Talanta 120 (2014) 325–330

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Development of variable pathlength UV–vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring

Baisheng Chen [a], Huanan Wu [b], Sam Fong Yau Li [a,b,c,*]

[a] Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore
[b] NUS Environmental Research Institute (NERI), #02-01, T-Lab Building (TL), 5A Engineering Drive 1, Singapore 117411, Singapore
[c] Shenzhen Engineering Laboratory for Eco-efficient Polysilicate Materials, School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055, PR China

ARTICLE INFO

Article history:
Received 15 October 2013
Received in revised form
10 December 2013
Accepted 11 December 2013
Available online 17 December 2013

Keywords:
Data fusion
Slope-derived spectroscopy
Partial-least-squares regression (PLSR)
Chemical oxygen demand (COD)
Wastewater quality monitoring

ABSTRACT

To overcome the challenging task to select an appropriate pathlength for wastewater chemical oxygen demand (COD) monitoring with high accuracy by UV–vis spectroscopy in wastewater treatment process, a variable pathlength approach combined with partial-least squares regression (PLSR) was developed in this study. Two new strategies were proposed to extract relevant information of UV–vis spectral data from variable pathlength measurements. The first strategy was by data fusion with two data fusion levels: low-level data fusion (LLDF) and mid-level data fusion (MLDF). Predictive accuracy was found to improve, indicated by the lower root-mean-square errors of prediction (RMSEP) compared with those obtained for single pathlength measurements. Both fusion levels were found to deliver very robust PLSR models with residual predictive deviations (RPD) greater than 3 (i.e. 3.22 and 3.29, respectively). The second strategy involved calculating the slopes of absorbance against pathlength at each wavelength to generate slope-derived spectra. Without the requirement to select the optimal pathlength, the predictive accuracy (RMSEP) was improved by 20–43% as compared to single pathlength spectroscopy. Comparing to nine-factor models from fusion strategy, the PLSR model from slope-derived spectroscopy was found to be more parsimonious with only five factors and more robust with residual predictive deviation (RPD) of 3.72. It also offered excellent correlation of predicted and measured COD values with R^2 of 0.936. In sum, variable pathlength spectroscopy with the two proposed data analysis strategies proved to be successful in enhancing prediction performance of COD in wastewater and showed high potential to be applied in on-line water quality monitoring.

1. Introduction

Water quality monitoring is of growing importance all over the world with increasing demand of the real-time water quality information. To monitor wastewater quality during different treatment processes is crucial to ensure the treatment efficiency and comply with increasingly stringent regulations [1–3].

Chemical oxygen demand (COD) is a common indicator of organic matter concentration to assess wastewater quality. The standard COD test involves adding toxic chemicals and is time-consuming, requiring 2–4 h until the result is obtained. With the development of optical techniques, spectroscopic analysis, including UV–vis spectroscopy [4–7], fluorescence spectroscopy [8–10] and near-infrared spectroscopy [6,11], shows high potential in wastewater COD monitoring. They are fast, non-destructive and environment-friendly that requires no chemicals added. Among these spectroscopic techniques, UV–vis spectroscopy shows the most extensive application and exhibits the best correlation with COD by multivariate data analysis. In particular, partial least squares (PLS) regression is routinely adopted to generate regression model based on UV–vis spectral data to estimate the water quality parameters [12,13].

Conventionally, an optimal pathlength is required to be selected and fixed for UV–vis spectroscopic measurements. According to Beer–Lambert law, absorbance is proportional to pathlength. It is a crucial parameter for measurements as it defines the distance that light travels through a sample. However, the selection of the optimized pathlength is a difficult task because it varies from a few millimeters for wastewater influent to dozens of millimeters for wastewater effluent [4]. Furthermore, different treatment processes usually result in different matrixes and concentrations of wastewaters, thus

* Corresponding author at: Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore. Tel.: +65 65162681; fax: +65 67791691.
E-mail address: chmlifys@nus.edu.sg (S.F.Y. Li).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.12.026

requiring different pathlengths for different treatment processes. Even for the same treatment process, the wastewater concentrations vary significantly which would consequently compromise the performance of the monitoring technique. Therefore, UV–vis spectroscopic system with variable pathlength would be a more desirable solution.

