

Fluorescence coupled with chemometrics for simultaneous monitoring of cell concentration, cell viability and medium nitrate during production of carotenoid-rich

*Dunaliella salina*

Marta Sá, Joana Monte, Carla Brazinha, Claudia F Galinha\*, João G Crespo

LAQV-REQUIMTE, Chemistry Dept, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

\*Corresponding author: [cf.galinha@fct.unl.pt](mailto:cf.galinha@fct.unl.pt)

**Abstract**

Two-dimensional (2D) fluorescence spectroscopy was investigated as a monitoring tool for cultivation, harvesting, and effluent treatment of *Dunaliella salina* with high carotenoid concentration; aiming to improve the production process and minimise costs. Chemometric analysis, namely Principal Component Analysis (PCA) and Projection to Latent Structures (PLS), were used to build models for estimation of cellular concentration, cellular viability, and nitrate concentration in media. The estimations were based on fluorescence excitation-emission matrices (EEMs) acquired directly from algal suspensions. Cell concentration during cultivation and harvesting can be predicted by a single model capturing 92.0% of the variance, and with  $R^2$  of 0.92 and 0.97, for training and validation, respectively. Cell viability during harvesting by ultrafiltration was modelled with 79% of variance and  $R^2$  of 0.79 for training and 0.73 for validation. Nitrate concentration was successfully predicted during cultivation and permeate treatment using a single model with 81.8% of variance and  $R^2$  of 0.82 for training and 0.80 for validation. Therefore, this work demonstrates the strong potential of combining 2D fluorescence and chemometrics for monitoring different processes during microalgae production.

**Keywords:** *Dunaliella salina*; fluorescence EEMs; Carotenogenesis; Membrane harvesting; Monitoring.

## 1. Introduction

*Dunaliella salina*, a green unicellular microalga, is remarkably halo-tolerant due to its ability of *de novo* synthesis of glycerol with increasing saline stress. Additionally, under salt and light stress conditions, nutrient deprivation or extreme temperatures, *D. salina* accumulates a high concentration of carotenoids. Although *D. salina* is considered a food source [1–3], its industrial production is still focused on niche markets, such as carotenoids (high-value pigments). To increase the economic interest of microalgal cultivation, the biomass cost needs to be reduced while the production scale increases [4].

One of the challenges of microalgal cultivation is the dilute suspension that characterises these cultures [5]. Increasing the concentration of biomass during the cultivation can lead to a decrease in photosynthetic yields. Therefore, high volumes of aqueous media must be processed, and the cost of dewatering algal biomass was identified as the major factor hindering their production [5]. In that sense, membrane-based harvesting systems offer a potential solution to concentrate biomass. Since membrane filtration is a highly reliable and directly scalable technology, it enables a hybrid harvesting process, due to the possibility of coupling with other technologies, such as centrifugation. Moreover, membrane filtration has low energy consumption and low operational costs [5–10]. As described by Monte et al. [11], including a membrane pre-concentration step before the usual centrifugation procedure reduces the number of needed centrifuges, thus reducing the energy consumption (up to 63%) and capital costs (up to 60%).

Furthermore, the use of a membrane dewatering system enables the recovery of the permeate, suitable as microalgae cultivation medium, a critical requirement to minimise the water use and new nutrients [5]. The costs associated with the cultivation medium make the reuse and recycling of this permeate a key parameter to reduce water footprint [12–14]. Although this permeate is usually clear, free of cells, and high molecular weight contaminants, it can contain some remaining nutrients, cell debris, and organic compounds [12]. Therefore, before reusing the permeate, a further intermediate treatment step, like ozone or ultraviolet (UV) light, needs to be assessed [15].

The biorefinery can enhance the potential of microalgal technologies by extracting multiple products from the same biomass. This enhancement can be achieved with the development of an *in situ* and online monitoring tool

capable of simultaneously detecting multiple analytes and/or inferring about other relevant biological parameters. Nowadays, biomass concentration is a parameter that can be estimated online using optical density (OD) probes at several wavelengths. However, it is known that pigments and organelles affect the light absorbance significantly and that microalgal cells may change composition during cultivation; thus, requiring a careful selection of the right wavelengths [16,17].

Recently, a technology based on *in situ* microscopy (ISM) was developed for monitoring cell concentration, as well as cell morphology and cell size distribution, based on the correlation between cell concentration and macroscopic colour intensity of the microalgal suspension [18]. Being able to monitor cell concentration and integrity online is of extreme importance during filtration steps of *D. salina* biorefinery. The absence of a rigid cell wall coupled with the shear stress inherent to a filtration process may lead to severe cell damage and, consequently, to the loss of highly valuable compounds to the aqueous medium.

As mentioned by Lamers et al. [19], nitrogen depletion, is a frequent strategy to enhance carotenoids concentration in *D. salina*. The possibility of monitoring nitrogen concentration during cell growth and induction of carotenoids will make it possible to improve the design and operation of these processes. Similarly, the monitoring could be applied during the treatment of the permeate recovered from the harvesting step.

Fluorescence spectroscopy is an extensively used methodology due to its high sensitivity, specificity, simplicity, and ability to detect simultaneously several compounds (extracellular and intracellular metabolites); it is suitable for applications such as online monitoring of complex samples [20,21]. After molecular excitation over a range of wavelengths, the excess energy will be emitted for longer wavelengths due to the transition between the excited singlet and the ground state. The large matrices of data obtained, named excitation-emission matrices (EEMs), encode valuable information of the presence of natural fluorophores, the interactions between them, and the interactions with the medium. Scattering can be also a source of information on the present suspended material. The use of an *in situ* fluorometry technique was reported by Shin et al. [22] to collect fluorescent signals of specific pigments (Chlorophyll *a* and Phycocyanin) of algal cells (*Chlorella vulgaris* and *Spirulina* sp., respectively). Also, a previous study of the authors was performed to develop a tool to monitor cell concentration and viability of *D.*

*salina* in its "non-stressed" biological state ("green" *D. salina*) using the complex information contained in fluorescence EEMs (regarding not only pigments but also other intracellular compounds, such as amino acids) [23].

To validate fluorescence spectroscopy as a monitoring tool for *D. salina* biorefinery, independently from its biological state, a real-time monitoring tool was developed in this study for stressed "orange" culture. EEMs obtained during cell cultivation (induction from "green" to "orange"), harvesting by membrane filtration, and permeate treatment were correlated with off-line measurements, such as cell concentration and viability, and nitrate concentration. Models based on fluorescence EEMs were constructed using the chemometric tools PCA (Principal Component Analysis) and PLS (Projection to Latent Structures).

## **2. Material and Methods**

### ***2.1. Growth conditions of *Dunaliella salina* and carotene induction***

*D. salina* DF40, collected from open ponds of Monzón Biotech (Spain), was isolated by the Marine Biological Association (United Kingdom).

The inoculum was maintained and scaled-up in the facilities of A4F – Algae for Future (Lisbon, Portugal). Cultures were maintained in artificial saltwater (ASW), an A4F Industrial Medium [15] and successively scaled-up volumetrically in round flasks up to 5 L, at 25°C and under continuous illumination (150  $\mu\text{mol}/\text{m}^2\cdot\text{s}$ ) provided by fluorescent tubes. Continuous aeration (0.1-0.3 % vvm) was provided through 0.2  $\mu\text{m}$  filtered air supplemented with 2% of CO<sub>2</sub>, enabling the mixing of the culture.

For pilot-scale carotenogenesis cultivation experiments (n=6) the culture was scaled-up to outdoor flat-panel photobioreactors (supplied by A4F – Algae for Future, Portugal), with an optical pathlength of 7 to 10 cm, and adjustable volume through the adjustable length. The culture was exposed to outdoor light and temperature conditions, where only the maximum temperature was controlled (maximum temperature of 25°C). Carotenoid induction was achieved by depletion of nitrogen and increasing the medium salinity, and monitored since the inoculation with non-stressed "green" cells until a stressed "orange" culture was obtained. These experiments were performed through ten months, between January and October 2017.

