Marta Filipa Batista de Sá

Mestre em Engenharia Alimentar

Monitoring of biological processes in microalgae production using Fluorescence Spectroscopy

Dissertação para obtenção do Grau de Doutor em
Engenharia Química e Bioquímica – Especialidade Engenharia Bioquímica

Orientadora: Cláudia Galinha Loureiro
Investigadora
Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa

Co-orientadores: João Goulão Crespo
Professor Catedrático
Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa

Maria Barbosa
Professora Catedrática
Wageningen University & Research

Júri

Presidente: Prof. Doutora Maria da Ascensão M. Reis

Arguentes: Prof. Doutor Olivier Gonçalves
Prof. Doutor João A. Lopes

Vogais: Doutor Luis da Costa
Doutora Cláudia Galinha Loureiro

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2019

Monitoring of biological processes in microalgae production using Fluorescence Spectroscopy

"*(…) It matters not how strait the gate,*
How charged with punishments the scroll,
I am the master of my fate,
I am the captain of my soul."

Invictus in *Book of Verses*,
William Ernest Henley (1888).

Acknowledgements

O doutoramento foi uma etapa na minha vida que contribuiu para o meu crescimento profissional e pessoal. São anos de muito trabalho e dedicação, altos e baixos, mas o que se recorda são as conquistas e alegrias e o sentimento de realização. Este percurso teria sido insuportável sem o apoio de diversas pessoas às quais gostaria de expressar o meu profundo agradecimento.

Antes de mais, gostaria de agradecer aos meus orientadores Cláudia Galinha e João Crespo, por terem acreditado em mim e me terem dado a oportunidade de realizar este trabalho. Mas sobretudo, ficarei para sempre grata por me terem dado a liberdade que precisei para crescer como investigadora, para seguir as minhas ideias e traçar o percurso do meu doutoramento. Gostaria de agradecer ao João pela oportunidade que me proporcionou de participar em conferências e pelo incentivo de fazer um estágio no estrangeiro. Cláudia, espero que, como primeira aluna, não tenha sido muito difícil de aturar! Muito obrigada por estares sempre disponível para mim, obrigada pelo apoio e por teres sempre uma palavra amiga.

À minha orientadora Maria Barbosa, faltam-me as palavras para expressar o quão grata estou por me ter acolhido! Obrigada pelo voto de confiança, por acreditares em mim, pelo carinho e preocupação com o meu bem-estar. És uma enorme inspiração para qualquer mulher na ciência, e estou extremamente grata pela aprendizagem e experiência que adquiri por trabalhar lado-a-lado contigo.

Um agradecimento a todos os colaboradores da A4F – Algae for Future, em especial ao Dr. Luís Costa, à Celina Parreira, ao Pedro Fonseca e à Joana Galante, pela partilha de conhecimentos e boa disposição com que sempre me receberam.

Aos meus colegas do BPEG agradeço a boa disposição, entreajuda e oportunidade de partilhar bons momentos. Ao longo destes anos várias pessoas passaram pelo grupo que deixaram um carinho muito especial. Um muito obrigado pela boa disposição e companheirismo, à Catarina Oliveira, ao Christophe Roca, à Carla Daniel e à Joana Cassidy.

Luísa Neves, obrigada pelo apoio incondicional que me deste nas fases mais difíceis destes últimos anos. Obrigada por estares sempre disposta a ouvir, pelo teu ombro amigo, pelos teus conselhos. Mónica Carvalheira, obrigada por estares sempre disponível para me ajudares no que fosse preciso. Nunca esquecerei as nossas infindáveis conversas no infernal trânsito de Lisboa! Bruno Oliveira, Tiago Costa e Eliana Guarda (tuxinha!) obrigada pela vossa boa disposição. Meus queridos, vocês têm uma energia positiva incrível. Partilhar laboratório convosco é uma alegria, mas ainda melhor é poder fazer parte das vossas vidas e ver-vos "crescer"!

Rita Ferreira, muito obrigada por tudo! Em ti encontrei uma verdadeira amiga, para os maus momentos, mas sobretudo para os momentos espetaculares. Obrigada pela tua amizade, pelo exemplo de mulher forte e independente que és!

Ricardo Marques, recordo tantos momentos que passámos juntos. Contigo partilhei caminhadas, campismos, cervejinhas ao fim do dia… festejámos santos populares, passagens de ano, festivais de verão... contigo partilhei frustrações, indecisões, profissionais e pessoais, e

tantas alegrias! Irei recordar-te para sempre, pelas tuas teorias malucas, mas sobretudo pela tua enorme amizade. Muito obrigada amigo!

In Wageningen I found a second home and I would like to thank all my friends and colleagues for make me feel so welcome. Special thanks to my “work-husband” Narcis Ledo, for all the knowledge you shared with me and the immense patience; to Robin Barten for all the fun moments and conversations, even when you called me old!; to Pieter Oostlander, for listening to my worries, for the amazing company in our road trip in Scotland, for all the amazing gaming nights; to Barbara Guimarães, pela tua alegria contagiante, pelos teus abracinhos! To Snezana Gegic and Fred van den End, the best technicians a PhD student could ask for! Thank you for your big patience and giant humour.

Iago Teles, obrigada por me aturares, especialmente na fase intensa de escrita. É bom ter alguém com quem posso rir e brincar e resmungar, de porta fechada e em português! Obrigada pela tua boa disposição, e por partilhares comigo as expressões de sabedoria da tua vovó!

A very warm thanks to Wendy Evers and Jorijn Janssen, who helped me so much in so many moments. It’s amazing to know that I can always count on you, girls! Thank you both for the immense support during my stay in WUR, not only in the lab but also in the coffee corner! Your joy and company always lighten up my days!

Marek Wazynski, with you we stablish our home in this foreign country. With your help and support, every challenge it’s easier to endure. Thank you for brightening my every day, thank for your love.

Às minhas amigas do peito, Erica Cabral, Alzira Ramos e Isabel Rosa, que me acompanham nesta jornada há tantos anos! Não importa em que parte do mundo nos encontremos, o grupo “Rocking the 30s” é a melhor coisa de sempre… O grupo com os temas mais loucos e variados, de alguma sabedoria, às vezes todos em simultâneo! Adoro-vos! Sem vocês, esta etapa da minha vida teria sido muito mais aborrecida e pesada. Obrigada!

Aos meus meninos de Leiria, Tiago Gonçalves, João Órfão e Ana Ramos, obrigada por fazerem parte da minha vida desde os tempos de liceu! Obrigada pelos cafezinhos na praça e pelas conversas animadas. Ao Tiago Mora, não há palavras para descrever o quanto agradeço o teu apoio incondicional. Obrigada por teres estado sempre do meu lado, nos bons e nos maus momentos.

E por fim à minha família, sobretudo ao meu mano e à minha mãe. Obrigada mãezinha por seres o meu porto de abrigo, obrigada pelo teu apoio incondicional em todas as decisões da minha vida, obrigada por tudo o que fizeste por mim, para que eu tivesse todas as oportunidades possíveis. Amo-vos muito.

Abstract

Microalgae industrial production is nowadays viewed as a solution for environmental conscious and sustainable alternative production of fuel, feed, food and chemicals. Throughout the years, several technological advances have been studied and implemented that increased the competitiveness of microalgae production. However, online monitoring and a real-time process control of a microalgae production factory still requires development to support economic sustainability.

In this work, fluorescence spectroscopy coupled with chemometric modelling is studied as an online monitoring tool to be used in microalgae production. Fluorescence spectroscopy is a non-invasive and highly sensitive technique, able to detect instantaneously several natural fluorophores but also the interferences between them and the environmental media. Chemometric methods are often used to deconvolute the information within the fluorescence matrices, known as excitation-emission matrices (EMMs), and to determine the relationship between them and the parameters to be monitored.

