered COD was achieved by a global PLS calibration model included in the equipment software and improved through local adjustments based on grab samples (local calibration). The prediction ability of the PLS calibration models was only evaluated in terms of the coefficient of determination ($R^2$), which was always higher than 0.90. Nitrate, TSS and filtered COD values were also predicted in an activated sludge sequencing batch reactor (SBR) with $R^2$ values of 0.98, 0.995 and 0.90, respectively (Langergraber et al., 2004b). A similar approach was used for *in situ* estimation of COD and TSS in a stormwater separate sewer system, at the inlet of a storage-settling tank, with $R^2$ values higher than 0.91 and 0.99, respectively (Torres and Bertrand-Krajewski, 2008). The prediction ability of the PLS models was also assessed by the root mean squared error of prediction (RMSEP) and this indicator was considered more pertinent and discriminative than $R^2$. One study reported the use of submersible UV-Vis spectrometers installed at combined sewer overflow structures to estimate TSS and COD (Gruber et al., 2006). Two different installation solutions were compared, i.e., a floating device within the sewer structure and a by-pass measuring flume located outside the sewer. Different calibration results were obtained but quantitative model performance evaluation was not provided. Since ideal evaluation criteria are difficult to establish, several efficiency indicators should be applied to evaluate PLS model performance.

PLS calibration models for estimating TSS and COD in two different locations of a wastewater drainage system near Lisbon, Portugal, are presented in this paper, based on multi-wavelength spectroscopy in the UV-Vis range. Various statistics are applied and performance ratings are proposed.
for PLS model evaluation, which enable developed models to be classified from Unsatisfactory to Very Good. This study aims to evaluate the applicability of UV-Vis spectroscopy to drainage systems, and in particular to:

- Examine whether UV-Vis spectra can be used to develop robust estimators of TSS and COD, with statistical regression performed on spectra acquired off-line, in collected samples;
- Assess if in situ TSS and COD predictions provided by a submersible UV-Vis probe are reliable;
- Evaluate whether TSS and COD predictions provided by the submersible probe can be optimized by performing statistical regression on the in-line spectra;
- Test the application of several statistical parameters for PLS model evaluation.

2. METHODOLOGY

2.1 Monitoring stations and sampling campaigns

Campaigns were implemented in two sites of a drainage system in the Lisbon district, Portugal, which differ in terms of drainage area and land use. Sites 1 and 2 are located in the Rio da Costa Interceptor, which is operated by SIMTEJO – Saneamento Integrado dos Municípios do Tejo e Trancão, S.A.. This interceptor is 8.8 km long. Site 1 is located upstream, receives most domestic wastewater contributions of the whole basin and is located before major industrial connections. Site 2 is located downstream, a few manholes before the wastewater treatment plant. Both sites are located in a sewer 1500 mm in diameter. The hydrographic basin has densely populated urban areas side to side with large farms and an industrial campus. This interceptor receives wastewater from separate and combined systems. The velocity profile was not permanent and hydraulics changed rapidly in wet weather conditions. Unstable eddies involving circular motion were detected. Turbulence in the fluid tends to promote radial mixing,
which favors sample representativeness. However, air bubbles occasionally appeared which might cause interference on the optical signal and result in inaccurate in situ spectral acquisition. Sampling was performed with a 5L bucket, with a pole device, which aimed at the centerline of the flow. On site measurements were performed inside the sampling bucket and two 0.5 L bottles were filled from it, one for TSS and turbidity measurements and another for COD. The inherent sample representativeness issues of this procedure are addressed in section 3.5. The sampling campaigns occurred between March and December 2010. Nine campaigns were carried out, in different weather conditions. Dry weather campaigns occurred after at least 5 days without rainfall and wet weather campaigns occurred after at least 8 hours of rainfall. Campaigns lasted from 3 hours up to 7 hours and 4 to 14 samples were collected in each campaign. Site 1 was monitored in the first four campaigns and site 2 was monitored in all nine campaigns. A total number of 93 and 139 samples were available in sites 1 and 2, respectively. Off-line ex situ spectra were obtained from samples collected in all campaigns. In-line in situ spectra (in-line spectra) were only obtained from three wet weather campaigns, at site 2. Due to this specificity, this subset of samples (12, of which only 11 had TSS results) is further designated as belonging to site 2*. The number of samples available from each site and for each climatic condition is presented in the Results section.