A large quantity of spectral data, in this sense, would be generated thus requiring a suitable strategy to manage and analyze these data. Data fusion has become a popular method to deal with abundant amount of data in analytical chemistry recently [7,14–16]. It is the process to integrate multiple data from different sources with the goal of obtaining information of greater quality. However, simply fusing data does not generally deliver better results [17]. As far as UV–vis spectroscopy is concerned, data saturation occurs when too long a pathlength is implemented, whereas too short a pathlength would not be able to generate a sufficiently strong signal. Furthermore, not all the combinations of different pathlengths will yield better results than individual pathlength since data redundancy and poor information from some pathlengths might compromise the others. In order to overcome this problem, variable selection was implemented to select informative variables that contribute to multivariate regression. Two data fusion levels, low-level data fusion (LLDF) and mid-level data fusion (MLDF), combined with variable selection are presented and compared for analyzing variable pathlength spectral data.

Another novel strategy is also presented to handle informative features from different pathlengths by applying slope calculation to generate slope-derived spectroscopy. Slope spectroscopy™, developed by C Technologies, Inc. (Bridgewater, NJ, USA), is a spectroscopic technique that is applied in determining protein and antibody concentration with known extinction coefficients of proteins, particularly at high concentrations which traditionally requires a series of dilutions and is prone to preparation error and sample contamination [18,19]. This technique basically selects the peak wavelength for the slope calculation and successfully applies it in selecting the linear range for concentration calculation with a prior knowledge of the extinction coefficient of the analyte. But the slope information has not been reported for full-spectrum application or for regression purpose. Here, we propose a novel application of the slope calculation for each wavelength and incorporate calculated slope data for the corresponding wavelength to generate a slope-derived spectrum. As such, each sample measured at variable pathlength would be represented by a slope-derived spectrum. The slope-derived spectral data would be subsequently input for regression.

To the best of our knowledge, this is the first time to apply variable pathlength UV–vis spectroscopy to monitor wastewater quality. COD is chosen as the parameter to evaluate the application of the proposed variable pathlength spectroscopy. Two strategies, fusion analysis and slope-derived spectroscopy, are employed to analyze variable pathlength spectral data. This study focuses on investigating the prediction performance by the proposed variable pathlength spectroscopy compared with the conventional single pathlength spectroscopy.

2. Experimental

2.1. Sample preparation and UV–vis spectroscopic measurements

Wastewater samples were taken from four different treatment processes (clarification, flotation, activated sludge and effluent) of Singapore Airline Terminal Services (SATS) wastewater treatment plant. They are mostly composed of degraded food from airline catering and dishwashing detergents. All samples were filtered by 2.5 μm Whatman filter paper and kept at 4 °C before analysis. 98 wastewater samples were divided into two sets: 82 samples for calibration set and 16 samples for validation set. The COD range in calibration set was 112 mg L^{-1} to 1872 mg L^{-1}, whereas in validation set, COD values were from 130 mg L^{-1} to 1792 mg L^{-1}.

UV–vis spectra were measured by HACH DR/5000 spectrophotometer with wavelength range from 200 to 650 nm (Δ=2 nm). A home-made sample cell (Fig. 1) was designed and fabricated to adjust pathlength (0.5–18 mm). The adjustment of pathlength, defined by the distance of two quartz windows, was subjected to the movement of a threaded cylindrical tube sealed with O-ring. One revolution of the cylindrical tube equals to 1 mm adjustment. Each wastewater sample was subjected to UV–vis spectroscopic measurements at 19 different pathlengths (0.5 mm and 1–18 mm with 1 mm increment). The pathlength accuracy was checked by comparing with standard cuvettes of 1 mm, 5 mm and 10 mm. The differences of absorbance values were found to be negligible. Absorbance values of different pathlengths were set to zero according to MilliQ deionized water (18.2 MΩ cm at 25 °C) at the corresponding pathlength. Triplicate measurements for each wastewater sample at each pathlength were taken and the average spectrum was generated for further data processing.

COD measurements were conducted according to standard methods using COD digestion reaction and direct reading spectrophotometer (DR/5000, Hach Company, USA).

2.2. Multivariate data analysis

All chemometric data analysis was made with Matlab 2010b Software (The MathWorks, Natick, MA, USA) and PLS Toolbox 7.0 (Eigenvector Research Inc., Wenatchee, WA, USA). Contiguous block with 10 splits was selected as the cross-validation method the relatively large number of samples analyzed in random order in this study [20]. During the cross-validation process, 82 samples were split into 10 blocks and each block would be left out and predicted using the model built by the rest of the data from the other nine blocks. Samples from each block would be tested out once accordingly. Regarding validation, samples from validation set are predicted by the model built with the calibration set.

