Article

Monitoring Urban Wastewaters' Characteristics by Visible and Short Wave Near-Infrared Spectroscopy

Ignacio Melendez-Pastor *, M. Belén Almendro-Candel, Jose Navarro-Pedreño, Ignacio Gómez, Mariano García Lillo and Encarni I. Hernández

Department of Agrochemistry and Environment, University Miguel Hernández, Av Universidad s/n Edificio Alcudia, Elche 03202, Spain; E-Mails: mb.almendro@umh.es (M.B.A.-C.); jonavar@umh.es (J.N.-P.); ignacio.gomez@umh.es (I.G.); mariano.g@umh.es (M.G.L.); ehernandez@umh.es (E.I.H.)

* Author to whom correspondence should be addressed; E-Mail: imelendez@umh.es; Tel.: +34-96665-8736; Fax: +34-96665-8340.

Abstract: On-line monitoring of wastewater parameters is a major scientific and technical challenge because of the great variability of wastewater characteristics and the extreme physical-chemical conditions that endure the sensors. Wastewater treatment plant managers require fast and reliable information about the input sewage and the operation of the different treatment stages. There is a great need for the development of sensors for the continuous monitoring of wastewater parameters. In this sense, several optical systems have been evaluated. This article presents an experimental laboratory-based approach to quantify commonly employed urban wastewater parameters, namely biochemical oxygen demand in five days (BOD_5), chemical oxygen demand (COD), total suspended solids (TSS), and the ratio BOD_5:COD, with a visible and short wave near infrared (V/SW-NIR) spectrometer (400–1000 nm). Partial least square regression (PLSR) models were developed in order to quantify the wastewater parameters with the recorded spectra. PLSR models were developed for the full spectral range and also for the visible and near infrared spectral ranges separately. Good PLSR models were obtained with the visible spectral range for BOD_5 (RER = 9.64), COD (RER = 10.88), and with the full spectral range for the TSS (RER = 9.67). The results of this study show that V/SW-NIR spectroscopy is a suitable technique for on-line monitoring of wastewater parameters.

1. Introduction

In recent decades, a large number of wastewater treatment plants (WWTP) has been constructed in response to environmental demands by the societies and the legislation [1]. Conventional WWTP comprise a series of treatment stages designed to treat large volumes of wastewater to acceptable regulatory standards. Most of the WWTP include a succession of unitary operations, for example, sedimentation tanks, coagulation-flocculation chambers, biological reactors or even final disinfection systems. The complex succession of unitary operations needs to be sharply operated in order to achieve an effective reduction of pollution. In this sense, the most challenging issue is the great variability of wastewater volume and contamination level of the incoming sewage.

The operation of a WWTP needs the continuous monitoring of contamination level of incoming wastewater and the effective pollution reduction along each stage of the treatment process. On-line monitoring of wastewater quality parameters through the treatment process is a major technical challenge because of the spatial and time dependent variability of wastewater characteristics [2]. Some traditional wastewater quality parameters include the Biochemical Oxygen Demand within five days (BOD$_5$), Chemical Oxygen Demand (COD) and Total Suspended Solids (TSS). The quantification of these parameters is time consuming and implies a notable continuum cost. A fast, accurate and cost-effective system to continuously monitor wastewater parameters through the treatment process may allow operators to optimize the hydraulic retention times (and associated electrical energy cost) and the doses of reactants for some of the stages (e.g., coagulant-flocculent).

The employment of optical-electronic systems for the online monitoring of wastewater treatment process is a relatively recent approach [3] but with a promising future [2,4]. Many studies have employed optical-electronic systems to quantify the relationships among several wastewater parameters and several UV-VNIR (ultraviolet—visible and near infrared) spectral features. Thomas *et al.* [5] employed an UV detector to estimate total organic carbon (TOC), TSS, COD and BOD. Great effort has been focused in the application of fluorescence spectroscopy to improve the quantification of BOD$_5$ [6,7]. In this sense, the application of UV spectroscopy for monitoring wastewater parameters has been greatly improved and new statistical methods, such as neural networks [8], are currently being researched. Moreover, the application of VNIR spectroscopy for wastewater quality is an area of current research with great challenges ahead but with a great potential for on-line monitoring of wastewater parameters [9,10].