2.2 TSS and COD analyses

Standard TSS and COD laboratory analyses of collected samples were performed in every campaign. TSS was determined based on a standard method (APHA, 1995) using GF/C glass fiber filters and a HB43-S Moisture Analyzer (Mettler Toledo). COD was determined using test kits (COD Cell Test ref 14690WTW). The digestion step was performed in an Accublock digital dry bath (Labnet) and the measurements were conducted on a Spectro-flex 6600 photometer (WTW).
All TSS and COD determinations were done in duplicate, on two subsamples from the same sample. The difference between duplicates was calculated and 95% confidence intervals were determined. The samples that fell outside the 95% confidence interval were further examined through pH, turbidity and temperature measurements in order to determine whether those samples should remain in the database or could be classified as outliers.

2.3 UV-Vis spectral acquisition

2.3.1 Off-line (ex situ) spectra

UV-Vis spectra of the collected samples were acquired off-line in a Specord 200 laboratory bench spectrophotometer (Analytik Jena) between 190 and 800 nm with 1 nm interval using a quartz cell with a 10 mm path length. Spectra were acquired from raw samples and after appropriate dilution (1:10) in order to overcome signal saturation in the UV region. In raw samples, records on wavelengths lower than 220 nm were rejected, as in most samples absorbance in this region exceeded the equipment saturation limit.

2.3.2 In-line spectra

UV-Vis spectra were acquired in situ with a UV-Vis portable submersible probe (Spectro::lyser, S::can), between 200 and 750 nm with a 5 mm optical path length and 5 nm interval. A global calibration model for COD and TSS estimation in municipal wastewater is a part of the available probe software (ana::pro, from S::can).

2.4 Partial Least Squares (PLS) calibration models

2.4.1 General aspects

PLS is a statistical multivariate method used to find the underlying relations between two matrices, X and Y. This method is based on a guided decomposition model where the dependent variables, Y,
intervene directly in the decomposition of the independent variables, X. PLS searches for a set of components, named latent values (LV), that provides a simultaneous decomposition of both X and Y, so that those components explain the covariance between X and Y.

2.4.2 Calibration models available from the submersible equipment software

Global calibration is performed through the introduction, by the manufacturer, of several spectra and correspondent COD or TSS obtained from similar wastewater samples. In this specific project, a global calibration for domestic sewage was selected. Local calibration is advised by the manufacturer as part of model development. Local calibrations were performed using in-situ spectra of a limited number of site specific samples and the corresponding TSS and COD laboratorial results. Local calibration consists of an adjustment to global calibration, as the information that is manually introduced into the probe’s software (spectra and laboratorial results) refers specifically to the site under study.

2.4.3 Proposed PLS Models

The proposed monitoring strategies rely on two sequential steps: the acquisition of spectra and PLS regression of the data. Two types of models were developed separately:

- Model A: PLS model based on UV-Vis spectra acquired off line
- Model B: PLS model based on UV-Vis spectra acquired in line

Comparing Models A and B addresses the robustness of models based on different spectral acquisition procedures. In-line spectra are determined in a more aggressive, heterogeneous and variable environment. In this study, in-line spectra also provided less information, because wavelength range is slightly lower and the wavelength interval is higher than in laboratory spectra. Besides all the logistic aspects of an in situ installation inside a sewer, comparing models A and B acknowledged whether the collected spectra were reliable and enabled the development of dependable PLS models.
The results of model B were further compared to the results of the model included in the submersible probe software (Model C). Spectra acquired in-line are used in Model B, but also in Model C as a local calibration. In Model C, local calibration is solely an adjustment to global calibration. Global calibration relies on many spectra, available in the probe’s software, which are unknown to the user; model parameters, validation procedure and outlier management are also not identified. In that sense, the user is not thoroughly in control of what has been decided, accepted or discarded in Model C. Comparing the results from models B and C addresses these issues.

2.4.4 PLS model development

2.4.4.1 Data adequacy pre-tests

Standard regression methods usually fail when applied to time series data due to correlations between sequenced observations. Even though UV-Vis spectra and water quality data have been collected in a time sequence, data did not present characteristics of a time series, such as a repetitive pattern or a consistent trend. On the other hand, spectral data present autocorrelation between absorbance values at different wavelengths. However, PLS is particularly suited for application to correlated data since a small number of new uncorrelated variables (LV) is determined and used whenever data is normally distributed. In this project, data adequacy for the PLS model development was verified through the Shapiro-Wilk test (Shapiro and Wilk, 1965) for normal distribution evaluation. This test was performed on TSS, COD and spectra from these campaigns. As far as spectra were concerned, three wavelengths were chosen, namely, 250, 460 and 550 nm. The wavelength choice depended on the PLS results concerning variable selection (see section 2.4.4.3).