Spectral data involve continuous information, so interval window instead of individual data is exploited [21]. Interval partial-least squares (iPLS) is suitable for this application. It involves selecting equidistant subintervals on spectra aiming to produce superior PLSR model compared to full spectral model [22]. The selection criterion of intervals is based on those that provide the lowest RMSECV generated by the selected variables. Two expansion methods of iPLS, designated as backward/forward iPLS (BiPLS/FiPLS), have been developed and performed sucessfully to optimize the infromative regions [23,24]. In FiPLS, PLSR models are built using successively improving intervals with respect to RMSECV measure [25]. If only one interval is desired, the algorithm would stop calculating at that point and produce the results. If, however, more than one interval is requsted, additional calculation cycles would be performed until no improvement of RMSECV is achieved.

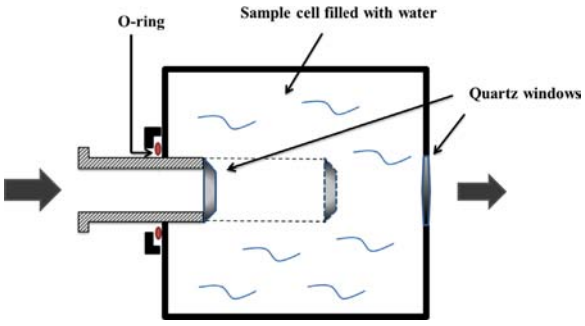

Fig. 1. Design of home-made sample cell with adjustable path length.

BiPLS works in an opposite manner to FiPLS with a different algorighm. For BiPLS algorithm, PLS models are built with each interval left out in a sequence and intervals are eliminated from the one with the worst RMSECV until the pre-determined interval number is achieved. FiPLS was chosen to perform variable selection in this study as it yielded better models (results not shown). The size of interval window from two to 30 intervals was varied and optimized to give the best regression model in terms of the lowest RMSECV obtained.

2.3. Variable pathlength spectral data fusion

To handle variable pathlength spectral data, two data fusion strategies are suitable: (a) data level fusion or low level data fusion (LLDF) and (b) feature level fusion or mid-level data fusion (MLDF) [26], as shown in Fig. 2.

LLDF is the most straightforward which involves concatenating all raw data from variable pathlength measuremnents into a single matrix called a "meta-spectrum" [27]. Variable selection was subsequently performed and finally input for clasification or regression analysis. In this study, PLS regression was applied to generate regression model for prediction purpose.

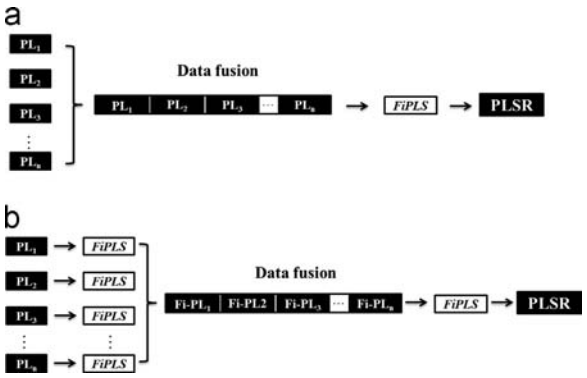

Fig. 2. Two different levels of data fusion of variable pathlength spectral data. (a) LLDF, (b) MLDF. PL: spectral data from each pathlength; Fi-PL: variables selected by FiPLS; PLSR: partial-least-squares regression.

Unlike LLDF, MLDF required a previous variable selection step to extract the most relevant features from different spectra before data concatenating. As a total of 19 different pathlength spectra were contatenated, a relatively large number of variables were still included. Therefore, another round of variable selection by FiPLS before regression was performed to further refine the relevant features and minimize number of variables.

2.4. Generation of slope-derived spectrum

To generate a slope-derived spectrum, a plot of absorbance against pathlength at each wavelength was created, as shown in Fig. 3. Absorbance tends to saturate as pathlength becomes excessively long. The saturation data points (data points in red in Fig. 3) would affect the linearity of the linear regression line by lowering the coefficient of determination R^2. Since these saturation data points would eventually influence the performance of PLS regression model, it was advisable to eliminate them in order to avoid the destruction and so increase the R^2 value. The R^2 value of 0.998 was determined as the lowest acceptable value for slope calculation [19]. Data points were eliminated in sequence from the longest pathlength until the acceptable R^2 value was reached. Noisy data points from short pathlength were also eliminated to suit the acceptable R^2. As such, slope obtained at each wavelength was produced and subsequently combined to generate the slope-derived spectrum.