The culture for the harvesting experiments was collected from the flat-panel photobioreactors described above, in a semi-continuous harvesting regime. Each experiment started at a slightly different initial biological state, with cell concentrations varying between 1.6 to  $9.3 \times 10^5$  cells/mL and Car/Chl (carotenoids/chlorophylls) ratio between 3 to 14 (Table 1).

## **2.2. Operation conditions for membrane harvesting**

A small pilot filtration set up was used for the harvesting experiments [15]. Hollow fibres modules of polyethersulfone with 63.5 cm of length and 0.48 m<sup>2</sup> of membrane area were used (UFP-100-C-6A and UFP-500-C-6A, GE Healthcare, UK). The unit was operated under permeate flux control and with a 0.35 m/s recirculation cross-flow velocity. Three parameters were tested: membrane type – ultrafiltration membrane with a Molecular Weight Cut-off (MWCO) of 100kDa, and with an MWCO of 500kDa; imposed permeate volumetric flux – 10, 15 and 25 L/(m<sup>2</sup>.h); and membrane relaxation procedure (cycles of operation for 9 minutes and relaxation for 1 minute).

In total, eleven experiments were performed, as shown in Table 1. Three replicates were performed at a permeate flux of 25 L/(m<sup>2</sup>.h), using the 100 kDa membrane and applying relaxation. For an imposed permeate flux of 10 L/(m<sup>2</sup>.h), and the same membrane type and relaxation, two duplicate experiments were performed. To enhance the total biomass concentration factor and, at the same time, keeping the maximum cell integrity, in two of the experiments performed at 10 L/(m<sup>2</sup>.h), 100 kDa and with membrane relaxation, a semi-continuous feeding approach was tested by adding two consecutive loads of fresh biomass, as represented in Figure 1. The experiment started with an initial culture load of 20 kg, comprising biomass and water medium, (concentration factor (CF) of 1) and was filtered until reaching 10 kg (CF 2) when a load of 10 kg of new fresh culture was added to the concentrate (CF 1.5). This second load with a total of 20 kg was filtered until reaching 10 kg again (CF 3), and the third load of fresh culture (10 kg) was added again (CF 2). After this last load, the filtration was performed until 1.1 kg of concentrate was reached (CF 36.7). Further detail about the methodology of the membrane harvesting and the equipment used is provided by Monte et al. [15].

## **2.3. Permeate treatment: oxidation, photodegradation and membrane purification**

The permeate obtained during harvesting of *D. salina* by ultrafiltration was treated by different techniques (oxidation, photodegradation and nanofiltration) aiming at reusing it in new cultivation batches of *D. salina*. The

permeate from harvesting used in these experiments was collected during an experiment using a 100 kDa membrane, at a cross-flow velocity of 0.35 m/s, under controlled permeate flux at 10 L/(m<sup>2</sup>.h) and with membrane relaxation.

For the oxidation and photodegradation assays, three different procedures were tested: with no pH correction, with no pH correction and addition of 120 mg/L of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), and with pH correction to 9. Both pH correction and H<sub>2</sub>O<sub>2</sub> addition are known to accelerate the ozone decomposition and the oxidation process [14]. Further details about the treatment of the permeate by oxidation and photodegradation are provided in Monte et al. [15].

Three different nanofiltration membranes were used for the permeate treatment: a polyethersulfone membrane with a MWCO of 400 Da (NP030, Microdyn Nadir, Germany) and controlling the permeate flux at 27.9 L/(m<sup>2</sup>.h); a polyamide membrane, with a MWCO between 150 and 300 Da (Desal 5DK, GE Osmonics, USA) and controlling the permeate flux at 20.9 L/(m<sup>2</sup>.h); and a NF90 membrane (polyamide selective layer and polysulfone supporting layer) with a MWCO of 100 Da (Dow Filmtec, USA) and controlling the permeate flux at 5.7 L/(m<sup>2</sup>.h). Each experiment processed 250 mL of permeate at a transmembrane pressure of 30 bar.

#### ***2.4. Sampling procedure and analysis***

The sampling of cultivation batch experiments was performed every other day, until reaching a stable cell concentration and Car/Chl ratio. Samples collected during membrane harvesting experiments were withdrawn from the culture and the permeate tanks, simultaneously. For ozonation, each experiment lasted one hour, samples were taken every twenty minutes; one last experiment with optimised conditions lasted six hours, with samples taken every hour. Each UV optimisation treatment lasted three hours, samples were taken every hour; the last experiment with the optimised conditions lasted six hours, with samples taken every hour. During membrane filtration of the permeate, samples were taken at concentration factors of 1, 1.25, 1.5, 2, 3 and 3.6. Figure 2 shows the sampling points of the different processes. All the samples collected during the different processes were assessed by fluorescence spectroscopy. Cell concentration, cell viability and nitrate concentration analyses were performed for the samples collected as described below.

#### ***2.4.1. Cell concentration and viability***

Cell concentration was measured during the cultivation growth experiments and for samples acquired from the concentrate stream of the membrane harvesting experiments, using a Muse® Cell Analyser (Merck Millipore, Germany). For each sample, three replicates were performed. To acquire a linear response of the equipment, all samples were diluted with ASW to a range between 10 and 500 cells/mL (variance between replicas lower than 10%). Propidium iodide (PI) is a fluorescent dye with maximums of excitation at 535 nm and emission at 617 nm, used to assess cell viability. Intact cells with intact membrane are impermeable to PI, thus this dye is used to stain non-viable cells. After mixing the PI with the culture sample, the sample was immediately analysed with Muse® Cell Analyser. The concentration of viable cells is given by the total cell concentration and viability percentage measured by the equipment. Since non-viable cells were measured based on the permeability of the cell membranes, the viability percentage was used as cell integrity indicator.

#### ***2.4.2. Nitrate concentration***

Nitrate concentration, measured as ppm of nitrogen, was determined for samples acquired during cultivation batches and from the permeate of the harvesting experiments. All samples were centrifuged, to eliminate cells or cell debris, and stored at -20°C until analysed by a colourimetric method based on the cadmium reduction method implemented in a segmented flow analyser (Skalar San++, Skalar Analytical, The Netherlands) according with standard methods [24].

#### ***2.5. Fluorescence spectroscopy***

All the samples in this study were analysed by 2D fluorescence spectroscopy. Cell culture samples acquired from cultivation batches and from the concentrate stream of membrane harvesting experiments were analysed directly in a stirred sampling tube to avoid cell sedimentation. Cell-free samples, like permeate of the harvesting experiments and supernatant of the cultivation batches, did not require agitation. All samples were measured with the immersion of an optical fibre probe directly in the samples without any further treatment. A fluorescence spectrophotometer (Agilent Cary Eclipse, California, USA), equipped with monochromators for excitation and emission and an optical fibre bundle probe, was used to acquire the fluorescence spectra. The conditions of spectra acquisition were described previously by Sá et al. [23]: excitation wavelength between 250 and 690 nm (10 nm of

excitation slit and 5nm increment) and emission wavelength between 260 and 700 nm (20nm excitation slit and 5 nm increment).

## ***2.6. Multivariate data analysis***

In order to reduce the noise and co-linearity characteristic of the excitation-emission matrices (EEMs), principal component analysis (PCA) was used to compress and deconvolute them into principal components (PCs). Briefly, PCA reduces the dimension of the data while extracting meaningful information from the spectra. For a certain  $n$  number of initial variables,  $n$  linear combinations will be obtained; the PCs will be characterised for being uncorrelated, orthogonal, and ordered according to the variance explained (the first PC explains the higher part of the variance, and so on) [25,26].