Spectroscopic analysis of complex media, such as wastewater, makes the task of implementing optical-electronic systems for the online monitoring of wastewater treatment more difficult. Thus, researchers are encouraged to exploit the full potential of spectral information, to analyze the absorbance that describes the chemistry and the scatter that is related with the particles size and distribution in complex dispersive media [11]. The scatter is generally considered as a "parasitic" phenomenon, complicating the spectroscopic analysis of complex dispersive media, but the detected

intensities of scattered light at different wavelengths depend on the number and sizes of colloidal particles, and consequently, on the respective component content [12]. Scattering in the visible and near infrared tends to be lower than for the UV spectral range (*i.e.*, scattering is inversely proportional to the wavelength). The effect of suspended particles in wastewater samples for UV spectroscopy systems without an integrating sphere assembly that take into account for scattered UV light, is the over-prediction of absorbance [13,14] and lower predictions accuracy. Scatter-based spectroscopic analysis of multi-component mixtures in the visible and short wave near infrared (SW-NIR) has been previously studied in several industrial applications, thus providing suitable quantitative information of the studies [11,12,15–17].

This study focused on the use of visible and short wave near infrared (V/SW-NIR) spectroscopy to quantify wastewater quality parameters (*i.e.*, BOD_5, COD, and TSS) in samples from an urban WWTP, collected at different treatment stages (*i.e.*, input sewage, after a physical-chemical treatment, and after a biological treatment).

2. Materials and Methods

Wastewater samples were collected at an urban WWTP serving Alicante in the Southeast of Spain during the spring season. Samples were obtained along four months and at different weekdays in order to obtain a larger variability of sewage inputs. Three different locations within the treatment processes were sampled: (1) at the entrance to the WWTP, where water continuously flowed through a set of sieves with a minimum aperture of around 1 cm; (2) after the primary treatment, consisting of a series of aeration tanks for sand-fat removal and coagulation-flocculation chamber and subsequent decantation; (3) after the biological treatment, consisting of a biological reactor and subsequent decantation. A total of 84 samples were obtained (*i.e.*, 28 samples per treatment stage) and analyzed in this experiment. Water samples were collected and immediately stored under cold conditions (~4 °C) to minimize water degradation.

Water analyses were conducted within hours from the sampling (less than 4 hours). Selected wastewater quality parameters were the biochemical oxygen demand within five days (BOD_5), chemical oxygen demand (COD), and total suspended solids (TSS) that were analyzed with standard laboratory methods. BOD5 was determined after 5-days of incubation at dark constant-temperature (20 ± 1 °C) conditions. Water samples were placed in incubation bottles on agitation racks, and included a manometer to quantify oxygen consumption. Chemical oxygen demand was determined with a closed reflux, colorimetric method, and total suspended solids were determined by gravimetry of the increase of weight by the residue retained on a filter dried to a constant weight at 103–105 °C [18].

These parameters are the existing standard and most widely applied methods for organic load monitoring [2]. Also, the ratio BOD_5:COD was computed as a proxy of wastewater biodegradability. As a reference value, an ideal biodegradability index is a BOD_5:COD ratio close to 1.0 [19]. Four replicates per water sample were done for all wastewater parameters and the average values of the replicates were used for further analyses.

2.1. Spectral Measurements

An ASD Field Spec Hand Held VNIR radiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) was utilized to measure wastewater spectra. This radiometer covers the wavelength range of 325–1075 nm, which approximates with the visible (V) and short wave near infrared (SW-NIR) spectral regions, an accuracy of ±1 nm and a resolution of <3 nm at 700 nm. The radiometer was connected through a fiber optic cable to an Ocean Optics (Leesburg, FL, USA) cuvette holder where water samples were placed. Another fiber optic cable was connected from the cuvette holder to an ASD Fiber Optic Illuminator® as light source. This systems enables the illumination of a 10 mm cuvette from one of its faces and the transmittance record from the opposite cuvette face.

