

Document downloaded from:

<http://hdl.handle.net/10251/117575>

This paper must be cited as:

Martínez-Guijarro, MR.; Paches Giner, MAV.; Ferrer, J.; Seco Torrecillas, A. (2018). Model performance of partial least squares in utilizing the visible spectroscopy data for estimation of algal biomass in a photobioreactor. *Environmental Technology & Innovation*. 10:122-131. <https://doi.org/10.1016/j.eti.2018.01.005>



The final publication is available at

<https://doi.org/10.1016/j.eti.2018.01.005>

Copyright Elsevier

Additional Information

Model performance of partial least squares in utilizing the visible spectroscopy data for estimation of algal biomass in a photobioreactor

Remedios Martínez-Guijarro^{1*}, Maria Pachés¹, Jose Ferrer¹, Aurora Seco²

¹ Research Institute of Water and Environmental Engineering (IIAMA). Universitat Politècnica de València. Valencia, Spain. Camino de Vera s/n 46022. Valencia, Spain

² Departament of Chemical Engineering. Escola Tècnica Superior d'Enginyeria. Universitat de València, Avinguda de la Universitat s/n. 46100 Burjassot, Valencia, Spain

*Corresponding author: mmarting@hma.upv.es

Abstract

Spectroscopy technology and statistical methods (Partial Least Squares) have been integrated to develop a model that allows estimating the microalgal biomass in a photobioreactor. The model employing PLS combines the absorption spectrum measurements in the visible range (400-750 nm) with a microalgae cell density in a water sample. First, a calibration model was constructed using a calibration data set, and then, the predictive capacity of the model was determined by cross validation. Finally, an external validation of the predictive performance of the model was carried out with an independent data set. To test the accuracy of the model it was applied to different culture conditions yielding a predictive capacity of 96.7 %. The results achieved are highly satisfactory due to the good lineal adjustment between observed cell densities vs. predicted ones obtained.

According to the results obtained, the application of the model is a useful tool for the management and the decision-making process when operating a photobioreactor. Moreover, this model may boost the real-time measurements and may represent a previous step for further technical development in the “internet of things” applied to the management of the photobioreactor.

Keywords

Microalgae biomass, Absorption spectra, Partial Least Squares, photobioreactor, *Scenedesmus* sp

1. Introduction

The use of microalgae culture for different purposes has been widely applied and nowadays is one of the modern biotechnologies. In natural ecosystems, a high cell density of microalgae beside the presence of specific taxa may indicate eutrophication and may cause toxic microalgae blooms (Smayda 2008). For that, monitoring the microalgae behavior in order to prevent these episodes is of interest. For that purpose, ecosystem optical properties may be analyzed. The optical properties are of interest to detect changes in the composition and structure of microalgae community. In fact, several investigations have reported the use of these optical properties for the study of primary productivity in natural ecosystems (oceans, lakes, rivers, estuaries, etc.) (Bricaud et al., 1998; Ciotti et al., 2002; Millán-Núñez et al., 2004; Barocio-León et al., 2006; Mercado et al., 2006; Barlow et al., 2008; Le et al., 2008; Organelli et al., 2013; Ali et al., 2014). An additional use of these optical properties has been shown to be useful in the detection of harmful algal blooms (Cullen et al., 1997; Millie et al., 1997). Moreover, in the last decades the use of microalgae culture applicate to wastewater treatment and mass production has been widely expanded (Abdel-Raouf et al., 2012). Microalgae use energy from the sun to grow while consuming inorganic nutrients (nitrogen and phosphorus) and CO₂. This ability to remove nutrients has been proposed as an alternative way to treat wastewater. In addition, these organisms can grow in wastewater and simultaneously produce valuable biomass while removing organic carbon (Mallick, 2002; Arbid et al., 2014). According to these studies, several authors have demonstrated that different species of microalgae are capable of attaining high levels of efficiency in the removal of nitrogen and phosphorus. The main species used for this purpose are *Scenedesmus obliquus* (Martínez et al., 2000; Shi et al., 2007; Park et al., 2010), *Spirulina platensis* (Lodi et al., 2003; Olguín et al., 2003) and *Chlorella* sp. (Hernandez et al., 2006; Wang et al., 2010). However, the microalgae culture should be checked and some parameters need to be monitored on a daily basis. One of these parameters is microalgae cell density, since a high nutrient removal rate is related to large amounts of biomass.

For that, microalgae cell density and biomass are to some extent the main parameters to evaluate in both, natural and bioreactor systems. In fact, for the latter system, in order to evaluate cell density and microalgae biomass several methods have been reported. These are based on optical density measurements to wavelengths within the maximal absorbance range of chlorophyll *a* (400-460 nm, 650-680), or dry cell weight measurements (Martínez et al., 2000; Rocha et al., 2003; Chiu et al., 2008; Simionato et al., 2011; Sacristan et al., 2013; Sforza et al., 2014; Tan et al., 2014; Delgadillo et al., 2016). However, using these methods errors can be introduced due to a pigment content in cells varies across the growth cycle and under different culture conditions. In addition, selecting a wavelength within the maximal absorbance range the largest errors are introduced (Griffins et al., 2011). Therefore, an alternative way to obtain quantitative information about microalgae is to analyze sample absorption through the entire visible spectrum combined with the Partial Least Squares multivariate method. These optical methods are based on the study of the whole absorption spectrum generated by two spectral components: material containing pigments (microalgae) and material that does not (detritus). The absorption properties of microalgae are influenced by the composition of intracellular pigments. These photosynthetic pigments, mainly chlorophylls and carotenoids, absorb visible light between wavelengths from 400 nm (violet) to 750 nm (red). Chlorophyll *a* is the main pigment involved in algal photosynthesis. Secondary pigments such as chlorophylls *b* and *c* and various carotenoids extend the range of absorbed light to other zones of the spectrum, thus ensuring optimal absorption efficiencies. Other pigments such as carotenoids work to protect microalgae cells against the effects of high irradiances. (Kirk et al., 1994; Barlow et al., 2008).

The whole absorption spectrum yields large amounts of data that require mathematical and statistical transformation to become useful information. The Partial Least Squares (PLS) is the statistical tool applied to establish statistically significant correlation between the spectral data matrix (X-data) and the response data matrix (Y-data) of the samples set (Brown, 1982; Martens and Naes, 1989). The PLS method yields a model that enables the quantification of cell density

speeding up or even replacing the tedious and time-consuming traditional microscopic cell counts.