To prove the potential of fluorescence spectroscopy coupled with chemometric modelling techniques, different strategies are studied. Firstly, the EEMs of the spectra are used as raw data, without pre-treatment for removal of water scatter and inner-filter effects. Principal Component Analysis (PCA) is used to extract the meaningful information from the spectra, resulting in Principal Components (PCs). Through Projection to Latent Structures (PLS) modelling, prediction models are developed using the PCs from the fluorescence EEMs as inputs, to find linear correlations with the parameters to be monitored, the outputs. A second strategy is studied with pre-treated EEMs. With these EEMs, two input strategies in the PLS models are tested: using directly the EEMs in PLS or compressing the EEMs into PCs though PCA prior to PLS.

Two marine microalgae are used in these studies, *Dunaliella salina and Nannochloropsis oceanica*. Five parameters are monitored – cell concentration, cell viability, pigments concentration, fatty acids composition and nitrogen concentration – in four different processes – cultivation, product formation (carotenoids and lipids), harvesting by membrane filtration and permeate recover.

The combination of fluorescence spectroscopy, with its high sensitivity and resolution, coupled with chemometric analysis for data pre-treatment and development of prediction models, enhances the knowledge and the possibility for real-time process control in microalgae production systems.

Keywords: online monitoring, fluorescence spectroscopy, Principal Component Analysis, Projection to Latent Structures, microalgae production, *Dunaliella salina*, *Nannochloropsis oceanica*.

Resumo

A produção industrial de microalgas é atualmente vista como uma solução para a produção ambientalmente responsável e sustentável nas áreas dos combustíveis, alimentação humana e animal, e na produção de diversos compostos químicos. Ao longo dos anos, foram estudados e implementados vários avanços tecnológicos que permitiram aumentar a competitividade da produção de microalgas. No entanto, ainda é necessário algum desenvolvimento na monitorização *in situ* e em tempo real de forma a assegurar um controlo mais preciso da produção de microalgas e a sua sustentabilidade económica.

Nesta tese é estudada uma ferramenta de monitorização em tempo real para a produção de microalgas, tendo por base a espectroscopia de fluorescência acoplada a quimiometria. A espectroscopia de fluorescência é uma técnica não invasiva, com grande sensibilidade de deteção e especificidade. Permite detetar instantaneamente vários fluoróforos naturais, assim como as interferências entre eles e o meio de cultura. Frequentemente são usados métodos de quimiometria para extrair e deconvoluir a informação contida nas matrizes de fluorescência, conhecidas como matrizes de excitação-emissão ("*excitation-emission matrices*", EEMs), e estabelecer a relação entre elas e os parâmetros a monitorizar.

Foram estudadas várias estratégias para demonstrar o potencial da espectroscopia de fluorescência acoplada a métodos quimiométricos. Em primeiro lugar, as EEMs foram usadas em bruto, sem pré-tratamento para a remoção do efeito de dispersão da radiação no meio contínuo (água) e do efeito de "filtro interno" no meio de cultura ("*inner filter effect*"). Análise de componentes principais ("*Principal Component Analysis*", PCA) é então usada para extrair informação das matrizes, resultando em componentes principais ("*Principal Components*", PCs). Por fim, foram desenvolvidos modelos de previsão usando as variáveis observadas como informação fornecida ("*inputs"*) (os PCs das EEMs da fluorescência) através de projeção de estruturas latentes ("*Projection to Latent Structures*", PLS), para encontrar correlações lineares com os parâmetros a monitorizar ("*outputs*"). Foi estudada uma segunda estratégia com o uso de EEMs pré-tratados, ou seja, sem efeito de dispersão da água e de filtro interno. Com estes EEMs pré-tratados, foram estudadas duas estratégias de *fornecimento de informação ("inputs")* nos modelos PLS: usando os EEMs directamente no PLS ou depois de serem compactados em PCs através de PCA antes do PLS.

Nesta tese foram estudadas duas microalgas marinhas, *Dunaliella salina* e *Nannochloropsis oceanica*. Foram monitorizados cinco parâmetros – concentração celular, viabilidade celular, concentração de pigmentos, composição em ácidos gordos e concentração de azoto – em quatro processos diferentes – cultivo, formação do produto (carotenoides ou lípidos), colheita por filtração através de membranas e recolha de permeado.

A combinação de espectroscopia de fluorescência, de alta sensibilidade e resolução, associada à quimiometria para o pré-tratamento e desenvolvimento de modelos de previsão, aumenta o conhecimento e a possibilidade de controle em tempo real de processos na produção de microalgas.

Palavras-chave: monitorização em tempo real, espectroscopia de fluorescência, Análise de componentes principais, projecção de estruturas latentes, produção de microalgas, *Dunaliella salina*, *Nannochloropsis oceanica*.

Contents

List of Figures

List of Tables

List of Acronyms

Abbreviations

2D – Two Dimensional

AM – Analytical Method

ANN – Artificial Neural Networks

ASW – Artificial Sea Water

Car – Carotenoids

Cc – Cell concentration

Chl – Chlorophyll; Chl a – Chlorophyll a, Chl b – Chlorophyll b

Cumul. – Cumulative

Cv – Cell viability

CV – Cross-Validation

DW – Dry Weight

ECMWF – European Centre for Medium-Range Weather

EEM – Excitation-Emission Matrix

EPA – Eicosapentaenoic Acid

FA – Fatty Acids

Fluor.PCs – Fluorescence PCs

FPAR – Fraction of Photosynthetically Active Radiation

IR – Infrared Spectroscopy

ISE – Iterative Stepwise Elimination

LOOCV – Leave-One-Out Cross-Validation

LV – Latent Variable

MAPLASS – Multivariate Additive PLS Splines

MF – Microfiltration

MLR – Multiple Regression Regression

MODIS – Moderate Resolution Imaging Spectroradiometer

MTBE – Methyl-tert-butylether

MWCO – Molecular weight cut-off

NIR – Near-to short-wave Infrared Spectroscopy

NL – Neutral Lipids

n-PLS – multiway version of PLS

N-starv – Nitrogen depletion (“starvation”)

PAHs – Polycyclic Aromatic Hydrocarbons

PAM – Pulse Amplitude Modulated fluorometry

PC – Principal Component

PCA – Principal Component Analysis

PCR – Principal Component Regression

pCO_2 – CO_2 Partial pressure

PI – Propidium Iodide

PL – Polar Lipids

PLS – Projection to Latent Structures

pO_2 – O_2 Partial pressure

PUFA – Polyunsaturated Fatty Acids

R^2 – Coefficient of determination

RMSECV – Root Mean Square Error of Cross-Validation

RMSEP – Root Mean Square Error of Prediction

ROS – Reactive Oxygen Species

Sl – Slope

TAG – Triacylglycerols

UF – Ultrafiltration

Introduction

I.1. Challenges in monitoring biological systems

Due to the increased interest for greener and renewable bio-based economies, biomass is being valorised for the sustainable production of food and feed, but also chemicals, fuels and power. Through the years, several technological advances have been studied and implemented that increased the competitiveness of biomass production. The introduction of closed cultivation systems, like bioreactors, enabled the production under defined and controlled conditions, leading to optimised viability, reproducibility and higher productivities. Following this trend for more competitive production systems, the need of online monitoring emerged.

Within the biotechnological processes involved in a cell-based biorefinery, three types of parameters need to be monitored: physical, such as temperature and conductivity; chemical, like pH, O_2 and CO_2 partial pressure (pO_2 and pCO_2); and biological, as cell concentration and viability and products concentration. It is important to take into consideration that interactions between those three classes of parameters can occur, and that these interactions are usually complex. Also, in these biological systems, the measured parameters are usually less than the parameters that need to be controlled, increasing the difficulty to monitoring and control the process [1,2]. Currently most of the monitoring is based on off-line analysis of samples withdrawn from the cultivation or biorefinery processes. Usually the sampling frequency is conservative to not disturb the system or contaminate it. All metabolic activities in the sample need to be stopped to reflect the cell status at that specific time. In addition, some results depend on time consuming analytical procedures, making this sampling strategy not appropriate for direct process control [3].