Radiometric measurements were also conducted within hours from the sampling in order to minimize the alteration of the wastewater samples. All spectra were acquired in transmittance mode using distilled water as blank. Five radiometric measurements (with 15 automatic replicate spectra per measurement) were taken for each water sample. The dark current (detector background) and reference spectra were taken immediately before each spectral measurement. The five radiometric measurements were visually inspected and then averaged to obtain a single spectrum per water sample ($n = 84$). Random noise was minimized by applying a Savitzky-Golay algorithm across a moving window of 10 nm with a 3rd order polynomial [20]. Spectra were acquired at room temperature (24 ± 1 °C). All samples were homogenized (15 s) by mechanical mixing, and then immediately placed in the cuvette. Homogenization is a useful sample pretreatment, because it does not remove the scatter but simplifies its effect on spectra, thus, facilitating the data analysis [11].

2.2. Spectroscopic Analyses

Partial Least Squares Regression (PLSR) was the selected statistical technique to relate the wastewater quality parameters with the water spectra. PLSR has been designed to confront the situation that there are many, possibly correlated, predictor variables, and relatively few samples [21]. In this sense, PLSR provides feasible quantitative multivariate modeling methods for chemometrics [22] where highly detailed spectra data (*i.e.*, high spectral resolution or number of bands) are employed to quantitatively predict a limited number of problem samples. In this study, PLSR models were developed for a sensor-noise-free spectral range. Models were developed for the wavelength range 400–1000 nm and also for two smaller spectral ranges of 300 nm each one. These spectral subsets were identified with the visible (400–700 nm) and the near infrared (700–1000 nm) spectral ranges [23]. PLSR models and further statistical analyses were developed with the R statistical programming language [24].

The selection and evaluation process of the PLSR models [21] was based on the following methodological procedure: (1) the original 84 samples dataset was randomly divided into an initial 60 samples dataset for model cross-validation (with 20 samples per treatment stage), and the 24 remaining samples for independent test; (2) a leave-one-out (LOO) cross-validation (CV) procedure was used for the development of PLSR models with a dataset of 60 samples [25]; and (3) selected models were tested with the independent validation dataset in order to assess the predictive capabilities of the selected PLSR models.

Several diagnostic statistics were employed for PLSR models assessment. Cross-validation Pearson correlation coefficient (R^2) was employed as an illustrative diagnostic statistic. The root mean squared error (RMSE) was the fundamental statistical parameter used to guide the number of model components or latent variables (LV) selection. RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z(x_i) - Z^*(x_i))^2} \tag{1}$$

where N is the sample size, $Z(x_i)$ is the observed value at location i and $Z^*(x_i)$ is the predicted value at location i. In addition, the bivariate RMSE is suitable for overall measurements of model performance [26]. The number of optimal components was determined based on the lowest RMSE values for the adjusted CV with the LOO procedure (*i.e.*, RMSECV). In addition, the range error ratio (RER) was used to determine the practical utility of the models [27]. The RER is computed as:

$$RER = \frac{Range}{RMSEP} \tag{2}$$

where the numerator is the range of the dataset and RMSEP in denominator is the model error of prediction. The RER was employed in the prediction test stage in order to compare the practical utility of the spectroscopic approach to predict the different wastewater parameters.

3. Results

Several descriptive statistical parameters of the wastewater parameters were computed (Table 1). A great reduction of the magnitude of wastewater parameters was observed along the wastewater treatment process. Average BOD_5 was reduced from 461.0 mg/L of the raw sewage, to 202.1 mg/L and 17.7 mg/L after the primary and secondary treatments respectively. Similarly, average COD was reduced from 946.1 mg/L of the raw sewage, to 407.4 mg/L and 57.6 mg/L after the primary and secondary treatments respectively. The TSS greatly reduced from the raw sewage (471.8 mg/L) to the primary treatment outlet (131.0 mg/L) and also after the secondary treatment (19.3 mg/L). The ratio BOD_5:COD changed from about 0.5 for the raw sewage (0.483) and primary treatment (0.502), to 0.307 for the secondary treatment effluent. All of these parameters provide an overview of the main characteristics of the wastewaters analyzed.