On the other hand, the adequacy of a model to a specific set of data may be perceived after model development, through the analysis of the residuals series. Although this verification is not a pre-test on data, it allows confirmation of whether all valuable information in data was adequately explained by
the selected model. Residuals should be independent and follow a normal distribution, with a constant
mean and variance. The assumption of residual independence requires that the ordered residual terms
display no autocorrelation, as autocorrelation in residuals is a symptom of a systematic lack of fit in
data (Ravagan, 2008). Moreover, serially correlated residuals are also a symptom of linear dependence
between observations (Box, et al., 1976). If the residuals display the ideal behavior, then testing them
becomes a tool for the validation of the chosen model.

Independence was checked through the autocorrelation Box-Ljung test (Ljung and Box, 1978). Both
this and the Shapiro-Wilk test were performed using SPSS Statistics 19 (IBM). If normal distribution
and lack of autocorrelation are not to be rejected, for a 0.05 significance level, the p-value of the
Shapiro-Wilk and Box-Ljung tests are higher than 0.05.

2.4.4.2 Data pre-processing and outlier detection

The Savitzky-Golay filter (Savitzky and Golay, 1964) was applied to spectral matrices and all matrices
were mean-centered. Analyzing score and loading plots enables a better understanding of the data
structure and also detection of outliers in the measured data (Rosen and Lennox, 2001). Outliers were
identified in score plots (scatter graph of the score of each sample in the two first LV) and in the plots
of Hotelling $T^2$ and $Q_{\text{residuals}}$ statistics. Samples that fell outside the 95% confidence interval in one of
these plots were identified as outliers. These outliers were further studied in order to evaluate if they
corresponded to underrepresented normal samples, in which case they were kept in the model.

2.4.4.3 PLS model calibration and internal validation

PLS was used to predict TSS and COD values from UV-Vis spectral data. In this project, the
independent $X$ matrix was composed of one UV-Vis sample spectrum in each row, with the respective
absorbance for each wavelength, in every column. The dependent $Y$ matrix was composed of
laboratory results for TSS or COD for each sample. Models are usually built on a training set (2/3 of available data) and then validated with an external test set (remaining data).

In the present study, each training set was submitted to full cross-validation (CV), through which the models were both calibrated and internally validated. A cross-validation procedure using “contiguous blocks” was adopted in matrices containing more than 15 samples. With this procedure, a group of sequential samples from the training set is associated to internal validation, while all the remaining samples are used for model calibration. In an iterative sequence, this group is then associated to calibration, and a next block is left out for internal validation. This sequence goes on until every block was used once in validation. A cross-validation “leave-one-out” procedure was implemented for smaller matrices (with up to 15 samples). With this procedure, isolated samples are sequentially selected and left out of the calibration set. The error of each run was calculated by RMSECV (root mean squared error of cross-validation) according to Equation 1 (see section 2.4.5). The optimal number of latent values (LV) in the PLS models was based upon minimum RMSECV.

In the smaller matrices, all the available data was used as the training set and an external validation was not performed. With cross-validation, a prediction ability of the model is readily available from the training set (Tobias, 1997).

Given the whole UV-Vis spectrum, information in some intervals of wavelengths may actually be irrelevant to COD or TSS, and it may be better to discard these intervals, in order to provide a better prediction model (Siegrist, 2006, Lourenço, et al., 2010). Interval PLS (IPLS) selects a subset of variables that will result in a model with superior prediction (lower RMSECV) compared to using all the variables in the data set. IPLS does a sequential, exhaustive search for the best interval within the variables. The interval in IPLS can either be a single variable or a “window” of adjacent variables. For each model, the interval selected by IPLS varies, so this information is displayed in the Results section (see tables 3 and 4 in section 3.4).
Both PLS and IPLS models were developed in Matlab 7.4.0 (*The Matworks Inc.*) with the specific package PLStoolbox 3.0 (*Eigenvector Research Inc.*).

### 2.4.5 PLS model performance evaluation

Model evaluation allows one to determine if the best model provided by PLS is good enough for the intended purposes. Model evaluation may be performed by a variety of statistics and such statistics must be rated.