Fluorescence spectra were acquired in 5 streams: the cultivation broth and its supernatant from carotenogenesis, concentrate and permeate from harvesting, and treated permeate samples (Figure 2).

Fluorescence spectra compression into PCs was made for each stream individually, resulting in 10 PCs for each stream that captured always more than 99.8% of the variance. For models where the fluorescence of two streams was used (for example data from cultivation and harvesting), fluorescence spectra of those streams were compressed together resulting in 10 or 15 PCs that captured more than 99.6% of the variance. The information about the number of PCs used for each model is shown in Table 2 and Figure 2.

Projection to latent structures (PLS) modelling was used to describe multilinear correlations between the fluorescence PCs (inputs) and the offline measures of cell concentration, cell viability and nitrate concentration (outputs). More detailed information about PLS modelling can be found here [20,27]. Table 2 summarises the different strategies used for the PLS model construction of cell concentration, cell viability and nitrate concentration prediction.

For cellular concentration models, two processes were studied: cultivation experiments and harvesting. Three input strategies were attempted: one using as input ten PCs from the cultivation batches (41 fluorescence spectra used); a second one using only PCs from the concentrate stream ( $n=138$ ); and a third one using cultivation and concentrate PCs of fluorescence spectra compressed together ( $n=179$ ).



Cell viability prediction models used only data acquired during membrane harvesting experiments. Two input strategies were attempted: using ten PCs of the fluorescence acquired on the concentrate stream (n=138) and using the same ten PCs acquired in the concentrate coupled with the ten PCs acquired from the permeate stream (n=138).

Nitrate concentration models were studied in two different processes, cultivation and permeate treatment, using three different input strategies: using only the supernatant of the cultivation samples (n=41), using the samples from the permeate treated stream (n=89) and using the samples of the two processes combined (n=130). Ten PCs were used as inputs in the first two approaches, while 15 PCs were used in the model with combined data from both processes.

To develop the PLS models, 75% of the experimental data was randomly attributed for training purposes and the remaining 25% for model validation. The normalisation of the data was achieved by subtracting their average and dividing by their standard deviations. The selection of the predictors from the inputs initially provided was performed by the following mathematical methods: iterative stepwise elimination, stepwise elimination and the Martens uncertainty test using jack-knife standard deviations [28–30]. To compare the quality of the models, several parameters are discussed, such as the variance captured, the coefficients of determination ( $R^2$ ) and slopes for both training and validation sets, and the root mean square errors of cross-validation (RMSECV) and of prediction (RMSEP).

The n-way toolbox of Matlab was used to implement PARAFAC and nPLS [31].

### **3. Results and Discussion**

Figure 3 shows an emission-excitation fluorescence spectrum acquired during a carotenogenesis cultivation experiment. It is noteworthy in two main regions, a protein-like region (excitation between 250 and 350 nm, and emission below 300 nm), and the pigment band (excitation above 650 nm). More detail description of the fluorescence signal acquired for *D. salina* microalgae is provided in Sá et al. [23]. Although these regions are the main fluorescence response in fluorescence spectra of *D. salina* samples, due to interferences of medium and cell composition, their intensities do not change directly with cell concentration/composition. Therefore, chemometric tools were used to translate the complex signal obtained in these spectra with quantitative parameters.

### 3.1. Cell concentration modelling

In the present study, cell concentration prediction models during cultivation growth experiments were based on 10 PCs acquired from PCA of the fluorescence EEMs (Fig.4).

Fig.4(a) shows the prediction model of cell concentration, where it is plotted the observed values on the y-axis against the predicted values on the x-axis. The presence of a line corresponding to two times the standard deviation of the observed values helps to identify the presence of outliers. Models with good fitting will have a coefficient of determination ( $R^2$ ) and a slope near to 1, which indicates that the same trend is followed by observed and predicted values. The  $R^2$  of training and validation sets were both high (0.82 and 0.86, respectively), and the root mean square error of prediction (RMSEP) was  $3.60 \times 10^5$  cells/mL, which represents the distance, on average, from a data point to the predicted value. From the ten PCs used as inputs, only eight were used, explaining 85.7% of the captured variance (Fig.4(b)).

Cell concentration is a vital factor for the economic feasibility of microalgae biorefinery. A similar study was performed by Sá et al. [23] to develop an online tool able to monitor *D. salina* cell concentration in its non-stressed biological state ("green" *D. salina*). To develop a tool able to be used in different biorefinery scenarios, the present work is focused on the development of a model for cell concentration during membrane harvesting using stressed *D. salina*. This stressed biomass is orange due to its high concentration of carotenoids.

The model developed for cell concentration during harvesting experiments used ten PCs from samples of the concentrate stream only (Fig.4(c) and (d)). Training and validation  $R^2$  were higher (0.93 for both) than the ones obtained for the same parameter in cell growth models (Fig.4(a)). In a similar way to what was achieved before, for the model developed for the culture growth experiments, eight PCs, from the ten used as initial inputs, were used to explain 93.4% of the variance (Fig.4(d)) and an RMSEP of  $1.35 \times 10^6$  cells/mL was obtained.

To analyse a model performance, the root mean square error of cross-validation (RMSECV) can be used as an internal validation method for the training data set. Values of RMSECV close to RMSEP indicate a good fitting of the model to both training and validation data sets. Contrary to what was observed in the models for cell concentration in culture growth experiments, where RMSEP values were close to RMSECV, for the harvesting

model this difference is higher, indicating a less robust model. It is noteworthy that in harvesting experiments, the diluted *D. salina* culture was, sometimes, concentrated more than ten times. Therefore, when measuring cell concentration with the Muse® Cell Analyser, a higher dilution is needed to achieve between 10 to 500 cells/mL, an equipment requirement, leading to a higher analytical error.

The use of fluorescence spectroscopy as a monitoring tool to determine *D. salina* cell concentration could be more useful as an industrial technique if there was no need to have distinct models for distinct processes. For that reason, the fluorescence spectra acquired during both processes, cultivation and harvesting, were combined and compiled together. The ten PCs resulting from this new PCA were used for the PLS modelling. The  $R^2$  for training and validation were similar to the ones found with the harvesting data (0.92 and 0.97, respectively) as well as the variance captured, 92.0% (Fig.4(e)). From the ten new PCs given to the model, nine were selected (Fig.4(f)). The difference between RMSEP ( $8.51 \times 10^5$  cells/mL) and RMSECV ( $1.04 \times 10^6$  cells/mL) was still higher than desired. As observed from the data dispersion in Fig.4(e), more experimental data at the high cell concentration range could increase the robustness of the model, since only a few points were observed for concentrations above  $1.50 \times 10^7$  cells/mL. However, the addition of cultivation data to the harvesting model increased the prediction quality of the model. However, the use of an extended range of concentrations (namely in harvesting) results in higher errors, thus it can be advantageous to use a model calibrated only for cultivation (for the lower concentration range) to have a better prediction while monitoring *D. salina* growth.

2D Fluorescence spectroscopy was already described as a powerful and sensitive technique to monitor several biological processes [32]. The use of this technique in microalgae biorefinery was first described by Sá et al. [23], where a model to predict "green" *D. salina* cellular concentration was developed. The results presented herein enhance the potential of this technique to monitor *D. salina* biomass production and recovery, regardless of the biological state of the culture, stressed ("orange") or non-stressed ("green"). It is noteworthy that the cultivation batches monitored in the present study started with "green" *D. salina* and ended "orange", meaning that the models developed now for cell concentration can be used despite cellular variations in composition along the cultivation time.

### 3.2. Viability modelling

Harvesting *D. salina* cells can result in cellular damage since this microalga lacks a rigid cell wall, and this would represent the loss of valuable intracellular compounds to the medium. Thus, monitoring cell viability during harvesting is important for the biorefinery process.