The characteristic spectra for the three different wastewater treatment stages are shown (Figure 1). The relative transmittance spectra were highly different according to the wastewater treatment stage. The wastewater treatment process promoted a notable increase of the relative transmittance of the samples, especially for the near infrared spectral range. Raw sewage samples were dark brown with low relative transmittance values, ranging from 0.03 at 400 nm to 0.26 at 1000 nm. Primary treatment relative transmittance ranged from 0.05 at 400 nm to 0.49 at 1000 nm. Secondary treatment relative transmittance ranged from 0.66 at 400 nm to 0.93 at 1000 nm. Secondary treatment promoted the most accentuated clarification of the wastewater samples, as observed by the greater relative transmittance increment as compared with the primary treatment samples. No other accentuated characteristic spectral features were observed.

Table 1. Summary statistics of the calibration dataset for Biochemical Oxygen Demand within five days (BOD5), Chemical Oxygen Demand (COD), Total Suspended Solids (TSS), and BOD_5:COD at selected treatment stages.

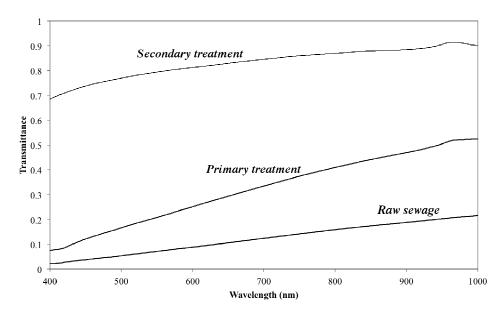| Variables	Statistics	Treatment stage		
		Raw sewage	Primary	Secondary
BOD_5 (mg/L)	Mean	461.0	202.1	17.7
	St. Dev.	147.3	46.9	8.1
	Minimum	280	120	6
	Maximum	920	260	34
COD (mg/L)	Mean	946.1	407.4	57.6
	St. Dev.	242.6	106.2	13.6
	Minimum	615	197	35
	Maximum	1522	572	83
TSS (mg/L)	Mean	471.8	131	19.3
	St. Dev.	154.8	27.9	8.5
	Minimum	290	79	6
	Maximum	902	172	35
BOD_5:COD	Mean	0.483	0.502	0.307
	St. Dev.	0.048	0.041	0.076
	Minimum	0.437	0.451	0.178
	Maximum	0.613	0.609	0.410

PLSR Models Calibration and Validation

The number of latent variables or model components was selected based on the minimization of the RMSECV. The number of components ranged from 2 to 4 for all variables (Table 2). The Pearson coefficient (R^2) was also computed for a general overview of the models performance. Pearson coefficient values ranged from about 0.87 for BOD_5, COD, and TSS, to about 0.6 for the BOD_5:COD ratio. RMSECV values for BOD_5 were about 78 mg/L that is about 8.5% of the range of the BOD_5 calibration values. The absolute values of the RMSECV for COD were slightly higher (~139 mg/L), but were very similar to the previous variable as expressed relatively to the range of the COD calibration values (9%). TSS calibration results were also very similar to the previous variables, with a RMSECV about 79 mg/L (8.8% of the range). Calibration results for the BOD_5:COD ratio reported a RMSECV for the absolute value about 0.06, with an RMSECV value about 14%–15% relative to the range of the variable. The best PLSR were obtained for the visible spectral range (400–700 nm) for all wastewater parameters with the exception of the TSS that performed better for the full spectral range (400–1000 nm).

Models selected from the cross-validation stage were employed with the independent test dataset in order to assess their predictive capability and generalization (Table 2). The magnitude of the RMSE values of the predictions (RMSEP) was similar to the RMSECV. RMSEP for the BOD_5 independent validation dataset was 77.81 (10.37%) and RER was 9.64 for the visible spectral range (400–700 nm). RMSEP for the COD independent validation dataset was 128.40 (9.19%) with a RER of 10.88 for the visible spectral range (400–700 nm). RMSEP for the TSS independent validation dataset was 83.26

(10.34%) with a RER of 9.67 for the full spectral range (400–1000 nm). RMSEP for the BOD$_5$:COD independent validation dataset was 0.059 (17.90%) with a RER of 5.59 for the visible spectral range (400–700 nm).

Figure 1. Characteristic relative transmittance spectra for the raw sewage, primary treatment and secondary treatment wastewater samples.

Table 2. Results of the partial least square regression (PLSR) models cross-validation (CV) and prediction (P) models for the BOD5; COD; TSS and the ratio BOD$_5$:COD wastewater parameters.

