Biochemical Engineering Journal 233 (2026) 110249

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular article

Integration of flow cytometry and chemometric methods to monitor nitrogen-starvation-induced stress in *Chlorella sorokiniana*

Eugenio Sandrucci [a,b], Riccardo Boemio [a], Francesca Pagnanelli [a], Federico Marini [a], Fabrizio Di Caprio [a,*]

[a] Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, Rome 00185, Italy
[b] Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, UK

ARTICLE INFO

Keywords:
Chemometrics
Single-cell analysis
Bioprocesses monitoring
Nitrogen-starvation
Carbohydrate accumulation
Chlorella sorokiniana

ABSTRACT

Nitrogen-starvation is a widely exploited stress condition for inducing the accumulation of valuable molecules in microalgae. However, because two-stage batch cultivation is the most commonly adopted configuration, accurately predicting when cells are in nitrogen-starvation remains challenging and strongly affects process performance. This study aims to develop an innovative, rapid and reliable method to determine when cells are in nitrogen-starvation condition, and the associated changes in microalgal biochemical profiles. The method is based on flow cytometric analysis requiring no sample pretreatment other than dilution, followed by chemometric data processing using Partial Least Squares Discriminant Analysis (PLS-DA). The model was developed using 50 samples collected under nitrogen-replete (N^+) and 52 under nitrogen-starvation (N^-). It efficiently discriminated cells exposed to N^- from N^+, achieving excellent classification performance (AUC = 0.98) with cross-validated sensitivity, specificity and precision between 0.86 −0.90. PLS-DA was also applied to classify changes in carbohydrate, chlorophyll and carotenoid content, resulting from nitrogen-starvation. The model achieved AUC = 0.95 for carbohydrates and chlorophylls, and 0.93 for carotenoids. Cross-validated sensitivity, specificity and precision were 0.81 −0.95 for carbohydrates and chlorophylls, and 0.69 −0.87 for carotenoids. Classifications based on metabolite content showed greater overlap than those based on N^+ versus N^-. The proposed method offers a promising tool for monitoring nitrogen-starvation in industrial-scale bioreactors, with potential application in predicting carbohydrate accumulation or chlorophyll reduction.

1. Introduction

Microalgae are photosynthetic microorganisms that have attracted increased scientific and industrial interest due to the environmental benefits associated with using them for the production of food, feed, biomaterials and other bio-based products [1]. Nevertheless, their large-scale application in many economic sectors is limited by the high production costs, primarily attributed to low energy efficiency of the industrial facilities [2]. Significant improvements could be achieved thorough the adoption of advanced reactor control systems, which enable the optimization of operating conditions and thereby minimize the energy demand of bioreactors [3]. To have efficient control systems, efficient reactor monitoring methods are mandatory. Therefore, it is important to develop and implement monitoring methods that can provide reliable information about the physiological state of microalgal cells and about environmental parameters inside the reactor. Microalgae are usually cultivated in batch or in two-phase cultivation systems, in which the physiological state of microalgal cells change over time. These systems are widely used because many microalgae are cultivated for the production of specific molecules that accumulate only under stressful conditions, such as carotenoids [4], fatty acids [5], and polysaccharides [6]. Nitrogen depletion is the most widespread method used for inducing stressful conditions in microalgae cultures. It is easy to obtain, and its depletion has strong effects on the metabolisms because it stops protein and nucleic acids synthesis, inducing cell exit from the cell cycle to a G_0 phase, and accumulation of reserve compounds as starch and triacylglycerols [7]. The arrest of the cell cycle leads to a decrease in biomass productivity, which occurs alongside the increased accumulation of target compounds. Verifying, prior to harvesting, whether microalgae are actually experiencing N-starvation stress is crucial for determining the optimal harvesting time. In the absence of N-starvation, the accumulation of target compounds may still be insufficient, despite

* Corresponding author.
E-mail address: fabrizio.dicaprio@uniroma1.it (F. Di Caprio).

https://doi.org/10.1016/j.bej.2026.110249
Received 25 March 2026; Received in revised form 8 May 2026; Accepted 10 May 2026
Available online 12 May 2026

high biomass productivity. Conversely, if cells remain under N-starvation condition for an excessively long period, the resulting decline in biomass productivity may significantly compromise process sustainability [6,7]. It is therefore of vital importance to have monitoring tools to monitor whether cells are in optimal or stressful conditions, since it allows to determine the right moment for biomass harvesting. Ideally, frequent analysis of nitrogen concentration in the culture medium and of the biomass composition with conventional chemical methods would be the best way. But these methods are slow, labor intensive and costly. Some molecules, such as lipids, can be analyzed quite easily by alternative methods based on fluorescence spectroscopy, by exploiting fluorescent dyes that specifically bind to lipids, such as BODIPY and Nile Red [8,9]. Fluorescence spectroscopy can allow fast, accurate and sensitive analyses, and, if integrated with flow cytometry, can also allow online monitoring of bioreactors [10]. However, lipids are usually accumulated only in the late phase of starvation conditions (typically after 4–5 days), while the primary compounds typically stored by microalgae immediately after the starting of nitrogen starvation is starch [11]. The accumulation of carbohydrates is not currently analyzable with online monitoring systems because fluorescent dyes are not commercially available. Other methods, such as Raman flow cytometry, although promising for this application, are currently only feasible with custom-made equipment [12]. A possible solution could be the application of conventional flow cytometry combined with chemometric methods. Although carbohydrate (starch) accumulation in microalgae under nitrogen starvation is not expected to be directly associated with a specific signal measured by flow cytometry, the complete set of recorded parameters may contain indirect correlations that can be identified through multivariate analysis. This integrated approach may enable discrimination between cells under optimal growth conditions (with nitrogen in the medium) and stressed cells subjected to nitrogen starvation.

Classification using chemometric models is one of the most effective approaches for analyzing complex data in fields such as chemistry, food science, and environmental studies. These models allow for the assignment of samples to predefined categories (classes) based on multivariate data obtained from different analytical techniques. Supervised multivariate statistical methods, including Partial Least Squares Discriminant Analysis (PLS-DA), are commonly employed to develop predictive models capable of recognizing patterns and discriminating between groups with high accuracy [13]. Due to their ability to handle high-dimensional data and their versatility across various application domains, chemometric classification models have become indispensable tools for the analysis and interpretation of experimental data [14,15]. PLS-DA is a multivariate classification method widely applied in the analysis of high-dimensional biological data, including flow cytometry, where it is used to discriminate among predefined cell populations or clinical phenotypes based on complex multichannel fluorescence data [16,17]. Flow cytometry datasets are typically characterized by a high number of correlated variables (e.g., fluorescence intensities, scatter signals) and a relatively low number of samples or individuals, making PLS-DA an attractive choice due to its ability to handle multicollinearity and reduce dimensionality through latent variables. Flow cytometry has been widely applied to microalgae to rapidly assess the physiological status of individual cells. However, a major limitation of this technique lies in the development of reliable models capable of extracting meaningful information from complex datasets. In flow cytometry applications, PLS-DA has been successfully employed to classify immune cell subpopulations, distinguish between healthy and pathological samples, and detect subtle phenotypic changes in response to treatment. [18–20]. Recent works have also highlighted the utility of sparse PLS-DA (sPLS-DA) in biotechnological applications, where it facilitates the identification of the most informative markers contributing to class separation, enhancing interpretability and reducing redundancy [21–23].

The aim of this study is to discriminate, stressed and not-stressed microalgae, as result of nitrogen starvation, by a simple and quick flow cytometric analytical method, requiring minimum pre-treatment (only sample dilution). To this aim, flow cytometric analysis was performed on microalgae samples after just a dilution, without any staining pre-treatment, to minimize costs, labor and time of the analysis, making it more applicable for online bioreactor monitoring [24].

2. Materials and methods

2.1. Microalgae cultivation and sample collection

Chlorella sorokiniana SAG 211/8k was cultivated into two 500 mL (h = 35 cm; d = 5 cm) column glass photobioreactors (PBRs) under controlled conditions. Cultures were grown in modified M-8 medium (described in a previous work [24]), under either nitrogen-replete (N^+) conditions, supplemented with KNO_3, and nitrogen-deprived (N^-) conditions, without KNO_3. These conditions were chosen to induce distinct physiological states (stressed and not-stressed) and to generate a dataset suitable for PLS-DA. The culture medium was maintained at pH 6.7 ± 0.5 (by CO_2 supply). Cultures were inoculated at an initial biomass concentration of 0.5 g/L (dry weight). Illumination was provided by LED panels (GROWSTAR L-QB1, sunlight full spectrum) at a photon flux density of 500 μmol m^{-2} s^{-1} , by positioning the lamps at 42 cm from the reactors. The temperature was maintained at 28 °C via a thermostatic water bath. Continuous aeration and mixing was provided by sterile-filtered air (1 L/min) supplemented with CO_2 (20 mL/min), which maintained stable the pH and served as the sole inorganic carbon source.