Therefore, online monitoring and a closer process control of bioreactors is a basic requirement for the development of efficient biological processes [4,5]. With online sensors it is possible to acquire a continuous stream of information with short response times. In this way, biological systems can be monitored faster and more efficiently, allowing an immediate response, if needed. This increases the control on production, leading to an improved production processes with a high-quality control.

The selection of the right sensor for the online monitoring of biological processes must match the needs of the bioprocess in question. The most general requirements are selectivity, sensitivity and response time [2,6,7]. Accordingly to Glindkamp *et al.* [2] three configurations have been explored so far: *in situ,* where the medium is monitored directly; online, when the sensor is moved to a particular part of the bioreactor, like a bypass, so the bubbles from aeration do not interfere (for example); and off-line*,* removed from the bioreactor.

Nowadays, the most common *in situ* and online sensors used in biotechnology are based in electrochemical principles like pH, pO2, conductivity, and are well-known in the field. However, the need of an efficient tool to monitor simultaneously a wider range of substrates and products,

and to control the cultivation environment, increased the urgency for better solutions, and the use of optical sensors for *in situ* and online monitoring is emerging.

I.2. Optical sensors - Techniques available

With the development of better optical fibres, able to be used in larger distances and in longer communication ranges, the optical sensor technology became very promising for online monitoring. The principle of spectroscopic optical fibre sensors is based on the interaction between light waves (absorption or luminescence) with the molecules. This technique presents several advantages for online monitoring of complex biological systems, as for example industrial microalgae production and refinery. These sensors are non-invasive, non-destructive and allow the detection of several different molecules simultaneously. The need for sampling the system is low or inexistent, so it decreases the culture contamination, and no time delay is observed when acquiring the data, making them a great solution for real-time monitoring. Also, as non-invasive techniques, they will not interfere with the biological material, allowing to monitor *in vivo* cells and obtain information about their intracellular state, in addition to information about extracellular media [1,2,4,8].

The spectroscopic optical fibre sensors can be divided in four categories based on the way they interact with the sample. They can be used combined with an indicator for specific molecules (usually fluorescent dyes) [1], coupled to a biological receptor like a catalytic effect (enzyme-based) or comprise immune and gene sensors (antibodies) [4], or, in the most simple case, analyse the optical properties of the sample [9]. This last approach is more adequate when aiming to monitor microalgae cultivation and biorefinery online, since no interaction with the culture is required and several metabolites can be analysed at the same time.

Most spectroscopic methods, including infrared, Raman and fluorescence spectroscopies, have been study for (bio)processes monitoring, often in combination with optical fibres [9].

I.2.1. Infrared spectroscopy

Infrared (IR) spectroscopy has been used in bioprocess monitoring for the measurement of several metabolites, such as ethanol, glucose and fructose. Since most bioprocesses take place in aqueous medium, and due to the huge IR absorbance of water (for wavelengths higher than 2500 nm), this technique can only be applied with a short optical path length or in the range of near-to short-wave IR range (NIR), from 780 to 2500 nm [2].

NIR spectroscopy provides information about the structure composition of the molecules (position, shape and size) through the detection of the biological bonds C-H, N-H, O-H and S-H. The absorption bands can be weak and similar, can also be broad or overlap. Each molecule in the sample can absorb at more than one wavelength, and each wavelength can have multiple contributions from the different molecules present in a sample. For this reason, a direct spectra interpretation is not possible, being necessary the use of advanced data analysis to extract the information embedded in spectra. However, the restrictions of this technique also bring advantages. The low absorption coefficients enable a higher penetration depth, allowing the use

of this technique in solids or turbid liquids, such as culture broths [1–4]. NIR spectroscopy, coupled with an optical probe, was used to monitor cell density online through light absorption (turbidity) or scattering, in the visible and/or NIR range. However, the inability to distinguish viable from non-viable cells, the narrow cell concentration range where can be applied, and the sensitivity towards different morphologies, are some of the restrains that need to be improved for the use of this spectroscopy as an online monitoring tool [4].

I.2.2. Raman spectroscopy

Unlike IR spectroscopy, that is being used in industrial proposes for several years, Raman spectroscopy is still in the stage of academic research [10].

Raman spectroscopy is based on the phenomena of shifted wavelength scattering of molecules excited with monochromatic light due to inelastic collisions of photons with the molecule [1]. This technology does not require clear samples, increasing the spectrum of usage in biotechnology industries. Additionally, with the development of Fourier-transform, the acquisition times and the photodecomposition were considerably decreased. Nevertheless, the intensity of the signal acquired is very weak and difficult to separate from the scattering [10].

Nowadays, several Raman methodologies are being studied to replace the standard procedures, such as GC or HPLC, for quality control and confirm adulterations of several products, like microorganisms, food products or even pesticides [10].

I.2.3. Fluorescence spectroscopy: principles

Luminescence, the phenomenon of light emission, can be divided into two categories, phosphorescence and fluorescence. Fluorescence occurs when an excited singlet state, in which an electron is paired by opposite spin to another electron from the ground-state orbital, returns to its ground state by the rapidly emission of a photon [11,12].

The first fluorescence sensors developed enabled only one wavelength of excitation and emission, meaning that only one fluorophore could be measured, restricting the use of this technology in complex processes. Later on, the development of multi wavelength fluorescence sensors make it possible to detect simultaneously several fluorophores in the same measurement, boosting the use of fluorescence spectroscopy as a scanning technique [1,2,11,13]. The measurement of several emission wavelengths over a range of excitation wavelengths creates a 2 dimensional (2D) excitation-emission matrix (EEM), that can be plotted in 3D graphs through the intensity recorded for each excitation-emission pair [4,12,14–16].

Fluorescence spectroscopy is a non-invasive technique, with high sensitive detection and specificity, able to detect instantaneously several natural fluorophores. These fluorophores can be classified into two main categories, extrinsic and intrinsic. Extrinsic fluorophores are the ones that need to be added to a sample that naturally does not possess fluorescence, such as fluorescein and rhodamine, among others. Intrinsic fluorophores are compounds that occur naturally in nature, such as NAD(p)H, chlorophyll, amino acids, cofactors and vitamins [4,8,11,12,14,17].

This technique is sensitive to intracellular and extracellular metabolites and media composition. In biological systems, characterised for having rich media composition, the interaction fluorophore-media is rather complex, since parameters like the polarity of the media, fluorophores structure and the interaction between them and with nearby molecules can cause a shift in the spectra. The phenomenon of masking the response of the fluorescence intensity of natural fluorophores is called quenching. Quenching can be collisional or static, if the decrease in the fluorescence is achieved by contact with another fluorophore or when forming a non-fluorescent complex, respectively [1,12]. For this reason, it was reported the use of fluorescence to indirectly determinate non-fluorescent compounds, since their presence can interfere with the fluorescence captured, and therefore, create a finger-print in the fluorescence spectra [1]. A different phenomenon is the inner-filter effect, also mentioned as self-absorption, where the fluorescence light is absorbed by the fluorophore itself [18]. It is also noteworthy that very dilute solutions will present a higher scatter interference in the EEMs due to the interaction between water molecules and the fluorophores. Two types of scatter can be notice: 1st and 2nd order Rayleigh, when the emission equals the excitation or two times the excitation, respectively, and Raman, with shift to longer wavelengths called red shift [19,20].

I.2.4. Fluorescence spectroscopy: biochemical applications

The urge for a rapid and online technology in the bioprocess industry, but also in environmental or clinic applications, accelerated the research and development of a system that could be coupled to a bioreactor, via optical fibre, and that could give an answer at real time.