The aim of this study is to develop a statistical model based in PLS to track the microalgae cell density of a batch culture in a PBR. The monitoring of cell density in the PBR is carried out by visible spectroscopy technique combined with multivariate statistical methods. The absorption spectrum is first characterized and studied, and then is related to microalgae cell density through a statistical models. This study presents the development of a simple, fast, low cost method based on visible spectroscopy combined with statistical methods for the quantification of microalgae cell density.

2. Materials and Methods

2.1. Experimental start up and sample collection

The photobioreactor (PBR) is a homogeneous batch culture of 457 L of feed and 4.7 L inoculum. The feed introduced into the PBR comes from the effluent of a Submerged Anaerobic Membrane Bioreactor (SAnMBR) pilot plant (Giménez et al., 2011). This effluent exhibits low organic matter content, furthermore, it is solid free due to a membrane filtration process (Ruiz-Martinez et al., 2012), in addition it has high concentrations of ammonium (N-NH_4^+) and orthophosphate (P-PO_4^{-3}). The PBR was fed once at the starting point and no further addition of feed was made during the experiment. The inoculum came from another previously run PBR, which maintained a free-cell microalgae culture. Further details of the characteristics of the PBR may be found in Ruiz Martínez et al., (2012) and Viruela et al., (2016).

Daily samples from the PBR were taken during 25 days period (T0 the first and T25 the last culture day). Samples were collected from the beginning (low cell density) until the steady state (high cell density) with the aim of span a big range of cellular density values. Since samples were measured daily the microalgae concentration was increasing gradually. The data related to

biological and optical properties of these samples were used to generate the model calibration data set.

Another set of samples from four different PBR cultures (R1, R2, R3-dec, R4) were collected and analyzed, to gather the external validation data set. These PBR used to validate the model were maintained under different operational condition (hydraulic and cellular retention times, feed, purge, pH control, light conditions, etc.) to assure the usefulness of the model.

2.2. Analytical techniques

To determine the absorption spectra, a volume of 20 ml of PBR sample (calibration and validation samples) was filtered with a glass-fiber filter (Whatman GF/F). The filter with the retained biomass was placed on a petri slide and stored at -20° C until analysis.

The absorption spectrum for each filter, with the biomass particles retained, was determined using a Perkin Elmer Lambda 35 spectrophotometer, fitted with an integrating sphere (Labsphere RSA-PE-20). The sphere device in the spectrophotometric measures allows determining the fraction of light absorbed, reflected and transmitted by a solid surface (AQ-00073-000, Rev.7, Perkin Elmer). This solid surface is formed by the biomass particles retained in the glass fiber filter. The filter with the solid biomass was placed in the reflectance port position and absorbance option was selected in the software. A pre-filtered sample was passed through glass-fiber filter and this wet filter was used as a blank. The absorbance value was determined at 1-nm intervals, between 400 and 750 nm wavelengths. All experiments were conducted under controlled lighting conditions in order to prevent pigment photodegradation. This experimental procedure was determined according to the procedure outlined by Cleveland and Weidemann (1993).

Regarding to the microalgae abundance determination, the epifluorescence microscopic count method was used. The samples contained in 250 ml glass bottles were fixated with

glutaraldehyde until reaching a final concentration of 2%. Samples were filtered with Millipore GTTP membranes (pore size 0.2 μm). The filters were then dehydrated with successive washes of 50%, 80%, 90% and 99% ethanol. Finally, a cover glass was placed on top of the filter (Fournier, 1978). The counts were performed by epifluorescence microscopy (Vargo, 1978) with a Leica DM 2500, using 100x-oil immersion, the most abundant genera were counted with an error below 20% (Lund et al., 1958).

2.3.Data processes

The data set obtained to perform the statistical model are made up by two matrices. The first one is a predictive data matrix (X-variables, absorbances) with 350 variables/sample (7700 predictors); and the second one is a response data matrix (Y-variables, cell L-1 *Scenedesmus*) with 1 variable/sample (22 responses). Table I shows the raw data matrix (since there is a large number of X variables, only absorbance values at 12 wavelengths are shown).