Microalgae were cultivated in these PBRs over a total period of 8 months. The reactors were operated in sequential batch mode, with a total of 26 different batches operated with the two PBRs. Each batch lasted between 4 and 30 days. Overall, 14 batches were carried out under N^+ condition and 12 under N^- conditions. At the end of each batch, microalgae were harvested by centrifugation (3000 g x 5 min), the supernatant was removed, the pellet was washed, and cells were resuspended in fresh medium. Microalgae cultivated under N^+ conditions were used to inoculate subsequent batches under either N^+ or N^- conditions, whereas microalgae from N^- batches were used exclusively to inoculate a new batch under N^+ condition. This experimental design, with a separation between N^+ and N^- obtained by harvesting algae and resuspending them in a new media, was intended to ensure a precisely defined distinction between N^+ and N^- conditions. During cultivation water loss due to evaporation was compensated by adding distilled water. Samples were collected from different independent batches at various cultivation times in order to obtain representative samples from different phases in nitrogen-deprived and nitrogen-replete conditions.

2.2. Chemical analysis

For each sample collected and processed by flow cytometric analysis, an aliquot was centrifuged (3000 g x 5 min) and rinsed with distilled water. Pellets were freeze-dried, manually ground and stored at −20 °C until analysis. Pigments were extracted from lyophilized biomass using 96% (v/v) ethanol, at a ratio of 10 mL per gram of dry biomass equivalent inside glass tubes, heated at the boiling temperature. After 15 min, the suspension was centrifuged (3000 g x 5 min), the solvent removed, and the pellet was suspended again in fresh ethanol for a new extraction cycle. The procedure was repeated until the solvent was colorless. Absorbance of the extract was read at 470, 649, and 665 nm with a spectrophotometer (Varian Cary 50 UV-Vis) to quantify total carotenoids, chlorophyll b, and chlorophyll a, respectively with equation reported by Lichtenthaler [25].

The residual biomass was used for the determination of carbohydrate content. This analysis was performed by acid saccharification with 4% H_2SO_4 in autoclave, following the protocol described by NREL [26], and using the phenol-sulfuric acid method (Dubois) [27] for total sugar

quantification, using a standard curve obtained with pure glucose. Biomass concentration inside PBRs was determined by filtering a known volume through 0.7 μm glass fiber filter, which was then dried at 105 °C.

2.3. Flow cytometric analysis

For flow cytometric analysis, microalgal samples collected from PBRs were divided into three aliquots (C1, C2 and C3). C1 (fixed cells) was diluted in PBS (pH 7.4) with 1% glutaraldehyde, to $OD_{750} = 0.50$. C2 was diluted in PBS (pH 7.4) to $OD_{750} = 0.50$. C3 was obtained by further diluting C2–1:50 in PBS ($OD_{750} = 0.01$). Unfixed samples (C2 and C3) were analyzed quickly as soon as they were collected, while fixed sample (C1) were analyzed after being stored in the fridge at 4 °C for a period between 1 – 5 months. Samples were analyzed using an Attune™ NxT flow cytometer with a 488 nm laser. All samples cells were measured at 100 μL/min. C1 and C2 were acquired for 10,000 events; C3 for 2000 events. Parameters recorded included Forward Scatter (FSC), Side Scatter (SSC), and two fluorescence channels: BL3 (chlorophyll autofluorescence, 695/40 nm) and BL4 (far-red autofluorescence, 780/60 nm). For each one, the height (H), width (W) and area (A) of the signal was recorded. For each sample, the sample cell mean value of the fluorescent or scatter signal was used for data elaboration with chemometric methods. Detector voltages were set to 80 mV (FSC), 260 mV (SSC), and 300 mV (BL3 and BL4). To exclude noise, a threshold was applied to record only events with BL3 signal > 200 mV, in order to include only cells with relevant autofluorescence given by chlorophyll.

2.4. PLS-DA model

PLS-DA is an extension of Partial Least Squares Regression (PLSR) adapted for categorical outcomes. It models the relationship between a predictor matrix **X** (samples × variables) and a response matrix **Y** (samples × classes), often encoded using dummy variables. In fact when the problem only involves two classes, as in the present study, the response is a binary-coded vector, **y**, where the value of 1 is associated to one category and the value of 0 to the other.

The method projects **X** onto a reduced set of latent variables that are constructed to maximize covariance with **y**. The projection of X into the latent space is defined by the PLS weight vectors **R**, according to:

$$\mathbf{T} = \mathbf{XR} \tag{1}$$

where **T** is the score matrix containing the latent variables, and **R** is the matrix of PLS weights that define the directions in the original variable space along which covariance between X and y is maximized.

For two-class PLS-DA, the response vector is modeled as a linear function of the latent scores:

$$\mathbf{y} = \mathbf{Tq} + \mathbf{f} \tag{2}$$

where **q** is the vector of Y-loadings linking the latent variables to the response, and **f** is the residual vector. The latent variables are extracted such that the covariance between **T** and **y** is maximized:

By combining the score definition with the regression model, the predicted response can be expressed directly as a function of the original predictor variables:

$$\hat{\mathbf{y}} = \mathbf{Xb} \tag{3}$$

where **b** is the vector of PLS-DA regression coefficients, obtained from the weights, and the Y-loadings:

$$\mathbf{b} = \mathbf{Rq}. \tag{4}$$

This formulation makes explicit that class prediction is driven by the original flow cytometry variables, rather than solely by the latent scores.

For a new observation $\mathbf{x}_{new}$, the predicted response is computed as:

$$\hat{\mathbf{y}}_{new} = \mathbf{x}_{new}\mathbf{b}. \tag{5}$$

Class membership is then assigned by comparing $\hat{\mathbf{y}}$ to a predefined decision threshold. In this framework, the PLS weights **R** define the construction of the latent variables, while the regression coefficients **b** quantify the contribution and direction of each original predictor to class discrimination, enabling both classification and biological interpretability.

Its ability to reduce data dimensionality while preserving class discrimination is particularly useful when dealing with complex datasets containing numerous variables. In addition, the analysis of weights and PLS-DA regression coefficients allows for the identification of variables that contribute most to class discrimination, thus providing interpretative insight into the model. The regression coefficients quantify the magnitude and direction of each variable's contribution to the predicted response, providing insight into the variables driving class separation in the model. Typically, variables with large absolute coefficient values are considered highly influential. These values are often visualized using coefficient plots, which graphically represent the relative impact of each variable on the model.

PLS-DA models were developed using the full dataset of 102 samples, without removal of any observations. Prior to modelling, data were preprocessed by autoscaling (mean-centering and variance scaling). The final classification model was built using 8 latent variables (LVs), selected as the optimal balance between information retention and overfitting.

Internal validation was performed using Venetian Blinds cross-validation with 3 segments, in which the dataset is divided into three interleaved subsets. Two segments are iteratively used for model calibration, while the remaining segment serves as the validation block; this rotation is repeated until all groups have been used for validation once. Classification scores from each iteration were averaged to derive final performance metrics.

Beyond interpretability, the predictive performance of the PLS-DA models was assessed using standard classification metrics derived from the confusion matrix, including sensitivity (true positive rate), specificity (true negative rate), and precision (positive predictive value). These metrics provide a quantitative evaluation of the model's ability to correctly classify samples into their respective classes, based on the empirically defined thresholds. Such evaluation is essential to ensure that the model not only identifies relevant features, but also performs robustly in distinguishing between biologically meaningful groups.

The metrics are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Where TP = True Positives (Class 1 correctly predicted), TN = True Negatives (Class 2 correctly predicted), FP = False Positives (Class 2 misclassified as Class 1) and FN = False Negatives (Class 1 misclassified as Class 2).Classification accuracy was calculated as the ratio between correctly assigned samples (TP + TN) and the total number of samples used for prediction, whereas error rate was computed as 1 – accuracy, corresponding to the fraction of misclassified samples (FP + FN) over the total. Both metrics were derived from the confusion matrix obtained for each cross-validation iteration and subsequently averaged.

3. Results

3.1. Analysis of carbohydrates and pigments

The dataset used for this study was derived from 26 different batches of *C. sorokiniana* cultures grown under two different conditions: a

complete M8 medium containing all essential nutrients (N^+) and a modified M8 medium under nitrogen-starvation conditions (N^-), namely: without KNO_3. In total, 102 samples were collected, comprising 50 from cultures grown in N^+ and 52 from N^- conditions. Inside PBRs biomass concentration ranged between 0.5 and 4 g/L, representative of typical concentrations achieved inside PBRs. Cell concentration inside PBRs was determined by flow cytometric analysis and ranged between $15 \cdot 10^6$ and $390 \cdot 10^6$ cells/mL. This sampling strategy was designed to be representative of the variability of microalgae real samples that would be collected under different environmental conditions, as for instance from bioreactors used for the production of microalgal lipids (e.g., omega-3 fatty acids for food and feed, or triacylglycerols for biodiesel production) [5,28]. The analysis of data obtained from independent reinoculations of *C. sorokiniana* under both N^- and N^+ culture conditions enabled the identification of recurring patterns across different growth experiments.