Two different input strategies were studied to monitor cell viability during membrane harvesting experiments. Firstly, ten PCs of the PCA applied to the fluorescence matrices of the concentrate stream samples were used to develop the model shown in Fig.5 (a) and (b). The data dispersion is high, and it is possible to notice some outliers (data points plotted outside the two-times-standard-deviation lines of the experimental data). This resulted in low training and validation  $R^2$  (0.36 and 0.23, respectively). The RMSEP was 12.8%, which represents around 13% distance between the observed data point and the predicted value. According to Sá et al. [23], values of loss of viability above 10% are considered high when aiming to harvest *D. salina*, since it corresponds to noticeable cell damage that can be observed under the optical microscope.

A second input strategy was attempted to improve the predictability of the cell viability model: adding ten PCs of the fluorescence acquired in the permeate stream to the ten PCs of the concentrate stream. The addition of the permeate stream information in models to determine cell viability during *D. salina* harvesting is of great importance, as can be seen by the weight of the PCs in the model (Fig.5(d)). This phenomenon was previously reported by Sá et al. [23] when monitoring "green" cells of *D. salina* during harvesting by membrane filtration. It is expected that small intracellular compounds (released due to cell disruption) can pass through the membrane during the processing of the biomass, accumulating on the permeate stream and, for that reason, changing the fluorescence profile of the permeate matrix. A significant improvement was achieved with the addition of information captured in the permeate stream, with the increase of both training and validation  $R^2$  (0.76 and 0.69, respectively) (Fig.5(c)). Both slopes were close to 1, and the RMSEP decrease to 9.5%. By using five PCs from the fluorescence EEMs acquired in the concentrate and three PCs acquired in the permeate, the model can explain 76.3% of the captured variance, and the data points that were outliers in the previous model are now well predicted. Even expecting that low concentrations of such intracellular compounds would pass through the membrane, it is noteworthy the sensibility of the 2D fluorescence technique to detect these small concentration changes.

The range of cell viability in the models presented is wider than would be needed for a harvesting step by ultrafiltration membranes if a biorefinery approach is followed. When aiming for a higher yield of intracellular compounds recovered from the biomass, high cell integrity is desired to guarantee that there is no loss of intracellular compounds to the concentrate and permeate streams. The model obtained predicts better (with less dispersion) at higher viability values than at lower values. However, the accuracy of such models may be improved, if required, by the use of physical parameters, such as pH or salinity, that can be related to these compounds. Nevertheless, the ability of fluorescence spectroscopy to sense such small variabilities in matrices so different as cell broth and clean permeate boosts the potential of this technique to be used as a monitoring tool.

### ***3.3. Nitrate concentration modelling***

*D. salina* capability of high production of carotenoids is achieved by imposing stress factors such as the depletion of nitrogen. Monitoring of nitrogen concentration in pilot-scale cultivations, usually nitrate, improves the operator knowledge about the biological state of the culture and real-time decision making. Additionally, with the inclusion of a membrane harvesting step, a large amount of aqueous medium can be recovered and recycled. Although this medium may have diverse compositions, depending on the membrane and the operating conditions used, some cell debris and intracellular compounds may be released to the permeate. For this reason, three different treatments were applied to the permeate media before recycling it for cultivation (oxidation, photodegradation, and nanofiltration), and nitrate concentration was monitored as a potential oxidation by-product formed during these treatment steps.

To monitor nitrate concentration (expressed as ppm of nitrogen), three different modelling strategies were studied: using data acquired only from the cultivation experiments samples, using data acquired only from samples of permeate treatment, and combining data from both stages (Table 3).

For nitrate concentration during the cultivation experiments, ten PCs of the samples' supernatant fluorescence were used for the PLS model (n=41). All ten PCs were used to explain 77.7% of the variance. Looking more carefully to the model's parameters, a better  $R^2$  was achieved for validation data (0.95) when compared to the training set (0.78). The presence of a higher validation slope (1.14) and the difference between RMSEP (4.1 ppm of nitrogen) and RMSECV (9.5 ppm of nitrogen) values indicate that the model to determine nitrate concentration

was not robust. Based on data dispersion (data not shown), only 15% of the data given to the model was higher than 20 ppm of nitrogen. To improve the quality of the model, more experiments should be performed, with more samples with higher concentrations of nitrate or more sampling points through time, so the consumption of nitrate could be more closely monitored.

During harvesting experiments, samples of the permeate were collected for further treatment, aiming to decrease the organic load. Nitrate concentration was monitored in these permeate samples before, during, and after the oxidation and membrane treatments were applied (oxidation, photodegradation, and nanofiltration). Ten PCs from each fluorescence spectra were used for the model construction (n=89). Although a higher explained variance (82.9%) was achieved when compared to the previous model (using nine of the ten PCs), lower values of validation  $R^2$  and slope were observed (0.65 and 0.73, respectively). Again, a significant difference was observed between RMSEP (9.9 ppm of nitrogen) and RMSECV (12.1 ppm of nitrogen) values, indicating poor quality of model fitting. When analysing the data dispersion (data not shown), it is possible to observe that the model predicted nitrate concentrations is worst in the range between 0 and 20 ppm, usually underestimating the concentration.

As observed in the two previous models, the compression of the EEMs in ten PCs is probably not enough to contain all patterns concerning nitrate concentration. Thus, for the development of the models with cultivation and permeate treatment data combined, fifteen PCs resultant from the PCA of the fluorescence EEMs were used.

Since the nitrate concentration in most of the culture experiment samples is lower than 20 ppm, and for treated permeate there are more samples with higher concentrations (> 20 ppm), both data sets were joined in the same model to increase the variability, the range of prediction, and so the applicability of this model. Also, 2D fluorescence spectroscopy would be a more robust monitoring tool if the same model for the same biological parameter could be applied in different processes.

Due to the high number of samples with nitrate concentrations near zero, two matrices were used to develop the nitrate concentration models, one with all the results (n=130) and a second one without those values close to zero concentration (n=58). The presence of negative values, close to zero, is explained by the calibration curve of the analytical equipment. These values were used directly in the model, instead of zero, since they reflect small differences in nitrate concentrations.

For the data matrix with all values, a linear model using fifteen PCs from the PCA of all fluorescence samples was studied. The  $R^2$  for training and validation sets were 0.80 and 0.82, with slopes of 1.04 and 1.00, respectively. Eleven of the fifteen PCs were used to explain 81.8% of the captured variance. Furthermore, the difference between RMSEP and RMSECV is lower in this model than in the other (Table 3).

For the matrix with only nitrate concentration values above zero, also a linear model using fifteen PCs from the PCA of all fluorescence samples was studied. The validation  $R^2$  was lower (0.68) than the training one (0.79), though both values were lower than expected. All fifteen inputs were used to explain 79.5% of the variance, and the RMSEP was higher (9.7 ppm of nitrogen) than the one found in the previous linear model, meaning that there is a difference of 9.7 ppm from the data point to the predicted value.

Therefore, when analysing the quality of the overall models using data from both processes (carotenogenesis and permeate treatment), with and without the data points of zero ppm of nitrogen, it is more advantageous to include the concentration values close to zero in the calibration of the correlating models (even if they are negative, due to the calibration of the analytical equipment). This is an advantage from an industrial perspective: considering that the induction of carotenoids in *D. salina* is reached by depletion of nitrogen, values near zero will be better predicted. However, the samples used for nitrate prediction models required a pre-treatment before being analysed by fluorescence spectroscopy, since the presence of cells needs to be removed by centrifugation. Nevertheless, the possibility of predicting nitrate concentration without the need for an elaborate and time-consuming protocol is an advantage in the daily routine of microalgae production.

