

Multivariate modeling for microalgae growth in outdoor photobioreactors

A. Mazzelli^{a,b,*}, A. Cicci^a, F. Di Caprio^b, P. Altimari^b, L. Toro^a, G. Iaquaniello^a,
F. Pagnanelli^b

^aBIO-P srl, Via Di Vannina, 88, 00156, Roma, Italy

^bDipartimento di Chimica, Università Degli studi di Roma "La Sapienza",
Piazzale Aldo Moro 5, 00185, Roma, Italy

Abstract

An empirical model for prediction of microalgal growth in outdoor photobioreactors cultivation, using Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression method, is implemented. Experimental data of biomass production were collected over 1 year of operation of a bubble column prototype, monitoring light and temperature and changing cultivation's conditions. PCA isolates 2 Principal Components that explain 80% of the variance and are associated with Environmental Conditions and Cultivation Conditions. Moreover, the PLS regression model showed positive results in term of responses ($R^2=0.84$) and residuals, following the experimental trends of outputs as specific growth rate ($\mu(d^{-1})$) and productivity calculated at C_{max} ($P_{max}(gL^{-1}d^{-1})$), giving also good prediction results in its validation test. This method could be easily used for other purpose, by changing the input values of the specific cultivation used (including CO_2 uptake or wastewater dilution ratio in the culture medium), obtaining as outputs the desired variables (lipid production rate, etc.).

Keywords: PCA, PLS, Multivariate statistics, outdoor cultivation, microalgae, empiric model

*Corresponding author

Email address: alessio.mazzelli@uniroma1.it (A. Mazzelli)

1. Introduction

Recently, the global market is shifting the focus towards new green products, drawing the attention to microalgae as a valid alternative source of high added value products [1, 2] and renewable biofuels, for example: biodiesel, biohydrogen and biomethane [3]. A large number of microalgal species can attain a lipid content between 20-50% DCW (Dry Cell Weight) depending on the cultivation conditions [3–9]. With the aim of reducing costs and environmental impact [10], several studies have focused on using microalgal cultivations for CO₂ capture from industrial processes [7, 11], and also on the addition of wastewater to algal nutrient media, to reduce both the cost of biomass production and wastewater treatment [12–15]. For photoautotrophic production, the most popular cultivation systems are open ponds and closed photobioreactors. There are however a large number of factors that could affect the growth of microalgae and these can be divided in three categories [7]:

- Abiotic factors such as light, temperature, nutrients (CO₂, N, P, K, etc.), O₂, pH, salinity and toxins.
- Biotic factors including bacteria, viruses, fungi, and other species in competition with microalgae.
- Operational factors such as mixing and stirring conditions, dilution ratio, vessel width and depth, and harvest frequency.

For photoautotrophic cultivations, light availability is the most important factor affecting cell growth; its control is difficult in outdoor cultures, due to the variation of solar radiation during day and season, and its non-homogeneous distribution inside the photobioreactors. This heterogeneous light distribution can cause photolimitation and/or photoinhibition, depending on both light irradiance intensity and microalgal concentration, thus significantly affecting photoproduction. Light distribution, temperature is also an important limiting factor for growing algae in both indoor or outdoor systems. Many microalgae can tolerate temperature increases or decreases, but diverging the optimum

temperature by only 2 or 4 °C may lead to culture loss [16]. Moreover, overheating problems may occur in outdoor systems, making it necessary to use cooling systems to keep temperatures below 28-30°C [17]. Given these considerations, in the last two decades an increasing number of scientists have tried to predict microalgal growth and metabolites production under transient conditions of light intensity and temperature using empirical models [18, 19] or semi-empirical models [20, 21]. With respect to light modeling and its connection to microalgal growth, there is a differentiation based on the models' ability to take into account light gradients [18, 22], light cycles [23, 24] and also physical phenomena (as scattering) [22, 25] that occur in outdoor cultures. In the case of temperature, the modeling approaches can be divided into coupled and uncoupled, depending on whether the models take into account or not the potential interdependence of light and temperature on growth [26]. Predictive power of such model is usually achieved by a large number of adjustable parameters difficult to relate to physical and chemical phenomena, and thus with scarce identifiability [27]. These problems inevitably affect the model validation for an end user, making them useless for an accurate prediction of microalgal growth in slightly different outdoor conditions. For these reasons, there is a need to develop models that an end user would be easily able to handle, without losing accuracy [25]. In this work an empirical model to predict microalgae growth in outdoor cultivation system, using multivariate statistics, is implemented to overcome the problems of complex mathematical models. In particular, the multivariate statistical projection methods PCA (Principal Component Analysis) and PLS (Partial Least Square) are used for the purpose. PCA and PLS were initially used in process control for their ability to compress multidimensional data and to extract the most useful informations, by projecting these data into a low-dimensional space having as new reference system: the principal components [28]. In particular, for algae production both techniques have been used to analyze the water chemistry conditions in three wastewater stabilization ponds with excessive algae growth and fluctuating pH, finding correlations between variables (pH, temperature, light, dissolved oxygen, etc), and developing a multivariate regressions

model for pH as a dependent variable [29]. Other similar applications were used to study and identify correlations between different algal species present in Lake Wingra [30] and also to illustrate the influence of environmental variables on phytoplankton composition in the Vaal River [31]. Unlike previous works, in this paper an innovative use of PCA and PLS methods is developed, not only to reduce data redundancy but also to predict microalgal growth in a specific period of the year with defined weather and cultivation conditions. 1 year experimental data from a pilot scale phototrophic plant were analyzed with PCA and a PLS multivariate regression method, to obtain the specific growth rate ($\mu(\text{d}^{-1})$) and the productivity calculated at C_{max} ($P_{max}(\text{gL}^{-1}\text{d}^{-1})$).

2. Materials and Methods

2.1. Microalgal outdoor cultivation

In this work two different strains of algae named *Tetradismus obliquus* and *Graesiella emersonii* were selected and cultivated in an outdoor photobioreactors (PBRs) pilot plant [17, 32]. Each inoculum was prepared using local tap water [33] in place of distilled water [34, 35]. The pilot plant was installed in Rome (Italy) at the Bio-P s.r.l. site (N 41°55' 5" E 12°35' 35"); it was fitted with 10 column photobioreactors with an operative volume of 21 L each (internal diameter=14 cm, height=150 cm), anchored to a metal support structure. Each reactor was connected to air (for mixing purposes) and to CO₂ lines (for pH control). The air flux was generated by a membrane compressor (AIRMAC 40W). Mixing inside the reactors was achieved by using toroidal spargers, while CO₂ was injected as pure on demand from cylinders. When a certain concentration was reached (1-3 gL⁻¹), the microalgal suspension was collected and sent to a 95 Lh⁻¹ bucket centrifuge (Raw Power Centrifuge). An aliquot of the concentrated suspension was used for the subsequent inoculum. The model is designed with the aim of predicting microalgae growth at specific outdoor cultivation conditions. Inputs (variables for PCA and predictors for PLS) and required outputs (responses for PLS) are summarized in Tab. 1. The experimental data covered a 9 months period (March 2017-December 2017) and to each experiment

Tab. 1: Input and Output Values of the model

Input Values	Output Values
Daily Illumination Time (h)	
Microalgal Inoculum Concentration: $C_{init}(gL^{-1})$	Specific Growth Rate:
NaNO ₃ Initial Concentration: $NaNO_{3init}(gL^{-1})$	$\mu(d^{-1})$
Maximum Internal Temperature Averaged: $T_{imax avg}(^{\circ}C)$	
Maximum External Temperature Averaged: $T_{emax avg}(^{\circ}C)$	
Minimum Internal Temperature Averaged: $T_{imin avg}(^{\circ}C)$	Productivity calculated at C_{max} :
Minimum External Temperature Averaged: $T_{emin avg}(^{\circ}C)$	$P_{max}(gL^{-1}d^{-1})$
Photosynthetic Photon Flux Density: PPF (μE m ⁻² s ⁻¹)	

(totally 10) is assigned a label (A-L) for a better clarity. In Tab. 2 the experiments with the corresponding label and the period in which were conducted are resumed. All the cultivations periods were ended when the microalgae showed a decrease of their growth rate due to nutrients depletion. Internal and External refer to the positions of the probes relatively to the reactors; furthermore the “averaged” terms are referred to the variables’ collected data, averaged over all their values for each experiment.