Box plot analysis of carbohydrate, chlorophyll, and carotenoid content provided insights into the metabolic profile of *C. sorokiniana* under different nitrogen regimes (Fig. 1). Carbohydrate content showed a significant increase ($p = 6 \cdot 10^{-26}$) in nitrogen-starved cultures (N^-), with a median of 47.68% and a mean of 49.42%, compared to a median of 26.13% and a mean of 23.62% in nitrogen-replete conditions (N^+) (Fig. 1A). The interquartile range (IQR) for N^- samples was narrower, indicating a more uniform accumulation of carbohydrate reserves in response to nitrogen deprivation. Cultures grown under N^+ conditions exhibited lower carbohydrate levels and greater variability ($p = 0.01$), suggesting more heterogeneous metabolic activity under nutrient replete conditions, possibly due to higher metabolic adaptations to variations in environmental factors (e.g., light supply per cell, concentration of other nutrients). In contrast, N-starvation induces cells to exit the cell cycle, which may lead to a slowdown of metabolism toward a more homogeneous state. The distribution of samples shifted toward higher concentrations of carbohydrates in N^-, with a greater proportion of values exceeding 55% w/w. This trend is a phenomenon induced by metabolic reprogramming in *C. sorokiniana*, due to the stop of the cell cycle induced by nitrogen deficiency. Cell cycle stops because when cells have no nitrogen they cannot proceed with protein and nucleic acid synthesis, exit from cell cycle and enter in G_0 phase [7]. Since cell duplication is stopped, the excess in energy and organic carbon resulting from photosynthetic activity is directed towards the accumulation of reserve compounds, mainly in the form of starch [29].

As is shown in Fig. 1B, carotenoid content was higher in N^+ cultures ($p = 2 \cdot 10^{-5}$), with a mean of 4.51 mg/g and a median of 4.07 mg/g, although accompanied by considerable higher variability ($p = 1 \cdot 10^{-12}$). In N^- samples, both the mean (2.88 mg/g) and median (2.87 mg/g) were reduced. This decline could be the result of a long-term stress response, in which the capacity for carotenoid biosynthesis might be reduced, or, alternatively, can be a result of dilution of carotenoids inside microalgal cells, resulting from the accumulation of starch and increase in cell size.

Chlorophyll levels decreased markedly ($6 \cdot 10^{-11}$) under N^- (Fig. 1C). In N^+ conditions, the median and mean chlorophyll content were 24.7 mg/g and 26.89 mg/g, respectively, with a broader data dispersion ($p = 6 \cdot 10^{-13}$) indicative of higher physiological heterogeneity, likely tied to higher variability in light exposition per cell under rapid growth phase. Under N^-, the median dropped to 11.52 mg/g and the mean to 10.59 mg/g, accompanied by reduced variability, indicating a physiological response to a more uniform condition. This was likely due to the decline in chlorophyll content under N^- conditions to a minimum level, a well-established phenomenon [30]. This can be attributed in part to dilution, due to carbohydrate accumulation, and in part to the degradation of photosynthetic complexes, which contain nitrogen, allowing its recycling for protein synthesis and enabling the cell to adapt to the stress conditions encountered. Although this adaptation leads to decreased photosynthetic capacity, it represents a strategic response to environmental stress.

The average dry weight per cell across all samples ranged between 10 and 50 pg/cell. Despite significant biochemical differences between N^+ and N^- conditions, mean cell dry weight were comparable: 23.6 pg/cell for N^+ and 23.4 pg/cell for N^- cultures.

Overall, data on biochemical composition (carbohydrates, chlorophylls and carotenoids) showed that the samples collected in this study from cultures under N^+ and N^- conditions are related to substantial metabolic reprogramming of *C. sorokiniana* in response to nitrogen availability. This confirms nitrogen limitation as a strategy to modulate algal biochemical composition, with significant applications in bioenergy, bioethanol production, and the food industry. The variation in the biochemical composition observed in such samples is expected to result in a variation of signals detected by flow cytometry. Indeed, chlorophyll has red autofluorescence properties when excited with blue light [31]. To the best of our knowledge, there is currently no information in the scientific literature about the possibility to correlate intracellular carbohydrate content with flow cytometry-derived signals. However, as intracellular carbohydrate accumulation affects cellular internal complexity, it may alter light scattering and other flow cytometry-derived signals in a non-intuitive manner. These preliminary data were therefore further analyzed using chemometric approaches methods to infer the physiological status of microalgae from rapid flow

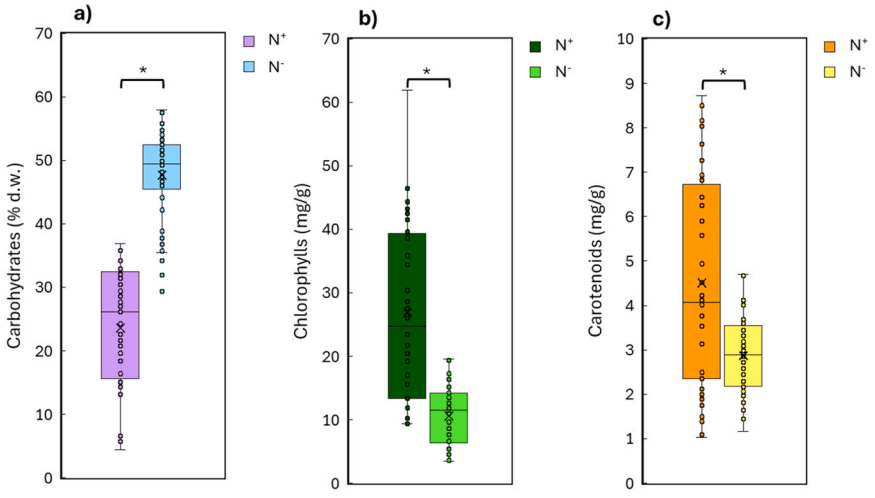

Fig. 1. Box showing carbohydrate content (A), total chlorophyll content (B) and total carotenoid content (C) in samples grown under nitrogen rich (N^+) and in nitrogen deplete (N^-) conditions. * indicates statistically significant difference ($p < 0.01$, *t*-test).

cytometric measurements.

Proteins and lipids are other major cellular components known to change as a result of N-starvation; however, they were not analyzed in this study. Lipid content was not analyzed because previous experiments conducted in our laboratory have shown that *C. sorokiniana* has only a limited capacity for lipid accumulation ($\leq$ 20% w/w) and exhibits such variations only after a prolonged period from the onset of N-starvation (typically after 4–5 days). Protein content was also not considered, as its variation mainly results from dilution due to the rapid accumulation of intracellular carbohydrates, concomitant with the cessation of protein synthesis and cell division. Protein content is typically inversely proportional to carbohydrate content; therefore, its analysis was expected to provide "redundant" information.

3.2. Classification for stressed (N-starved) and not-stressed cells

To distinguish *C. sorokiniana* cells grown under stress condition induced by N-starvation (N^-, class 2) from cells grown in medium fully replete of nutrients (N^+, class 1), a PLS-DA model was developed using flow cytometry data. The predictors included scatter parameters and fluorescence channels (FSC, SSC, BL3, BL4), each represented by their area (A), height (H), and width (W) signal components.

PLS-DA analysis was performed for the three different protocols used, namely: C1 (fixed sample at $OD_{750} = 0.5$), C2 (unfixed sample at $OD_{750} = 0.5$) and C3 (unfixed sample with $OD_{750} = 0.01$).

For all the three protocols, the model exhibited excellent discriminative power. The predicted response plot (Fig. 2 and supplementary material) always showed a clear separation between the two classes, with positive values corresponding to class 1 (N^+) and negative values corresponding to class 2 (N^-).

To elucidate the structure of the PLS-DA model and identify the variables contributing most strongly to class separation, the regression coefficients derived from the full model were examined. These coefficients describe the magnitude and direction of each predictor's contribution to the discrimination between classes, reflecting how variations in individual variables influence the predicted class response. Coefficient plots were used to visualize variable importance, providing a complementary tool to the assessment of model discrimination and class separation given by the plots of the predicted response values. Together, these representations enabled interpretation of both the variables driving the model and its overall classification behavior, supporting a rational reduction of model complexity without loss of predictive performance.

Analysis of the model coefficients revealed that the variables positively or negatively associated with class 1 (N^+) and class 2 (N^-) varied slightly among the different protocols used for sample processing before flow cytometric analysis (Fig. 3).

FSC-W was highly positively correlated with class 1 (N^+) for C1 and C3, but slightly negatively associated for C2. BL4-H was positively associated with class 1 for all protocols C1, C2 and C3. BL3-A and BL3-W were always negatively associated to class 1 for all protocols C1, C2 and C3, while FSC-A was negatively associated to class 1 for C3 and C2, but slightly positively associated for C1. These data show that the fixation phase with glutaraldehyde and cell concentration of analyzed samples affect the flow cytometric response for different parameters. Glutaraldehyde denatures cell proteins and can therefore affect cell morphology and roughness. It is also known to affect cell fluorescence signals [32]. Instead, cell concentration can affect the likelihood of multiple cells being analyzed simultaneously at the interrogation point.

To enable a more rigorous and objective assessment of model performance, Receiver Operating Characteristic (ROC) curves were generated separately for each PLS-DA model, allowing direct comparison of discriminant ability across the different preparation protocols. The Area Under the Curve (AUC) values derived from these analyses provided a quantitative measure of the models' classification capabilities. ROC curves (Fig. 4 and supplementary material) confirmed the high accuracy of the model, for all the three protocols, with AUC values very similar, between 0.98 and 0.99, indicating outstanding predictive performance of the three protocols.