It is advisable not to rely solely on correlation indicators. The coefficient of determination, $R^2$, frequently used, may produce high values in poor models, is insensitive to additive and proportional differences and is over sensitive to high extreme values (Legates and McCabe, 1999; Krause, et al., 2005; Harmel, et al., 2006). Krause, et al. (2005) tested the application of several efficiency criteria to watershed models, namely RSR (root mean squared error standard deviation ratio of observations), NSE (Nash-Sutcliffe coefficient of efficiency) and PBias (percentage bias). None of them performed ideally, so a combination of criteria should be complemented by absolute or relative errors. The inclusion of graphical information is also recommended (Legates and McCabe, 1999; Moriasi, et al., 2007). RER (range to error ratio) has been applied for PLS model evaluation in pharmaceutical processes (Sarraguça, et al., 2011). The error associated with predicted values in PLS models may be determined either by RMSECV (RMSE in cross-validation) or with RMSEP (RMSE in prediction), whenever an independent test set is available.

An independent test set was available for matrices with more than 15 samples. In such cases, it was possible to submit the models to new data that was not involved in model development. An external validation was possible and RMSEP was determined. Even for Model C, only a subset of the available *in situ* data was used for local calibration. Estimated TSS and COD values for the remaining local spectra may be considered predictions, allowing an external validation.
Given the recommendations on model evaluation, the following criteria were used (Table 1): RMSE (Equation 1, either RMSECV or RMSEP); RMSE$_{rel}$ (Equation 2); RSR (Equation 3); NSE (Equation 4); PBias (Equation 5); and RER (Equation 6).

{Insert Table 1}

Both NSE and PBias are recommended by ASCE (1993) and PBias has the ability to clearly indicate a poor model (Gupta, et al., 1999). NSE is largely used in watershed models and is widely reported (e.g. Moriasi et al., 2007). Relative RMSE (either RMSEP or RMSECV) determines the magnitude of the error. RSR standardizes RMSE and allows for the establishment of guidelines for model evaluation (Sing et al., 2004).

The NSE determines the relative magnitude of the residual variance compared to the observed data variance (Nash and Sutcliffe, 1970). NSE equal to 1.0 is the best value, whereas NSE less than or equal to 0.0 means that using the mean observed value is a better estimate than the predicted values.

PBias measures the average tendency of the predicted data to be larger or smaller than the corresponding observed data. The optimal value of PBias is 0.0. RER evaluates the importance of RMSE regarding the amplitude of measured data.

In order to rate these statistics, Moriasi, et al. (2007) established guidelines for model evaluation (namely for RSR, NSE and PBias). These models were based on detailed watershed description for the simulation of streamflow and transport of sediments and nutrients. For this purpose, it can be assumed that the processes in combined sewer systems are roughly the same, as far as streamflow, sediments or organic matter are concerned.

Moriasi, et al. (2007) proposed performance ratings for RSR and NSE that are the same for streamflow or quality modeling, but proposed less stringent PBias for water quality. The studied models described physical, chemical and biological processes, subject to parameter uncertainty. However, as in the
present study a description of the processes was not made, it was assumed that performance ratings for TSS and COD had to be as rigorous as the ones proposed for streamflow.

The presented ratings were determined based on daily or monthly time steps; the authors underlined that less accurate estimates were expected for shorter time steps. In the present study, even though the time step is shorter, performance ratings proposed by Moriasi, et al. (2007) were used.

No recommendations were found in the literature for rating RMSE_{rel}. The authors used as guidelines the values presented in Table 2.

{Insert Table 2}

3 RESULTS AND DISCUSSION

3.1 TSS and COD analysis

TSS and COD were determined in laboratory, as defined in section 2.2. TSS ranged from 120 mg/L to 410 mg/L and COD ranged from 220 mg/L to 800 mg/L, as shown in Figure 1 (the upper and lower quartiles are represented in light and dark grey, respectively, and the upper and lower extreme in light lines). These box-and-whisker plots illustrate data distribution and provide a visual presentation of the information given by the Shapiro-Wilk test.

{Insert Figure 1}

For most situations, data obviously did not have a normal distribution. In both sites, due to campaign restrictions and weather conditions, more samples were available from dry weather than from wet weather. The number of samples (n) available for each group is presented in Figure 1. The ranges of measured data values were wider in the former, probably due to the absence of the rainwater flushing or dilution effects, which tend to dampen the daily or seasonal variations in the composition of the collected sewage. Average TSS and COD were higher in dry weather, which was expected. All
determinations in Sites 1 and 2 were done in duplicate. The average difference between duplicates was 22.1 mg/L for TSS and 31.5 mg/L for COD. Outliers were analyzed as previously mentioned.