2.2. Monitoring of microalgae growth

During each experiment, microalgae concentration was determined daily both by cell count and dry weight for both species. For dry weight measures, each sample was firstly washed with 1 mL of sodium acetate buffer solution (sodium acetate 0.5 M, pH=4.8) in order to dissolve any salt that could have misrepresented the measure. After that, 10 mL of each sample was filtrated using 0.70 μm microfiber filters (VWR). The filters were then dried at 105 °C for half an hour and weighed. Cell counting was performed with an optical microscope (Motic EF-N PLAN) in a 10⁻⁴ mL Thoma chamber. The values of specific growth rates were obtained as:

$$\mu = \frac{1}{x} \frac{dx}{dt} \quad (1)$$

Where x are the cell concentration values (10⁶cell mL⁻¹) during the exponential growth phase. The productivity calculated at C_{max} , was obtained as:

$$P_{max} = \frac{C_{max} - C_{init}}{t_{max} - t_{init}} \quad (2)$$

Tab. 2: Experiments' Periods during 2017

Experiment	Period of 2017	Notation
A	02-March/14-March	March I
B	21-March/31-March	March II
C	4-April/14-April	April
D	19-April/10-May	April-May
E	15-May/22-May	May
F	20-June/07-July	June-July
G	10-July/27-July	July
H	22-September/17-October	Sep-Oct
I	27-October/22-November	October-November
L	24-November/22-December	Nov-Dec

Where C_{max} and C_{init} are the maximum observed biomass concentration (gL^{-1}) reached in each experiment and the biomass concentration (gL^{-1}) at the beginning of each experiment respectively; likewise t_{max} and t_{init} are the time at which the maximum observed biomass concentration was reached and the experiment start time (time zero). Consequently, P_{max} is not referred to the final/total productivity that is actually related to that is actually related to the batch duration.

2.3. Monitoring and control of variables

During the experiments pH and temperature were continuously monitored using probes inside reactors, controlled by an active feedback control system. Both pH and temperature data were continuously registered for analysis and displayed on a PC interface using LABVIEW software. The pH was maintained at its set point (pH=8) with the use of CO_2 , that was injected on demand directly inside the reactors. The temperature was kept below the maximum temperature threshold setting ($T=30\text{ }^\circ\text{C}$) by using a water spray system, designed and built for the purpose. Two probes were placed inside and outside the reactors, to measure internal and external temperature. The illuminance was measured every day (at 10 am, at 2 pm and at 5.30 pm) by using a luxmeter (LM-8000, LT-Lutron) and transformed to the corresponding value of Photosynthetic Photon Flux Density (PPFD) ($\mu\text{Em}^{-2}\text{ s}^{-1}$) through multiplication with the conversion factor (0.0185) for sunlight light source [36]. Every measure was taken at three points for each reactor,

at different heights from the ground: at the bottom (20 cm), at a middle height (80 cm) and at the top (140 cm). Light measurements at the bottom and middle were normalized with respect to light on the top, being always lower than this value. As a reference of light fluctuations, light was also measured at a fixed point not affected by any shadowing (named as “unshaded reference point”).

2.4. Multivariate Statistical Analysis

In the present study two multivariate methods, PCA and PLS respectively, were used, firstly to investigate the effects of variables on microalgal growth, and then to develop an empirical model for growth estimation using *MINITAB*[®] and OriginPro (OriginLab Corporation) software. For a short explanation of PCA and PLS methods see Appendix A.1-A.2.

3. Results and Discussion

3.1. Input & Output Data

Before showing PCA and PLS results, the values of the two inputs and outputs, representing the most important outdoor variables (light and temperature) influencing the outdoor cultivation in closed photobioreactors, are reported in this paragraph. Temperature is a key outdoor factor regulating the metabolism of each microorganism, by affecting reaction kinetic and biomolecules stability. Light is the energy source of microalgae, crucial for the photosynthesis and the whole metabolism. Both these outdoor factors can change several folds in photobioreactors in response to seasons, day/night cycle and weather, in a range known (from laboratory experiments) to affect microalgae metabolism. Other possible outdoor factors, such as rain and wind, were excluded because by working with photobioreactors their influence could be assumed negligible. In the case of other reactors configuration, as for example for open ponds, such other factors could be relevant because they could produce phenomena as water evaporation (by wind) and rainwater addition in the culture (by rain). Besides the environmental variables, in this study two factors

(considered as key factors), connected to the cultivation conditions, are investigated and tuned during the various experiments: the initial biomass concentration (C_{init}) and the initial NaNO_3 concentration (NaNO_{3init}) of each test. These two factors are essential to understand the experiment's duration and the biomass production. Other factors, such as CO_2 and pH, can also influence the microalgae growth, but in the pilot plant used in this study, the pH was controlled by a control system which maintained it always constant around 8 (which is also the optimal pH for such species). For such reason pH was assumed to be not influent on our tests. Furthermore, CO_2 was supplied on demand by the pH control system and thus was assumed to be sufficient, and not a limiting nutrient, for microalgae growth throughout the cultivation. In Fig. 1 maximum and minimum values of Internal and External Temperature during the experiments are shown. It is evident that the plots of Minimum Temperatures (Fig. 1.b and Fig. 1.d) have the same trend and similar values, in contrast to the Maximum Temperatures (Fig. 1.a and Fig. 1.c). This difference is due mainly to two reasons: firstly, the water contained in a closed vessel directly irradiated by sun, without an active cooling system, heats up to a temperature higher than that of the air ($T_{imax} > T_{emax}$) for physical reasons related to heat transfer efficiency and the thermal capacity of water. Moreover, these effects are improved by the trigger of the Non-Photochemical Quenching (NPQ) mechanism. This particular defence mechanism protects microalgae from the negative effects of high solar light absorption, dissipating the excess amount of light energy to heat and giving the appearance of an exothermic reaction. The second reason is the presence of the temperature control system with water spray cooling that maintains, when it is active in June-September period, the Internal Temperature at its set point $T = 30^\circ\text{C}$. In all four subplots, anyway, the seasons temperature trend is visible, and temperature reaches maximum value at about $T_{max} = 38^\circ\text{C}$ in July, and minimum value at about $T_{min} = 4^\circ\text{C}$ in December.