Finally, confusion matrices (Table 1 and Supplementary Material) were computed for the full dataset under both training (autoprediction) and cross-validation conditions for each of the three preparation protocols. These matrices represent a fundamental tool for evaluating the performance of classification models, particularly in the context of cross-validation, where model robustness and generalization capacity are critically assessed. These matrices summarize the number of correctly and incorrectly classified instances, providing a detailed view of classification behavior across the two classes (N^+ and N^-). The metrics allow the calculation of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). On the training data, the model achieved sensitivity, precision and specificity between 92% and 100%, for the three protocols, for both N^+ and N^-. Higher values were obtained for protocol C3 (Table 1, supplementary material). Cross-validation results also demonstrated strong and comparable performance for all the three protocols. Sensitivity values were between 0.86 and 0.96 for class 1 (N^+) and between 0.88 and 0.92 for class 2 (N^-). Specificities values were between 0.88 and 0.92 for class 1 (N^+) and between 0.86 and 0.96 for class 2 (N^-). Precision values were between 0.89 and 0.92 for class 1 (N^+) and between 0.87 and 0.96 for class 2 (N^-).

Accuracy on detecting N^+ and N^- cells was 0.92 for training (0.08

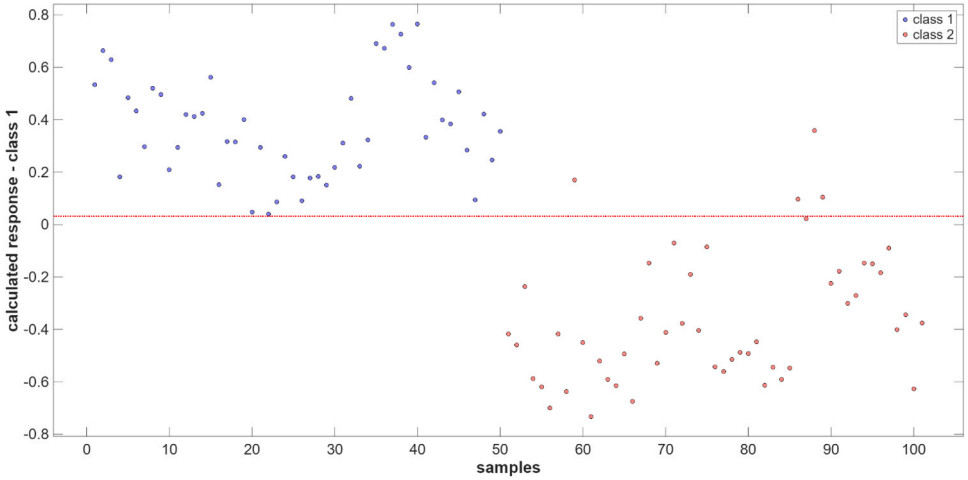

Fig. 2. Predicted response plot for class 1 (N^+) for data obtained with C3 protocol. Each dot represents a sample. Values above the red threshold line are classified as class 1 (N^+), while those below are assigned to class 2 (N^-), showing clear class separation.

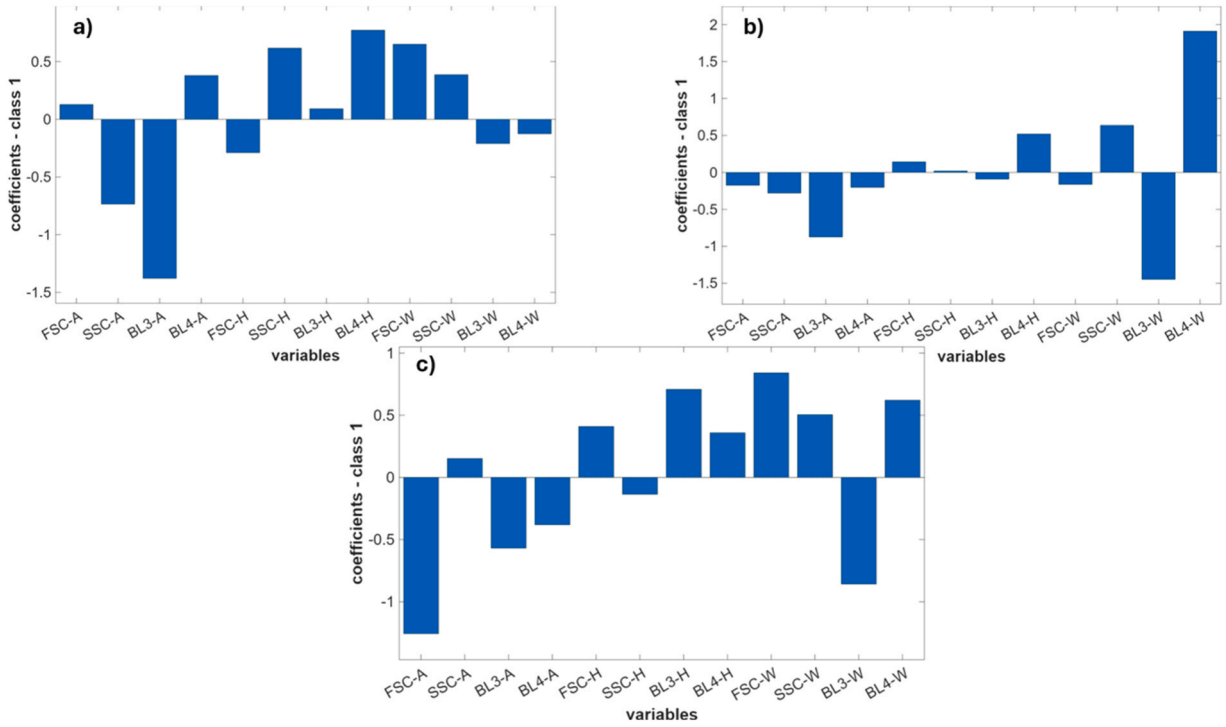

Fig. 3. Plots of the regression coefficients of PLS-DA models calculated on for data obtained with protocol C1 (A), C2 (B) and C3 (C). Bars represent the weight of each cytometric variable in the discrimination between cells grown under nitrogen-replete (N^+) and nitrogen-starvation (N^-) condition. Positive values indicate a stronger association with class 1 (N^+) while negative values indicate inverse contribution.

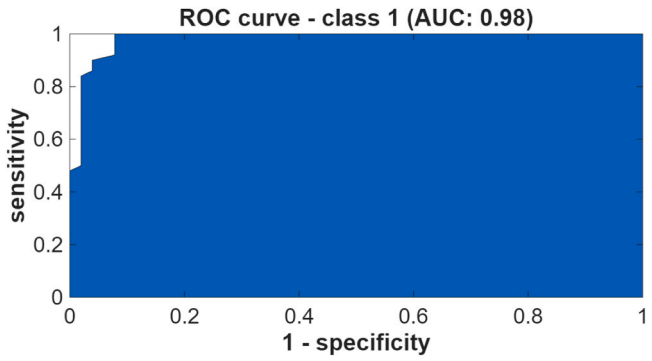

Fig. 4. ROC curve for the PLS-DA model calculated on data obtained with C3 protocol. The model results in an AUC of 0.98, indicating excellent classification performance.

error rate) and decreased slightly to 0.89 (0.11 error rate) in cross validation (Table 1). Overall, the PLS-DA model proved to be highly effective in classifying cells according to nitrogen availability, showing for the first time as flow cytometry can be a fast and reliable approach for discrimination microalgae cells under nitrogen starvation stress from cells growing in replete nutrient medium, without the need of any staining pre-treatment. The method is effective using both unfixed fresh cells, if analysis is conducted as soon as they are collected, or using fixed cells, if samples have to be stored for weeks or months. The method remains effective even when applied at different dilutions (OD = 0.01 and OD = 0.5), enabling the monitoring of bioreactors after a simple dilution within this range. This range of dilution considered is sufficiently broad to allow the analysis of virtually any microalgal cultivation. Some previous studies about microalgae status classification were performed using PLS-DA applied on data collected with other analytical techniques. PLS-DA applied to spectra collected by FT-IR spectroscopy demonstrated to be effective on predicting both nitrogen limited and phosphorous limited microalgae cells [33]. However, the FT-IR method requires dried biomass, that would take a few hours per sample preparation. Yet, PLS-DA was reported to be effective on classifying microalgal samples as viable or non-viable from spectra data collected with Raman spectroscopy analysis [34]. However, Raman spectroscopy requires cell coating on Petri dished covered with L-lysine polymer, and data acquisition with a microscope, resulting in a labor intensive and slow method. In fact, each acquisition of single spectra, for every single cell, takes about 2–3 min. Globally, such a method requires approximately 2–3 h to analyze about 50 cells in a single sample. The flow cytometric analysis described here requires only sample dilution, which takes just a few seconds, and allow the analysis of 2000 - 10,000 cells in 5 – 30 s. Such dilution can be performed in a few minutes or even

Table 1
Confusion matrices and classification metrics for the PLS-DA model, reported for both the training and cross-validation phases. Class 1 corresponds to nitrogen-replete samples (N^+), and class 2 to nitrogen-starvation samples (N^-). Metrics include sensitivity (Sens.), specificity (Spec.) and precision (Prec.). Data shown are obtained with C3 protocol.