Wavelength (400-750 nm)	Predictors (absorbances), X-variables															Responses <i>Scenedesmus</i> , Cell L ⁻¹ Y-variables
	400	401	402	...	500	501	502	...	600	601	602	...	747	748	750	
T0	0.2215	0.2210	0.2207	...	0.1780	0.1764	0.1748	...	0.1451	0.1453	0.1454	...	0.1022	0.1018	0.1013	1.26E+08
T1	0.4112	0.4114	0.4118	...	0.3111	0.3082	0.3053	...	0.2364	0.2363	0.2362	...	0.1372	0.1365	0.1358	2.04E+08
T2	0.5445	0.5482	0.5522	...	0.4142	0.4048	0.3955	...	0.2604	0.2615	0.2628	...	0.0570	0.0566	0.0560	8.43E+08
T3	1.2536	1.2593	1.2653	...	1.0622	1.0398	1.0164	...	0.6501	0.6532	0.6566	...	0.0744	0.0738	0.0731	1.82E+09
T4	1.6127	1.6095	1.6062	...	1.4136	1.3932	1.3713	...	0.9533	0.9578	0.9626	...	0.1112	0.1103	0.1094	6.56E+09
T5	1.6733	1.6688	1.6633	...	1.5490	1.5334	1.5167	...	1.0934	1.0983	1.1036	...	0.1219	0.1205	0.1190	6.45E+09
T6	1.4564	1.4574	1.4580	...	1.4570	1.4481	1.4380	...	1.1502	1.1549	1.1599	...	0.1358	0.1343	0.1328	6.45E+09
T7	1.5504	1.5491	1.5481	...	1.5498	1.5437	1.5370	...	1.2878	1.2928	1.2980	...	0.1475	0.1460	0.1444	8.39E+09
T8	1.4593	1.4604	1.4605	...	1.4742	1.4690	1.4632	...	1.2616	1.2660	1.2708	...	0.1534	0.1518	0.1502	8.81E+09
T9	1.4626	1.4637	1.4644	...	1.4667	1.4605	1.4531	...	1.2292	1.2340	1.2390	...	0.1293	0.1276	0.1259	1.04E+10
T10	1.4711	1.4723	1.4730	...	1.4781	1.4712	1.4634	...	1.2196	1.2246	1.2298	...	0.1315	0.1298	0.1282	1.09E+10
T11	1.3900	1.3916	1.3932	...	1.4226	1.4132	1.4022	...	1.0325	1.0369	1.0416	...	0.0950	0.0936	0.0920	9.01E+09
T12	1.4792	1.4805	1.4817	...	1.4841	1.4764	1.4675	...	1.2169	1.2218	1.2267	...	0.1447	0.1429	0.1411	9.97E+09
T14	1.7203	1.7149	1.7080	...	1.6212	1.6084	1.5944	...	1.2254	1.2307	1.2365	...	0.1418	0.1401	0.1386	1.04E+10
T15	1.4191	1.4208	1.4212	...	1.4327	1.4260	1.4182	...	1.1489	1.1537	1.1584	...	0.0837	0.0824	0.0811	1.05E+10
T16	1.4310	1.4322	1.4335	...	1.4277	1.4198	1.4114	...	1.1426	1.1472	1.1520	...	0.1051	0.1037	0.1024	1.09E+10
T17	1.4182	1.4193	1.4204	...	1.4100	1.4018	1.3924	...	1.1241	1.1286	1.1334	...	0.0856	0.0841	0.0827	1.08E+10
T21	1.4084	1.4097	1.4109	...	1.3507	1.3386	1.3255	...	1.0191	1.0236	1.0284	...	0.0938	0.0922	0.0908	1.15E+10
T22	1.4060	1.4078	1.4090	...	1.3279	1.3129	1.2967	...	0.9635	0.9680	0.9727	...	0.0671	0.0656	0.0641	1.01E+10
T23	1.4148	1.4165	1.4180	...	1.3521	1.3380	1.3231	...	0.9994	1.0039	1.0087	...	0.0806	0.0791	0.0777	1.09E+10
T24	1.4033	1.4054	1.4073	...	1.3166	1.3013	1.2851	...	0.9536	0.9581	0.9627	...	0.0676	0.0660	0.0645	1.07E+10
T25	1.3928	1.3946	1.3967	...	1.2855	1.2687	1.2510	...	0.9115	0.9160	0.9206	...	0.0341	0.0326	0.0312	9.99E+09

Table 1. Matrix of raw data (X-variables and Y-variables).

Before applying the statistical techniques to the raw data, a data-processing was applied:

2.3.1. Processing of Spectrophotometric data

Although the absorption spectrum was performed for every single sample with 1 nm wavelength intervals, due to the requirements of the statistical software used, the number of data was reduced from 350 to 320. Therefore, one absorbance was eliminated every 12 nm of wavelength. However, it was verified that equal results were obtained with or without these 30 absorbance data. The absorbance value obtained at 750 nm was added to each spectrum to minimize differences between the samples and reference filters, assuming a negligible absorption at this wavelength. Furthermore, the absorbance data of every sample were normalized, dividing the absorbance values in the visible spectrum by the maximum absorbance value obtained (which is generally produced at wavelengths of 400-440 nm), yielding a whole data set of absorbances between 0 and 1. Figure 1 shows the absorption spectra of three samples (T1, T9 and T25) without normalization (left), and after normalization (right).

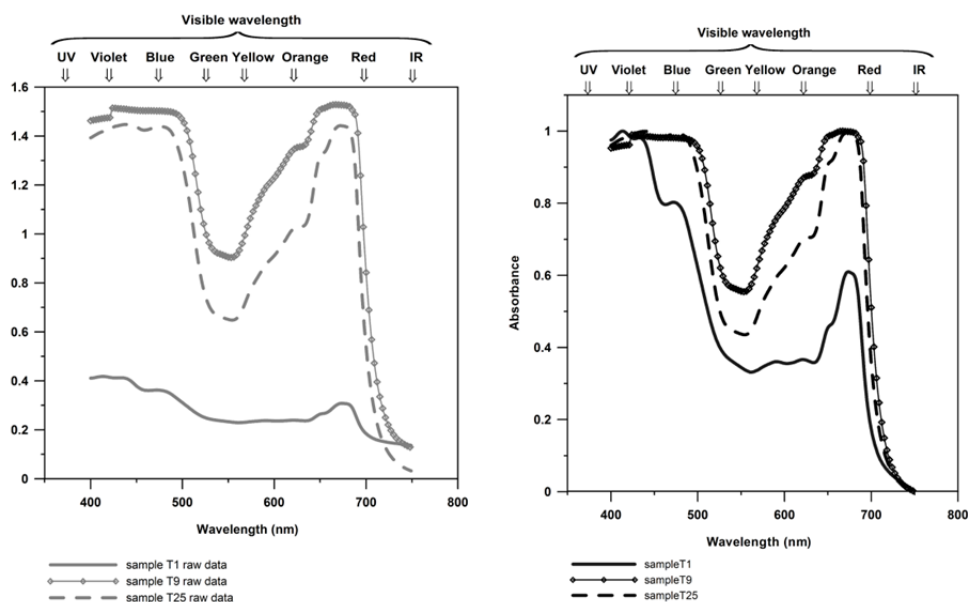


Fig. 1: Absorbance values in the visible spectrum for samples T1, T9 and T25, before normalization (left) and after normalization (right)

2.3.2. Processing of Biological Data (*Scenedesmus* sp (cell L⁻¹))

All cell density values were transformed by Square Root (Sqrt-Y) to achieve approximately normally distributed data, before the multivariate statistical analysis.

2.3.3. Data sets

Two data sets were generated to develop the statistical model with processed data; the first one is the calibration data set, (Calibration Data set) and the second one the further data validation set (Validation Data set).

The Calibration Data set contains 22 samples (T1-T25) represented by two matrices data; predictive and response. The first one is the predictive matrix data that include 320 X-variables (absorbances) by sample, and the second one is the response matrix data that include one Y-variable (square root cellular densities of *Scenedesmus* sp.) by sample (Fig. 2).

The Validation Data set comprises 40 samples belonging to different PBR's. These 40 samples are used to generate the predictive matrix data with 320 X-variables by sample, and the response matrix data with one Y-variable by sample (Fig. 2).