2.5. Evaluation of model performance

The basic principle of PLS regression is to decompose spectral data into their most common variations, known as factors or latent variables, that represent the variations over the measurement range. The optimal number of PLS factors was selected to provide the minimum value for the predicted residual error sum of squares (PRESS).

The evaluation of the prediction performance of the model was based on several performance indexes, such as correlation of determination (R^2), root-mean-squared error of cross-validation (RMSECV), root-mean-squared error of prediction (RMSEP) and residual predictive deviation (RPD). The equations of these performance indexes are

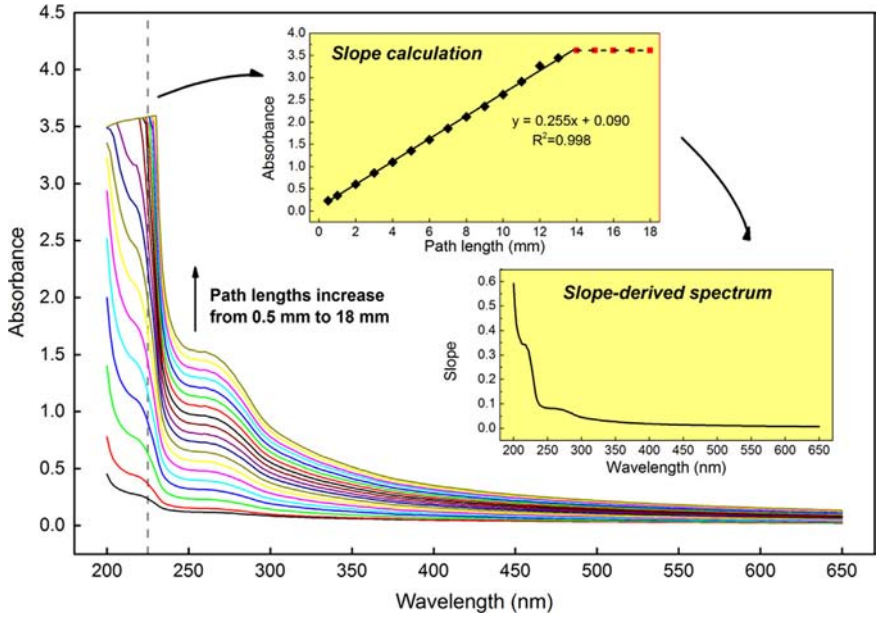

Fig. 3. Typical wastewater (COD 1190 mg/L) UV–vis spectra by variable pathlength measurements and example of the derivation of slope-derived spectroscopy. Inset graphs: the slope calculation plot of absorbance against pathlength at wavelength of 226 nm (top); the plot of slope-derived spectrum (bottom). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

shown as follows:

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

$$\text{RMSE (CV or } P) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (2)$$

$$\text{RPD} = \text{SD}/\text{RMSEP} \quad (3)$$

y_i is the measured (actual) value by reference method; $\hat{y}_i$ is the predicted value by calibration model; y_m is the mean of the actual values; n is the number of samples used in each data set; SD is the standard deviation in each data set.

3. Results and discussion

3.1. Wastewater UV–vis spectral characteristics

Generally, the UV–vis spectral shape was found to be featureless with no evident peaks (see Fig. 3). But two absorption shoulders could be observed at around 220 nm and 275 nm. The absorption at around 275 nm was related to high content of organic substances from degraded food, whereas the absorption at around 220 nm originates mainly from the dishwashing detergents in wastewater [28].

From spectra in Fig. 3, one can also observe that as the pathlength becomes longer, absorption becomes stronger until it goes beyond the detection limit of detector, resulting in signal saturation. Absorption signals of shorter wavelengths tend to be more readily saturated than those of longer wavelengths. This problem occurs when the pathlength for measurement is not chosen appropriately. If, on the other hand, absorption becomes too low, signal resolution would become insufficient, particularly when too short a pathlength is chosen at wavelengths longer than 350 nm (see Fig. 3). In order to simultaneously circumvent the above-mentioned problems, UV–vis spectroscopy with variable pathlength is presented and compared with single pathlength spectroscopy in the subsequent sections.

3.2. Performance of single pathlength spectroscopy

Variable selection was performed to discard irrelevant information and retain quality-relevant variables. Fig. 4 summarizes the variables selected by FiPLS computation. Variables in UV region (200–300 nm) were generally selected for each pathlength, suggesting that this region represented the most important features for organic substances of degraded food and detergents [28].