3.2 UV-Vis spectra

For some UV-Vis spectra acquired in laboratory, the maximum detection limit of the spectrophotometer was exceeded and samples had to be previously diluted, with a dilution factor of 10. Spectra were submitted to preliminary analysis before PLS, so as to identify different patterns, based on principal components analysis (Booksh, 2000; Lourenço, et al. 2010). This previous step identified that spectra followed a pattern for most samples, but that patterns differed from samples collected in dry weather, as opposed to wet weather. Preliminary analysis also identified different patterns for sites 1 and 2 (sites in the same sewer line). In site 2, for some campaigns, spectra were acquired both in-line, with the submersible probe, and off-line, with a bench spectrophotometer. For all samples, spectra acquired in-line had slightly higher values of absorbance than spectra acquired off line. However, this wavelength difference was not constant throughout the entire spectral range, which might prevent the use of spectra from both sources in the same PLS model.

Figure 2a) illustrates dry weather spectra. The first image on the left illustrates a collection of spectra, acquired off-line in site 2. The vertical lines correspond to wavelengths identified in IPLS. The central image displays a comparison between two off-line spectra, one from site 1 and one from site 2. The image on the right shows two spectra from site 2, one acquired in-line and the other off-line. Figure 2b) illustrates wet weather spectra, with the same information as the above mentioned.

{Insert Figure 2}

The differences between spectra, observed in Figure 2, supported the decision to develop separated PLS models for each site, for different hydrological conditions and for spectra acquired either off-line (Model A) or in-line (Model B).
3.3 Data adequacy pre-tests

TSS, COD and spectra normal distribution were examined to check the premises of PLS regression. The wavelengths used (see section 2.4.1) were selected because, in most IPLS models, relevant intervals surrounded these values (Figure 2). The application of the Shapiro-Wilk normality test revealed that data did not present a normal distribution, failing on this PLS premise. Nevertheless, after model development, residuals were tested to verify if all valuable information was explained by the PLS models (see section 3.4).

3.4 PLS model development

Models A (off-line spectra) were developed both for raw and diluted samples. In the present study, raw samples provided better PLS results than diluted samples. Only results on raw samples are discussed for Models A.

Both PLS and IPLS were performed on all matrices for Models A and B (in-line spectra), and the best combination of wavelengths was determined for COD and TSS. IPLS provided better results than PLS, so any further reference to PLS regression actually refers to the IPLS technique.

Tables 3 and 4 summarize PLS reference information for each model, such as the number of samples (n), the number of outliers that were identified and removed (out), the number of latent values (LV) that provided the best model, the correspondent value of RMSECV and the wavelength interval identified in IPLS (λ). RMSECVrel is also presented, to put errors into context. Table 3 refers to TSS models and Table 4 refers to COD models. Whenever information is not available, as for Model C, “-” has been inserted.

As previously mentioned, in situ spectra were only available from a subset of samples (designated site 2*). These samples were used for the three models (Models A, B, and C). For Model C, the results
refer to models using all the spectra available for global calibration (in an unidentified number, u.n.) and 5 samples in local calibration.

{Insert Table 3}

{Insert Table 4}

The distinctive pattern illustrated in Figure 2 in spectra acquired in-line and off-line justifies why PLS characteristics (determination of outliers, number of required LV, selected intervals in IPLS, just to name a few) may be different from Model A to Model B, even if the models were developed for the same site and the same weather characteristics.

Considering all the models, the average RMSE for TSS was 18.0 mg/L, and for COD it was 45.4 mg O₂/L. These results were in line with the differences between duplicates obtained in laboratory for TSS and COD, as presented in section 3.1. A detailed model evaluation is provided in section 3.5.

Model residuals were studied. The results from the Shapiro-Wilk normality test showed that although data normality was not assured, residuals followed a normal distribution. This enabled to establish the conclusion that no relevant information had been disregarded by the models and left out in the residuals.

All in all, it can be concluded that, although PLS premises were not all fulfilled by data in these wastewater campaigns, this did not compromise the use of PLS models. There was no evident need to evolve to more complex algorithms.

3.5 Model evaluation

As previously mentioned, TSS and COD prediction was only performed when more than 15 samples were available. This was the case of Models A, in sites 1 and 2, and Model C, in site 2*.