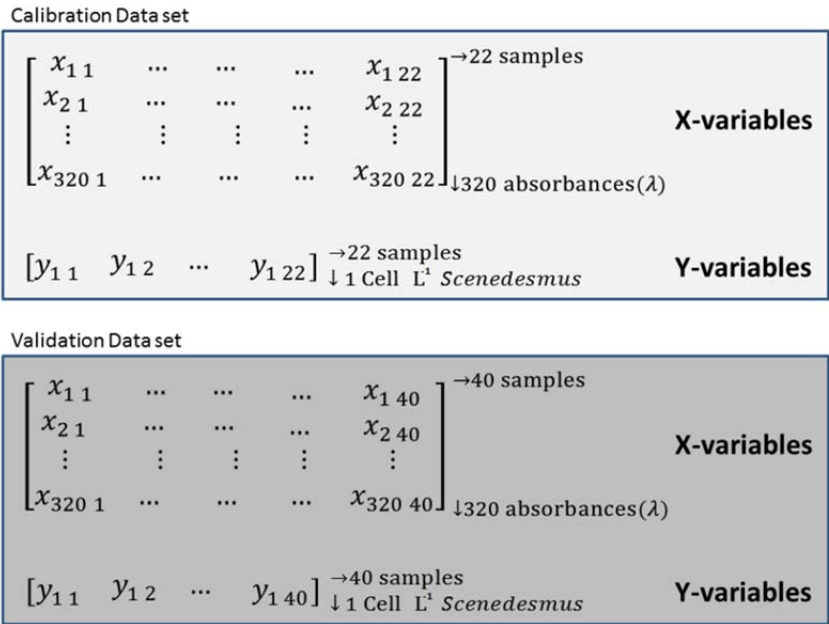


Fig. 2: Calibration and Validation Data set

2.4. Statistical techniques

The software tool used for this statistical process were Simca-P 9.0 (mainly used for Principal Component Analysis method (PCA), and Partial Least Square regression(PLS), and Statgraphics Centurion XVI.

The statistical technique PLS was used for the multivariate analysis of the experimental data obtained (Geladi, 1986). This method was used to discern the quantitative relationship between the spectral data (absorbances) and the biological responses (cell L⁻¹) of the samples of interest (Brown, 1982; Martens and Naes, 1989).

A statistical model was developed for the major algae group (*Scenedesmus* sp.) grown in the PBR. The quality of the model can be estimated by the following model parameter: R² is the cumulative sum of squares of all the Xs and the Ys respectively explained by the extracted components in the PLS models. Q² is the cumulative fraction of the total variation of the variables that can be predicted by a component. The values for both R² and Q² range between 0 and 1, where 1 means a perfect model and 0 an irrelevant model, in a manner similar to the coefficient of determination (r²) in a linear regression analysis.

The predictive capacity of the models (Q²) was determined using internal validation or cross-validation, a method of repeated sampling where multiple subsets are left out of the calculation of the sum of the squared difference between the observed values and estimated values. As this approach, deals with repeated sampling, cross-validation can be carried out in various ways (Geladi, 2002). The cross-validation method selected for the all models consisted of leaving out each observation once from the regression process (“leave one out”) and calculating the model without it. The resulting model was subsequently run to predict this observation. Then, the predicted data are compared with the original data and the Predicted Residual Sum of Squares (PRESS) errors calculated for the whole data set.

For convenience, it then converts PRESS into Q² to resemble the scale of the R². PRESS is divided by the sum of the squared deviation of the y-values (y_i) from their mean (\bar{y}) and

subtracted from 1. Good predictions will have low PRESS and therefore high Q^2 .

$$Q^2 = 1 - \frac{PRESS}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

For the evaluation of the predictive ability of a multivariate calibration model, the Root Mean Square Error of Calibration (RMSEC) and Root Mean Square Error of Prediction (RMSEP) can be used.

3. Results and discussion

3.1. Microalgae culture composition

PBR samples, from which the statistical model is built, exhibit a microalgae composition of mainly *Scenedesmus* sp. (*Chlorophyceae*) and to a lesser extent pennate diatom. Other genera belonging to the *Chlorophyceae* and Dinoflagellates were also detected in some isolated samples. In fact, the latter group appeared in the last days of the PBR experiment. Since the cell density reached by these other groups (pennate diatoms, Dinoflagellates) is low and non-significant (<1%) they were excluded from the statistical model. For this reason, the PBR culture may be considered a *Scenedesmus* microalgae monoculture.

3.2. Statistical PLS modeling

A PLS model that related the absorbances (factors/predictors, X-data) of the sample to the square root of the cellular density (response, Y-data) was developed (Table 2).

To decide the adequate number of components required to diminish the possibility of overfitting the model, the general model fit must be taken into account. This is defined by parameter $R^2Y(\text{cum})$, and the prediction ability thereof, which is defined by parameter Q^2 . High Q^2 and R^2Y values and low PRESS value indicate the validity of the multivariate calibration and the quantitative method. For this reason, the best model appears to be the one using 7 components (Table 2) since it fulfills the requirements.

Comp	R ² X	R ² X (cum)	Eigen.	R ² Y	R ² Y (cum)	Q ²	limit	Q ² (cum)	PRESS
1	0.786	0.786	17.3	0.625	0.625	0.526	0.05	0.526	4.69668E8
2	0.132	0.918	2.9	0.214	0.839	0.484	0.05	0.755	2.08529E8
3	0.0399	0.957	0.879	0.0713	0.91	0.292	0.05	0.827	1.64453E8
4	0.0283	0.986	0.623	0.0638	0.974	0.644	0.05	0.938	7.25648E7
5	0.00765	0.993	0.168	0.0115	0.985	0.31	0.05	0.957	5.4538E7
6	0.00252	0.996	0.0554	0.00494	0.99	0.143	0.05	0.964	5.32291E7
7	0.00317	0.999	0.0698	0.00177	0.992	0.105	0.05	0.967	2.78644E7
8	0.000264	0.999	0.00581	0.00171	0.994	-0.251	0.05	0.964	3.70488E7
9	0.000105	1	0.00232	0.00168	0.995	-0.492	0.05	0.961	4.95326E7
10	0.000124	1	0.00273	0.00094	0.996	-0.228	0.05	0.957	6.95879E7

Table 2. Partial Least Square (PLS) results.

The statistical significance of the model was assessed through the Analysis of Variance (ANOVA). The P-Values in the ANOVA (Table 3) indicate that model 7-components is significant at the 5.0% level of significance. Therefore, there is statistical evidence for the existence of a relationship between absorbance and cellular density of microalgae.