The ability of 2D fluorescence spectroscopy for simultaneous detection of natural fluorophores in a complex culture medium and clean permeate samples emphasises the prospect of this technique as a multi-parameter online monitoring tool in distinct processes.

### ***3.4. Application perspectives***

The use of microalgae as a source of several high-value compounds has been studied for several authors and several constraints were already identified that make this industry less competitive than others, such as plant-based refineries. The use of better monitoring methodologies increases the knowledge about the system and helps the operator to take decisions at real-time. Several advantages are pointed when using spectrophotometric techniques,

such as fluorescence spectroscopy. The direct use of fluorescence is already a reality, although restricted to measuring only chlorophyll content through a specific pair of excitation-emission wavelengths [33]. As shown in this work, the potential of this technology is wider than that. The possibility of obtaining a fluorescence matrix increases the amount of information acquired from the system, and since each analysis can take up to 5 minutes, it is possible to have real-time control of the processes. It is possible to couple an optical probe directly immersed in the system for monitoring, like a bioreactor or a harvesting vessel, similar to temperature or pH probes. Several probes can be connected to the same spectrofluorometer (through a switch box), enabling the acquisition of fluorescence signal from different processes simultaneously. Some spectrofluorometer companies already provide software that allows the operator to choose the desired sampling times. This is already an advance in the use of fluorescence spectroscopy as a monitoring tool, but from the author's point of view, this technology would benefit from a user-friendly software where the results shown to the operator was already processed through the model.

#### **4. Conclusions**

This work demonstrates that 2D fluorescence spectroscopy is a strong tool to monitor cell concentration during cell cultivation and during harvesting by membrane filtration, even when the microalga cells change their internal products from “green” to “orange”, due to the production of carotenoids. Since the range of concentrations between these two processes is wide, two separate models (one for each process) show a better prediction at lower concentrations, although a single model was also robust. Measuring cell concentration online will improve the operational optimisation and control of *D. salina* cultivation, leading to the possibility of making decisions over the optimal point to start the induction of carotene production or to harvest.

Cell viability monitoring can lead to higher product recovery yields, since a higher percentage of viability is desirable at the end of the biomass recovery. The cell viability model based on fluorescence data showed a higher prediction ability for viabilities higher than 70%, which increases its potential in a biorefinery context.

The capability of 2D fluorescence spectroscopy to monitor nitrate concentration during different stages of the biorefinery, in complex matrices and clean permeates simultaneously, show the versatility of this technique in a *D. salina* pilot-scale production system.



Moreover, the possibility of acquiring a continuous information flux with the use of a fluorescence optical probe may contribute to better control of the operating conditions throughout the entire production chain, including the induction and recovery of biomass and added-value compounds.

### **Acknowledgments**

This work was supported by the Associate Laboratory for Green Chemistry- LAQV which is financed by national funds from FCT/MCTES (UID/QUI/50006/2019), by the European KBBE FP7 project “D-Factory”, by KAUST OSR award no. OSR-2016-CPF-2907-05, and by the follow Fellow grants of FCT/MCTES: SFRH/BPD/95864/2013, SFRH/BPD/79533/2011 and SFRH/BD/108894/2015. The authors would like to thank The Marine Biological Association (Devon, UK) and NBT Ltd (Israel). The authors would also like to thank the company A4F – Algae for future (Portugal), who performed all the pilot-scale cultivation trials and provided the biomass needed to develop this work.

### **Declaration of authors contributions**

MS was involved in the conception and design of the work, collection and assembly of samples and data, as well as in mathematical analysis and interpretation. JM was involved in the collection and assembly of samples and data and provided technical support. MS, JM and CB were involved in drafting the article. CFG, JGC and CB were involved in the conception and design of the work, CFG was involved in the mathematical interpretation of data, and CB provided administrative support. CG and JGC made the critical revision of the article for important intellectual content. All authors approved the final version of the article.

### **Declarations of interest: none**

No conflicts, informed consent, human or animal rights applicable

## References

- [1] A. Ben-Amotz, M. Avron, The Potential Use of *Dunaliella* for the Production of Glycerol,  $\beta$ -Carotene and High-Protein Feed, in: *Biosaline Res.*, Springer, 1982: pp. 207–214. doi:10.1007/978-1-4899-4998-1.
- [2] A. Ben-Amotz, M. Avron, Accumulation of Metabolites by Halotolerant Algae and its Industrial Potential, *Annu. Rev. Microbiol.* 37 (2003) 95–119. doi:10.1146/annurev.mi.37.100183.000523.
- [3] M.A. Borowitzka, L.J. Borowitzka, D. Kessly, Effects of salinity increase on carotenoid accumulation in the green alga *Dunaliella salina*, *J. Appl. Phycol.* 2 (1990) 111–119. doi:10.1007/BF00023372.
- [4] R.H. Wijffels, M.J. Barbosa, M.H.M. Eppink, Microalgae for the production of bulk chemicals and biofuels, *Biofuels, Bioprod. Biorefining.* 4 (2010) 287–295. doi:10.1002/bbb.
- [5] R. Bhave, T. Kuritz, L. Powell, D. Adcock, Membrane-based energy efficient dewatering of microalgae in biofuels production and recovery of value added co-products, *Environ. Sci. Technol.* 46 (2012) 5599–5606. doi:10.1021/es204107d.
- [6] J. Milledge, S. Heaven, A review of the harvesting of micro-algae for biofuel production, *Rev. Environ. Sci. Biotechnol.* 12 (2013) 165–178. doi:http://dx.doi.org/10.1007/s11157-012-9301-z.
- [7] R. Baker, *Membrane technology and applications*, Wiley Online Library, 2000. doi:http://dx.doi.org/10.1002/0471238961.1305130202011105.a01.
- [8] M.R. Bilad, H.A. Arafat, I.F.J. Vankelecom, Membrane technology in microalgae cultivation and harvesting: A review, *Biotechnol. Adv.* 32 (2014) 1283–1300. doi:10.1016/j.biotechadv.2014.07.008.
- [9] J. Trivedi, M. Aila, D.P. Bangwal, S. Kaul, M.O. Garg, Algae based biorefinery - How to make sense?, *Renew. Sustain. Energy Rev.* 47 (2015) 295–307. doi:10.1016/j.rser.2015.03.052.
- [10] R.J. Radmer, B.C. Parker, Commercial applications of algae: opportunities and constraints, *J. Appl. Phycol.* 6 (1994) 93–98. doi:10.1007/BF02186062.
- [11] J. Monte, M. Sá, C.F. Galinha, L. Costa, H. Hoekstra, C. Brazinha, J.G. Crespo, Harvesting of *Dunaliella salina* by membrane filtration at pilot scale, *Sep. Purif. Technol.* 190 (2018) 252–260. doi:10.1016/j.seppur.2017.08.019.
- [12] J. Fret, L. Roef, R. Blust, L. Diels, S. Tavernier, W. Vyverman, M. Michiels, Reuse of rejuvenated media