In Fig. 2, the light trend both in term of Daily Illumination Time (dotted) and of PPFD (yellow bars) is plotted; the Daily Illumination Time is averaged on the light hours during each experiment. The irradiance measured values can be positioned between

those of Amsterdam (N 52°22' 13" E 4°53' 43") and Ankara (N 39°55' 32" E 32°51' 59") [37]. It can be seen, as expected, that the maximum value of PPFD ($260 \mu\text{Em}^{-2}\text{s}^{-1}$) is reached during the summer period, as also for the Daily Illumination Time (15.6 h) that begins to decrease with autumn. In Fig. 3 the output values during each experiment are plotted in terms of specific growth rate ($\mu(\text{d}^{-1})$) and productivity calculated at C_{max} ($P_{max}(\text{gL}^{-1}\text{d}^{-1})$), both averaged on the replicates. These values are referred to both *Tetradismus* and *Graesiella*, but for the empiric model's development are considered as belonging to a single microalgal strain. These assumptions are verified analyzing the results obtained running PCA/PLS methods, with the experimental values of each

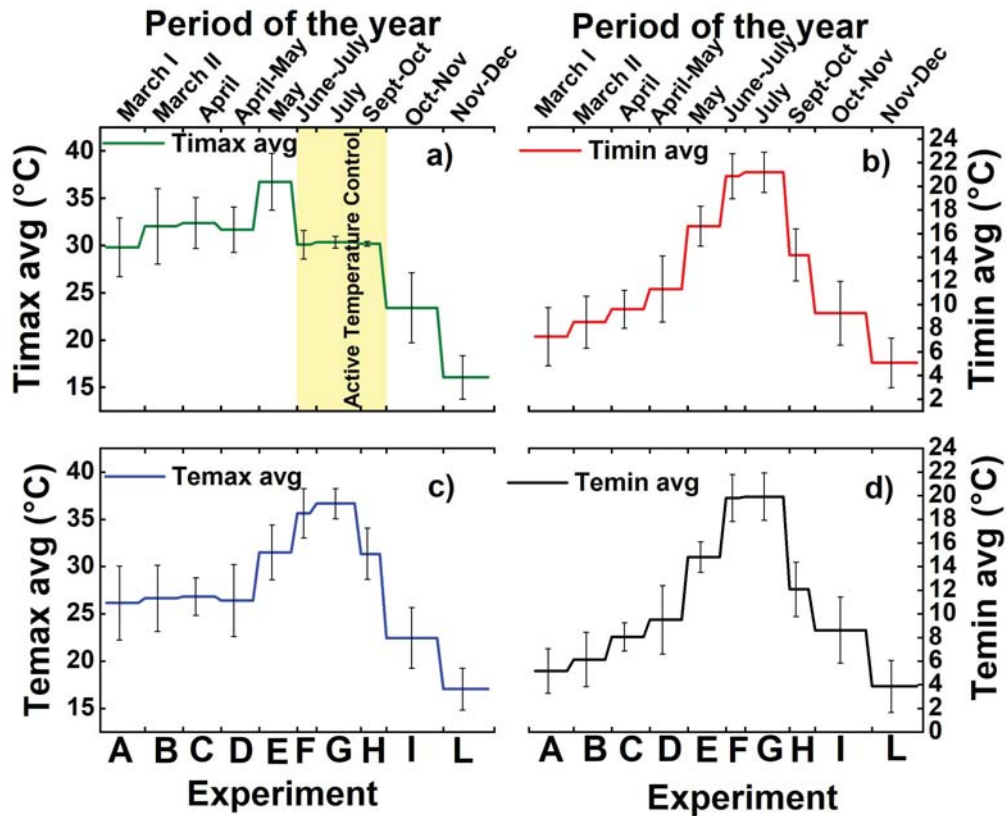


Fig. 1: Averaged maximum/minimum values of Internal (a/b) and External (c/d) Temperature during the experiments. The yellow bar indicates the cooling system activation during the summer periods. The bars represent the Standard Deviations of the replicates. For all the 4 subplots (a-d) the replicates consist in the duration, in terms of days, of each experiment (n=13 for A; n=11 for B; n=11 for C; n=23 for D; n=8 for E; n=3 for F; n=18 for G; n=26 for H; n=27 for I; n=29 for L).

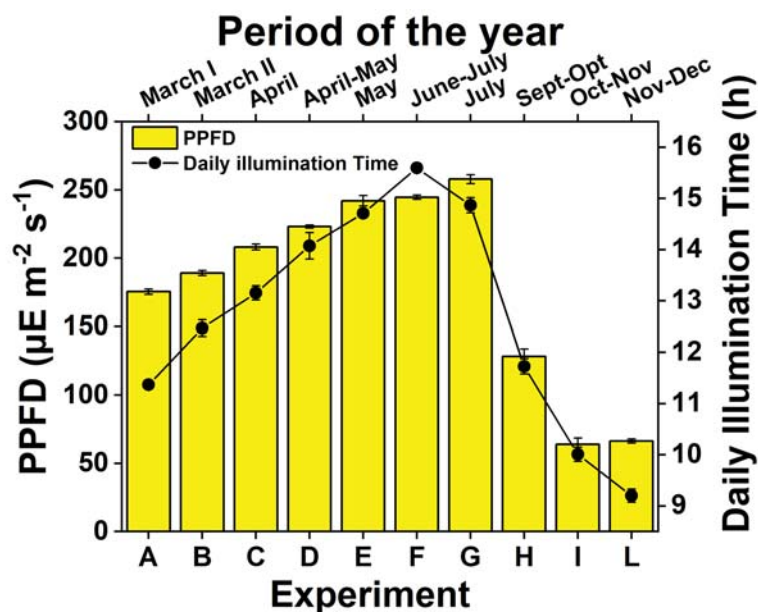


Fig. 2: Photosynthetic Photon Flux Density (PPFD) ($\mu\text{E m}^{-2} \text{s}^{-1}$) and Daily Illumination Time (h) averaged trends during each experiment. The higher values are obtained in summer periods, having more daylight hours and a better solar irradiance. The bars represent the Standard Deviations of the replicates. About PPFD the replicates' number (n) consist in the number of reactors used in each experiment (n=6 for A,B,C,D,E; n=3 for F; n=7 for G; n=4 for H; n=9 for I; n=8 for L). Concerning to Illumination Time the replicates consist in the duration, in terms of days, of each experiment (n=13 for A; n=11 for B; n=11 for C; n=23 for D; n=8 for E; n=3 for F; n=18 for G; n=26 for H; n=27 for I; n=29 for L).

microalga considered singularly (See Appendix A.3 for further informations). The results obtained are very similar, highlighting thus that these slight differences, in terms of the experimental values of μ and P_{\max} between *Tetradesmus* and *Graesiella*, are too small to be detectable from these multivariate statistics methods. Fig. 3 shows a positive trend of P_{\max} moving towards the summer period (from experiment A to G), with the highest value ($0.17 \text{ gL}^{-1} \text{d}^{-1}$) obtained in the experiment G, and a negative trend moving towards the winter months (from experiment H to L). These P_{\max} values are fully in line with the outdoor variables plotted in Fig. 1 and Fig. 2 showing better results with the increasing of temperature and light availability. The only exception of P_{\max} increasing trend towards summer period, is shown in the experiments D and E and is due to the different NaNO_3 concentration used. Indeed, in D a doubled concentration of NaNO_3 respect to others

experiments was used, leading to an increase of biomass and therefore of productivity; on the other hand in E experiment using an half of normal NaNO_3 concentration lead to a low productivity. These results highlight the relevance of C_{init} and NaNO_{3init} factors, in order to understand the experiments duration and the biomass production; indeed, the D and E experiment showed different values of P_{max} (therefore intrinsically also of duration), even if were conducted at similar environmental conditions. As regards the specific growth rate μ , in Fig. 3 there isn't a clear trend because of high standard deviations between the replicates, especially in the experiments C and E. These variability of μ data are caused by the several factors influencing the outdoor microalgal growth and for this reason PCA and PLS methods were chosen to isolate the different factors affecting the growth.