To evaluate the prediction performance of models built by single pathlength UV–vis spectroscopy, R^2, RMSECV, RMSEP and RPD are compared, as summarized in Table 1. The fitness between predicted COD by PLSR model and measured COD by reference

Table 1
Comparison of PLSR model performances for single path length UV–vis spectroscopy, variable-path length data fusion and slope-derived spectroscopy.

PL (mm)	Variable #	Factor #	Calibration		Validation		
			RMSECV (mg/L)	R^2 CV	RMSEP (mg/L)	R^2 Pred	RPD
0.5	40	6	155	0.808	176	0.912	2.58
1	32	7	153	0.829	170	0.895	2.68
2	50	7	186	0.723	160	0.915	2.85
3	24	6	210	0.647	214	0.814	2.13
4	15	5	152	0.814	159	0.918	2.86
5	24	8	199	0.692	156	0.914	2.91
6	64	7	226	0.592	189	0.843	2.41
7	132	5	247	0.525	172	0.855	2.64
8	30	5	213	0.641	171	0.851	2.67
9	55	5	201	0.678	160	0.868	2.84
10	15	6	153	0.817	158	0.910	2.87
11	92	5	193	0.702	159	0.873	2.87
12	70	5	185	0.727	174	0.851	2.62
13	150	5	187	0.719	168	0.862	2.70
14	48	7	220	0.643	153	0.879	2.97
15	20	8	191	0.713	179	0.865	2.54
16	110	6	172	0.766	175	0.848	2.61
17	90	6	195	0.696	164	0.855	2.78
18	16	6	184	0.732	182	0.833	2.49
LLDF	77	9	150	0.824	142	0.920	3.22
MLDF	72	9	153	0.815	138	0.922	3.29
Slope	36	5	152	0.823	122	0.936	3.72

PL: path length; variable #: number of variables selected by FiPLS; factor #: number of PLS factors; R^2 CV: R^2 of cross validation; R^2 Pred: R^2 of prediction; Slope: slope-derived spectroscopy.

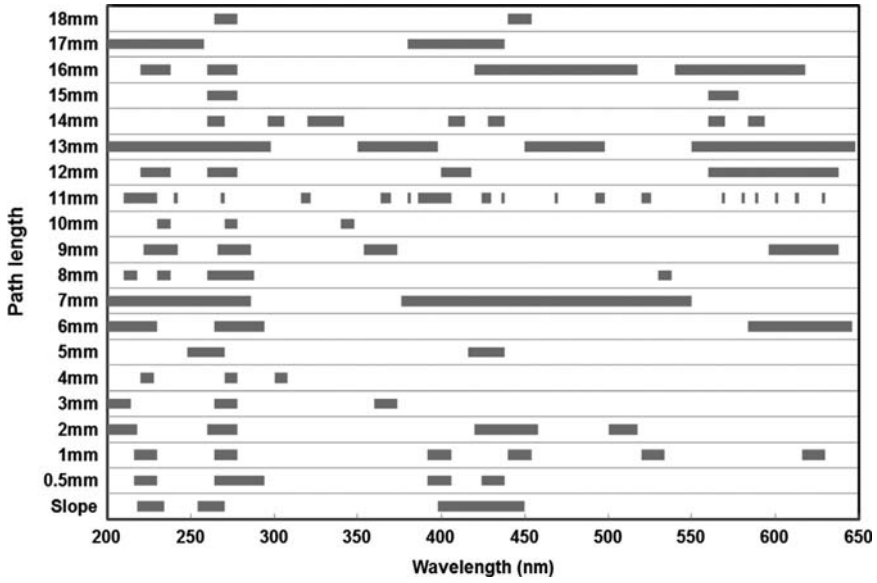

Fig. 4. Variables selected by FiPLS for each single pathlength and the proposed slope-derived spectroscopy. Slope: slope-derived spectroscopy.

method is represented by the correlation of determination R^2 of prediction. Generally, it was found that a moderately good correlation with prediction R^2 of 0.8–0.9 could be obtained by single pathlength UV–vis spectroscopy [4,6,7]. From Table 1, one can observe that analogous correlations were obtained by single pathlength spectroscopy. The best correlation was from pathlength of 4 mm (R^2=0.918). Comparable correlations could also be found from pathlengths of 0.5 mm, 2 mm, 5 mm and 10 mm with R^2 value of 0.912, 0.915, 0.914 and 0.910, respectively. However, if the pathlength was not appropriately chosen, the correlation would drop substantially.