- during laboratory and pilot scale cultivation of *Nannochloropsis* sp., *Algal Res.* 27 (2017) 265–273. doi:doi:10.1016/j.algal.2017.09.018.
- [13] V. Discart, M. Bilad, L. Marbelia, I.F.J. Vankelecom, Impact of changes in broth composition on *Chlorella vulgaris* cultivation in a membrane photobioreactor (MPBR) with permeate recycle., *Bioresour. Technol.* 152 (2014) 321–328. doi:doi:10.1016/j.biortech.2013.11.019.
- [14] U. von Gunten, Ozonation of drinking water: part I. Oxidation kinetics and product formation, *Water Res.* 37 (2003) 1443–1467. doi:doi:10.1016/S0043-1354(02)00457-8.
- [15] J. Monte, M. Sá, C. Parreira, J. Galante, A.R. Serra, C.F. Galinha, L. Costa, V.J. Pereira, C. Brazinha, J.G. Crespo, Recycling of *Dunaliella salina* cultivation medium by integrated membrane filtration and advanced oxidation, *Algal Res.* 39 (2019) 101460. doi:10.1016/j.algal.2019.101460.
- [16] R. Bosma, E. van Zessen, J.H. Reith, J. Tramper, R.H. Wijffels, Prediction of volumetric productivity of an outdoor photobioreactor, *Biotechnol. Bioeng.* 97 (2007) 1108–1120. doi:10.1002/bit.21319.
- [17] J. Quinn, L. de Winter, T. Bradley, Microalgae bulk growth model with application to industrial scale systems., *Bioresour. Technol.* 102 (2011) 5083–5092. doi:10.1016/j.biortech.2011.01.019.
- [18] M. V. Córdoba-Matson, J. Gutiérrez, M.Á. Porta-Gándara, Evaluation of *Isochrysis galbana* (clone T-ISO) cell numbers by digital image analysis of color intensity, *J. Appl. Phycol.* 22 (2010) 427–434. doi:10.1007/s10811-009-9475-0.
- [19] P.P. Lamers, C.C.W. van de Laak, P.S. Kaasenbrood, J. Lorier, M. Janssen, R.C.H. De Vos, R.J. Bino, R.H. Wijffels, Carotenoid and fatty acid metabolism in light-stressed *Dunaliella salina*, *Biotechnol. Bioeng.* 106 (2010) 638–648. doi:10.1002/bit.22725.
- [20] C.F. Galinha, G. Carvalho, C. a M. Portugal, G. Guglielmi, M. a M. Reis, J.G. Crespo, Multivariate statistically-based modelling of a membrane bioreactor for wastewater treatment using 2D fluorescence monitoring data., *Water Res.* 46 (2012) 3623–36. doi:10.1016/j.watres.2012.04.010.
- [21] E. Benito-Peña, M.G. Valdés, B. Glahn-Martínez, M.C. Moreno-Bondi, Fluorescence based fiber optic and planar waveguide biosensors. A review, *Anal. Chim. Acta.* 943 (2016) 17–40. doi:10.1016/j.aca.2016.08.049.

- [22] Y.S. Shin, H. Il Choi, J.W. Choi, J.S. Lee, Y.J. Sung, S.J. Sim, Multilateral approach on enhancing economic viability of lipid production from microalgae: A review, *Bioresour. Technol.* 258 (2018) 335–344. doi:10.1016/j.biortech.2018.03.002.
- [23] M. Sá, J. Monte, C. Brazinha, C.F. Galinha, J.G. Crespo, 2D Fluorescence spectroscopy for monitoring *Dunaliella salina* concentration and integrity during membrane harvesting, *Algal Res.* 24 (2017) 325–332. doi:10.1016/j.algal.2017.04.013.
- [24] W.E. Federation., W.P.C. Federation., A.W.W. Association., A.P.H. Association., Standard methods for the examination of water and wastewater., (1980).
- [25] R. Leardi, Chemometric Methods in Food Authentication, in: *Mod. Tech. Food Authentication*, Elsevier, 2018: pp. 687–729. doi:10.1016/B978-0-12-814264-6.00017-7.
- [26] J. Sádecká, J. Tóthová, Fluorescence spectroscopy and chemometrics in the food classification &minus; a review, *Czech J. Food Sci.* 25 (2008) 159–174. doi:10.17221/687-CJFS.
- [27] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130. doi:10.1016/S0169-7439(01)00155-1.
- [28] R. Boggia, M. Forina, P. Fossa, L. Mosti, Chemometric study and validation strategies in the structure-activity relationships of new cardiogenic agents, *Quant. Struct. Relationships.* 16 (1997) 201–213. doi:10.1002/qsar.19970160303.
- [29] T.P. Ryan, *Modern Regression Methods*, John Wiley & Sons, Inc., 1997.
- [30] C. Duchesne, J.F. MacGregor, Jackknife and bootstrap methods in the identification of dynamic models, *J. Process Control.* 11 (2001) 553–564. doi:10.1016/S0959-1524(00)00025-1.
- [31] C. a Andersson, R. Bro, The N-way Toolbox for MATLAB, *Chemom. Intell. Lab. Syst.* 52 (2000) 1–4. doi:10.1016/S0169-7439(00)00071-X.
- [32] C.F. Galinha, G. Carvalho, C. a. M. Portugal, G. Guglielmi, M. a. M. Reis, J.G. Crespo, Two-dimensional fluorescence as a fingerprinting tool for monitoring wastewater treatment systems, *J. Chem. Technol. Biotechnol.* 86 (2011) 985–992. doi:10.1002/jctb.2613.
- [33] Y.-H. Shin, J.Z. Barnett, E. Song, M.T. Gutierrez-Wing, K.A. Rusch, J.-W. Choi, A portable fluorescent

**Tables**

Table 1: Operational conditions of membrane filtration of *D. salina*. Parameters tested: permeate volumetric flux imposed; type of membrane according to the pore size; and membrane relaxation procedure (N - without relaxation, y - 9 minutes of operation and 1 minute of relaxation). A consecutive increment of biomass was performed twice (Y) (n - no biomass increment).

Permeate Flux (L/m <sup>2</sup> .h)	Memb. pore size (kDa)	Memb. relaxation (y/N)	Consecutive biomass increment (Y/n)	Car/Chl*	Cell Concentration (cells/mL)
<b>30</b>	100	y	n	14.2	1.6E+05
<b>25</b>	100	N	n	12.5	7.9E+05
<b>25</b>	500	y	n	7.7	7.2E+05
<b>25</b>	100	y	n	4.6	9.3E+05
<b>25</b>	100	y	n	3.0	1.6E+05
<b>25</b>	100	y	n	2.2	5.5E+05
<b>15</b>	100	y	n	5.1	5.9E+05
<b>10</b>	100	y	Y	5.5	6.1E+05
<b>10</b>	100	y	Y	4.5	3.3E+05
<b>10</b>	100	y	n	5.6	6.3E+05
<b>10</b>	100	y	n	7.3	4.7E+05

\* carotenoids to chlorophylls ratio

Table 2: Input strategies used for models construction of cellular concentration, cellular viability and nitrate concentration prediction, and the corresponding prediction model representation.

Output	Inputs			Prediction model
	Operational Process	<i>n</i>	Strategy PCs	
Cellular concentration	Carotenogenesis	41	10 PCs	Fig 4 (a)(b)
	Harvesting	138	10 concentrate PCs	Fig 4 (c)(d)
	Carotenogenesis+ Harvesting	179	10 PCs	Fig 4 (e)(f)
Cellular viability	Harvesting	138	10 concentrate PCs	Fig 5 (a)(b)
			10 concentrate + 10 permeate PCs	Fig 5 (c)(d)
Nitrate concentration	Carotenogenesis	41	10 PCs	.
	Permeate treatment	89	10 PCs	-
	Carotenogenesis + Permeate treatment	130	15 PCs	-
		58	15 PCs	-

PCs – principal components;

*n* – number of samples.

Table 3: Statistical parameters of the selected models for nitrate concentration prediction.

	Nr of inputs	<i>n</i>	Var (%)	RMSECV (ppm N)	RMSEP (ppm N)	Validation		Training		Selected Fluorescence PCs
						R <sup>2</sup>	Slope	R <sup>2</sup>	Slope	
Carotenogenesis	10	41	77.7	9.5	4.1	0.95	1.14	0.78	1.00	1 2 3 4 5 6 7 8 9 10
Permeate treatment	10	89	82.9	12.1	9.9	0.65	0.73	0.83	1.01	1 2 3 4 6 7 8 9 10
Carotenogenesis + Permeate treatment	15	130	81.8	8.6	7.0	0.80	1.04	0.82	1.00	1 2 4 5 8 9 10 11 12 13 15
	15	58	79.5	11.4	9.7	0.68	1.05	0.79	1.00	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

*n* – number of samples;

Var – variance captured by the model;

RMSECV – root mean square error of cross-validation;

RMSEP – root mean square error of prediction;

PCs – principal components.