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	1.9379E10	7	2.76842E9	249.351	0.0
Residual	1.55435E8	14	1.11025E7		
Total (corr.)	1.95344E10	21			

Table 3. Analysis of Variance for Y-data (PLS model).

As Table 2 shows, the first three components accounted for most of the variability of the Y-data ($R^2Y(\text{cum})=91\%$) and the capacity of prediction is 82.7%. This parameter increases to 0.664 when adding a fourth component ($Q^2(\text{cum})=93.8\%$) although the variability of the Y-data added 6% and $R^2Y(\text{cum})$ remains similar.

Figure 3 shows the standardized regression coefficients that are obtained by subtracting the mean and dividing by the standard deviation. As it can be seen in Figure 3, for 3 and 4 component models the X-variables that most contribute to the predicted variable (Y-variable) are those placed at the edges of the absorption spectrum (400-500 and 650-750 nm).

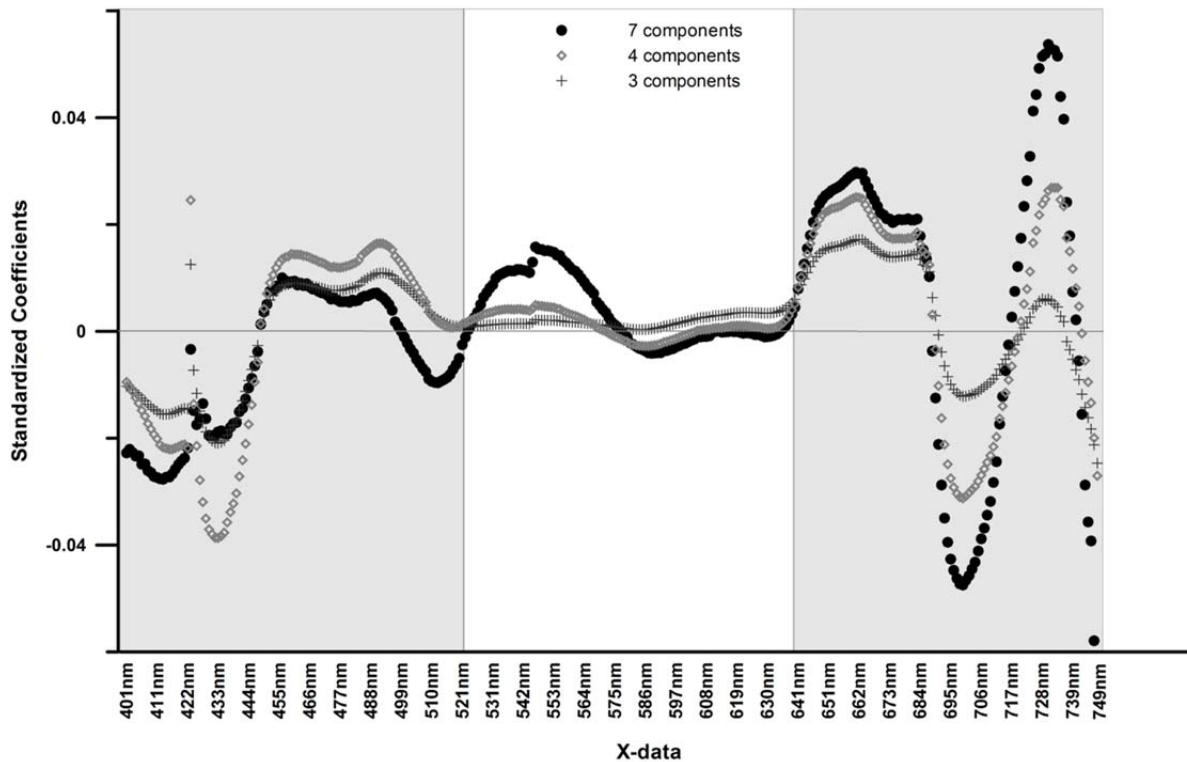


Fig. 3: Standardized coefficients

It is important to highlight that all photoautotrophic organisms contain chlorophyll a as a pigment in the thylakoids membrane. This photosynthetic pigment has its maximum absorption preferentially in these wavelength (Lichtenthaler and Buschmann, 2001). This means that the 3 and 4 component models only provide raw information of biomass value.

However, the 7-components model shows the highest regression coefficients in the middle spectrum zone ($\approx 510\text{-}560\text{ nm}$). This wavelength range has maximum absorption for carotenoid pigments (Lichtenthaler and Buschmann, 2001). These accessory pigments are used to discriminate between different taxa in the microalgae composition.

Therefore, to achieve accurate cell density predictions the model has to include the variability due to pigments, chlorophylls and carotenoids allowing to determined shift in the microalgae community.

The 7-component model is highly recommended for all mentioned before (high R^2Y ; Q^2 values and low PRESS value) and furthermore, this model considers the variability due to carotenoids which may represent the community composition.

The PLS calibration model with 7 components, explained 99.2% ($R^2Y=0.992$) of the variation of Y-data (Sqrt *Scenedesmus*) and its ability to predict is 96.7% ($Q^2=0.967$). To evaluate the efficiency of the prediction-model, predicted Sqrt *Scenedesmus* values were plotted against the observed Sqrt *Scenedesmus* values (Fig. 4a). As can be seen from Figure 4a, satisfactory results are obtained for the predicted values of cell density of the studied microalgae. In addition, outliers in the regression model were detected using standardized residuals (Fig. 4b). For samples in the calibration set, standardized residuals with absolute values in excess of ± 2 should be examined closely. As Figure 4b shows, the residuals are scattered around the horizontal line, there are no outliers in the calibration samples and the constructed model is reliable.