Parameters	Spectral range (nm)	LV	Cross-validation			Prediction test		
			R^2	RMSECV	RMSECV (%)	RMSEP	RMSEP (%)	RER
BOD$_5$	400–1000	2	0.849	78.81	8.62		-	
	400–700	3	0.854	77.53	8.48	77.81	10.37	9.64
	700–1000	2	0.849	78.81	8.62		-	
COD	400–1000	2	0.877	139.1	9.35		-	
	400–700	3	0.878	138.5	9.31	128.40	9.19	10.88
	700–1000	2	0.877	139.4	9.37		-	
TSS	400–1000	4	0.864	78.00	8.71	83.26	10.34	9.67
	400–700	3	0.857	80.09	8.94		-	
	700–1000	3	0.861	79.00	8.82		-	
BOD$_5$:COD	400–1000	3	0.643	0.062	14.25		-	
	400–700	3	0.644	0.062	14.25	0.059	17.90	5.59
	700–1000	2	0.599	0.066	15.17		-	

4. Discussion

One major limitation of VNIR spectroscopy is the reduced or absent number of diagnostics bands for chemometric analysis [28]. This implies that single band regression or derivative analyses are really difficult to establish by the absence of accentuated characteristic spectral features. With this

premise in mind, this study applied wide spectral ranges to quantify the wastewater parameters based on the general shape or "area under the curve" of the relative transmittance spectra. The effect of the wastewater parameters variability on spectra can be described as spectrum offset and slope differences (Figure 1). This effect has been previously described in complex dispersive media in the visible and short wave NIR by the multiple scattering by the presence of particles with different composition, size and distribution [11,17]. This method is based on an indirect scatter effect and its practical applicability scope depends on the model robustness in the whole range of natural sample variability [12]. In this sense, previous studies have been able to successfully quantify the composition of wastewater sludge with a transflectance probe for the spectral range 900–1700 nm [29].

PLSR allows a feasible modeling of the different wastewater parameters with VNIR spectroscopy by taking into account the full spectral range of the models instead of single bands. The main advantage of PLSR is that it can be used with any number of explanatory variables, generally providing regression models with highest predictive ability with the smallest number of factors as compared with other regression methods such as ordinary least squares estimator or ridge regression [30]. The PLSR modeling approach allowed the development of a full cross-validation scheme with a moderate number of samples and large number of relative transmittance records (*i.e.*, 600 records for the full spectrum). The number of samples in this study (60 cross-validation + 20 prediction) was in accordance to previous studies such as Thomas *et al* [5] that used 86 samples, Reynolds and Ahmad [6] analyzing up to 54 samples from the same wastewater treatment plant, or even higher than other studies than employed around 40 samples [8,9].

PLSR models selection was based on the lowest RMSECV values (Table 2). Many authors reported their model performance solely based on the Pearson correlation coefficient; however, several computed statistics (*i.e.*, RMSE, RER, and R^2) were applied as a guide of this discussion. Additionally, many studies have been based on the UV spectroscopy instead of VNIR spectroscopy but could be valuable to compare the model performance with different optical-electronic systems. Our best BOD_5 model was determined with three latent variables for the visible spectral range (400–700 nm) and a RMSECV of 8.48% ($R^2 = 0.854$). This correlation coefficient was slightly better than the reported by Thomas *et al.* [5] with an UV system ($R^2 = 0.73$). Reynolds and Ahmad [6] also employed an UV system for BOD_5 spectroscopy and reported a slightly better correlation coefficient ($R^2 \sim 0.94$). They concluded that a linear relationship between the BOD_5 of wastewater and their corresponding fluorescence intensities at 340nm exists using an excitation wavelength of 280 nm.

The best COD model was also obtained with three latent variables for the visible spectral range (400–700 nm) and a RMSECV of 9.31% ($R^2 = 0.878$). Although the RMSECV of the COD was slightly higher than for the BOD_5, the correlation coefficient was better as previously reported by Thomas *et al.* [5] that obtained a R^2 of 0.92 for the chemical oxygen demand instead of the previously noted 0.73 for the biochemical oxygen demand. Fogelman *et al.* [8] predicted COD of wastewater with an artificial neural network for the spectral range 190–350nm. They reported a correlation coefficient $R = 0.96$ (*i.e.*, $R^2 = 0.92$), which is slightly better than our visible spectral range PLSR model. They also noted that the addition of covariates, like the turbidity information, did not significantly improve the accuracy of the artificial neural network. Sarraguça *et al.* [9] monitored COD of an activated sludge reactor with an UV-visible spectrometer with the spectral range 230–700 nm and a NIR spectrometer with the spectral range 900–1700 nm. They also employed PLSR analysis and reported a

low correlation coefficient (0.28) with the NIR system and comparable correlation coefficient results (0.82) for the UV-visible system.