## Figures

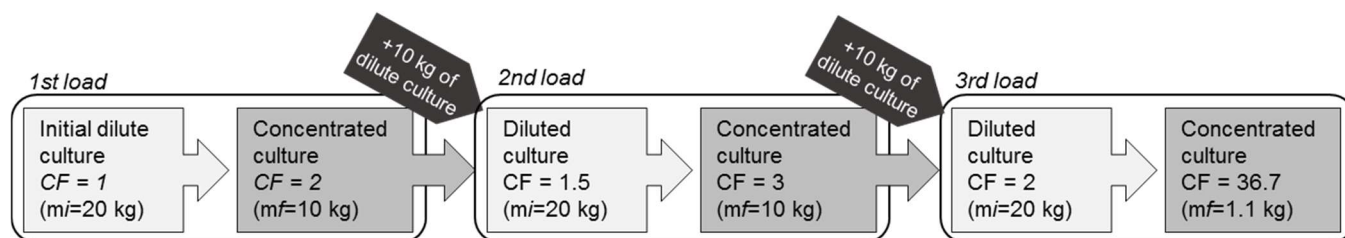
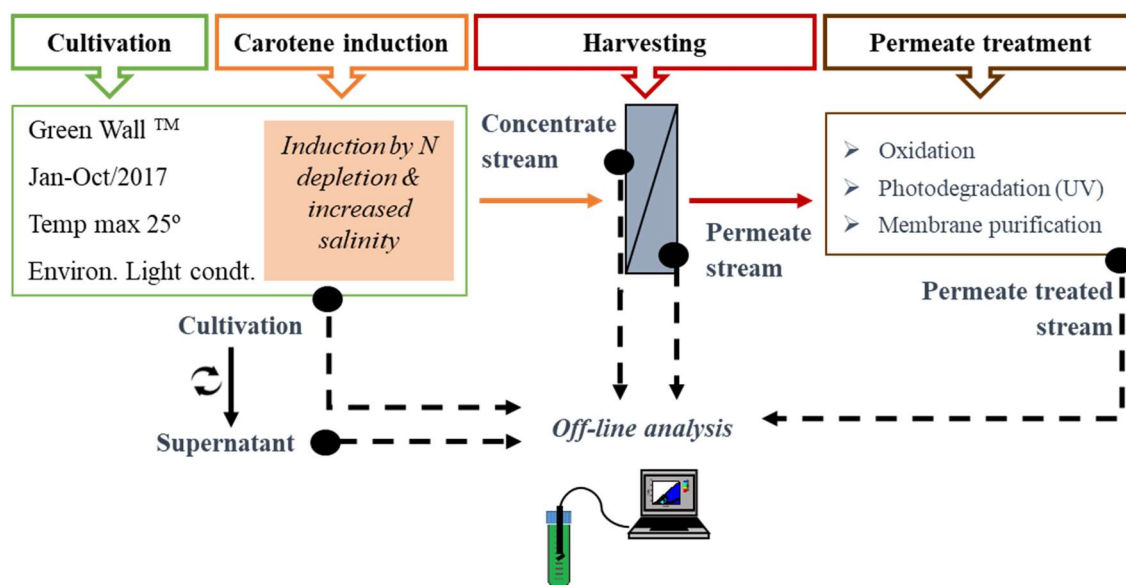


Figure 1: Schematic representation of the procedure used during the consecutive membrane harvesting experiments. Concentration factors (CF) during the experiment were calculated in function of mass of culture, in the beginning of the load ( $m_i$ ) and in the end of the load ( $m_f$ ).



Modelling strategies		
<p><b>Cell concentration</b></p> <ol style="list-style-type: none"> <li>10 PCs from cultivation (n=41)</li> <li>10 PCs from concentrate stream (n=138)</li> <li>10 PCs from cultivation and concentrate stream combined (n=179)</li> </ol>	<p><b>Cell viability</b></p> <ol style="list-style-type: none"> <li>10 PCs from concentrate stream (n=138)</li> <li>10 PCs from the concentrate + 10 PCs from the permeate (n=138)</li> </ol>	<p><b>Nitrogen concentration</b></p> <ol style="list-style-type: none"> <li>10 PCs of the supernatant of cultivation (n=41)</li> <li>10 PCs of the permeate treated stream (n=89)</li> <li>15 PCs of the supernatant and permeate stream combined (n=130)</li> </ol>

Figure 2: Schematic representation of the overall process, with indication of the sampling points, and of streams used in different modelling strategies. Samples were withdrawn from cultivation broth and its supernatant from



carotenogenesis, from the concentrate and permeate streams during membrane filtration, and from permeate treatment. The *PCs* are the principal components used from the principal component analysis performed to the fluorescence spectra;  $n$  are the number of samples used for each model.

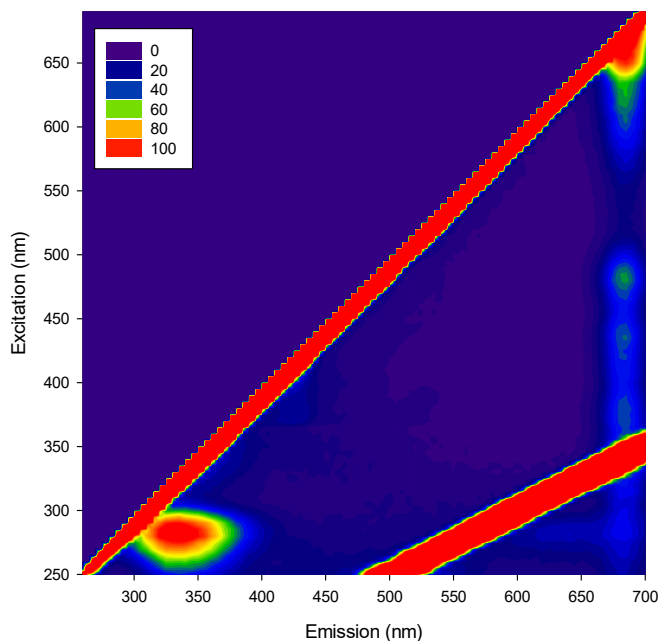


Figure 3: Fluorescence spectrum of *Dunaliella salina* biomass, acquired during carotenogenesis experiments. The emission wavelengths is in the x-axis, the excitation on the y-axis, and the fluorescence intensity in colour-grade scale. Two distinct areas can be identified: protein-like region for excitation wavelengths between 250 and 350 nm, and emission below 300 nm; and the pigment band, for excitation wavelengths above 650 nm through the entire range of emission wavelengths.