Several authors have been studying the use of fluorescence spectroscopy to monitor online conventional fermentations. Most of the reported cases investigate the fluorescence of the reduced form of NAD(P)H, which fingerprint is detected at an excitation wavelength of 340 nm and emission wavelength of 460 nm. NADH is a highly fluorescent molecule, and it is known that the fluorescence intensity signal has a good correlation with the biomass concentration and its metabolic state [4,5,14,21]. For that reason, fluorescence spectroscopy was been used to track physiological changes, such as the transition between bacterial aerobic and anaerobic metabolisms [2], and also the change between oxidative and oxidoreductive metabolism in yeasts [8]. Fluorescence spectroscopy was also used to monitor biomass and substrates or products concentrations in *E. coli* and *S. cerevisiae* fermentations [4,17].

It is well reported in the literature the versatility of fluorescence spectroscopy to monitor several analytes, through direct or indirect correlations, such as proteins, vitamins, co-enzymes, and several metabolites such as glucose, ethanol, ATP, pyruvate, nitrate or succinate [1,2,8,11,22]. A system developed by the company DELTA Light & Optics (Lyngby, Denmark), called BioView™, was used to predict several cultivation parameters, such as enzyme activity, product formation and substrate consumption, with the final goal of defining an optimal harvest time [13,21].

Fluorescence spectroscopy has also been studied as a monitoring tool in several fields of the food industry such as the industrial downstream processing of sugar beet molasses [1], the characterization and classification of honeys [15], or the detection of orange juice frauds [23]. Other applications under study include *in situ* characterization of polycyclic aromatic hydrocarbons (PAHs) [24], or even as a sensor to detect illegal drugs (cocaine and marijuana) in street samples [25], which shows the versatility of this technique. In the pharmaceutical field, fluorescence

spectroscopy was studied as a monitoring tool for the physiological state of mammalian cell cultures, a platform for the production of anti-bodies, blood and growth factors or cytokines [26]. Also, several studies reported the potential of using this technique in wastewater treatment plants [27–32]. Ranzan *et al.* [11] stated that the application of this technology as a monitoring tool in biological systems proved to improve the ecological and economic management of the overall process under study. More applications can be found in the review of Pons *et al*. [21].

In the scientific fields of terrestrial plants and microalgae, the use of fluorescence was studied through Pulse Amplitude Modulated (PAM) fluorometry. This technique enables the measurement of chlorophyll *a* fluorescence and photosynthetic performance, both in growth and stress phases [33].

Although a major advance was observed in the quality of the optical sensors used in spectroscopy methodologies, these sensors are still rarely used in biotechnology industry, and most of the devices developed have been only study for research purposes [1,4]. The main concerns reported for the limited application of this technology are mostly due to interferences that affect the quality of the fluorescence spectra acquired, such as turbidity, gas bubbles and fouling [1,4,14].

I.3. Chemometrics: mathematical treatment of the spectra information

The first records of chemometrics were reported by Mandel in 1949, but only nearly 25 years later the term “chemometrics” itself was invented [34]. According to the Chemometrics Society, chemometrics is “the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analysing chemical data”.

A full fluorescence spectra can contain in-depth information about a biological system under monitoring and it is usually complex, since it contains information about the natural fluorophores present in the sample, but also the interferences between them and the environmental media [4,5]. For that reason, chemometric methods are often used to deconvolute the information within the fluorescence matrices, and to determine the relationship between them and the concentration of substrates and products to be monitored [3,26,29].

Multivariate analysis, like Principal Component Analysis (PCA), is often used to extract the meaningful information from the spectra. PCA reduces the dimension of the data set by extracting the most relevant information. For a n number of initial variables (X_1, X_2, …, X_n), n linear combinations are obtained, the so-called Principal Components (PCs), characterized for being uncorrelated and ordered according to the variance explained (the first PC explains the higher part of the variance, and smaller parts of variance are explained with subsequent components) [16,35]. The principle that no correlations exist between PCs is of most importance, since in that way it is guaranteed that each PC describes different patterns of the original data. After the identification of the first PC, the second one is calculated according to the direction that explains more residual variance and with a constrain of being orthogonal to the first one. The process continues until n PCs are defined. In the end, the PCA defines the initial data set (the matrix X) as $X=T.P^T + E$, where T matrix is the scores matrix (represents the objects in the new orthogonal

space), P matrix is the loadings matrix (represents the importance of each variable), and E is the error (residuals) matrix (contains the difference between the observed values and the ones predicted by the PCA). An important step to define the new set of variables through the PCA is the normalization, achieved by the subtraction of the mean value to the variable followed by the division by the standard deviation [16,35,36]. This procedure enables the comparison of variables that originally present different magnitudes and variances in a new reduced orthogonal space.

Most of the chemometric applications in analytical chemistry and biochemistry are prediction models, since the use spectral measurements to replace time-consuming chemical and biological lab analysis is very appealing. In other words, chemometric models are often used to find the correlations between a bio-/chemical variable (called dependent variable, y) and the data acquired from instruments such as a spectrofluorometer (independent variable, *x*) [37]. This relationship established between the dependent variable and a set of independent variables will be used as a linear polynomial prediction model, defined as $y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + f$, where b_0 is an offset, b_k ($k=1, \ldots, K$) are regression coefficients and f is the residual. To calculate b, the vector of the regression coefficients, multiple linear regression (MLR) is often used. However, the use of MLR requires uncorrelated spectral variables (*x*-variables), and in fluorescence spectra it is frequent to find collinearity in the *x*-variables and, also, matrices are often constituted by more variables than objects. A solution often used is to apply MLR analysis to PCA scores, technique denominated principal component regression (PCR), where PC's that explain high variability in the spectra are correlated with the target properties (*y*). Another multivariate regression model used frequently for fluorescence data analysis is the projection to latent structures (PLS). This differs from the PCR analysis since it find the correlations that best describe the highest variation between spectral (*x*-variables) and target properties (*y*) with a new set of axes (latent variables – LV) [2,4,11,14,16,35,38].

When is not possible to establish a linear relation between inputs and outputs, the regression methods fail to develop prediction models. Thus, the use of nonlinear methodologies may be helpful to model more complex relationships, such as multivariate additive PLS splines (MAPLASS) or artificial neural networks (ANN) [39]. Briefly, ANN embodies a "machine-learning" algorithm that attempts to mimic the information processing found in the human brain, using artifical "learning-from-experience". Detailed information about ANN can be found elsewhere [32,39–41].

When evaluating the quality and suitability of a prediction model, several parameters need to be assessed:

1) Variables selection: Not all the variables of the data set contribute with useful information for the model. The use of too many variables leads to overfitting, which means a false increase of the variance explained; the use of few variables leads to under fitting, which can result in the loss of relevant information. For that reason, several mathematical methods can be used to find the number of variables that better describe a model, such as iterative stepwise elimination (ISE) [42], stepwise elimination [43] and the Martens uncertainty test [43] using jack knife standard deviations [44].

2) Calibration (Training): From all the data set acquired, a part of the data is select to build the model, called the calibration or training data set. This data set should be representative of the range of all data acquired, and for this, random choice is often the best strategy. A

model can be evaluated through the percentage of variance explained, which means, how much variability it captures; and through the root mean square error of calibration (RMSEC), providing information about the magnitude of the error. To test the quality of the model developed, a cross-validation (CV) strategy can be used. Basically, several models are developed using another independent set of data, or leave one or *n* data points out (LOO and LNO, respectively), and repeatedly evaluate the error of the models (RMSECV). The importance of the calibration relies on the future usage of the model, when a new set of values will be computed through the model developed [35].

3) Validation (Test): To determine if a model is reliable, validation it's a crucial step of the process. It is expected that the model fit the initial assumptions and that defines the parameters under study, that is stable when a new data set is used for the prediction, and that the correct number of PCs were chosen. To validate the model developed with the calibration data set previously described, a new data set is selected, that was not used to build the model itself (validation set or test set), and the correspondent residuals are calculated. The distance, on average, from the data point observed and the prediction obtained with the model can be evaluated through the root-mean-square error of prediction (RMSEP) [37].