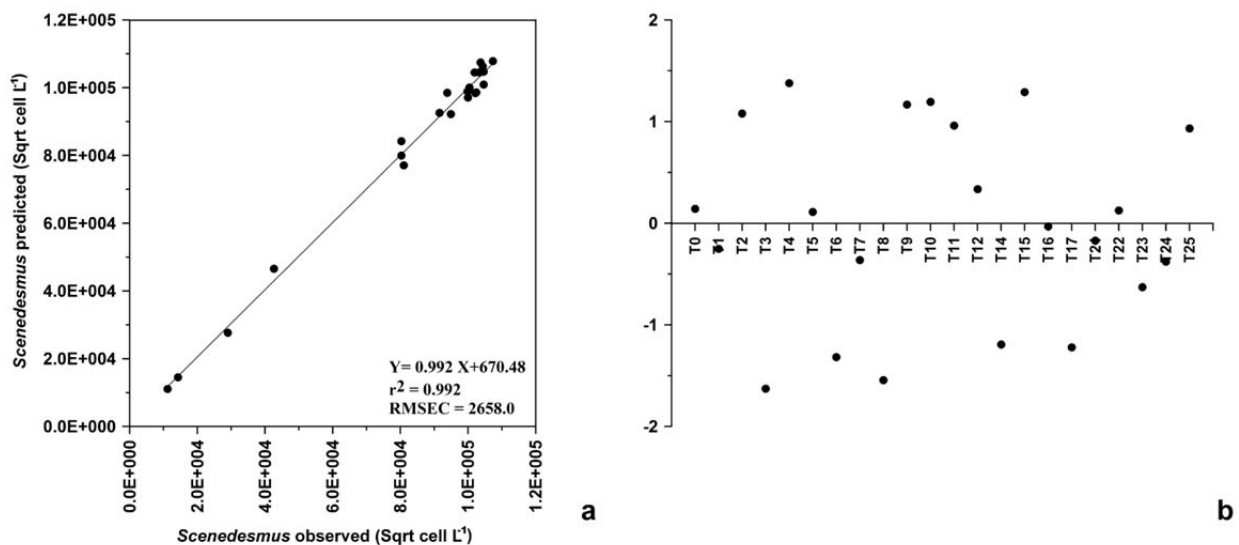


Fig. 4: (a) Square root of cellular density of *Scenedesmus* sp. observed versus predicted (Calibration Data set). (b) Standardized residuals for the Y variables

In order to evaluate the performance of the PLS model, a prediction test was carried out with external test samples (Validation Data set). As it can be observed in Figure 5, there is a high correlation between the measured and predicted Sqrt *Scenedesmus* values ($r^2=0.78$). For that, the

assessment of the PLS model through and independent samples representing different operating conditions is successful in quantifying the microalgae density.

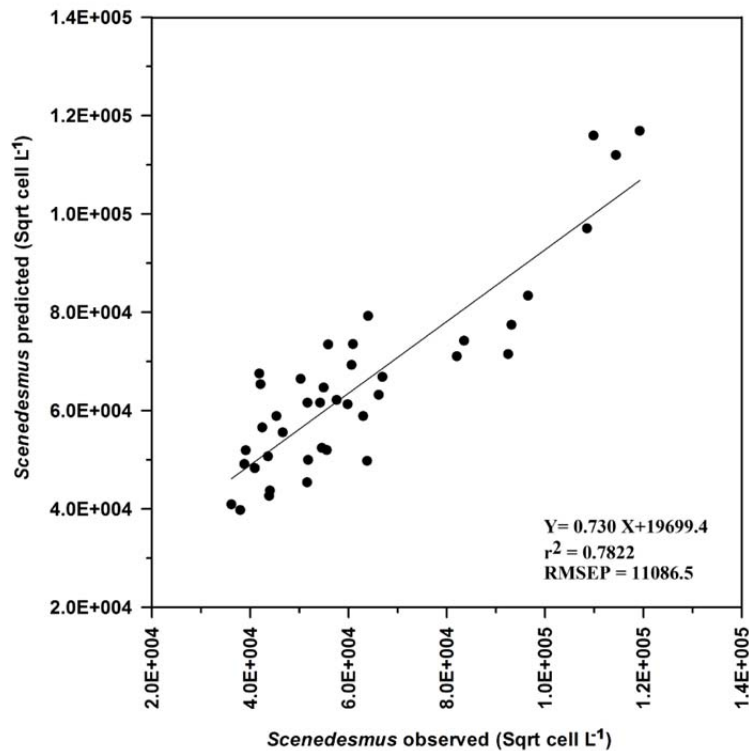


Fig. 5: Validation data set. Square root of cellular density of *Scenedesmus* sp. observed versus predicted

The statistical parameters RMSEC and RMSEP displayed in Figure 4a and 5 permit evaluation of the predictive capacity of the performed model. The difference between RMSEC and RMSEP values might be due to the fact that most of the calibration samples show higher cell density values than those corresponding to the validation set (R1, R2 y R4 PBR's). This suggests that the model could be strengthened by feeding it with similar cell density validation samples.

Therefore, the model will be able to estimate the cell density gradient more accurately from the initial cell concentration up to optimal cell density for wastewater treatment.

Finally, when representing the measured and predicted cell density of *Scenedesmus* for the Validation Data set (Fig. 6) it can be seen that the estimated values are close to the observed ones, especially the samples with low cell density ($\approx 2 \cdot 10^9 - 3 \cdot 10^9$ cell L⁻¹). There has been

verified through statistical parametric methods (One-way ANOVA) that the differences between the predicted and observed values are not statistically significant, (p -values > 0.05).