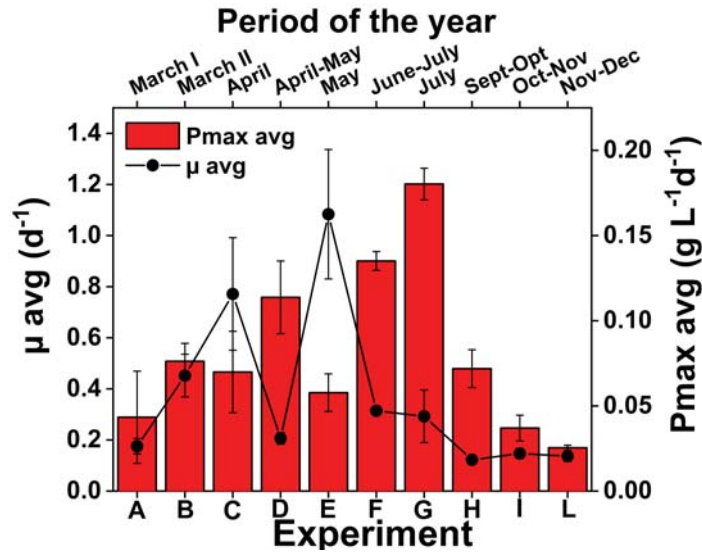


Fig. 3: Output Values Profiles during the experiments (March 2017-December 2017) in terms of Specific growth rate ($\mu(\text{d}^{-1})$) and Productivity calculated at C_{max} ($P_{max}(\text{gL}^{-1}\text{d}^{-1})$). A positive trend of P_{max} towards summer months (with the exception of E conducted at an half of normal NaNO_3 concentration) is shown. μ doesn't show a clear trend because of high standard deviations between the replicates (especially for C and E) due to the several factors influencing the outdoor microalgal growth. The bars represent the Standard Deviations of the replicates for both microalgae, in each experiment, considered separately. For both outputs the replicates' number (n) consist in the number of reactors used in each experiment (n=6 for A,B,C,D,E; n=3 for F; n=7 for G; n=4 for H; n=9 for I; n=8 for L).

The P_{max} and μ values in Fig. 3 are comparable with those reported in Tab. 3, obtained in similar environmental conditions in outdoor cultivations, both in Italy and in other

locations of the world. In particular during the cultivation period of work described by Camacho-Rodríguez et al. [38] we obtained $P_{max}=0.025-0.25$ and $\mu=0.12-1.1$, during the cultivation period of work described by Bosma et al. [39] we obtained $P_{max}=0.04-0.135$ and $\mu=0.15-0.31$ and lastly during cultivation period of work described by Cheng-Wu et al.[40] we obtained $P_{max}=0.025-0.18$ and $\mu=0.14-0.3$. Only Italy values of Chini Zittelli et al.[41] are slightly different than our ($P_{max}=0.07-0.18$ and $\mu=0.12-0.3$, especially for P_{max}), due to their optimized pilot plant positioning that have increased the photosynthetic productivity.

Tab. 3: μ and P_{max} comparison with other similar environmental conditions in outdoor microalgal cultivations

Photobioreactor type	Strain	Location	Cultivation period	$P_{max}(\text{gL}^{-1}\text{d}^{-1})$	$\mu(\text{d}^{-1})$	Author
Bubble Column	Tetraselmis suecica	Italy	August-September	0.4-0.65	0.32-0.65	[41]
Closed Bags	Nannochloropsis gaditana	Spain	January-December	0.02-0.13	0.1-0.5	[38]
Bubble Column	Monodus subterraneus	Netherlands	July-October	0.05-0.2	0.6-0.4	[39]
Flat Panel	Nannochloropsis sp.	Israel	Winter-Summer	0.14-0.25	0.15-0.17	[40]
Bubble Column	<i>Tetradesmus obliquus</i> and <i>Graesiella emersonii</i>	Italy	March-December	0.025-0.25	0.12-1.1	This work

3.2. PCA Results

PCA analysis is implemented and the redundant variables are eliminated; for example variables related to light measures along reactor axis (top, medium, bottom) are deleted, leaving as light representig data, the measure taken on the “unshaded reference point”. PCA showed that only 2 Principal Components (PC) are needed to explain most of the data variance (88%). This result is carried out by using a Cross Validation (CV) method and the calculation of the Predicted Residual Error Sum of Squares (PRESS) [42]. The results obtained with PCA implementation, giving a PC=2, granted an easy interpretation of the variables’ effects on PCs in a bidimensional plot.

In Fig. 4 the Loading Plot (Fig. 4.a) and the Score Plot (Fig. 4.b) are reported. The Loading Plot is an illustration of how much each variable contributes to build (or ‘load’) each PC, showing how much each variable influence the PCs. In particular, Fig. 4.a evidences that the variables with a horizontal orientation ($T_{imax\ avg}$, $T_{emax\ avg}$, $T_{imin\ avg}$, $T_{emin\ avg}$, Light time, PPF) and similar lenght (which means similar influence on a PC) have a low effect on PC2 but a high effect on PC1. These variables also have good

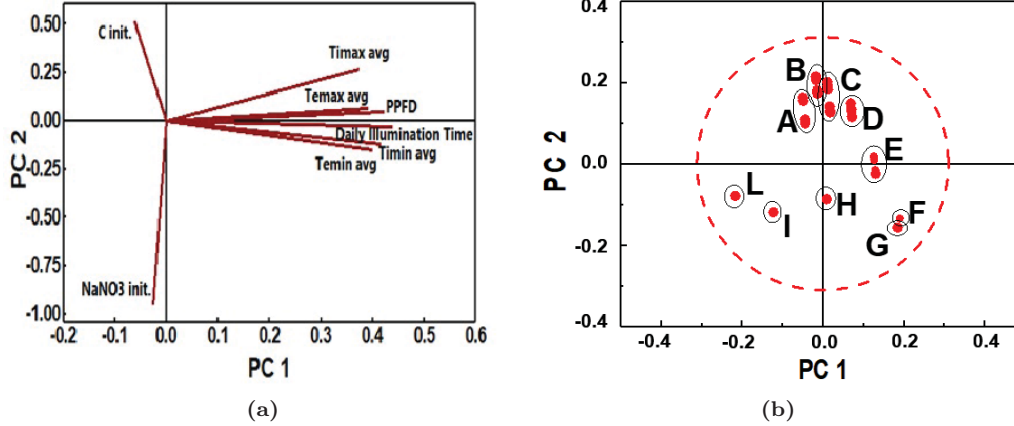


Fig. 4: Loading Plot (a) and Score Plot (b) relatively to the first two Principal Components (PC1 and PC2). In the Loading Plot the vectors with a horizontal orientation influence more PC1 (“Environmental Conditions”) and for the same reason the vectors with a vertical orientation influence more PC2 (“Cultivation Conditions”). The Score plot shows the individual observations, collected during the experiments, clustered in groups having as common features to belong to the same experiment(A-L) and moved near or far from the axis depending on the influence of PC1 or PC2.