With the development of new technologies and instrumentation able to produce a continuous flow of information, an expansion is being observed in several omics' fields. The possibility of having a clear insight in biochemical processes and reactions revealed to be a powerful tool for the industries. Workman Jr. [45] presented several advantages of using chemometrics, among them:

1) Allows the possibility of providing real-time and high-quality information from less data
2) Improves the existing knowledge of the processes under study and enables the improvement of measurements
3) Requires low capital investment

The combination of spectrophotometric instruments with their high sensitivity and resolution, like fluorescence spectroscopy, coupled with chemometrics analysis for comprehension, data pre-treatment and exploration, enhances the knowledge and control of the processes.

I.4. Case study: microalgae production and added-value products formation

Microalgae are well-known photosynthetic microorganisms with the ability to produce a wide variety of value-added metabolites with the use of sunlight and CO_2 fixation, from the atmosphere or fume gas. When comparing microalgae to other plant crops they present several advantages: they have higher photosynthetic conversion efficiencies and higher biomass production rates, they can grow in diverse and inhospitable environments, they don't compete for land and can be grown in sea water [46]. However, nowadays, microalgae production is far from the terrestrial crops, since the production capacity of microalgae is still done in niche markets for high value

products. Their industrial production costs need to be reduced and the scale needs to increase to make this industry more competitive [47,48]. Mostly of the industrial facilities are based in open ponds systems, being only a few produced in closed photobioreactors. Being the growth and productivity of microalgae cultivation so tightly correlated with light conditions, better engineering solutions have to be developed regarding light distribution, mass transfer and hydrodynamics [48,49]. Operational costs are still too high and dependable on the local conditions and the metabolites produced for commercialization [48].

By minimizing the production and processing costs of microalgae biomass, the production costs of microalgae products will decrease. New technologies developed need to be easily scalable, but they also need to be competitive in operational and investment costs. To increase the economical yield of this industry, a biorefinery approach has been proposed where, instead of focusing in a specific product, a portfolio of products can be extracted, separated and purified according to its end market [47].

In a microalgae biorefinery production, the entire value chain is connected. Higher product yields require both high biomass productivity and high efficiencies in harvesting and biomass processing. Microalgae cultivation is mostly done in very diluted biomass concentrations, from 0.2 to 2 g/l, to avoid a large dark zone in the cultivation system, which leads to low productivities [50,51]. This low biomass concentration with consequent large volumes, leads to high dewatering costs. The most common technology used for dewatering microalgae biomass is centrifugation, which is a very energetic demanding technology [52]. To reduce harvesting costs, two-step harvesting processes are being studied, as for example, using membrane filtration as a primary dewatering step before centrifugation [52,53]. The possibility of using this technology enables also the recovery and recirculation of the cultivation media, which represents a positive economic impact on the overall costs of microalgae biomass production. However, this harvesting step is particularly sensitive for several algae that lack rigid cell wall. Within a biorefinery concept, the importance of maintaining the cell integrity is directly proportional to the product yield that can be recovered from its biomass.

Microalgae cultivation is a complex production system in which several biological, fluid dynamics, environmental and nutritional conditions play a crucial role in the productivity and viability of the industrial process. For these reasons, the development of prediction models that allow online monitoring of substrates and products is fundamental for process control, enabling the possibility to take decisions and actions at real-time [49]. In the present work, two marine microalgae were studied, *Dunaliella salina* and *Nannochloropsis oceanica*. Both microalgae genus are currently produced industrially and are authorized for food and feed supplements. Thus, having an online monitoring tool would be of great interest, not only to monitor the cultivation process, but also the induction of their respective target compounds and their biomass harvest.

The industrial production of *Dunaliella salina* is one success case in the microalgae biorefinery. This halotolerant microalgae has the ability to produce high contents of carotenoids, around 10% of the total dry weight, under stress conditions as high salinity, high light intensity, nutrient depletion or extreme temperatures [54–56]. Carotenoids are light-harvesting pigments and reactive oxygen species (ROS) scavengers, they act as non-photochemical quenching, which means they absorb the excess light preventing damage on the photosynthetic apparatus. In the actual market, the global market of carotenoids reached nearly $1.5 billion in 2017 and should

reach $2.0 billion by 2022 [57]. The cultivation systems more often used for *D. salina* cultivation are large unstirred open ponds or paddlewheel stirred raceways. A two-step cultivation is commonly applied: the first stage, a “green” phase, where growth is done under optimal conditions; a second stage, the “orange” phase, where the culture is submitted to stress factors and the carotenoid production is enhanced [58]. Regarding *D. salina* harvesting, studies have been made to present more efficient processes, either economically or sustainably, since this microalga is highly sensitive to shear stress due to the lack of rigid cell wall, compromising the overall profit of the process [51,53].

Several *Nannochloropsis* species gained interest because of their known ability to produce and accumulate large amounts of lipids, between 25 and 45% of their total dry weight. This made *Nannochloropsis* sp. production of great potential for biofuel production but also for food supplements, due to the high concentration of omega-3 fatty acids like eicosapentaenoic acid (EPA) [59]. When this microalga is submitted to environmental stresses, like high light intensity and/or nitrogen depletion, it has the ability of storing triacylglycerol (TAG) compounds in lipid bodies. Several companies already produce *Nannochloropsis* sp. in outdoors facilities, usually following a two-step strategy: an initial growth phase under optimal conditions is followed by nitrogen depletion to induce the accumulation of fatty acids [60]. It is hypothesized that TAG accumulation helps the prevention of photo-oxidative stress, helps regenerate NADPH to NADP+, and storages chloroplast components to be used when favourable growing conditions are restored [61].

I.5. Thesis motivation and outline

I.5.1. Motivation

The main objective of this thesis is the development of an online monitoring tool able to be used in microalgae production, that enable the monitoring of several parameters simultaneously and different processes within the production. Two marine microalgae are studied, known for their capacity to produce compounds of interest for food and feed, under growth and stress conditions: *D. salina* for its capacity to accumulate carotenoids, and *N. oceanica* for its capacity to accumulate fatty acids. Five parameters are monitored – cell concentration, cell viability, pigments concentration, fatty acids composition and nitrogen concentration – in four different processes – cultivation, product formation (carotenoids and lipids), harvesting by membrane filtration and permeate recover. A schematic illustration of the questions and studies developed in this work are shown in Figure I.1.

To prove the potential of fluorescence spectroscopy coupled with chemometric modelling techniques, different approaches are studied. The EEM of the spectra are used raw, without pre-treatment for removal of the scatter and inner-filter effects, and with pre-treated EEM. Two input strategies are performed in the PLS modelling, using the PC’s resulting from a PCA performed on the EEM, or directly using the EEM. More detailed information about the different approaches are explained throughout the thesis.