The robustness of a regression model is indicated by RPD. The larger the RPD value, the more robust the regression model is. In contrast, relatively small value of RPD suggests that regression model having trouble and less reliable for the future prediction. RPD < 2 is regarded insufficient for applications, whereas value between 2 and 3 is considered good for approximate quantitative predictions. When RPD is larger than 3, it suggests excellent prediction [29]. From Table 1, models built by each pathlength give comparable results with 3 > RPD > 2, suggesting that the models were good for approximate quantitative predictions. It was found that when comparing the performances of models obtained by different single pathlengths, the RPD values did not vary substantially, which leads to the difficulty to select an optimal pathlength for UV–vis spectroscopic measurement.

3.3. Performance of variable pathlength data fusion

Owing to the limitations of single pathlength spectroscopy, variable pathlength spectroscopy was proposed to investigate whether complementary information and synergistic effect can be obtained. The analysis of variable pathlength spectral data was performed by data fusion, which had been successfully applied to extract complementary features and so as to improve predictive accuracy [14,30]. Two different levels (LLDF and MLDF) of data fusion were adopted to carry out this task.

Although models obtained by several single pathlengths (0.5 mm, 1 mm, 4 mm and 10 mm) gave comparable RMSECV and R^2 of cross validation, when applied to validation samples, variable pathlength data fusion showed its superiority. RMSEP values were found to be 142 mg/L and 138 mg/L for LLDF and MLDF, respectively. The corresponding RPD values were larger than 3, suggesting excellent predictive accuracy.

The FiPLS selected variables for LLDF and MLDF are shown in Fig. 5. Informative spectral regions at around 220 nm and 275 nm were both selected for LLDF and MLDF. Several visible regions were also chosen to cover the absorption of colloidal fraction of organic substances even after 2.2 μm filtration [31]. LLDF demanded variables measured from seven different pathlengths, while MLDF required only five, which might be due to an extra prior variable selection step by FiPLS in MLDF. It is noteworthy that overlapping spectral regions were selected by both fusion levels. By LLDF, spectral region of 212–222 nm was both selected from 1 mm and 9 mm measurements, while replicated selections were also performed in visible region (see Fig. 5a). Similar observations were made for MLDF with overlapping spectral region of 210–230 nm and 394–406 nm (see Fig. 5b), which highlighted the importance of these regions.

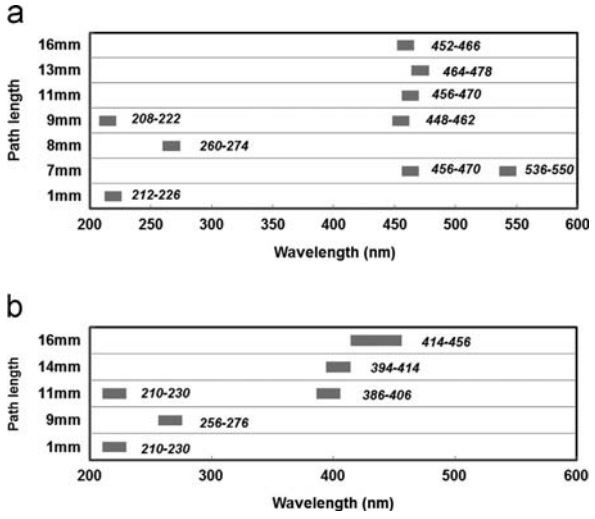

Fig. 5. Variables selected by FiPLS for variable pathlength with LLDF (a) and MLDF (b). The numbers next to the gray bar indicate the wavelength region selected.

Generally, by LLDF and MLDF, visible region was mostly selected from long pathlength measurements, whereas relatively short pathlength spectra contributed more in UV region. This observation implies complementary information can be obtained from variable pathlength measurements. By data fusion, the incorporation of complementary spectral data of variable pathlength measurements delivered more accurate and more robust prediction of COD than conventional single pathlength spectroscopy.

3.4. Performance of slope-derived spectroscopy

From Fig. 4, one can observe that by FiPLS, three wavelength regions (218–234 nm, 254–270 nm and 398–450 nm) were selected for slope-derived spectroscopy. A five-factor PLS regression model was built on these three selected wavelength regions with 36 variables. As shown in Table 1, comparing with single pathlength spectroscopy, the lowest RMSEP of 122 mg/L and the highest prediction R^2 of 0.936 were obtained. The prediction performance was found to be excellent, which was indicated by the largest RPD of 3.72. The improvements of RMSEP by slope-derived spectroscopy ranged from 20% (PL of 14 mm) to 43% (PL of 3 mm).