Tables 5 and 6 summarize the information for each model, such as the number of samples (n), the number of outliers (out) and the correspondent value of RMSEP and RMSEP_rel. Table 5 refers to TSS
models and Table 6 refers to COD models. Whenever information is not available, “-” has been inserted.

{Insert Table 5}

{Insert Table 6}

For TSS models, RMSEP values (Table 5) are similar to RMSECV values (Table 3), which was a sign of model robustness, since RMSEP is determined based on spectra that were not used in model development. For COD models, RMSEP (Table 6) and RMSECV (Table 4) values are also comparable, showing that the developed models also embraced the external validation spectra.

Performance criteria defined in section 2.4.5 for model evaluation were calculated and were rated according to the guidelines presented in Table 2. Performance criteria were applied to the predicted values mentioned in tables 5 and 6. When these were not available (Models A and B for site 2*), performance criteria were applied to estimated values from the cross-validation procedure. Results for Models A, B and C are presented in Figures 3 and 4.

{Insert Figure 3}

{Insert Figure 4}

The resulting PLS calibration models for TSS and COD estimation for site 2* are presented in Figure 5a and 5b, respectively.

{Insert Figure 5}

Most developed models were rated Good as far as RMSE$_{rel}$ is concerned, both for TSS or COD. All models were Very Good for RSR and NSE. The vast majority of the models were Very Good for PBias.

Considering the information required for each criterion (Equations 1 to 6), RMSE$_{rel}$ and PBias can be considered as measures of accuracy in TSS or COD predictions, stressing the relative error and the overall tendency of over or underestimation. RSR and NSE highlight the dispersion of predicted values
related to the dispersion of measured values. In this context, it can be stated that all the models provided accurate results and that the large majority provided estimates with an acceptable dispersion. RER evaluates whether RMSE is acceptable with regard to the data range. Some models were poorly evaluated for RER. This statistic was specially downgraded in Models A, where all the other statistics rated either Good or Very Good. Models A were only rated Good in RMSE$_{rel}$, which might mean that although RMSE was considered Good for wastewater monitoring, for more robust models a larger range of data should be gathered. On the other hand, this might be a sign that, for monitoring wastewater sewers, performance ratings for RER should probably be reviewed, given that those presented in Table 2 came from pharmaceutical processes.

Most Model A results were Very Good, which supports the fact that UV-Vis spectra is a robust estimator of TSS and COD in wastewater sewers, even when acquired off-line. Models in site 2*, which were related to the same samples, allowed the three strategies to be compared. It was evident that the performance of Models A or B was clearly higher than the performance of Model C, even if Model C relied on more spectra, due to global calibration. When spectra from in situ measurement were subjected to Model B, performance was better than when the same data were subjected to Model C.

In some cases, Model A proved to be superior to Model B; in others, the opposite occurred. This evidences that in-line spectra are an alternative for off-line spectral acquisition, in wastewater sewers, besides all the logistic constraints associated with in-line monitoring in such a harsh environment.

Models B and C were based on spectra acquired in situ, and in-line estimates were compared to laboratory measurements on collected samples. These models may be specially affected by sample representativeness because spectra were determined inside the flow and TSS and COD were determined from a collected sample. In Model A, both spectra and analytical parameters were determined in the collected sample, so sample representativeness was a smaller concern.
IPLS was carried out in Models A and B, so spectral information more relevant to TSS and COD results was analyzed and negligible spectral regions were rejected, which apparently overcame some difficulties present in Model C. This showed the benefit of performing IPLS rather than PLS, which had already been noticed for Models A and B alone, as mentioned in section 3.4.

Comparing the performance of Models A and B with the one achieved by Model C stresses the relevance of developing a PLS calibration model instead of using the calibration model provided by the sensor. The results and the improvement shown by the regression model are clearly illustrated in Figures 3 and 4. Model C relies on a global calibration where spectra came from several sites, which might not be representative of spectra in the site under study.

4 CONCLUSIONS

The use of in-line sensors based on UV-Vis spectra and PLS calibration models for real-time estimation of TSS and COD in wastewater is regarded as a step forward in quality control. However, its application in drainage systems has scarcely been reported and can be very complex, especially where there are stormwater contributions with consequent rapid variations in both hydraulic conditions and wastewater quality.

In the current study, PLS calibration models for TSS and COD prediction in a drainage system were developed using UV-Vis spectra acquired both off-line, with a bench spectrophotometer, and in-line, with a submersible spectrophotometer. The developed PLS models were compared to the calibration model included in the submersible spectrophotometer.