The best TSS model was for the full spectral range (400–1000 nm) with four latent variables for a RMSECV of 8.71% ($R^2 = 0.864$). Thomas *et al.* [5] reported a very similar correlation coefficient ($R^2 = 0.87$) for an UV system. Sarraguça *et al.* [9] also quantified TSS with their dual UV-visible and NIR spectrometers systems. They reported a good correlation coefficient (0.82) for the UV-visible system and even better for the NIR system (0.92). Their results indicates that TSS was better predicted with greater wavelengths, thus explaining our better performance of the full spectral range for TSS quantification instead of only the visible spectral range reported for the other wastewater parameters. The best BOD_5:COD ratio model was for the visible spectral range (400–700 nm) with three latent variables for a RMSECV of 14.25% ($R^2 = 0.644$). Although the performance of the BOD_5:COD ratio models is lower than for the other parameters, the novelty of this wastewater parameter quantification is a valuable topic for further research and for controlling operation techniques.

RER provided information about the practical utility of the PLSR models for the different wastewater parameters in order to determine which parameters were better modeled. As a general guideline, RER values of between 3 and 10 indicate limited to good practical utility, and values above 10 show that the model has a high utility value [27]. Range error ratio value for the COD parameter was higher than 10 (RER = 10.88) that indicates the capability of our visible and short wave near infrared spectroscopic modeling approach for monitoring chemical oxygen demand in wastewaters. Additionally, RER values very close to 10 were obtained for the BOD_5 (RER = 9.64) and the TSS (RER = 9.67). The modeling results for both parameters were also very promising for the online monitoring. A lower RER value (RER = 5.59) was obtained for the BOD_5:COD ratio.

5. Conclusions

This study employed visible and near infrared spectroscopy to quantify wastewater parameters, such as BOD_5, COD, TSS and the BOD_5:COD ratio. Partial least square regression modeling allowed the development of models suitable to quantify wastewater parameters from relative transmittance spectra. The better spectral ranges for wastewater parameters quantification were the visible spectral range (400–700 nm) for BOD_5, COD, and BOD_5:COD ratio, while the TSS were better predicted with the full spectral range (400–1000 nm). The performance of our models was similar to previous studies that used UV fluorescence or UV-VNIR spectrometers. This study provides some valuable information to promote the implementation of VNIR systems for online wastewater monitoring at wastewater treatment plants as a fast and feasible way.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. European Union. Council Directive 91/271/EEC of 21 May 1991 concerning urban waste-water treatment. L135/40–L135/52. Available online: http://eur-lex.europa.eu (accessed on 8 October 2013).

2. Bourgeois, W.; Burgess, J.E.; Stuetz, R.M. On-line monitoring of wastewater quality: A review. *J. Chem. Technol. Biotechnol.* **2001**, *76*, 337–348.

3. Dobbs, R.A.; Wise, R.H.; Dean, R.B. The use of ultra-violet absorbance for monitoring the total organic carbon content of water and wastewater. *Water Res.* **1972**, *6*, 1173–1180.

4. Reynolds, D. Prospects for "star wars" monitoring of water and wastewater quality. *Chem. Bus.* **1999**, *13*, 113.

5. Thomas, O.; Théraulaz, F.; Cerdà, V.; Constant, D.; Quevauviller, P. Wastewater quality monitoring. *Trends Anal. Chem.* **1996**, *16*, 419–424.

6. Reynolds, D.M.; Ahmad, S.R. Rapid and direct determination of wastewater BOD values using a fluorescence technique. *Water Res.* **1997**, *31*, 2012–2018.

7. Ahmad, S.R.; Reynolds, D.M. Monitoring of water quality using fluorescence technique: Prospect of on-line process control. *Water Res.* **1999**, *33*, 2069–2074.