This superior predictability of slope-derived spectroscopy was mainly attributed to its ability to discard non-relevant features and the synergy to incorporate relevant information from different pathlength spectral data. The slope information from the linear regression of absorbance (A) against pathlength (ℓ) was directly related to concentration (c) and extinction coefficient (ε) from the rearrangement of Beer–Lambert law ($A/\ell=\varepsilon c+$intercept). The linearity of absorbance against pathlength was strictly controlled by the correlation of determination R^2 larger than 0.998. As such, any non-linear features were able to be eliminated. In addition, synergistic effect from slope calculation was achieved by incorporating spectral data from variable pathlength measurements.

It has been demonstrated that highly accurate PLSR models can be obtained by fusing variable pathlength spectral data with both LLDF and MLDF. As compared to slope-derived spectroscopy, however, these two models were relatively complex, demanding up to nine PLS factors, which might not be preferable in practical application. The increased number of PLS factors might be attributed to some non-linear features which were still included in the regression and additional PLS factors were required to handle them. Effectively removing non-linear features, slope-derived spectroscopy produced a more parsimonious model with only five factors and exhibited much better prediction performance with lower RMSEP and higher prediction R^2. Fig. 6 shows the plot of predicted COD by PLS regression model from slope-derived spectroscopy against the measured COD by reference method. It can be observed that excellent fitness (R^2 of 0.936) was obtained with all data from validation set falling along the best-fit 1:1 line. Overall, slope-derived spectroscopy outperformed variable pathlength data fusion with higher predictive power while at the same time produced a much simpler PLSR model to be utilized.

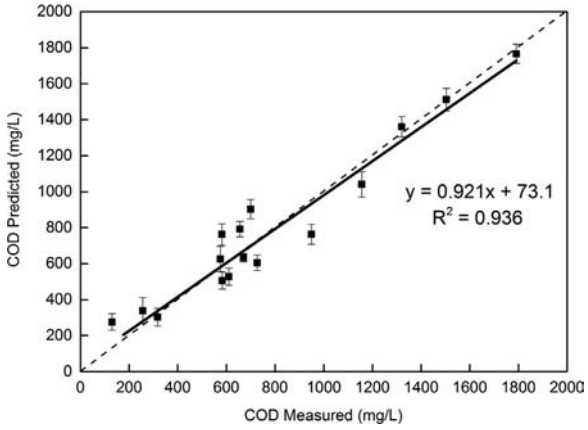

Fig. 6. Plot of predicted COD versus measured COD for 16 validation samples by slope-derived spectroscopy. The diagonal dash line represents the best fit of predicted COD and measured COD. Error bars indicate the standard deviations of triplicate measurements.

4. Conclusion

The application of variable pathlength UV–vis spectroscopy is presented for the first time in wastewater COD monitoring. Two different levels of data fusion were employed to analyze the variable pathlength spectra. Both LLDF and MLDF yielded more robust and more accurate PLS regression models than individual pathlength spectroscopy. Another proposed strategy slope-derived spectroscopy was successful in further enhancing the predictive accuracy (lowest RMSEP of 122 mg/L) while at the same time simplifying PLSR model with only five factors required.

By a simple pathlength adjustment set-up, the requirement of selecting appropriate single pathlength is eliminated. In combination with PLS regression, highly accurate and highly reliable PLRS model can be obtained to determine COD in the calibrated range of 112–1872 mg/L. It is expected that the dynamic range of COD determination would be broader with the variable pathlength approach compared with the single pathlength method. Since the PLS regression is based on full spectra rather than absorbance at individual wavelengths, and COD measures the total amount of organic compounds and hence is highly dependent on water matrix, it was not possible to make accurate comparison between the dynamic ranges for the whole spectra obtained using the variable pathlength approach with absorbance measured for individual pathlengths. Nevertheless, future studies will be conducted to determine the applicability of the variable pathlength approach in a broader COD measuring range than the calibrated range.

The proposed variable pathlength approach is also expected to be able to improve the prediction performance in monitoring other water quality parameters. Furthermore, high potential of applying this technique for on-line monitoring can be foreseen with automated control of the variations of pathlength. Future work will be focused on the investigation of multiple water quality parameter monitoring and the design of variable pathlength on-line sensor for real-time quality monitoring.

Acknowledgments

We acknowledge financial support from the National University of Singapore, National Research Foundation and Economic Development Board (SPORE, COY-15-EWI-RCFSA/N197-1) and Ministry of Education (R-143-000-519-112). We thank the Shenzhen Development and Reform Commission (SZ DRC) for supporting our collaborative project with Peking University Shenzhen Graduate School.