Data adequacy was previously tested in order to evaluate whether PLS regression was suitable for the available water quality data. Results showed that although PLS premises were not entirely met, there was no relevant information left out in the residuals and there was no evident need to evolve to more complex algorithms.
In addition to typical correlation indicators, such as $R^2$ and RMSEP, several efficiency criteria, including statistical parameters generally applied to watershed models, were tested to evaluate the performance of different PLS models based on UV-Vis spectra. Most statistics rated either Very Good or Good, showing that PLS models based on UV-Vis spectra, either acquired in-line or off-line, were robust estimators of TSS and COD, even in a wastewater sewer. Estimates provided by the submersible UV-Vis probe after local calibration delivered accurate TSS and COD results. However, the PLS models developed specifically for the studied drainage system performed better than the probe model. Furthermore, using in-line UV-Vis spectra for PLS model development was essentially equivalent to using off-line spectra, but with the advantage of providing real-time information.

A final remark has to be made on the statistics for performance rating. RMSErel and PBias provided a general overview of PLS model performance and PBias can be discarded from model evaluation unless one suspects of significant model over or underestimation.

The other criteria gave clues on which aspects could be improved in the models. RSR and NSE are quite equivalent, but performance ratings do not always match. Rating these criteria allowed putting results into context, but apparently RER application to wastewater monitoring in sewers could benefit from further research. Thus, RMSE_{rel} and both RSR and NSE can be recommended for PLS model evaluation.

In view of the obtained results, using PLS calibration models based on \textit{in situ} UV-Vis spectra for TSS and COD estimation in sewers is a useful approach for utilities, as it may provide real-time wastewater quality data with considerably low prediction errors. However, the establishment of adequate criteria for PLS performance evaluation should not be disregarded.

\textbf{Acknowledgements.}
R. Salgado Brito and N.D. Lourenço acknowledge the financial support from the Fundação para a Ciência e Tecnologia (FCT, Portugal) through a doctoral (SFRH/BD/46665/2008) and a post-doctoral (SFRH/BPD/31497/2006) research grant, respectively. The authors acknowledge the financial support from FCT, Portugal, for Project SIMAI - Monitoring and warning systems in urban drainage sewer infrastructures (PTDC/AAC-AMB/102634/2008). The authors would like to thank SIMTEJO, S.A. for operational support.

References


**FIGURE CAPTIONS**

Figure 1 – Box-and-whiskers plot of TSS and COD analyses.
Figure 2 – Example of UV-Vis laboratory spectra determined under wet and dry weather conditions.

Figure 3 – RMSErel, RSR, NSE, PBias and RER for TSS (Models A, B and C).

Figure 4 – RMSErel, RSR, NSE, PBias and RER for COD (Models A, B and C).

Figure 5 – TSS and COD measured and estimated values for site 2* (Models A, B and C).

**TABLE CAPTIONS**

Table 1 – Criteria for model evaluation.

Table 2 – Performance ratings for recommended criteria.

Table 3 – PLS model development details for TSS (see text for parameter definitions).

Table 4 – PLS model development details for COD (see text for parameter definitions).

Table 5 – PLS model predictions for TSS (see text for parameter definitions).

Table 6 – PLS model predictions for COD (see text for parameter definitions).
a) TSS

- Site 1, dry (n=67)
- Site 1, wet (n=26)
- Site 2, dry (n=78)
- Site 2, wet (n=61)

b) COD

- Site 1, dry (n=67)
- Site 1, wet (n=26)
- Site 2, dry (n=78)
- Site 2, wet (n=61)
a) dry weather UV-Vis spectra

b) wet weather UV-Vis spectra (dilution factor of 10)
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RMSE = \sqrt{\frac{\sum (O_i - P_i)^2}{n}}$</td>
<td>(Equation 1)</td>
</tr>
<tr>
<td>$RMSE_{rel} = \frac{RMSE}{\bar{O}} \times 100$</td>
<td>(Equation 2)</td>
</tr>
<tr>
<td>$RSE = \frac{RMSE}{S_O}$</td>
<td>(Equation 3)</td>
</tr>
<tr>
<td>$NSE = 1 - \frac{\sum (O_i - P_i)^2}{\sum (O_i - \bar{O})^2}$</td>
<td>(Equation 4)</td>
</tr>
<tr>
<td>$PBIas = \frac{\sum (O_i - P_i)}{\sum (O_i)}$</td>
<td>(Equation 5)</td>
</tr>
<tr>
<td>$RER = \frac{O_{\text{max}} - O_{\text{min}}}{RMSE}$</td>
<td>(Equation 6)</td>
</tr>
</tbody>
</table>