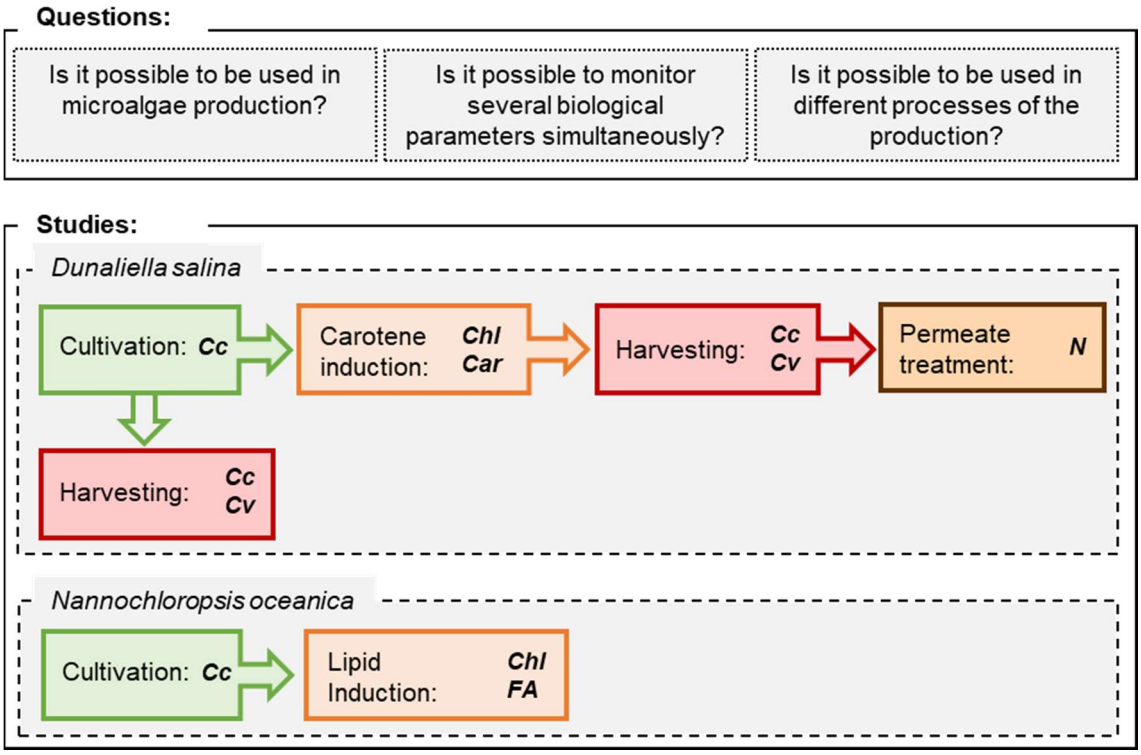


Figure I.1 Schematic illustration of the questions and studies performed in this thesis. Cc – Cell concentration; Cv – Cell viability; Chl – Chlorophyll; Car – Carotenoids; N – Nitrogen; FA – Fatty Acids.

I.5.2. Thesis outline

This PhD thesis comprise the following chapters:

Chapter 1 – Introduction: a state of the art is provided to present the current challenges of monitoring biological systems, such as microalgae, and the use of optical sensors coupled with chemometric tools as a solution.

Chapter 2 – *2D Fluorescence spectroscopy for monitoring Dunaliella salina concentration and integrity during membrane harvesting*: *D. salina* is well-known for the production of carotenoids, when in its "stress-phase". However, "non-stressed" cells are also rich in salt-tolerant proteins and reactive oxygen species (ROS) enzymes. Due to the lack of rigid cell wall, this microalga can easily disrupt during harvesting, losing valuable compounds to the saline water, affecting the downstream processing. Thus, monitoring cell concentration and integrity in real-time, of "green" *D. salina* cells, can assist the development and optimisation of harvesting methodologies.

Chapter 3 – *Development of a monitoring tool based on fluorescence and climatic data for pigments profile estimation in Dunaliella salina*: *D. salina* pigments profile were monitored, from a "green" (chlorophylls) to "orange" (carotenoids), in order to develop a tool able to be used online in the carotenogenesis step of *D. salina* production. Climatic parameters were also used as input variables due to their impact on the pigments profile in outdoor cultivations.

Chapter 4 – *One tool for simultaneous monitoring of cell concentration, cell viability and nitrogen during production and harvesting of stress-induced Dunaliella salina*: Fluorescence spectroscopy was studied as a potential monitoring tool in a biorefinery of *D. salina* for carotenoids

production. Three different parameters were monitored – cell concentration, cell viability and nitrogen concentration - in three different process – cultivation, membrane harvesting and permeate treatment.

Chapter 5 – *Monitoring of eicosapentaenoic acid (EPA) production in the microalgae Nannochloropsis oceanica*: An alternative source of omega-3 fatty acids is of great importance for the food and feedd industries. Eicosapentaenoic acid (EPA) accumulation by *N. oceanica* was evaluated through different environmental cultivation conditions – temperature, light cycle and nitrogen repletion/depletion – and a tool based on fluorescence spectroscopy was developed to monitor this fatty acid in the apolar (TAG) and polar (PL) fractions of the cell during cultivation and induction.

Chapter 6 – *Fluorescence spectroscopy coupled with chemometric modelling for the simultaneously monitoring of cell concentration, chlorophyll and fatty acids*: In a *N. oceanica* biorefinery, fatty acids are the main product of interest, that can be used for food, feed or biodiesel. During cultivation, the possibility to monitor several parameters simultaneously will contribute to better harvesting-time decision-making. Thus, prediction models for cell concentration, chlorophyll and fatty acids – as total, saturated and unsaturated – were developed.

Chapter 7 – Thesis overview and general conclusions: an overall conclusion of the results achieved throughout this work is presented, as well as suggestions for future work.

2D Fluorescence spectroscopy for monitoring *Dunaliella salina* concentration and integrity during membrane harvesting

II

Abstract

Dunaliella salina is able to produce simultaneously several valuable compounds (such as lipids, carotenes and functional proteins) within the biorefinery concept. However due to the lack of rigid cell wall, this microalga can easily disrupt during harvesting, losing valuable compounds to the saline water, affecting the downstream processing. Therefore, the development of non-invasive tools able to monitor cell concentration and integrity at real-time, can assist the development of harvesting methodologies. In the present work, a monitoring approach was developed based on two-dimensional (2D) fluorescence spectroscopy. Mathematical analysis of the monitoring data involved the use of Principal Component Analysis (PCA) and Projection to Latent Structures (PLS) modelling. For green *D. salina*, the models developed for prediction of cell number and percentage of viability captured 90.6 % and 86.3 % of variance, respectively. Both models have R^2 of 0.8 and 0.9, respectively for validation and training. Similar values were found for the prediction of cell number when using data from growth kinetics and harvesting combined. Orange *D. salina* rupture was also successfully modelled with 95 % of variance captured and R^2 of 0.9 for both training and validation. The combined approach using 2D fluorescence spectroscopy and the mathematical analysis proved to have the potential to monitor *D. salina* during cell growth and harvesting within a biorefinery concept.

Keywords: *Dunaliella salina*, 2D fluorescence spectroscopy, Ultrafiltration, Microfiltration, PLS modelling, Monitoring

Published as: Sá, M., Monte, J., Brazinha, C., Galinha, C. F., Crespo, J. G., 2017. 2D Fluorescence spectroscopy for monitoring Dunaliella salina concentration and integrity during membrane harvesting. Algal Research, 24:325–332.

II.1. Introduction

The growing interest on photosynthetic microorganisms, such as microalgae, is based on their ability to produce valuable compounds using sunlight and CO_2. The broad scope of the compounds produced, from pigments to proteins and lipids, is of high relevance in food, feed, bioenergy and biochemical sectors [47,62,63]. In this context, *Dunaliella salina* is a promising microalga for production within the biorefinery concept, since it enables the co-production of biofuel, energy and value-added products. *D. salina* is a green unicellular marine microalga well-known for its high contents in *β*-carotene and glycerol when induced by low nitrogen or high salinities conditions. Nevertheless, "green" *D. salina* also has other compounds of interest, such as antimicrobial compounds, salt-tolerant proteins and ROS (reactive oxygen species) scavengers [64–67].

When producing microalgae, the operational costs can be high due to the low cellular yields, which lead to high energy consumption to process the large volume of aqueous media during cell harvesting. The absence of rigid cell wall in *D. salina* is an extra challenge to be considered when harvesting the biomass. Therefore, cell production and separation require gentle operating conditions to avoid the loss of cellular integrity when maximal product recovery is intended. Nowadays several techniques have been applied for microalgae harvesting, such as flocculation, filtration, gravity sedimentation, flotation and centrifugation. The selection of the harvesting technique is mostly dependent on the microalgae characteristics, but the most used technique is centrifugation, since it is the fastest and most reliable method, despite the high investment and energy costs [46,51,67]. A two-step approach has been suggested by several authors for harvesting different microalgae, which include membrane filtration as a primary step, followed by centrifugation [51,52]. This first step of harvesting by membrane filtration can reduce significantly the aqueous volume to be processed and, consequently, the energy required during centrifugation. Harvesting by membrane filtration can be carried continuously, with relatively low energy consumption, and is easily scaled up, due to the possibility of assembling multiple membrane modules [51].