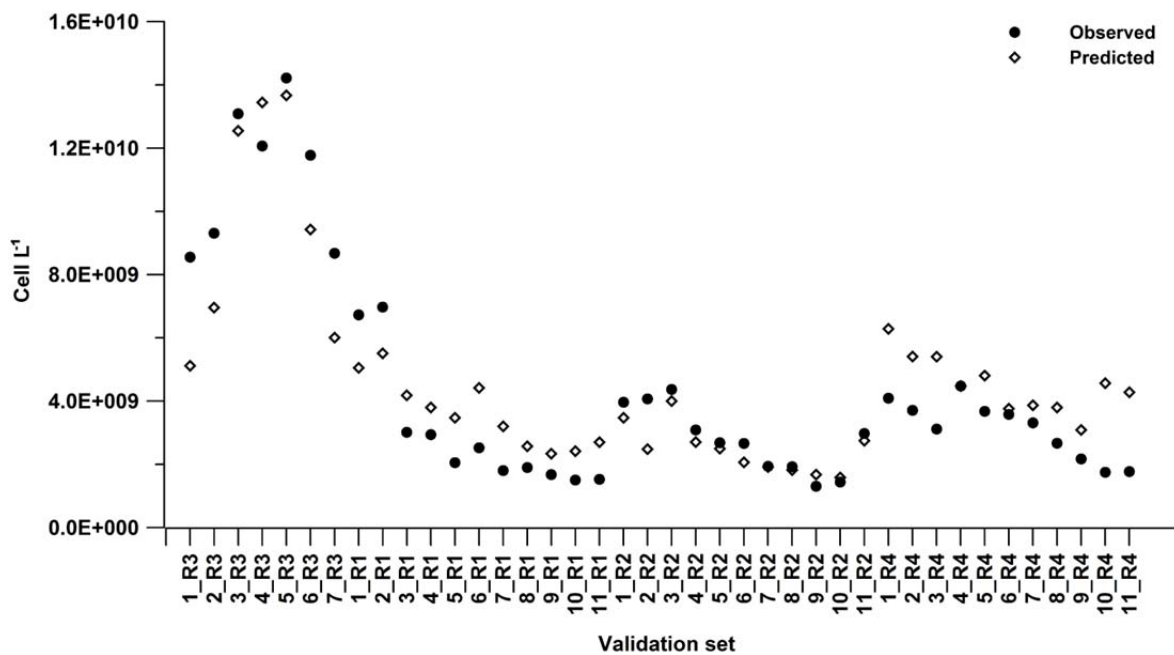


Fig. 6: Cell density of *Scenedesmus* sp observed and predicted in the validation samples

These results confirmed that the constructed PLS model has acceptable predictive abilities for estimating microalgae cell density in the PBR.

This model represents a baseline for further studies to foster the estimation of algal biomass, model prediction and data-driven management of photobioreactor.

Acknowledgements

This research project has been supported by the Spanish Research Foundation (CICYT, projects CTM2011-28595-C02-01 and CTM2011-28595-C02-02), whose support is gratefully acknowledged.

References

- Abdel-Raouf, N., Al-Homaidan, A.A., Ibraheem, I.B.M., 2012. Microalgae and wastewater treatment. *Saudi J Biol Sci*, 19, 257-275
- Ali, K.A., Witter, D.L., Ortiz, J.D., 2014. Multivariate approach to estimate colour producing agents in Case 2 waters using first-derivative spectrophotometer data. *Geocarto Int*, 29(2), 102-127, doi:10.1080/10106049.2012.743601
- Arbib, Z., Ruiz, J., Álvarez-Díaz, P., Garrido Pérez, C., Perales, J.A., 2014. Capability of different microalgae species for phytoremediation processes: Wastewater tertiary treatment, CO₂ bio-fixation and low cost biofuels production. *Water Res*, 49, 465-474, doi:10.1016/j.watres.2013.10.036
- Barlow, R., Kyewalyanga, M., Session, H., van den Berg, M., Morris, T., 2008. Phytoplankton pigments, functional types, and absorption properties in the Delagoa and Natal Bights of the Agulhas ecosystem. *Estuar Coast Shelf S*, 80, 201-211, doi:10.1016/j.ecss.2008.07.022
- Barocio-León, O.A., Millán-Núñez, R., Santamaría-del-Angel, E., González-Silvera, A., Trees C.C., 2006. Spatial variability of phytoplankton absorption coefficients and pigments of Baja California during November 2002. *J Oceanogr*, 62, 873-885
- Bricaud, A., Morel, A., Babin, M., Allali, K., Claustre, H., 1998. Variations of light absorption by suspended particles with the Chl a concentration in oceanic (case 1) waters: analysis and implications for bio-optical models. *J of Geophys Res*, 103 C13, 31033-31044
- Brown, P.J., 1982. Multivariate calibration. *J R Stat Soc*, 44(3), 282-321
- Chiu, S.Y., Kao, C.Y., Chen, C.H., Kuan, T.C., Ong, S.C., Lin, C.S., 2008. Reduction of CO₂ by a high-density culture of *Chlorella* sp. in a semicontinuous photobioreactor. *Bioresour Technol*, 99, 3389-3396

Ciotti, A.M., Lewis, M.R., Cullen, J.J., 2002. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnol Oceanogr*, 47(2), 404-417, doi:10.4319/lo.2002.47.2.0404

Cleveland, J.S., Weidemann, A.D., 1993. Quantifying absorption by aquatic particles: a multiple scattering correction for glass-fiber filters. *Limnol Oceanogr*, 38, 1321-1327

Cullen, J.J., Ciotti, A.M., Davis, R.F., Lewis, M.R., 1997. Optical detection and assessment of algal blooms, *Limnol Oceanogr*, 42, 1223-1239

Delgadillo-Mirquez, L., Lopes, F., Taidi, B., Pareau, D., 2016. Nitrogen and phosphate removal from wastewater with a mixed microalgae and bacteria culture. *Biotechnol Rep*, 11, 18-26

Fournier, R., 1978. Membrane filtering, in Sournia, A. ed. *Phytoplankton Manual. Monographs on Oceanographic Methodology*, Unesco, 197-201

Geladi, P., 2002. Some recent trends in the calibration literature. *Chemometr Intell Lab Syst* 60 (1), 211-224, doi 10.1016/S0169-7439(01)00197-6

Geladi, P., Kowalski, B.R., 1986. Partial Least-Squares Regression: A tutorial, *Anal Chim Acta*, 185, 1-17, doi:10.1016/0003-2670(86)80028-9

Giménez, J.B., Robles, A., Carretero, L., Durán, F., Ruano, M.V., Gatti, M.N., Ribes, J., Ferrer, J., Seco, A., 2011. Experimental study of the anaerobic urban wastewater treatment in a submerged hollow-fibre membrane bioreactor at pilot scale, *Bioresource Technol*, 102, 8799-8806, doi:10.1016/j.biortech.2011.07.014

Griffiths, M.J., Garcin, C., van Hille, R.P., Harrison, S.T.L., 2011. Interference by pigment in the estimation of microalgal biomass concentration by optical density, *J Microbiol Meth*, 85, 119-123, doi:10.1016/j.mimet.2011.02.005

Hernandez, J.P., de-Bashan, L.E., Bashan, Y., 2006. Starvation enhances phosphorus removal from wastewater by the microalgae *Chlorella* spp. co-immobilized with *Azospirillum brasilense*. *Enzyme Microb Tech*, 3, 190-198, doi:10.1016/j.enzmictec.2005.06.005.