8. Fogelman, S.; Zhao, H.; Blumenstein, M. A rapid analytical method for predicting the oxygen demand of wastewater. *Anal. Bioanal. Chem.* **2006**, *386*, 1773–1779.

9. Sarraguça, M.C.; Paulo, A.; Alves, M.M.; Dias, A.M.A.; Lopes, A.J.; Ferreira, E.C. Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy. *Anal. Bioanal. Chem.* **2009**, *395*, 1159–1166.

10. Ward, A.J.; Hobbs, P.J.; Holliman, P.J.; Jones, D.L. Evaluation of near infrared spectroscopy and software sensor methods for determination of total alkalinity in anaerobic digesters. *Bioresour. Technol.* **2011**, *102*, 4083–4090.

11. Bogomolov, A.; Melenteva, A. Scatter-based quantitative spectroscopic analysis of milk fat and total protein in the region 400–1100 nm in the presence of fat globule size variability. *Chemom. Intell. Lab. Syst.* **2013**, *126*, 129–139.

12. Bogomolov, A.; Dietrich, S.; Boldrini, B.; Kessler, R.W. Quantitative determination of fat and total protein in milk based on visible light scatter. *Food Chem.* **2012**, *134*, 412–418.

13. Mamane, H.; Ducoste, J.J.; Linden, K.G. Effect of particles on ultraviolet light penetration in natural and engineered systems. *Appl. Opt.* **2006**, *45*, 1844–1856.

14. Cantwell, R.; Hofmann, R. Errors in ultraviolet fluence calculations due to particles in wastewater samples. *Environ. Eng. Sci.* **2011**, *28*, 159–162.

15. Wu, D.; He, Y.; Feng, S. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. *Anal. Chim. Acta* **2008**, *610*, 232–242.

16. Tsenkova, R.; Meilina, H.; Kuroki, S.; Burns, D. Near infrared spectroscopy using short wavelengths and leave-one-cow-out cross-validation for quantification of somatic cells in milk. *J. Near Infrared Spectrosc.* **2009**, *17*, 345–351.

17. Kalinin, A.V.; Krasheninnikov, V.N.; Krivtsun, V.M. Short wave near infrared spectrometry of back scattering and transmission of light by milk for multi-component analysis. *J. Near Infrared Spectrosc.* **2013**, *21*, 35–41.

18. APHA-AWWA-WEF. *Standard Methods for the Examination of Water and Wastewater*, 22nd ed.; American Public Health Association (APHA)-American Water Works Association (AWWA)-Water Environment Federation (WEF): Washington, DC, USA, 2012; p. 1496.

19. Kumar, A.; Dhall, P.; Kumar, R. Redefining BOD:COD ratio of pulp mill industrial wastewaters in BOD analysis by formulating a specific microbial seed. *Int. Biodeterior. Biodegrad.* **2010**, *64*, 197–202.

20. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.

21. Mevik, B.-H.; Wehrens, R. The pls package: Principal component and partial least squares regression in R. *J. Stat. Softw.* **2007**, *18*, 1–23.

22. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

23. Curran, P.J. *Principles of Remote Sensing*; Longman: London, UK, 1985; p. 282.

24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013. Available online: http://www.r-project.org (accessed on 8 October 2013).

25. Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley and Sons: Chichester, UK, 1989; p. 419.

26. Willmott, C.J. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* **1982**, *63*, 1309–1313.

27. Williams, P.C. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd ed.; Williams, P.C., Norris, K., Eds.; American Association of Cereal Chemists: Saint Paul, MN, USA, 1987; pp. 143–167.

28. Pons, M.-N.; Le Bonté, S.; Potier, O. Spectral analysis and fingerprinting for biomedia characterisation. *J. Biotechnol.* **2004**, *113*, 211–230.

29. Páscoa, R.; Lopes, J.; Lima, J. *In situ* near infrared monitoring of activated dairy sludge wastewater treatment processes. *J. Near Infrared Spectrosc.* **2008**, *16*, 409–419.

30. Yeniay, O.; Göktas, A. A comparison of partial least squares regression with other prediction methods. *Hacet. J. Math. Stat.* **2002**, *31*, 99–111.