Harvesting *D. salina* can result in cell damage and consequent release and loss of valuable products into water. Therefore, it is of high relevance to monitor cell integrity to avoid excessive cell damage. Moreover, the ability to monitor cell concentration and cell integrity during the harvesting procedure may help process optimisation and control. Several sensor systems have been developed to measure physical parameters, but monitoring systems for chemical and biological quantities are not as well developed. Most of these systems require sample processing and, therefore, cannot be readily adapted to real-time monitoring [2,68]. Two-dimensional (2D) fluorescence spectroscopy is a fluorescence technique able to assess simultaneously several compounds by scanning through a range of excitation and emission wavelengths. 2D fluorescence spectroscopy is highly sensitive and non-invasive, suitable for in situ and on-line monitoring using an optical probe. Fluorescence spectra resulting from 2D fluorescence are large matrices of data (excitation-emission matrices - EEMs) that can be seen as fingerprints encoding valuable information concerning the presence of natural fluorophores as well as their interactions with the involving media. This fluorescence technique already proved its ability to monitor different biological processes, such as growth bioreactors of bacteria and yeast, and membrane bioreactors for wastewater treatment [14,17,29,69,70].

In the present work, a real-time monitoring approach was developed based on 2D fluorescence fingerprints. Fluorescence EEMs obtained during *D. salina* membrane harvesting and during cultivation in bioreactors were correlated with off-line measurements of cell number and percentage of viability. Statistical methods, such as Principal Component Analysis (PCA) and Projection to Latent Structures (PLS) were used to extract information from fluorescence spectra and develop models able to predict cell concentration and viability, based only on the fluorescence signal.

II.2. Material and Methods

II.2.1. Growth conditions of green and orange *Dunaliella salina*

D. salina DF17 was acquired from the Marine Biological Association Culture Collection (https://www.mba.ac.uk/culture-collection/) and it was isolated from the raceways at NBT Ltd (Natural Beta Technologies Ltd, Eilat, Israel). Batch cultivation was carried out by A4F – Algae for Future (Portugal) in artificial salt water (ASW) with an adjusted concentration of nitrate to 4 mM. The culture cells were continuously aerated and mixed by bubbling 0.2 µm-filtered atmospheric air enriched with 2 % of CO_2, at 25 °C and 150 µmol/(m^2.s) provided by daylight fluorescent tubes under continuous illumination. Growth lab scale bioreactors and scale-up to flat-panel (Green Wall™, GW) and tubular photobioreactors (PBRs) were operated with the culture described above. To monitor cellular growth, five lab-scale bioreactors of 1 L were inoculated with $2x10^5$ cells, and the pH was maintained at 7.5 through the injection of CO_2, with 18 h of light per day, provided by daylight fluorescent lamps with 140 µmol/(m^2.s), at 27 °C. The flat-panels and PBRs, placed in the A4F pilot plant, growth between September and November of 2015, with a mean daily irradiation of 106 W/m^2 and a mean daily temperature of 24 °C. The pH was maintained between 7 and 8 through CO_2 injection. For carotenogenesis induction of *D.salina* DF17, the culture was grown in the same ASW medium with depletion of nitrogen and increased salinity.

II.2.2. Membrane harvest equipment and operation conditions

The cell cultures produced in the flat-panels and PBRs were harvested using a pilot-size hollow-fibre filtration unit (polyethersulfone, 0.42 m of length and 2.4 m^2 of membrane area) operated in a crossflow mode. Several harvesting conditions were tested: type of membrane – ultrafiltration (Molecular Weight Cut-off (MWCO) of 150 kDa) and microfiltration (MWCO of 0.1 µm), pressure − 0.25 and 0.5 bar and recirculation flow velocity − 0.44 to 0.6 m/s (Figure II.1). Prior to each test, the permeate tank was filled with ASW, with the same composition as the medium used in microalgae cultivation. This procedure was needed to ensure the good practice of the backflush procedure during harvesting. Additionally, harvesting was also performed at lab-scale using carotene induced (“orange”) *D. salina*. This study was also performed in a crossflow configuration, using a microfiltration hollow-fiber membrane with a membrane area of 0.48 m^2 and a cut-off of 0.1 µm, at a constant recirculation flow velocity of 0.24 m/s and feed average pressure of 0.275 bar (Figure II.1). For “green” *D. salina*, four experiments were performed, three ultrafiltration (UF) and one microfiltration (MF). For “orange” *D. salina*, two harvesting experiments were performed, a pilot-scale UF and a lab-scale MF.

II.2.3. Sampling procedure and analysis

In cell growth bioreactor experiments, samples were taken every other day until the culture ran out of nitrogen. Nitrate concentration in the medium was determined by a method adapted from Collos *et al.* [71]. Cell concentration was determined by direct counting, using a light microscope and a Neubauer Modified cell (at least three measurements were made in order to achieve a variance lower than 10 % between replicas).

During harvesting experiments, samples were taken simultaneously from the culture and permeate tanks. Cell concentration and cell viability on the concentrate stream were monitored by flow cytometry using a Muse® Cell Analyser equipment (Merck Millipore, Germany). At least three measurements were performed using samples diluted with ASW to obtain a cell count between 10 and 500 cells/mL (variance lower than 10 % between replicas), in order to maintain linear response of the equipment. To determine cell viability, propidium iodide (PI) dye was used. PI is a fluorescent dye (excitation maximum at 535 nm and emission maximum at 617 nm) that does not permeate cells with functional and intact cytoplasmic membrane and, thus, it is excluded from viable cells. The number of viable cells is calculated from the total cell concentration and viability percentage values given by the cell analyser equipment. Cell viability was used in the present study as an indicator of cell integrity, since non-viable cells are measured based on cell membrane permeability.

All samples were analysed by 2D fluorescence spectroscopy by immersion of an optical fibre probe in a stirrer beaker to avoid cell sedimentation. 2D Fluorescence spectra were acquired with a fluorescence spectrophotometer Varian Cary Eclipse equipped with an excitation and emission monochromators and coupled to a fluorescence optical fibre bundle probe. Fluorescence spectra were generated in an excitation wavelength range of 250 to 690 nm (with an excitation slit of 10 nm and increment of 5 nm) and emission wavelength between 260 to 700 nm (with an excitation slit of 20 nm and increment of 5 nm).

II.2.4. Development of PCA and PLS models

The excitation-emission matrices (EEMs) obtained by 2D fluorescence measurements are characterised for being extremely large matrices, with more than four thousand values, composed by pairs of excitation-emission wavelengths and their correspondent emission intensity.

Principal Component Analysis (PCA) was used to compress and deconvolute the information in EEMs into Principal Components (PCs), with reduced co-linearity and noise. The fluorescence spectra acquired in concentrate and permeate streams during “green” *D. salina* harvesting were compressed separately, resulting in seven PCs for each, respectively, and captured more than 99 % of the variance. In “orange” culture harvesting models, ten PCs from the fluorescence spectra acquired in the concentrate and another ten from the spectra acquired in the permeate were used. For the bioreactor samples, also ten PCs were chosen explaining 99 % of the variance. For the models developed with harvesting and bioreactors data together, ten PCs were used from both the concentrate and the growth culture (Figure II.1). Multivariate statistical modelling was used to correlate the PCs of fluorescence (inputs) with the state of the culture (outputs), such as percentage of cell disruption and biomass concentration. Mathematical models were obtained using Projection to Latent Structures (PLS) modelling, where an output is

described by linear correlations of the inputs. For the models developed with the harvesting experiments of the “green” *D. salina* a total of 28 spectra (corresponding to 28 samples) were used. The models developed with the bioreactors comprised 32 fluorescence spectra from the bioreactor’s experiments and 28 from the harvesting. For “orange” culture 17 fluorescence spectra were utilised. All PLS models were trained using 75 % of the experimental data, and the remaining 25 % were used for validation. For “green” *D. salina* harvesting, two input strategies were attempt, using either only PCs from the concentrate or using PCs both from the concentrate and permeate, keeping the same training and validation sets. In validation with the bioreactors for “green” culture, only PCs from the concentrate were used as inputs, but two approaches were made, firstly using harvesting data as calibration (n=28) and bioreactors as validation (n=32), and secondly by using a random set of validation values (n=15) from both harvesting and bioreactors processes. For “orange” *D. salina* were also used PCs from the concentrate and permeate streams, with a random set of calibration and validation data set (Figure II.1).