Where:
- $n$: number of samples in the set
- $O_i$: observed value for sample $i$
- $P_i$: predicted value for sample $i$
- $\bar{O}$: average observed values for samples in the set
- $S_O$: standard deviation of the observed values
- $O_{\text{max}}$: maximum observed value
- $O_{\text{min}}$: minimum observed value
| Performance rating | RMSE<sub>rel</sub><sup>a</sup> | RSR<sup>b</sup> | NSE<sup>b</sup> | |Bias|<sup>b</sup> | RER<sup>c</sup> |
|-------------------|----------------|---------|-----------|-----------------|---|---|
| Very good         | < 5%           | 0.0 – 0.5 | 75 – 100 % | < 10%            | - | |
| Good              | 5 – 10 %       | 0.5 – 0.6 | 65 – 75 %  | 10 – 15%         | > 10 | |
| Satisfactory      | 10 – 20 %      | 0.6 – 0.7 | 50 – 65 %  | 15 – 25%         | - | |
| Unsatisfactory    | ≥ 20%          | ≥ 0.7    | ≤ 50 %     | ≥ 25%            | < 10 | |

<sup>a</sup>: proposed by the authors
<sup>b</sup>: adapted from Moriasi, et al, 2007
<sup>c</sup>: proposed by Sarraguça, et al, 2011
<table>
<thead>
<tr>
<th>Models \ Sites</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 2*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n out LV λ (nm) RMSECV (RMSE_rel)</td>
<td>n out LV λ (nm) RMSECV (RMSE_rel)</td>
<td>n out LV λ (nm) RMSECV (RMSE_rel)</td>
</tr>
<tr>
<td><strong>Dry weather</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>44 5 6 271-320; 470-520</td>
<td>31.8 (11.2%)</td>
<td>52 3 3 221-250; 281-340; 371-400; 521-550</td>
</tr>
<tr>
<td>Wet weather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>18 2 4 221-260; 461-500</td>
<td>18.1 (8.6%)</td>
<td>42 3 3 261-380; 421-460</td>
</tr>
<tr>
<td>Model B</td>
<td>-</td>
<td>-</td>
<td>11 0 3 400-447.5; 600-647.5</td>
</tr>
<tr>
<td>Model C</td>
<td>-</td>
<td>-</td>
<td>u.n. 0 - -</td>
</tr>
</tbody>
</table>

- : unavailable; u.n.: unidentified number
<table>
<thead>
<tr>
<th>Models \ Sites</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 2*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n out LV λ (nm) RMSECV (RMSE&lt;sub&gt;rel&lt;/sub&gt;)</td>
<td>n out LV λ (nm) RMSECV (RMSE&lt;sub&gt;rel&lt;/sub&gt;)</td>
<td>n out LV λ (nm) RMSECV (RMSE&lt;sub&gt;rel&lt;/sub&gt;)</td>
</tr>
<tr>
<td><strong>Dry weather</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>44 2 6 221-340; 541-580</td>
<td>76.6 (16.2%)</td>
<td>52 2 9 271-320; 421-470; 571-620</td>
</tr>
<tr>
<td>Model B</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model C</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Wet weather</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>18 2 4 221-260; 381-460</td>
<td>35.3 (9.1%)</td>
<td>42 3 4 221-260; 381-460</td>
</tr>
<tr>
<td>Model B</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model C</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- : unavailable; u.n.: unidentified number
<table>
<thead>
<tr>
<th>Models \ Sites</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 2*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>out</td>
<td>RMSEP (RMSE&lt;sub&gt;rel&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Dry weather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>23</td>
<td>1</td>
<td>27.0 (9.4%)</td>
</tr>
<tr>
<td>Wet weather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>8</td>
<td>0</td>
<td>14.0 (7.5%)</td>
</tr>
<tr>
<td>Model B</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Model C</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

- : unavailable
<table>
<thead>
<tr>
<th>Models \ Sites</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 2*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>out</td>
<td>RMSEP (RMSE&lt;sub&gt;rel&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Dry weather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>23</td>
<td>0</td>
<td>44.5 (8.6%)</td>
</tr>
<tr>
<td>Wet weather</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A</td>
<td>8</td>
<td>0</td>
<td>22.2 (7.1%)</td>
</tr>
<tr>
<td>Model B</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Model C</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

*: unavailable