		Confusion Matrix			Classification Metrics			
Training		Class 1	Class 2	n.a.	Sens.	Spec.	Prec.	Acc.
	Class 1	50	0	0	1.00	0.92	0.93	0.92
	Class 2	5	47	0	0.92	1.00	1.00	
Cross Validation	Class 1	43	7	0	0.86	0.90	0.90	0.89
	Class 2	5	47	0	0.90	0.86	0.87	

automated in an online monitoring system, as previously described for other flow cytometric analysis [35]. Both the methods previously reported are slower than the flow cytometric protocol here described. Flow cytometry may require a higher initial investment to purchase the instrument, compared to some other methods; however, because it requires less labor, it can make the analysis cheaper than more labor-intensive techniques, with the cost benefit depending on the number of samples processed [36].

3.3. Classification based on metabolite contents

The primary aim of this approach is to assess the discriminative capability of the model in classifying samples exhibiting varying concentration levels of biochemical components. Two-stage cultivation of microalgae, with a second stressing stage in nitrogen starvation, is generally employed to induce a metabolic switch to stimulate the accumulation of specific target compounds as carbohydrates, lipids and pigments [6,37]. A quick analytical tool that allows to quickly assess directly if the metabolic switch is active in microalgae is of high practical interest for industrial process monitoring. This classification approach reflects the view that, for the intended application, a chemometric model need not necessarily achieve optimal quantitative accuracy, but must reliably distinguish between samples with different biochemical states resulting from nitrogen starvation.

For each target analyte considered in this study — total carbohydrates, total chlorophylls, and total carotenoids – continuous reference values were binarized into two classes based on empirically derived thresholds. Reference concentrations for these compounds were determined via established analytical techniques (i.e., the Dubois colorimetric method for carbohydrates and spectrophotometric assays for chlorophylls and carotenoids).

The selection of class thresholds for the PLS-DA model was based on experimental data and corresponding mean values. The empirical thresholds applied were as follows: 35% (w/w) of cellular content for total carbohydrates, 4 mg/g for carotenoids, and 15 mg/g for chlorophylls. These thresholds were chosen as values in between the samples coming from N^+ and N^- conditions, with the idea to have different metabolic profiling as result of nitrogen starvation. In particular, for carbohydrates the threshold chosen was above the 90° percentile of N^+ samples and below the 90° percentile of N^- samples. This threshold also reflects the value attained within 1 day of nitrogen starvation as result of starch accumulation [6,11]. For carotenoids it was the opposite, while for chlorophylls it was above the 90° percentile of N^- samples and below the 50° percentile of N^+ samples.

Notably, across all constructed PLS-DA models, those obtained with protocol C3 gave better results as compared to protocols C1 and C2. For this reason, data discussed in the following are all regarding results obtained from PLS-DA analysis of samples processed with protocol C3. All the corresponding figures and table for protocols C2 and C1 are reported in supplementary material.

C3 corresponds to an acquisition rate between 50 and 200 events/s, whereas C2 and C1 correspond to higher rates, between 500 and 5000 events/s. Such higher rates could reduce performance as a result of cell overlap. The fluorescence parameter BL4-H consistently exhibited the largest absolute coefficient (Fig. 5), thereby confirming its status as the principal discriminant variable and a key marker for class prediction within the studied system. BL4-H measures the height of the peak of the emitted fluorescence signal in the wavelength range of filter BP 780/60 nm, in the near-infrared region. This region is slight out from the emission peaks of chlorophyll, but it can include a small fraction of the emission peak tail of chlorophyll a [38], which could be enough sensitive to allow a correlation with biochemical composition. BL4 may perform better than BL3 (BP 695/40 nm) because the latter corresponds to the emission peak of chlorophyll, and the high chlorophyll content in Chlorella may saturate the signal, making the analysis less sensitive to concentration variations. The height of the signal may perform better

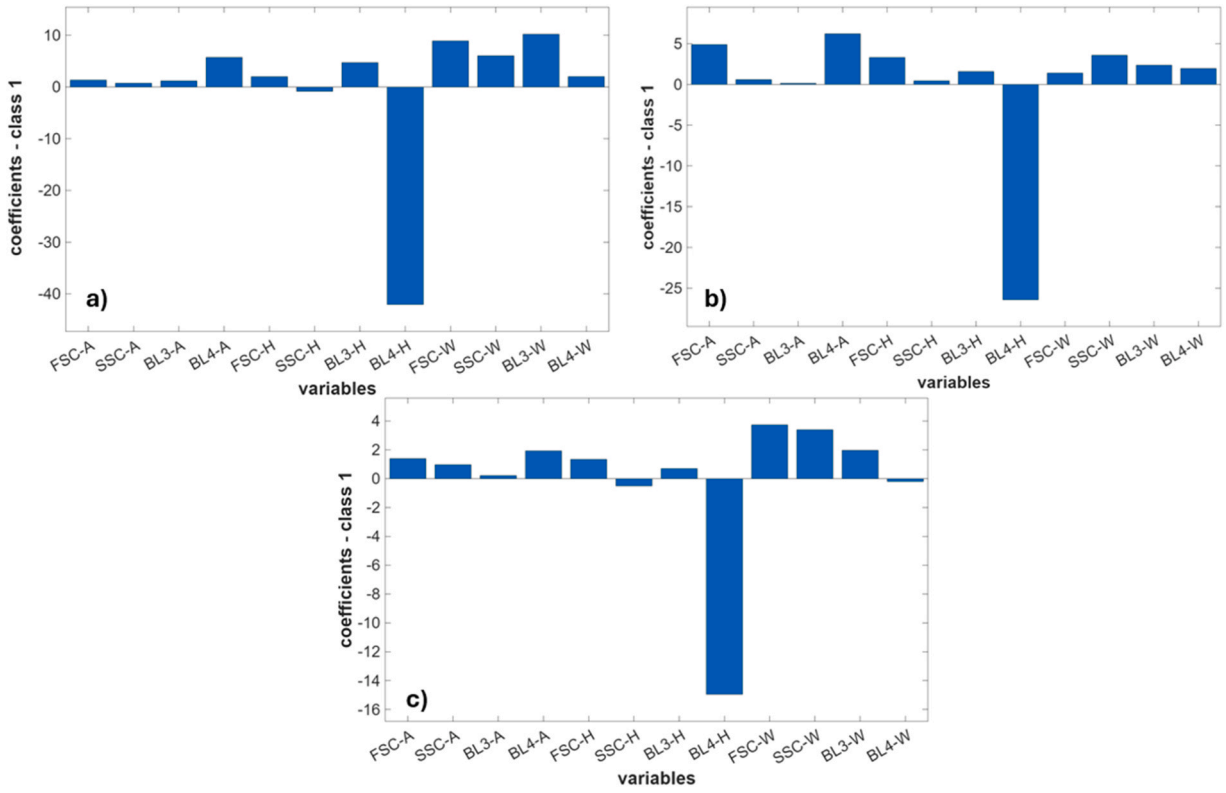

Fig. 5. PLS-DA regression coefficients for models built using binarized response variables, where Class 1 always represents samples with analyte concentrations above the selected thresholds. Panels show models for (a) total carbohydrates (>35% w/w), (b) total carotenoids (>4 mg g^{-1}), and (c) total chlorophylls (>15 mg g^{-1}). All results are shown for data obtained using the C3 preparation protocol.

than the area, as it typically correlates better with the maximum intracellular concentration.

The models yielded excellent results, with AUC values of 0.95 for total carbohydrates, and an AUC of 0.93 for both chlorophyll and carotenoids, as shown in Fig. 6. These values indicate a high discriminative power across all models, confirming their robustness and effectiveness in distinguishing between the defined concentration classes. For comparison, data from C1 and C2 protocol gave AUC between 0.65 and 0.86.

To further investigate model performance, confusion matrices were calculated for both training and cross-validation phases.

The absence of unclassified samples in both the calibration and cross-validation phases indicates that the PLS-DA decision boundaries are well defined in the multivariate space, supporting strong model robustness and generalization (Table 2). Across all three biochemical models, class 1 consistently exhibited higher classification frequencies, reflecting a more compact class structure and more reliable class assignment relative to class 2.

The model for carotenoids showed a moderate decrease in accuracy during cross-validation, reflecting a potential sensitivity to variations in sample composition or borderline cases near the threshold values.

In the case of carotenoid classification, the model demonstrated satisfactory training performance, with an accuracy of 86% and an associated error rate of 15%, indicating a fair ability to learn discriminative patterns between the two classes (Table 3). However, under cross-validation, the accuracy dropped to 81%, with the error rate increasing to 22%, suggesting partial instability in the model's capacity to correctly classify samples under less constrained, more realistic conditions. Although this decline in performance remains moderate, it may point to a degree of over-adaptation to the training data, highlighting the need for careful monitoring when considering real-world deployment of the model.