References

[1] A. Lynggaard-Jensen, Talanta 50 (1999) 707–716.
[2] A. Bonastre, R. Ors, J.V. Capella, M.J. Fabra, M. Peris, Trends Anal. Chem. 24 (2005) 128–137.
[3] C.M. Tsoumanis, D.L. Giokas, A.G. Vlessidis, Talanta 82 (2010) 575–581.
[4] G. Langergraber, N. Fleischmann, F. Hofstadter, Water Sci. Technol. 47 (2003) 63–71.
[5] H.S. Jeong, S.H. Lee, H.S. Shin, Environ. Monit. Assess. 133 (2007) 15–24.
[6] M.C. Sarraguca, A. Paulo, M.M. Alves, A.M.A. Dias, J.A. Lopes, E.C. Ferreira, Anal. Bioanal. Chem. 395 (2009) 1159–1166.
[7] X.S. Qin, F.R. Gao, G.H. Chen, Water. Res. 46 (2012) 1133–1144.
[8] D.M. Reynolds, J. Chem. Technol. Biotech. 77 (2002) 965–972.
[9] S. Lee, K.H. Ahn, Water Sci. Technol. 50 (2004) 57–63.
[10] J.N. Louvet, B. Homeky, M. Casellas, M.N. Pons, C. Dagot, Chemosphere 91 (2013) 648–655.
[11] A.C. Sousa, M.M.L.M. Lucio, O.F. Bezerra, G.P.S. Marcone, A.F.C. Pereira, E.O. Dantas, W.D. Fragoso, M.C.U. Araujo, R.K.H. Galvao, Anal. Chim. Acta 588 (2007) 231–236.
[12] J. Dahlen, S. Karlsson, M. Backstrom, J. Hagberg, H. Pettersson, Chemosphere 40 (2000) 71–77.
[13] N.D. Lourenco, J.C. Menezes, H.M. Pinheiro, D. Diniz, Environ. Technol. 29 (2008) 891–898.
[14] R.V. Haware, P.R. Wright, K.R. Morris, M.L. Hamad, J. Pharm. Biomed. 56 (2011) 944–949.
[15] C. Pizarro, S. Rodríguez–Tecedor, N. Pérez-del-Notario, I. Esteban-Díez, J.M. González-Sáiz, Food Chem. 138 (2013) 915–922.
[16] P.M. Ramos, I. Ruisanchez, Anal. Chim. Acta 558 (2006) 274–282.
[17] M. Forina, S. Lanteri, M. Casale, J. Chromatogr. A 1158 (2007) 61–93.
[18] S.V. Thakkar, K.M. Allegre, S.B. Joshi, D.B. Volkin, C.R. Middaugh, J. Pharm. Sci. 101 (2012) 3051–3061.
[19] B.D. Connolly, C. Petry, S. Yadav, B. Demeule, N. Ciaccio, J.M.R. Moore, S.J. Shire, Y.R. Gokarn, Biophys. J. 103 (2012) 69–78.
[20] E. Research, Eigenvector Research, Inc., 2009.
[21] C.M. Andersen, R. Bro, J. Chemom. 24 (2010) 728–737.
[22] L. Norgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Appl. Spectrosc. 54 (2000) 413–419.
[23] R. Leardi, L. Nørgaard, J. Chemom. 18 (2004) 486–497.
[24] X.B. Zou, J.W. Zhao, X.Y. Huang, Y.X. Li, Chemometr Intell. Lab. Syst. 87 (2007) 43–51.
[25] X.B. Zou, J.W. Zhao, M.J.W. Povey, M. Holmes, H.P. Mao, Anal. Chim. Acta 667 (2010) 14–32.
[26] P.M. Ramos, I. Ruisanchez, K.S. Andrikopoulos, Talanta 75 (2008) 926–936.
[27] C.V. Di Anibal, M.P. Callao, I. Ruisanchez, Talanta 84 (2011) 829–833.
[28] O. Thomas, F. Theraulaz, in: O. Thomas, C. Burgess (Eds.), Techniques and Instrumentation in Analytical Chemistry, Elsevier, The Netherlands, 2007, pp. 89–114. (Chapter 4).
[29] R. Zornoza, C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui, J. Mataix-Beneyto, Soil Biol. Biochem. 40 (2008) 1923–1930.
[30] C. Jiang, Y. Liu, H. Qu, Anal.l Methods 5 (2013) 4467–4475.
[31] N. Azema, M.F. Pouet, C. Berho, O. Thomas, Colloids Surf. A: Physicochem. Eng. Asp. 204 (2002) 131–140.