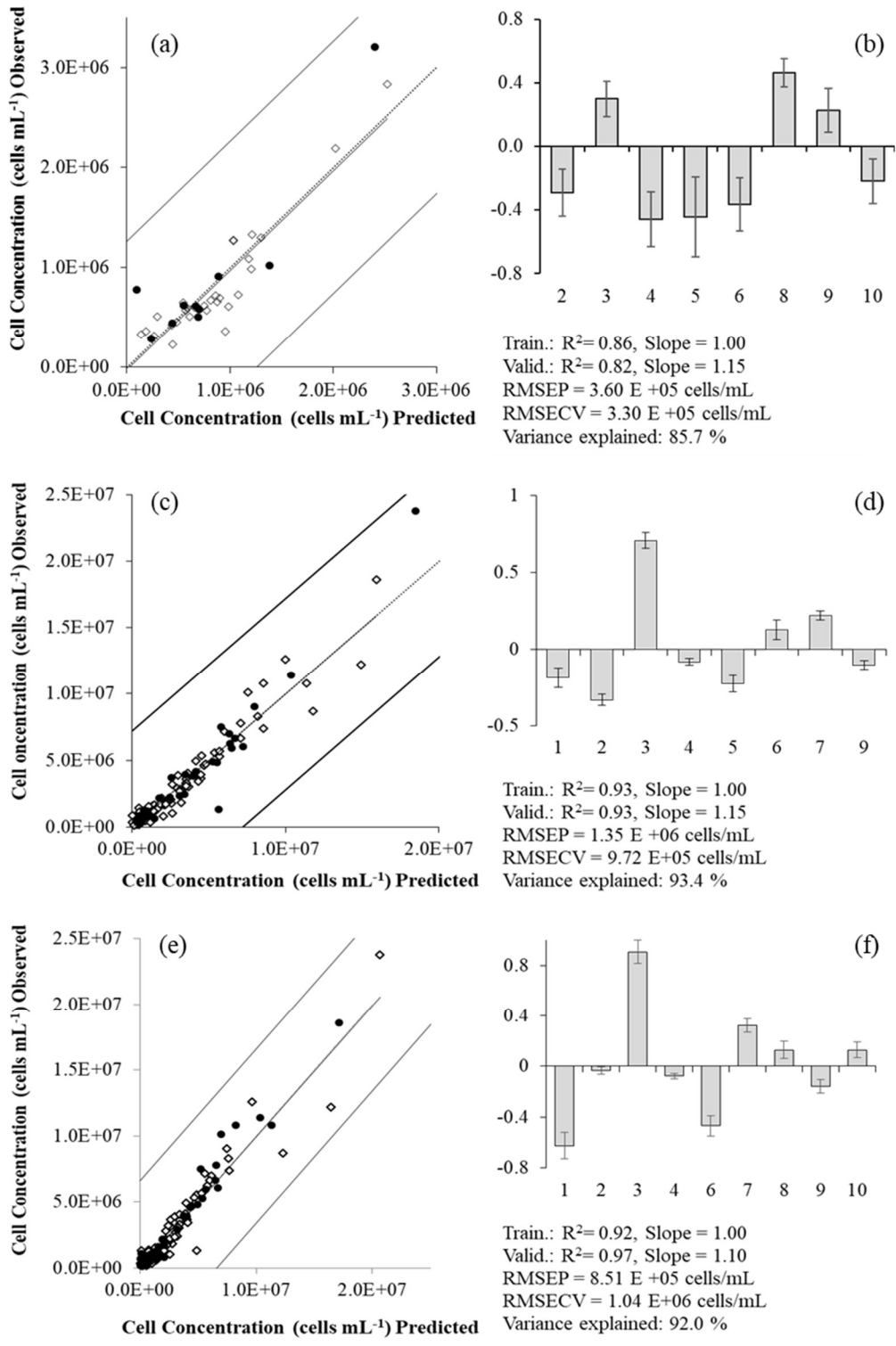


Figure 4: Prediction models for cell concentration obtained (a) only with cultivation experiments samples; (c) only with membrane harvesting experiments, using principal components (PCs) from fluorescence acquired in the concentrate stream; and (e) with cultivation and harvesting experiments, using fluorescence from both processes

compiled together with the same principal component analysis. Respective normalised regression coefficients are shown in (b), (d) and (f), in the same order (each number represent the PC used as input). Training ( $\diamond$ ) and validation ( $\bullet$ ) data are presented as cell/mL. Model (a) training set included 31 fluorescence spectra, and 10 for validation set. Model (c) training set included 104 fluorescence spectra, and 34 for validation set. Model (e) training set included 135 fluorescence spectra, and 44 for validation set. RMSECV – root mean square error of cross-validation; RMSEP – root mean square error of prediction.

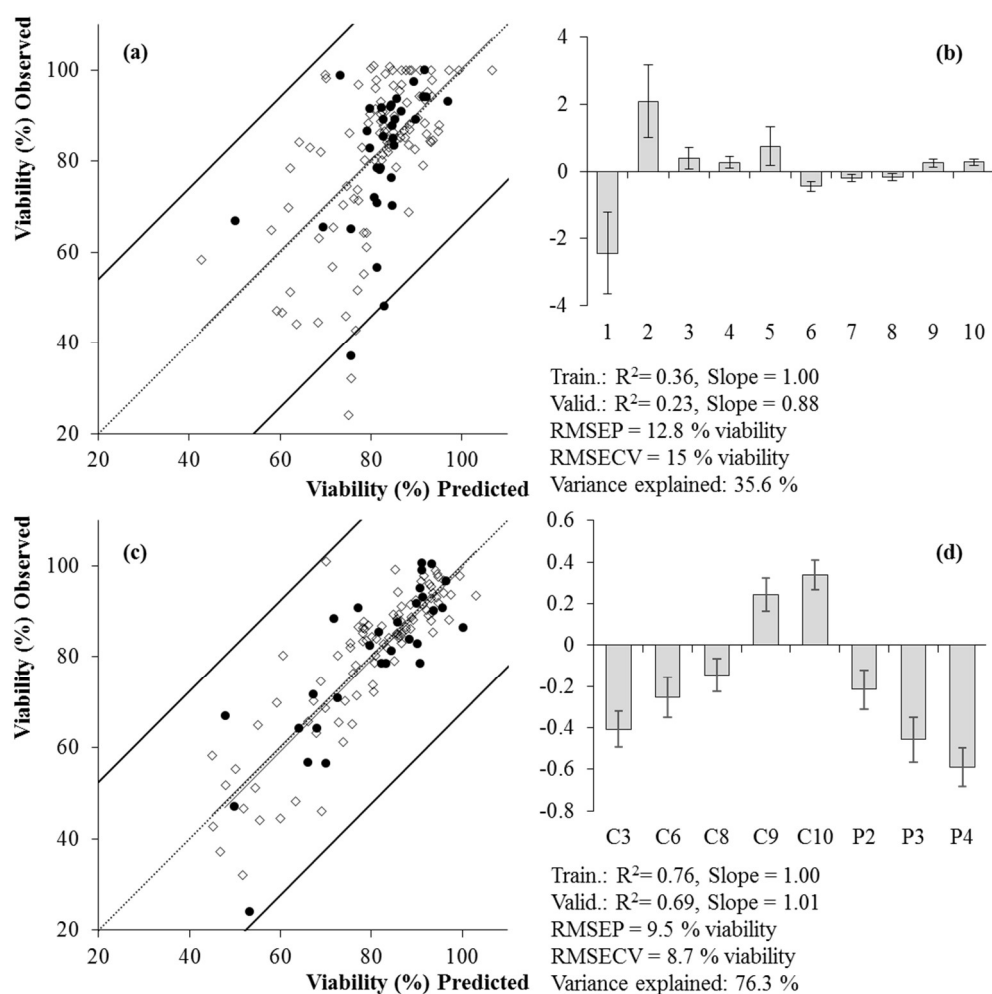


Figure 5: Prediction models for cell viability percentage during membrane harvesting experiments: (a) using ten principal components (PCs) of the fluorescence acquired in the concentrate stream and (c) adding ten PCs of the fluorescence acquired in the permeate stream to the ten PCs resulting from the concentrate stream. The respective

regression coefficients (b and d) are in normalised units. Each number represent the PC used as input, the letters "C" or "P" before the PC number indicates its origin, from concentrate or permeate streams, respectively. Training ( $\diamond$ ) (n=104) and validation ( $\bullet$ ) (n=34) data are presented as percentage of viability. RMSECV – root mean square error of cross-validation; RMSEP – root mean square error of prediction.