Conversely, the classification model for total carbohydrates and chlorophyll yielded considerably more robust performance. It achieved a training accuracy of 91% with a low error rate of 10%, and maintained strong generalization under cross-validation, reaching an accuracy of 89% and a reduced error of 12%. This stability across training and validation sets suggests both a higher internal consistency of the dataset and a stronger discriminative power of the selected instrumental variables. Data for carotenoids appear better separated among stressed and not-stressed algae than data for total carotenoids and chlorophylls (Fig. 1). For all three biochemical components, the PLS-DA models showed good discrimination ability, with consistently higher performance for chlorophyll and carbohydrate classifications compared to carotenoids. During training, sensitivity, specificity and precision ranged from 0.81 to 0.97, with class 1 generally exhibiting higher values than class 2. Cross-validation results followed the same trend, indicating stable generalization performance. In particular, chlorophyll and carbohydrate models maintained high and balanced metrics across both classes (sensitivity and precision up to 0.95–0.98), whereas carotenoids showed a modest decrease in cross-validated sensitivity and precision for class 2 (0.69–0.71), suggesting greater class overlap for this variable.

Overall, the concordance between training and cross-validation performance across both models indicates an absence of significant overfitting.

The lower classification performance observed for carotenoids, particularly for class 2, can be attributed to a greater overlap in the multivariate space, likely reflecting a more gradual and less pronounced response to nitrogen starvation compared to carbohydrates and chlorophyll.

From an application perspective, this level of performance may still be considered acceptable for trend monitoring purposes, where the objective is to detect general shifts in biochemical composition rather than to achieve strict class discrimination. However, further model refinement and larger datasets would be required for more accurate classification of carotenoid-related responses.

Nonetheless, the results emphasize the critical importance of the quality and structure of the input data in the development of reliable predictive models. The protocol used to process samples, in particular sample dilution and fixation, has relevant effects on the performances of the PLS-DA models. Moreover, the good findings for carbohydrates and

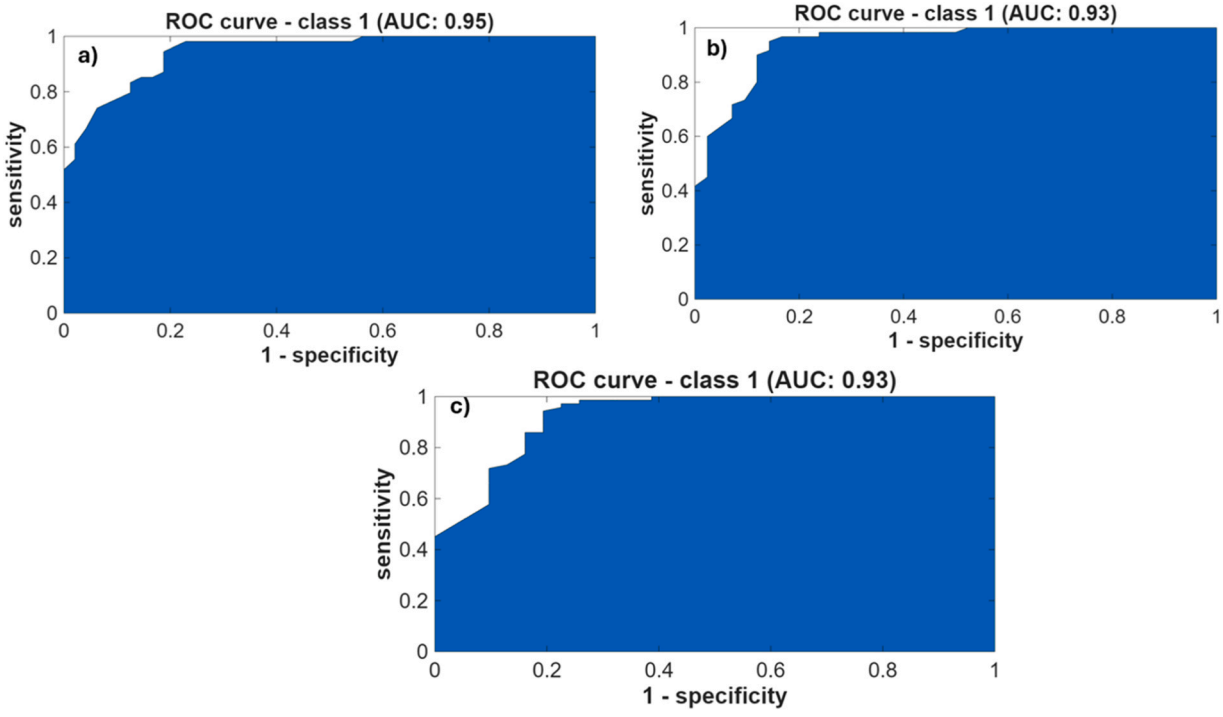

Fig. 6. ROC curves and AUC values for the PLS-DA models built using binarized response variables. Individual panels refer to: a) carbohydrates, b) chlorophyll and c) carotenoids. All results are shown for data obtained using the C3 preparation protocol and identifying Class 1 as the target for the definition of sensitivity and specificity.

Table 2
Confusion matrices for the PLS-DA classifications of total carotenoids, total carbohydrates, and total chlorophylls. Class 1 comprises samples with concentrations above the selected threshold, whereas Class 2 includes samples below the threshold. Results are shown for data obtained using the C3 preparation protocol.

		Confusion Matrix								
		Carotenoids			Chlorophyll			Carbohydrates		
Training		Class 1	Class 2	n.a.	Class 1	Class 2	n.a.	Class 1	Class 2	n.a.
	Class 1	63	8	0	58	2	0	46	8	0
	Class 2	6	25	0	7	35	0	6	42	0
Cross Validation	Class 1	61	10	0	57	3	0	45	9	0
	Class 2	9	22	0	8	34	0	9	39	0

Table 3
Figures of merit for the PLS-DA models predicting total carotenoids, total chlorophylls, and total carbohydrates, including sensitivity (Sens.), specificity (Spec.), precision (Prec.), and accuracy (Acc.). Results are shown for data obtained using the C3 preparation protocol.

		Classification metrics											
		Carotenoids				Chlorophyll				Carbohydrates			
Training		Sens.	Spec.	Prec.	Acc.	Sens.	Spec.	Prec.	Acc.	Sens.	Spec.	Prec.	Acc.
	Class 1	0.89	0.81	0.91	0.86	0.97	0.83	0.89	0.91	0.95	0.98	0.98	0.91
	Class 2	0.81	0.89	0.76		0.83	0.97	0.95		0.91	0.93	0.91	
Cross Validation	Class 1	0.86	0.71	0.87	0.81	0.95	0.81	0.88	0.89	0.93	0.91	0.93	0.89
	Class 2	0.71	0.86	0.69		0.81	0.95	0.92		0.98	0.95	0.94	

chlorophyll support the notion that the PLS-DA approach can serve as an effective and reproducible solution for classifying well-structured biochemical parameters. In contrast, for analytes such as carotenoids, further refinement in data preprocessing or the integration of more advanced modeling strategies may be required to achieve comparable levels of classification accuracy and generalizability. Here in this study, we used conventional spectrophotometric methods for the analysis of carbohydrates, chlorophyll and carotenoids, because these protocols are those typically employed for microalgae characterization.

However, we cannot exclude that more accurate analytical methods, are those based on HPLC, could provide better results. A previous study reported the application of PLS-DA on classifying microalgal species based on data of volatile chemical profiling collected with GC-MS [39], however, while effective, the analytical effort associated with acquiring such datasets may reduce the practicality of this approach for rapid or high-throughput applications.

To further explore the discriminant structure of the model, Potential of Class (PoC) plots among the carbohydrate content were generated along the first and third latent variables (LV1 and LV3), which together explain a cumulative variance of 86.63% (LV1: 48.4%, LV3: 38.23%) (Fig. 7). These bivariate density maps allow for a visual assessment of the separation between the two classes by representing the spatial distribution and density of samples in the latent variable space.

In the two panels of the Fig. 7, contour plots of the probability density for the two classes are shown overlapped to the sample projection onto the latent variable space; more intensely colored regions are associated to areas of highest probability density. While class 1 samples are concentrated primarily in the right quadrants, some overlap with class 2 (red) samples are observable, especially in the central area near the origin (Fig. 7a). Nevertheless, a tendency for class 1 samples to project positively along LV1 is evident, suggesting a significant discriminant gradient along this component.

On the other hand, Fig. 7b illustrates the density distribution for class 2 (red), revealing a higher accumulation of these samples on the negative side of the LV1 axis. The contour lines enclose a broader and more centralized distribution compared to class 1, indicating a greater within-class variance. Some overlap between the classes persists, especially near the center of the plot, but the distinct directional clustering reinforces the discriminant contribution of LV1 to class separation.