correlation each other (verified also physically), due to the small angles between the vectors. Fig. 4.a also shows the variables approaching a vertical orientation (NaNO_3 , C_{init}) that are irrelevant to PC1, but significantly influence PC2. In particular, the NaNO_3 vector is longer than C_{init} one, being the most influential in PC2 load. Furthermore these two variables do not correlate with each other; they physically represent two different conditions. Since the vectors that influence PC1 more are composed of environmental variables, the PC1 can be named: “Environmental Conditions” and for the same reason the PC2 is called “Cultivation Conditions”. In Fig. 4.b the Score Plot is represented, showing that the observations are clustered in groups having common features, that means belonging to the same experiment(A-L). Moreover, the experiments A-B-C-D are also clustered near the vertical axis and shifted from the origin, highlighting that the observations are more influenced by the variables with a vertical orientation (NaNO_3 and C_{init}) that define the PC2. This result is due to the changes, in terms of initial biomass concentration (C_{init}) and the initial NaNO_3 concentration (NaNO_{3init}), carried out during those experiments. Besides that, the experiments F-G-I-L are affected by both

Environmental conditions and Cultivation conditions, without either of these two Principal Components prevail over the other. Lastly for the experiments H and E there is almost no contribution of both variables groups, indicating that the variables have values near their average level, having the observation matrix been centered and scaled. In Fig. 4.b a dashed circle is reported, representing the Mahalanobis distance, denoting no outlier [28]. Like many other multivariate statistical methods, the results obtained (in terms of plots, tables and numbers) have to be interpreted and sometimes are not univocal. Indeed, even if the first principal component (PC1) is the most important one because it explains the maximum percentage of variability present in the data that can be represented in only one dimension, it is possible that the second one (PC2) is influenced by more important variables (maybe economically). In our case PC2, influenced by the Cultivation Condition variables, is less important for the explanation of data variance but is more important for the economical point of view. Indeed, the amount of nutrients (NaNO_3 in our study) can affect, besides the growth rate and productivity, also the Operative Expenditures (OPEX) of the process. For this reason, a correct and deeper interpretation of PCA results, without stopping to a shallow analysis, is essential.

3.3. PLS Results

3.3.1. Model Selection and Predictors Evaluation

PCA results showed the connections between variables and their effects on PCs, grouping both observations and variables. An empirical model to represent the effect of variables on biomass growth is developed by PLS.

As a first step, in Fig. 5.a the PRESS (Predicted Residual Error Sum of Squares) values of each predictor (the same meaning of component for PCA) are plotted. It is shown that the minimum value of PRESS is obtained at the 6th predictor and consequently the PLS model will need six predictors to describe most of the variance (84%). This is shown in Fig. 5.b, where the R-Squared (R^2) value for each predictor is plotted for μ response (the same trend is observed also for P_{\max}). The R-Sq value provides the

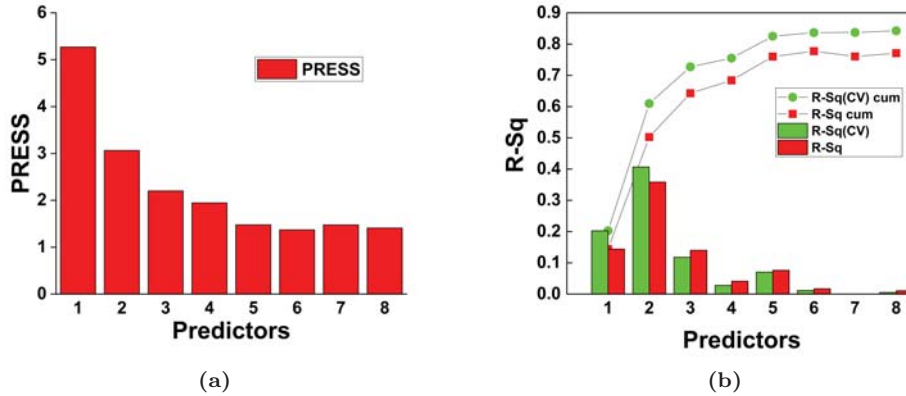


Fig. 5: PRESS (Predicted Residual Error Sum of Squares) Plot for each predictor (a); Model Selection Plot (b) representing the R-Squared trend for each predictor both with and without the Cross Validation (CV) method in absolute and cumulative values. The shows that the PLS model will need six predictors to describe most of the variance (84%) with Cross Validation method, as can be seen also in the Model Selection Plot.

proportion of variation in each response that is explained by the predictors, indicating how well each model fits data (the higher value obtained, the better the model fits the data). In particular in Fig. 5.b both absolute R-Sq (bars) and cumulative R-Sq (dots) are plotted, denoting the same result: i.e. maximum value of R-Sq of the 6th predictor (84%). As shown in Fig. 5.b using Cross Validation (CV), an improvement in variability representation can be achieved. The cross-validation technique (the leave-one-out case has been used in this study) works by omitting each observation one at a time, rebuilding the predictive model using the remaining data and then using this model to predict the omitted data, estimating at the end with PRESS the predicted residual error.

3.3.2. Response Analysis

In Fig. 6 predicted responses versus experimental data are reported for both outputs for direct fitting and CV procedure. Both plots in Fig. 6 show very little difference between the fitted and cross-validated fitted responses. Furthermore the points on both plots lie around the correlation line in a similar linear pattern, indicating that the model fits the data with a good level of accuracy. The analysis of residuals between fitted

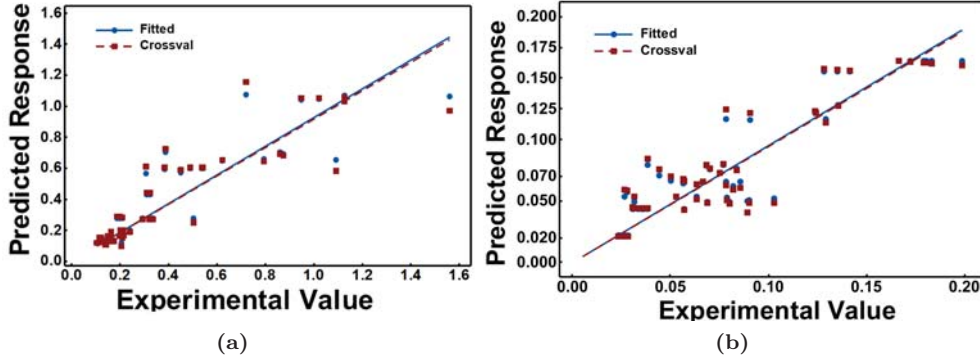


Fig. 6: Response Plot (plot of the fitted and cross-validated fitted responses versus the experimental values) for $\mu(d^{-1})$ (a) and $P_{max}(gL^{-1}d^{-1})$ (b). Both plots show very little difference between the fitted and cross-validated fitted responses with the points lying around the correlation line in a similar linear pattern, indicating the good level of accuracy of the model data fitting. In both plots the data are referred to the individual observations collected during the experiments.