Kirk, J.T.O., 1994. *Light and Photosynthesis in Aquatic Ecosystems*, 2nd ed. Cambridge University Press, Cambridge

Le, C., Li, Y., Zha, Y., Sun, D., 2008. Specific absorption coefficient and the phytoplankton package effect in Lake Taihu, China. *Hydrobiologia*, 619, 27-37, doi:10.1007/s10750-008-9579-6

Lichtenthaler, H.K., Buschmann, C., 2001. Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy. *Curr Protoc Food Analyt Chem*, F4.3.1-F4.3.8, doi:10.1002/0471142913.faf0403s01

Lodi, A., Binaghi, L., Solisio, C., Converti, A., Del Borghi, M., 2003. Nitrate and phosphate removal by *Spirulina platensis*. *J Ind Microbiol Biot*, 30, 656-660, doi:10.1007/s10295-003-0094-5

Lund, J.W.G., Kipling, C., Le Cren, E.D., 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia*, 11, 143-170

Mallick, N., 2002. Biotechnological potential of immobilized algae for wastewater N, P and metal removal: a review, *Biometals*, 15, 377-390

Martens, H., Naes, T., 1989. *Multivariate Calibration*, Wiley, New York

Martínez, M.E., Sánchez, S., Jiménez, J.M., El Yousfi, F., Muñoz, L., 2000. Nitrogen and phosphorus removal from urban wastewater by the microalga *Scenedesmus obliquus*. *Bioresource Technol*, 73, 263-72

- Mercado, J.M., Ramírez, T., Cortés, D., Sebastián, M., Reul, A., Bautista, B., 2006. Diurnal changes in the bio-optical properties of the phytoplankton in the Alborán Sea (Mediterranean Sea). *Estuar Coast Shelf S*, 69, 459-470, doi:10.1016/j.ecss.2006.05.019
- Millán-Núñez, R., Millán-Núñez, E., Álvarez-Borrego, S., Trees, C.C., Santamaría-del-Ángel, E., 2004. Variability of the phytoplankton community in San Quintín Bay based on pigment analysis. *Cienc Mar*, 30 1A, 145-153, Available in:
<http://www.redalyc.org/articulo.oa?id=48030111>
- Millie, D.F., Schofield, O.M., Kirkpatrick, G.J., Johnsen, G., Tester, P.A., Vinyard, B.T., 1997. Detection of harmful algal blooms using photopigments and absorption signature: a case study of the Florida red tide dinoflagellate, *Gymnodinium breve*. *Limnol Oceanogr*, 42, 1240-1251
- Olguín, E.J., Galicia, S., Mercado, G., Pérez, T., 2003. Annual productivity of *Spirulina* (*Arthrospira*) and nutrient removal in a pig wastewater recycling process under tropical conditions. *J Appl Phycol*, 15, 249-257
- Organelli, E., Bricaud, A., Antoine, D., Uitz, J., 2013. Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site). *Appl Optics*. 52(11), 2257-2273, doi:10.1364/AO.52.002257
- Park, J., Jin, H.F., Lim, B.R., Park, K.Y., Lee, K., 2010. Ammonia removal from anaerobic digestion effluent of livestock waste using green alga *Scenedesmus* sp. *Bioresource Technol*, 101(22), 8649-8657, doi:10.1016/j.biortech.2010.06.142
- Rocha, J.M.S., Garcia, J.E.C., Henriques, M.H.F., 2003. Growth aspects of the marine microalga *Nannochloropsis gaditana*. *Biomol Eng*, 20, 237-242
- Ruiz-Martinez, A., Martin Garcia, N., Romero, I., Seco, A., Ferrer, J., 2012. Microalgae cultivation in wastewater: Nutrient removal from anaerobic membrane bioreactor effluent. *Bioresource Technol*, 126, 247-253, doi:10.1016/j.biortech.2012.09.022

- Sacristán de Alva, M., Luna-Pabello, V.M., Cadena, E., Ortíz, E., 2013. Green microalga *Scenedesmus acutus* grown on municipal wastewater to couple nutrient removal with lipid accumulation for biodiesel production. *Bioresource Technol*, 146, 744-748
- Sforza, E., Ramos-Tercero, E.A., Gris, B., Bettin, F., Milani, A., Bertucco, A., 2014. Integration of *Chlorella protothecoides* production in wastewater treatment plant: From lab measurements to process design. *Algal Res*, 6, 223-233
- Shi, J., Podola, B., Melkonian, M., 2007. Removal of nitrogen and phosphorus from wastewater using microalgae immobilized on twin layers: an experimental study. *J Appl Phycol*, 19, 417-423, doi:10.1007/s10811-006-9148-1
- Simionato, D., Sforza, E., Corteggiani Carpinelli, E., Bertucco, A., Giacometti, G.M., Morosinotto, T., 2011. Acclimation of *Nannochloropsis gaditana* to different illumination regimes: Effects on lipids accumulation. *Bioresource Technol*, 102, 6026-6032
- Smayda, T.J., 2008. Complexity in the eutrophication–harmful algal bloom relationship, with comment on the importance of grazing. *Harmful algae*, 8, 140-151.
- Tan, X., Chu, H., Zhang, Y., Yang, L., Zhao, F., Zhou, X., 2014. *Chlorella pyrenoidosa* cultivation using anaerobic digested starch processing wastewater in an airlift circulation photobioreactor. *Bioresource Technol*, 170, 538-548
- Vargo, G.A., 1978. Using a fluorescence microscope, in Sournia, A. ed. *Phytoplankton Manual*. *Monographs on Oceanography Methodology*, UNESCO, 108-112
- Viruela, A., Murgui Mezquita, M., Gomez-Gil, T.A., Duran, F., Robles, A., Ruano, M.V., Seco, A., Ferrer, J., 2016. Water resource recovery by means of microalgae cultivation in outdoor photobioreactors using the effluent from an anaerobic membrane bioreactor fed with pre-treated sewage. *Bioresource Technol*, 218, 447-454, doi:10.1016/j.biortech.2016.06.116

Wang, L., Li, Y.C., Chen, P., Min, M., Chen, Y.F., Zhu, J., Ruan, R.R., 2010. Anaerobic digested dairy manure as a nutrient supplement for cultivation of oil-rich green microalgae *Chlorella* sp. *Bioresource Technol*, 101, 2623-2628, doi:10.1016/j.biortech.2009.10.062