Taken together, the PoC maps suggest that LV1 is the primary

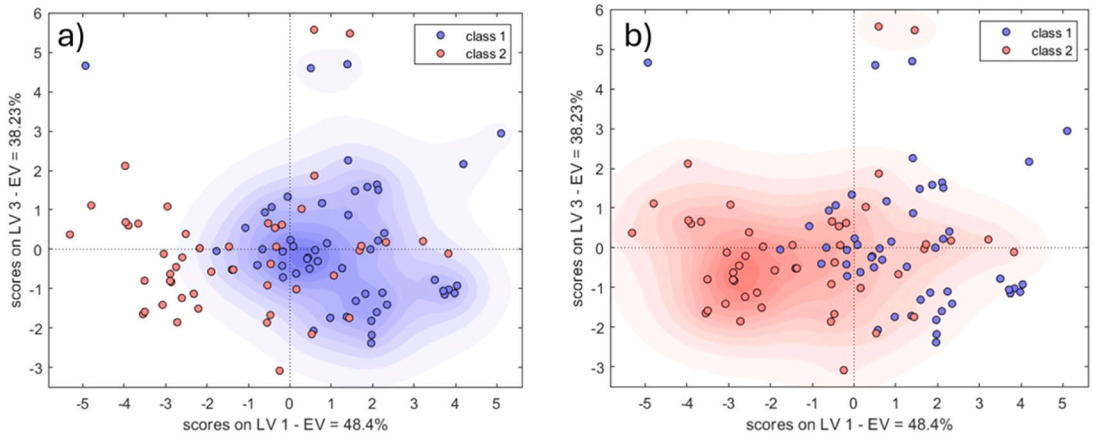

Fig. 7. Potential of Class (PoC) plots derived from the PLS-DA model. **a)** Distribution of samples belonging to Class 1 (samples with values above the selected threshold), and **b)** distribution of samples belonging to Class 2 (samples below the threshold), projected onto the latent variable space. The x-axis represents the latent variable responsible for class discrimination, while the explained variance (EV) indicates the percentage of Y-variance captured by the corresponding component. Contour lines represent the probability density of each class, with more intense regions corresponding to higher sample density. Data are obtained with the C3 protocol.

discriminant direction between the two classes, with LV3 contributing additional but less pronounced separation. The presence of overlapping regions underscores the complexity of the classification task, particularly for borderline samples, but also confirms the model's ability to capture relevant class-specific structure. These findings are consistent with the quantitative performance metrics obtained from the confusion matrices and AUC analysis and support the use of this latent space for robust class discrimination.

Although the method described here yielded very good results, it is important to note that all data used in this study were collected from laboratory-scale bioreactors operated under controlled conditions. Despite the inclusion of samples from several independent batches run at different times, real-world industrial cultivation is likely to exhibit much greater variability. Nitrogen starvation in large-scale systems may occur under different temperatures, pH levels, light regimes, nutrient composition and even in the presence of diverse biological contaminants. The influence of these factors was not systematically investigated here and should be addressed in future studies to determine how sensitive the classification method is to variations in environmental and operational conditions. The method is expected to be applicable even to different trophic regimes (such as heterotrophy and mixotrophy) and to other microalgal strains, as different algae and trophic regimes exhibit comparable biochemical changes when cells move from N^+ to N^- conditions [40].

Many different microalgal strains have been reported to be analyzable with flow cytometry, as *Tetradesmus obliquus* [41], *Tisochrysis lutea*, *Chaetoceros muelleri*, *Tetraselmis suecica* [42] *Chlorella vulgaris* [43], and others. All these species are known to have relevant variation in their biochemical composition as result of passage from N^+ to N^-, similarly to *C. sorokiniana*; therefore similar results may be expected. A major limitation may be represented by microalgal strains that grow in multicellular systems, which could generate issues in flow cytometric analysis, such as *Tetradesmus* sp. and *Arthrospira* sp. This issue can be addressed by adding a cell separation pretreatments, such as sonication [41]. In any case, every different application should be tested on a case-by-case basis by performing a specific calibration of the method for each kind of sample analyzed.

4. Conclusions

This study presents a rapid, simple and reliable method for monitoring microalgae cultivated in bioreactors and for determining the stress status induced by nitrogen starvation. The approach allows the identification of cells in nitrogen starvation phase in batch cultures with good sensitivity, specificity and precision. Nevertheless, its application to specific industrial configurations will require further validation using data collected directly under those operational conditions, in order to assess the method's robustness in the presence of broader environmental and process variability.

The same framework can also be applied to monitor the accumulation or degradation of carbohydrates, chlorophyll and carotenoids associated with nitrogen starvation, although in this case classification performance is comparatively lower due to the greater overlap in biochemical responses.

Overall, the method provides a promising foundation for real-time monitoring of physiological stress in microalgal cultures. Future studies may extend this approach to additional stress conditions and to the evaluation of other intracellular components, broadening its applicability in microalgal biotechnology and bioprocess control.

CRediT authorship contribution statement

Eugenio Sandrucci: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Riccardo Boemio:** Investigation, Formal analysis, Data curation. **Francesca Pagnanelli:** Writing – review & editing, Resources. **Federico Marini:** Writing – review & editing, Supervision, Software, Methodology. **Fabrizio Di Caprio:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.bej.2026.110249.

Data availability

Data will be made available on request.

References

[1] J.J. Milledge, Commercial application of microalgae other than as biofuels: a brief review, 2010 10:1, Rev. Environ. Sci. Biol. Technol. 10 (2010) 31–41, https://doi.org/10.1007/S11157-010-9214-7.
[2] J. Ruiz, G. Olivieri, J. De Vree, R. Bosma, P. Willems, J.H. Reith, M.H.M. Eppink, D. M.M. Kleinegris, R.H. Wijffels, M.J. Barbosa, Towards industrial products from microalgae, Energy Environ. Sci. 9 (2016) 3036–3043, https://doi.org/10.1039/c6ee01493c.
[3] C. Castaldello, A. Gubert, E. Sforza, P. Facco, F. Bezzo, Microalgae monitoring in microscale photobioreactors via multivariate image analysis, 5 (2021) 49, ChemEngineering 5 (2021) 49, https://doi.org/10.3390/CHEMENGINEERING5030049.
[4] S.N.H. Oslan, N.F. Shoparwe, A.H. Yusoff, A.A. Rahim, C.S. Chang, J.S. Tan, S. N. Oslan, K. Arumugam, A.Bin Ariff, A.Z. Sulaiman, M.S. Mohamed, A review on haematococcus pluvialis bioprocess optimization of green and red stage culture conditions for the production of natural astaxanthin, Biomolecules 11 (2021) 256, https://doi.org/10.3390/BIOM11020256.
[5] L. Qu, L.J. Ren, H. Huang, Scale-up of docosahexaenoic acid production in fed-batch fermentation by Schizochytrium sp. based on volumetric oxygen-transfer coefficient, Biochem. Eng. J. 77 (2013) 82–87, https://doi.org/10.1016/J.BEJ.2013.05.011.
[6] F. Di Caprio, S. Amenta, I. Francolini, P. Altimari, F. Pagnanelli, Microalgae biorefinery: optimization of starch recovery for bioplastic production, ACS Sustain. Chem. Eng. 11 (2023) 16509–16520, https://doi.org/10.1021/ACSSUSCHEMENG.3C04133.
[7] F. Di Caprio, A fattening factor to quantify the accumulation ability of microorganisms under N-starvation, N. Biotechnol. 66 (2022) 70–78, https://doi.org/10.1016/J.NBT.2021.04.001.
[8] F. Di Caprio, F. Pagnanelli, R.H. Wijffels, D. Van der Veen, Quantification of Tetradesmus obliquus (Chlorophyceae) cell size and lipid content heterogeneity at single-cell level, J. Phycol. 54 (2018) 187–197, https://doi.org/10.1111/jpy.12610.
[9] J. Rumin, H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J. P. Cadoret, G. Bougaran, The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae, Biotechnol. Biofuels 8 (2015) 1–16, https://doi.org/10.1186/S13068-015-0220-4.
[10] A.L. Heins, M.D. Hoang, D. Weuster-Botz, Advances in automated real-time flow cytometry for monitoring of bioreactor processes, Eng. Life Sci. 22 (2022) 260–278, https://doi.org/10.1002/ELSC.202100082.
[11] G. Breuer, L. De Jaeger, V.P.G. Artus, D.E. Martens, J. Springer, R.B. Draaisma, G. Eggink, R.H. Wijffels, P.P. Lamers, Superior triacylglycerol (TAG) accumulation in starchless mutants of Scenedesmus obliquus: (II) evaluation of TAG yield and productivity in controlled photobioreactors, Biotechnol. Biofuels 7 (2014) 70, https://doi.org/10.1186/1754-6834-7-70.
[12] Y. He, P. Zhang, S. Huang, T. Wang, Y. Ji, J. Xu, Label-free, simultaneous quantification of starch, protein and triacylglycerol in single microalgal cells, Biotechnol. Biofuels 10 (2017) 1–18, https://doi.org/10.1186/S13068-017-0967-X.
[13] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (2003) 166–173, https://doi.org/10.1002/CEM.785.
[14] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (2014) 2812–2831, https://doi.org/10.1039/C3AY41907J.
[15] Matthias. Otto, Chemometrics: statistics and computer application in analytical chemistry, (2024).
[16] N. Aghaeepour, G. Finak, H. Hoos, T.R. Mosmann, R. Brinkman, R. Gottardo, R. H. Scheuermann, D. Dougall, A.H. Khodabakhshi, P. Mah, G. Obermoser, J. Spidlen, I. Taylor, S.A. Wuensch, J. Bramson, C. Eaves, A.P. Weng, E.S. Fortuno, K. Ho, T.R. Kollmann, W. Rogers, S. De Rosa, B. Dalai, A. Azad, A. Pothen,

A. Brandes, H. Bretschneider, R. Bruggner, R. Finck, R. Jia, N. Zimmerman, M. Linderman, D. Dill, G. Nolan, C. Chan, F. El Khettabi, K. O'Neill, M. Chikina, Y. Ge, S. Sealfon, I. Sugár, A. Gupta, P. Shooshtari, H. Zare, P.L. De Jager, M. Jiang, J. Keilwagen, J.M. Maisog, G. Luta, A.A. Barbo, P. Májek, J. Vilček, T. Manninen, H. Huttunen, P. Ruusuvuori, M. Nykter, G.J. McLachlan, K. Wang, I. Naim, G. Sharma, R. Nikolic, S. Pyne, Y. Qian, P. Qiu, J. Quinn, A. Roth, P. Meyer, G. Stolovitzky, J. Saez-Rodriguez, R. Norel, M. Bhattacharjee, M. Biehl, P. Bucher, K. Bunte, B. Di Camillo, F. Sambo, T. Sanavia, E. Trifoglio, G. Toffolo, S.D. Slavica Dimitrieva, R. Dreos, G. Ambrosini, J. Grau, I. Grosse, S. Posch, N. Guex, M. Kursa, W. Rudnicki, B. Liu, M. Maienschein-Cline, P. Schneider, M. Seifert, M. Strickert, J. M.G. Vilar, Critical assessment of automated flow cytometry data analysis techniques, Nat. Methods 10 (2013) 228–238, https://doi.org/10.1038/NMETH.2365.