and experimental data can give further insight into the model goodness. In Fig. 7, the Residual Normal Probability Plot for both outputs $\mu(d^{-1})$ (a) and $P_{max}(gL^{-1}d^{-1})$ (b) are reported. These graphs show the standardized residuals versus their expected values when the distribution is normal. The residuals appear to follow a straight line for μ values, while some distortion can be noticed for P_{max} but always within the range of confidence intervals at 95%. In Fig. 8 the predicted and experimental results are compared for both μ and P_{max} outputs, confirming the quality of the PLS regression model for different sets of experiments. In both plots the model is able to follow the experimental trends, overcoming the variability specific of the environmental conditions and resembling the effect of changed cultivation condition. Indeed, in both plots are present some possible outliers for $\mu(d^{-1})$ in April and May and for $P_{max}(gL^{-1}d^{-1})$ during March. As regards μ , the outliers could be caused by outdoor variability as described in Paragraph 3.1. On the other hand, for P_{max} the presence of possible outliers can be explained by the different initial cultivation conditions, in terms of C_{in} and $NaNO_{3init}$ carried out in the first experiments, as described in Paragraph 3.2.

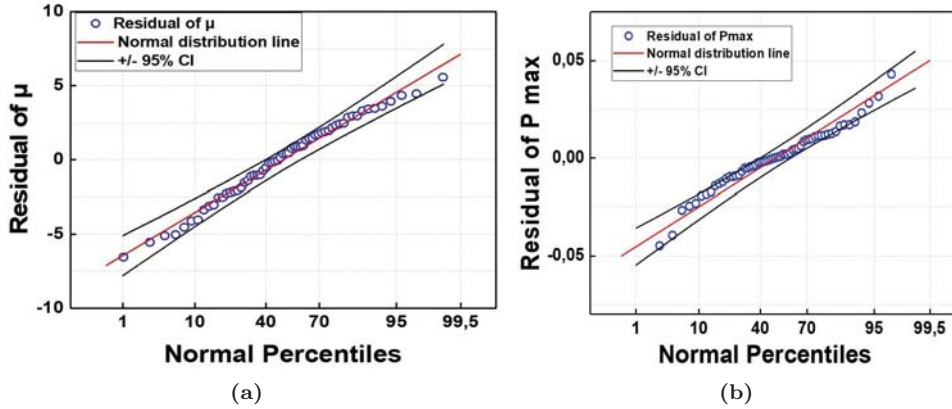


Fig. 7: Residual Normal Probability Plot (standardized residuals results versus their expected values when the distribution is normal) for $\mu(\text{d}^{-1})$ (a) and $P_{\text{max}}(\text{gL}^{-1}\text{d}^{-1})$ (b). The residuals appear to follow a straight line for μ values, while some distortion can be noticed for P_{max} but always within the range of 95% confidence intervals. In both plots the data are referred to the individual observations collected during the experiments.

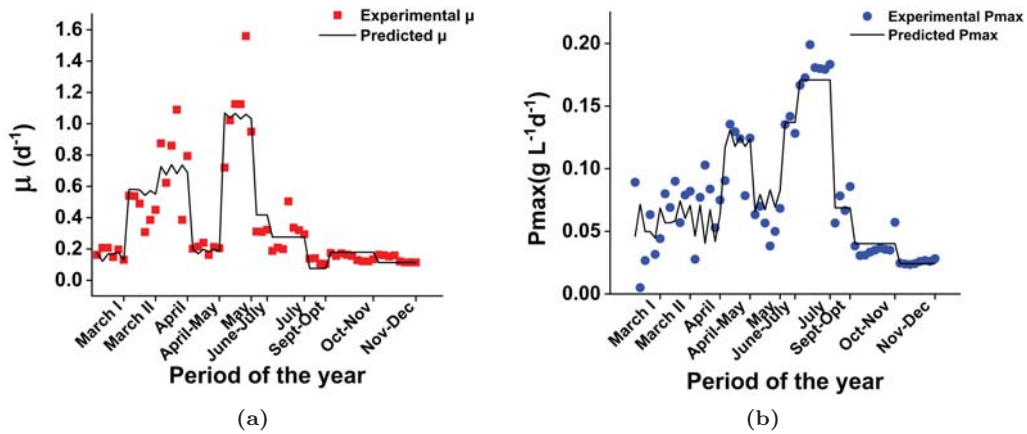


Fig. 8: Comparison between PLS predicted results and experimental ones for both outputs: (a) $\mu(\text{d}^{-1})$ and (b) $P_{\text{max}}(\text{gL}^{-1}\text{d}^{-1})$. In both plots the model is able to follow the experimental trends even with the presence of some possible outliers for $\mu(\text{d}^{-1})$ in April and May (due to the outdoor variability as described in Paragraph 3.1) and for $P_{\text{max}}(\text{gL}^{-1}\text{d}^{-1})$ during March (due to the different initial cultivation condition carried out). In both plots the data are referred to the individual observations collected during the experiments.

3.3.3. Empiric Model Prediction Results

In this subsection the prediction results using PLS model are shown. In particular, in Tab. 4 the experimental values obtained during period M (19-01-2018/22-02-2018) are listed; these values are used only for testing and validating PLS regression model's prediction

abilities and not for its developing. In Tab. 5 the Predicted Responses both for μ and

Tab. 4: Experimental values obtained during M period (19-01-2018/22-02-2018) used for testing and validating model prediction abilities. Responses: specific growth rate ($\mu(\text{d}^{-1})$) and productivity calculated at C_{max} ($P_{max}(\text{gL}^{-1}\text{d}^{-1})$). Predictors: maximum and minimum internal and external temperatures averaged during each experiment $T_{imax\ avg}$, $T_{imin\ avg}$, $T_{emax\ avg}$, $T_{emin\ avg}$ ($^{\circ}\text{C}$); initial inoculum concentration C_{init} (gL^{-1}) and initial nitrate concentration $\text{NaNO}_3\ init$ (gL^{-1}); photosynthetic photon flux density PPF($\mu\text{Em}^{-2}\text{s}^{-1}$); Daily Illumination Time (h).

EXPERIMENT	RESPONSE				PREDICTOR					
	μ	P_{max}	C_{init}	$\text{NaNO}_3\ init$	$T_{imax\ avg}$	$T_{imin\ avg}$	$T_{emax\ avg}$	$T_{emin\ avg}$	PPFD	Daily Illumination Time
M	0.20	0.031	0.30	0.35	19.56	5.43	18.9	4.74	4090	10.25

P_{max} , with their correspondig Standard Error (SE) of Fit and 95% Confidence Interval (CI), are listed and compared with the experimental ones. It can be seen that for both output variables, the regression values are similar but not identical to the experimental ones. This difference is mainly due to the absence of the experimental data collected during M period for PLS model developing, lowering its prediction ability. The values obtained in Tab. 5 are still an acceptable estimation, considering the high variability influencing the outdoor microalgal growth and underline the effective goodness of the PLS regression model. As a final step, in Tab. 6 the standardized regression coefficients, used with the predictors to calculate the fitted value of both response variables, are listed.