[17] G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A.K. Shalek, C.K. Slichter, H. W. Miller, M.J. McElrath, M. Prlic, P.S. Linsley, R. Gottardo, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data, Genome Biol. 16 (2015) 1–13, https://doi.org/10.1186/S13059-015-0844-5.

[18] S. Van Gassen, B. Callebaut, M.J. Van Helden, B.N. Lambrecht, P. Demeester, T. Dhaene, Y. Saeys, FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data, Cytom. A 87 (2015) 636–645, https://doi.org/10.1002/CYTO.A.22625.

[19] H.T. Maecker, J.P. McCoy, R. Nussenblatt, Standardizing immunophenotyping for the Human Immunology Project, Nat. Rev. Immunol. 12 (2012) 191–200, https://doi.org/10.1038/NRI3158.

[20] K.E. Diggins, A.R. Greenplate, N. Leelatian, C.E. Wogsland, J.M. Irish, Characterizing cell subsets using marker enrichment modeling, Nat. Methods 14 (2017) 275–278, https://doi.org/10.1038/NMETH.4149.

[21] Y. Shao, Y. Li, L. Jiang, J. Pan, Y. He, X. Dou, Identification of pesticide varieties by detecting characteristics of Chlorella pyrenoidosa using Visible/Near infrared hyperspectral imaging and Raman microspectroscopy technology, Water Res 104 (2016) 432–440, https://doi.org/10.1016/J.WATRES.2016.08.042.

[22] Y. Dong, X. Li, M. Gouda, Raman Spectroscopy for Tracking the Polysaccharides Chemical Composition and β-Carotene Accumulation of Microalgae Food Source, (2024) 57–70. ⟨https://doi.org/10.1007/978-1-0716-3938-2_4⟩.

[23] K.A. Lê Cao, S. Boitard, P. Besse, Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems, BMC Bioinforma. 12 (2011) 1–17, https://doi.org/10.1186/1471-2105-12-253.

[24] F. Di Caprio, F. Del Signore, L. Capobianco, F. Pagnanelli, P. Altimari, Feast–Famine in Cyclic Autotrophy/Heterotrophy Doubles Microalgal Productivity while Controlling Bacterial Contamination, ACS Sustain. Chem. Eng. 13 (2025) 16448–16458, https://doi.org/10.1021/ACSSUSCHEMENG.5C05990.

[25] H.K. Lichtenthaler, Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes, Methods Enzym. 148 (1987) 350–382, https://doi.org/10.1016/0076-6879(87)48036-1.

[26] S. Van Wychen, L.M.L. Laurens, Determination of Total Carbohydrates in Algal Biomass: Laboratory Analytical Procedure (LAP) (Revised), (2013). ⟨www.nrel.gov/publications⟩. (accessed September 26, 2025).

[27] M. Dubois, K.A. GILLES, J.K. HAMILTON, P.A. REBERS, F. SMITH, Colorimetric Method for Determination of Sugars and Related Substances, Anal. Chem. 28 (1956) 350–356, https://doi.org/10.1021/ac60111a017.

[28] Y. Maltsev, M. Kulikovskiy, S. Maltseva, Nitrogen and phosphorus stress as a tool to induce lipid production in microalgae, 2023 22:1, Microb. Cell Factor. 22 (2023) 1–22, https://doi.org/10.1186/S12934-023-02244-6.

[29] T. Li, M. Gargouri, J. Feng, J.J. Park, D. Gao, C. Miao, T. Dong, D.R. Gang, S. Chen, Regulation of starch and lipid accumulation in a microalga Chlorella sorokiniana, Bioresour. Technol. 180 (2015) 250–257, https://doi.org/10.1016/J.BIORTECH.2015.01.005.

[30] I.K. Blaby, A.G. Glaesener, T. Mettler, S.T. Fitz-Gibbon, S.D. Gallaher, B. Liu, N.R. Boyle, J. Kropat, M. Stitt, S. Johnson, C. Benning, M. Pellegrini, D. Casero, S.S. Merchant, Systems-Level Analysis of Nitrogen Starvation-Induced Modifications of Carbon Metabolism in a Chlamydomonas reinhardtii Starchless Mutant W, (n.d.). ⟨https://doi.org/10.1105/tpc.113.117580⟩.

[31] F. Di Caprio, F. Pagnanelli, R.H. Wijffels, D. Van der Veen, Quantification of Tetradesmus obliquus (Chlorophyceae) cell size and lipid content heterogeneity at single-cell level, J. Phycol. 54 (2018) 187–197, https://doi.org/10.1111/jpy.12610.

[32] D. Čertnerová, D.W. Galbraith, Best practices in the flow cytometry of microalgae, Cytom. Part A 99 (2021) 359–364, https://doi.org/10.1002/CYTO.A.24328.

[33] A.K. Bajhaiya, A.P. Dean, T. Driver, D.K. Trivedi, N.J.W. Rattray, J.W. Allwood, R. Goodacre, J.K. Pittman, High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation, 2015 12:1, Metabolomics 12 (2015) 9, https://doi.org/10.1007/S11306-015-0878-4.

[34] N.I. Novikova, H. Matthews, I. Williams, M.A. Sewell, M.K. Nieuwoudt, M.C. Simpson, N.G.R. Broderick, Detecting Phytoplankton Cell Viability Using NIR Raman Spectroscopy and PCA, 7 (2022) 5971. ⟨https://doi.org/10.1021/acsomega.1c06262⟩.

[35] F. Hammes, T. Broger, H.U. Weilenmann, M. Vital, J. Helbing, U. Bosshart, P. Huber, R. Peter Odermatt, B. Sonnleitner, Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis, Cytom. Part A 81A (2012) 508–516, https://doi.org/10.1002/CYTO.A.22048.

[36] S. Van Nevel, S. Koetzsch, C.R. Proctor, M.D. Besmer, E.I. Prest, J. S. Vrouwenvelder, A. Knezev, N. Boon, F. Hammes, Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring, Water Res 113 (2017) 191–206, https://doi.org/10.1016/J.WATRES.2017.01.065.

[37] A.K. Chowdhary, M. Kishi, T. Toda, Astaxanthin induction in Chromochloris zofingiensis by transition of nutritional modes, J. Appl. Phycol. 37 (2025) 1711–1718, https://doi.org/10.1007/S10811-025-03497-X.

[38] R.J. Strasser, W.L. Butler, Fluorescence emission spectra of Photosystem I, Photosystem II and the light-harvesting chlorophyll a/b complex of higher plants, Biochim. Et. Biophys. Acta (BBA) - Bioenerg. 462 (1977) 307–313, https://doi.org/10.1016/0005-2728(77)90129-3.

[39] J. Magpusao, I. Oey, B. Kebede, Chemical, rheological, and volatile profiling of microalgae Arthrospira, Isochrysis, Nannochloropsis, and Tetraselmis species (n. d.), Food Innov. Adv. (2024) 75–87, https://doi.org/10.48130/fia-0024-0007.

[40] F. Di Caprio, Cultivation processes to select microorganisms with high accumulation ability, Biotechnol. Adv. 49 (2021) 107740, https://doi.org/10.1016/J.BIOTECHADV.2021.107740.

[41] F. Di Caprio, S. Posani, P. Altimari, A. Concas, F. Pagnanelli, Single cell analysis of microalgae and associated bacteria flora by using flow cytometry, Biotechnol. Bioprocess Eng. 26 (2021) 898–909, https://doi.org/10.1007/s12257-021-0054-9.

[42] M. Kallau, H. Yang, Quantification of the total lipids in three aquaculture microalgae using BODIPY™ 505/515 stain and flow cytometry, J. World Aquac. Soc. 56 (2025) e70028, https://doi.org/10.1111/JWAS.70028.

[43] V. Pozzobon, W. Levasseur, E. Viau, E. Michiels, T. Clément, P. Perré, Machine learning processing of microalgae flow cytometry readings: illustrated with Chlorella vulgaris viability assays, 2020 32:5, J. Appl. Phycol. 32 (2020) 2967–2976, https://doi.org/10.1007/S10811-020-02180-7.