Tab. 5: $\mu(\text{d}^{-1})$ and $P_{max}(\text{gL}^{-1}\text{d}^{-1})$ PLS Regressed Values with relative Standard of Error (SE) and 95% Confidence Interval (CI) compared with the Experimental Values

OUTPUT VARIABLE	EXPERIMENTAL VALUE	REGRESSED VALUE	SE of FIT	95% CI
μ		0.20	0.037	(0.206; 0.357)
P_{max}		0.031	0.006	(0.012; 0.035)

Tab. 6: Regression standardized coefficients for both outputs: specific growth rate ($\mu(\text{d}^{-1})$) and productivity calculated at C_{max} ($P_{max}(\text{gL}^{-1}\text{d}^{-1})$)

PREDICTOR	μ STDCOEFF	P_{max} STDCOEFF
C_{init}	-0.050482	0.154351
NaNO_{3init}	-0.628423	0.402856
$T_{imax\ avg}$	0.664885	-0.366197
$T_{imin\ avg}$	-0.058646	0.281886
$T_{emax\ avg}$	-0.876960	0.271945
$T_{emin\ avg}$	0.009383	0.277537
PPFD	0.336034	0.295258
Daily Illumination Time	0.421541	0.124203

3.4. Conclusions

The innovative use of PCA and PLS for modeling and predicting microalgae growth in a phototrophic outdoor pilot plant, has shown many positive features including the ability to analyse numerous datasets, subject to high variability, without losing its predictive abilities. Model predictions results showed acceptable values for both responses μ and P_{max} , enabling the end user to establish how much biomass will be obtained in certain outdoor cultivation condition, with a good level of accuracy. This model could be easily adapted for prediction of other outputs (e.g. lipids accumulation for biodiesel production), by simply changing the input values. Indeed this model could be also applied for prediction of lipids accumulation for biodiesel production in outdoor cultivation in other geographical regions with different weathers conditions, only by changing the input values. All of this has to be done always limiting the sources of data alteration to a level that PCA and PLS could manage. For model improvement, the M period results could be used for model building (and not for its validation), covering then the entire experimental year (from march 2017 to february 2018). New sets of experiments could be carried out to increase the model's predictive strenght.

Conflict of interest, Informed Consent, Human/Animal Rights

No conflicts, informed consent, or human or animal rights are applicable to this study.

References

- [1] A. Mazzelli, D. M. Luzzi, G. Buonanno, A. Cicci, V. Piemonte, G. Iaquaniello, An optimized separation process of microalgal lipidic products by molecular distillation: Techno-economic analysis, *Chemical Engineering Science* 207 (2019) 1187–1195. doi:10.1016/j.ces.2019.07.043.
- [2] A. Mazzelli, G. Buonanno, D. M. Luzzi, A. Cicci, V. Piemonte, G. Iaquaniello, Multi-component extraction process of high added value compounds from microalgae with supercritical CO₂ : A technical and economic study, *Chem. Eng. Res. Des.* 150 (2019) 65–73. doi:10.1016/j.cherd.2019.07.023.
- [3] S. S. Oncel, Microalgae for a macroenergy world, *Renewable and Sustainable Energy Reviews* 26 (2013) 241–264. doi:10.1016/j.rser.2013.05.059.
- [4] S. Bellou, M. N. Baeshen, A. M. Elazzazy, D. Aggeli, F. Sayegh, G. Aggelis, Microalgal lipids biochemistry and biotechnological perspectives, *Biotechnology Advances* 32 (8) (2014) 1476–1493. doi:10.1016/j.biotechadv.2014.10.003.
- [5] X. Wen, K. Du, Z. Wang, X. Peng, L. Luo, H. Tao, Y. Xu, D. Zhang, Y. Geng, Y. Li, Effective cultivation of microalgae for biofuel production: A pilot-scale evaluation of a novel oleaginous microalga *Graesiella* sp. WBG-1, *Biotechnology for Biofuels* 9 (1) (2016) 1–12. doi:10.1186/s13068-016-0541-y.
- [6] T. M. Mata, A. A. Martins, N. S. Caetano, Microalgae for biodiesel production and other applications: A review, *Renewable and Sustainable Energy Reviews* 14 (1) (2010) 217–232. doi:10.1016/j.rser.2009.07.020.
- [7] L. Zhu, Microalgal culture strategies for biofuel production: a review, *Biofuels, Bioprod. Bioref* 9 (2015) 801–814. doi:10.1002/bbb.1576.
- [8] L. Brennan, P. Owende, Biofuels from microalgae-A review of technologies for production, processing, and extractions of biofuels and co-products, *Renewable and Sustainable Energy Reviews* 14 (2) (2010) 557–577. doi:10.1016/j.rser.2009.10.009.
- [9] H. H. Chen, J. G. Jiang, Lipid Accumulation Mechanisms in Auto- and Heterotrophic Microalgae, *Journal of Agricultural and Food Chemistry* 65 (37) (2017) 8099–8110. doi:10.1021/acs.jafc.7b03495.
- [10] R. Slade, A. Bauen, Micro-algae cultivation for biofuels: Cost, energy balance, environmental impacts and future prospects, *Biomass and Bioenergy* 53 (0) (2013) 29–38. doi:10.1016/j.biombioe.2012.12.019.
- [11] L. Meier, R. Pérez, L. Azócar, M. Rivas, D. Jeison, Photosynthetic CO₂ uptake by microalgae: An attractive tool for biogas upgrading, *Biomass and Bioenergy* 73 (2015) 102–109. doi:10.1016/j.biombioe.2014.10.032.
- [12] A. F. Miranda, N. Ramkumar, C. Andriotis, T. Höltkemeier, A. Yasmin, S. Rochfort, D. Wlodkovic,

- P. Morrison, F. Roddick, G. Spangenberg, B. Lal, S. Subudhi, A. Mouradov, Applications of microalgal biofilms for wastewater treatment and bioenergy production, *Biotechnology for Biofuels* 10 (1) (2017) 1–23. doi:10.1186/s13068-017-0798-9.
- [13] K. B. Cantrell, T. Ducey, K. S. Ro, P. G. Hunt, Livestock waste-to-bioenergy generation opportunities, *Bioresource Technology* 99 (17) (2008) 7941–7953. doi:10.1016/j.biortech.2008.02.061.
- [14] W. Zhou, P. Chen, M. Min, X. Ma, J. Wang, R. Griffith, F. Hussain, P. Peng, Q. Xie, Y. Li, J. Shi, J. Meng, R. Ruan, Environment-enhancing algal biofuel production using wastewaters, *Renewable and Sustainable Energy Reviews* 36 (2014) 256–269. doi:10.1016/j.rser.2014.04.073.
- [15] H. Chen, J. Wang, Y. Zheng, J. Zhan, C. He, Q. Wang, Algal biofuel production coupled bioremediation of biomass power plant wastes based on *Chlorella* sp. C2 cultivation, *Applied Energy* 211 (August 2017) (2018) 296–305. doi:10.1016/j.apenergy.2017.11.058.
- [16] M. Ras, J.-P. Steyer, O. Bernard, Temperature effect on microalgae : a crucial factor for outdoor production, *Reviews in Environmental Science and Bio/Technology* 12 (2013) 153–164. doi:10.1007/s11157-013-9310-6.
- [17] A. Mazzelli, A. Cicci, G. Franceschini, F. Caprio, G. Iaquaniello, P. Altimari, F. Pagnanelli, L. Toro, Investigation of effects of nutrients and external parameters on kinetic growth of outdoor microalgal cultivation, *Chemical Engineering Transactions* 64 (2018) 691–696. doi:10.3303/CET1864116.
- [18] T. Katsuda, T. Arimoto, K. Igarashi, M. Azuma, J. Kato, S. Takakuwa, H. Ooshima, Light intensity distribution in the externally illuminated cylindrical photo-bioreactor and its application to hydrogen production by *Rhodobacter capsulatus*, *Biochemical Engineering Journal* 5 (2) (2000) 157–164. doi:10.1016/S1369-703X(00)00054-1.
- [19] W. Blanken, P. R. Postma, L. de Winter, R. H. Wijffels, M. Janssen, Predicting microalgae growth, *Algal Research* 14 (2016) 28–38. doi:10.1016/j.algal.2015.12.020.
- [20] A. J. Klok, J. A. Verbaanderd, P. P. Lamers, D. E. Martens, A. Rinzema, R. H. Wijffels, A model for customising biomass composition in continuous microalgae production, *Bioresource Technology* 146 (2013) 89–100. doi:10.1016/j.biortech.2013.07.039.
- [21] A. Bernardi, A. Nikolaou, A. Meneghesso, B. Chachuat, T. Morosinotto, F. Bezzo, Semi-empirical modeling of microalgae photosynthesis in different acclimation states Application to *N. gaditana*, *Journal of Biotechnology* 259 (2017) 63–72. doi:10.1016/j.jbiotec.2017.08.002.
- [22] F. G. Ación Fernández, F. García Camacho, J. A. Sánchez Pérez, J. M. Fernández Sevilla, E. Molina Grima, A model for light distribution and average solar irradiance inside outdoor tubular photo-bioreactors for the microalgal mass culture, *Biotechnology and Bioengineering* 55 (5) (1997) 701–714. doi:10.1002/(SICI)1097-0290(19970905)55:5<701::AID-BIT1>3.0.CO;2-F.
- [23] A. Solimeno, F. Gabriel, J. García, Mechanistic model for design, analysis, operation and control of microalgae cultures: Calibration and application to tubular photobioreactors, *Algal Research* 21

- (2017) 236–246. doi:10.1016/j.algal.2016.11.023.
- [24] J. Quinn, L. de Winter, T. Bradley, Microalgae bulk growth model with application to industrial scale systems, *Bioresource Technology* 102 (8) (2011) 5083–5092. doi:10.1016/j.biortech.2011.01.019.
- [25] A. Mazzelli, A. Cicci, G. Sed, M. Bravi, Development of a Semi-Theoretical Light Radiation and Photosynthetic Growth Model for the Optimal Exploitation of Wastewaters by Microalgae, *Chemical Engineering Transactions* 64 (2018) 685–690. doi:10.3303/CET1864115.
- [26] Q. Béchet, A. Shilton, B. Guieysse, Modeling the effects of light and temperature on algae growth: State of the art and critical assessment for productivity prediction during outdoor cultivation (2013). doi:10.1016/j.biotechadv.2013.08.014.
- [27] A. Solimeno, R. Samsó, E. Uggetti, B. Sialve, J. P. Steyer, A. Gabarró, J. García, New mechanistic model to simulate microalgae growth, *Algal Research* 12 (2015) 350–358. doi:10.1016/j.algal.2015.09.008.
- [28] W. K. Härdle, L. Simar, *Applied Multivariate Statistical Analysis* (2015). doi:10.1007/978-3-662-45171-7.
- [29] J. Wallace, P. Champagne, G. Hall, Multivariate statistical analysis of water chemistry conditions in three wastewater stabilization ponds with algae blooms and pH fluctuations, *Water Research* 96 (October 2017) (2016) 155–165. doi:10.1016/j.watres.2016.03.046.
- [30] T. F. H. Allen, J. F. Koonce, Multivariate Approaches to Algal Stratagems and Tactics in Systems Analysis of Phytoplankton, *Ecology* 54 (6) (1973) 1234–1246. doi:10.2307/1934186.
- [31] S. J. van Vuuren, A. Pieterse, The use of multivariate analysis as a tool to illustrate the influence of environmental variables on phytoplankton composition in the Vaal River, South Africa, *African Journal of Aquatic Science* 30 (1) (2005) 17–28. doi:10.2989/16085910509503830.
- [32] F. Di Caprio, A. Visca, P. Altimari, L. Toro, G. Iaquaniello, F. Pagnanelli, Two Stage Process of Microalgae Cultivation for Starch and Carotenoid Production, *Chemical Engineering Transactions* 49 (2016) 415–420. doi:10.3303/CET1649070.
- [33] Acea, Tap water composition (2017).
URL <https://www.gruppo.acea.it/it/acqua/ato-2/mappa-qualita-acqua-acea-ato-2>
- [34] F. Di Caprio, P. Altimari, F. Pagnanelli, Integrated biomass production and biodegradation of olive mill wastewater by cultivation of *Scenedesmus* sp., *Algal Research* 9 (2015) 306–311. doi:10.1016/j.algal.2015.04.007.
- [35] F. Di Caprio, P. Altimari, F. Pagnanelli, Effect of Ca^{2+} concentration on *Scenedesmus* sp. growth in heterotrophic and photoautotrophic cultivation, *New Biotechnology* 40 (2017) 228–235. doi:10.1016/j.nbt.2017.09.003.
- [36] A. Instrument, Conversion lux to ppfd (2019).
URL <https://www.apogeeinstruments.com>

- [37] M. Boxwell, The solar electricity handbook (2009).
URL <http://www.solarelectricityhandbook.com/>
- [38] J. Camacho-Rodríguez, A. M. González-Céspedes, M. C. Cerón-García, J. M. Fernández-Sevilla, F. G. Acién-Fernández, E. Molina-Grima, A quantitative study of eicosapentaenoic acid (EPA) production by *Nannochloropsis gaditana* for aquaculture as a function of dilution rate, temperature and average irradiance, *Applied Microbiology and Biotechnology* 98 (6) (2014) 2429–2440. doi:10.1007/s00253-013-5413-9.
- [39] R. Bosma, E. van Zessen, J. Reith, J. Tramper, R. Wijffels, Prediction of Volumetric Productivity of an Outdoor Photobioreactor, *Aquaculture* 97 (3) (2006) 1108–1120. doi:DOI10.1002/bit.21319.
- [40] Z. Cheng-Wu, O. Zmora, R. Kopel, A. Richmond, An industrial-size flat plate glass reactor for mass production of *Nannochloropsis* sp.(Eustigmatophyceae), *Aquaculture* 195 (3) (2001) 35–49. doi:10.1016/S0044-8486(00)00533-0.
- [41] G. Chini Zittelli, L. Rodolfi, N. Biondi, M. R. Tredici, Productivity and photosynthetic efficiency of outdoor cultures of *Tetraselmis suecica* in annular columns, *Aquaculture* 261 (3) (2006) 932–943. doi:10.1016/j.aquaculture.2006.08.011.
- [42] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems* 2 (1987) 37–52. doi:10.1016/0169-7439(87)80084-9.