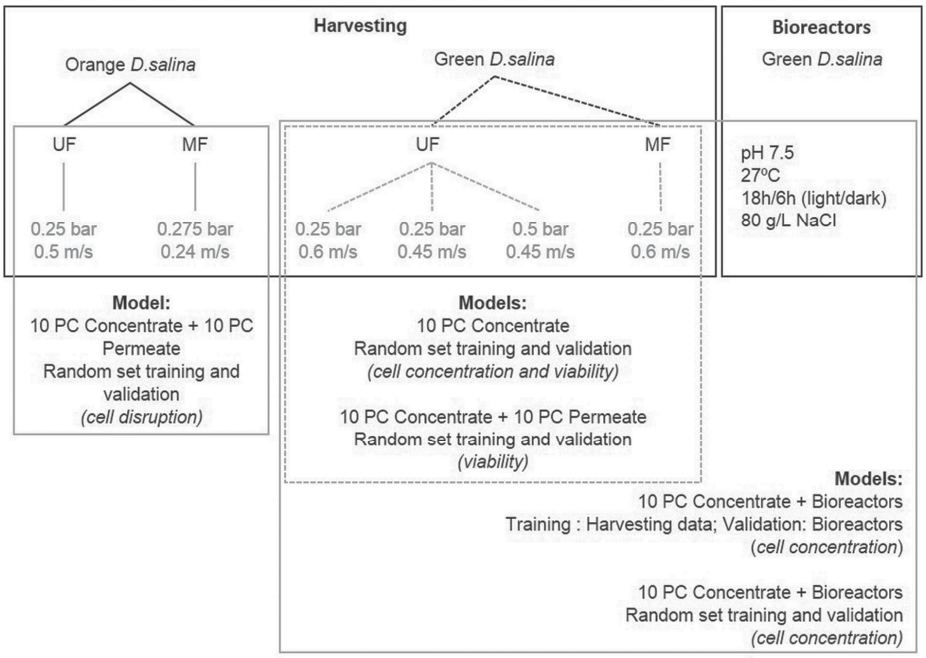


Figure II.1 Black boxes show the processes and conditions performed with D. salina, for both “orange” and “green” cells. For “orange” D. salina, two harvesting experiments were performed, a pilot scale ultrafiltration (UF) and a lab scale microfiltration (MF). For “green” D. salina, four experiments were performed, three UF and one MF. Five lab scale bioreactors were performed with the same growth conditions. Grey boxes indicate the experiments used to develop the predictive models.

All data used in PLS models were previously normalised by subtracting the respective average values and dividing by their standard deviations. Not all the parameters initially used as inputs are correlated with the outputs. Therefore, the useful predictors were selected by mathematical methods for input elimination such as iterative stepwise elimination (ISE) [42], stepwise elimination and the Martens uncertainty test [43] using jack-knife standard deviations [44]. The resulted models were compared based on the root mean square error of cross-validation (RMSECV), the root mean square of prediction (RMSEP), the variance captured, the coefficients of determination (R^2) and the slopes between prediction and experimental data for both training

and validation sets. PLS models and PCA were implemented in Matlab using the n-way tool box, respectively through nPLS and PARAFAC functions [72].

II.3. Results and discussion

Six different cell harvesting procedures, based on membrane separation, were tested in a pilot scale equipment. During cell concentration, off-line measurements of cell concentration and viability were performed in culture samples using Muse® Cell Analyser equipment. Additionally, fluorescence EEMs were obtained for both culture and permeate samples. Principal Component Analysis (PCA) and Projection to Latent Structures (PLS) modelling were used to compress fluorescence data and correlate it with the off-line measurements (cell concentration and viability). Furthermore, five lab-size growth bioreactors were monitored using 2D fluorescence spectroscopy and off-line measurements of cell concentration by direct counting, using Neubauer Modified cell. This data was used to verify the correlation found for cell harvesting experiments.

II.3.1. Harvesting

II.3.1.1. "Green" culture

Figure II.2 shows the typical fluorescence spectra obtained during cell harvesting of "green" *D. salina*, where the x-axis indicates the emission wavelength, the y-axis indicates the excitation wavelength, and the intensity is represented by a colour gradient defined in the legend.

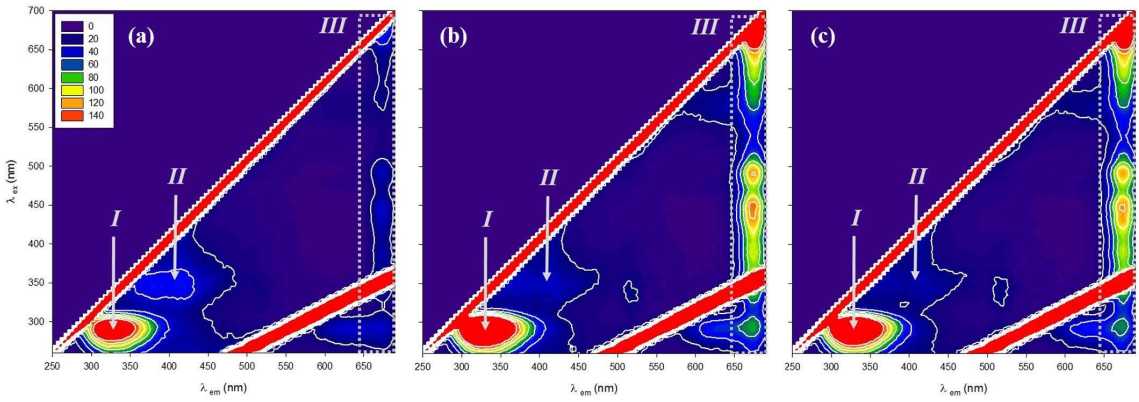


Figure II.2 2D Fluorescence spectra of the culture at the beginning of the membrane harvesting (a), immediately before (b) and after (c) the onset of cell disruption, defined as 10 % decreased of viability when compared to the initial biomass. Three distinct fluorescence regions were identified as I (excitation 275 nm, emission 300-350 nm), II (excitation 350 nm, emission 400 nm) and III (emission higher than 650 nm).

Three different regions can be observed in these fluorescence spectra. The region at an excitation wavelength of 275 nm and emission between 300 and 350 nm (Figure II.2 - region I) is assigned to the aromatic aminoacid tryptophan, also called protein-like fluorescence. This region has been previously identified by several authors in wastewater treatment systems [69] or in the identification of cyanobacterial blooms [73].

A second region (Figure II.2 - region II), with lower intensity, is present at an excitation wavelength of 350 nm and emission around 400 nm, and has been identified by several authors as humic-like fluorescence region [69,73,74]. This denomination can be misleading since this region also includes the presence of substances originated from quinone moieties, such as those present in extracellular polymeric substances (EPS) and that has been described to derived from algae [73].

In the third identified region (Figure II.2 - region III), for emission wavelengths higher than 650 nm, it is possible to verify the presence of several pigments with natural fluorescence [73]. As described previously by Moberg *et al.* [75], chlorophyll *a* and their accessory pigments (chlorophyll *b* and *c*), and also their degradation products (pheophytin and pheoporphyrin), can be detected in emission wavelengths between 640 and 670 nm, and excitation wavelengths between 350 and 450 nm. Some of these pigments can be identified on Figure II.2, namely pheophytin *a* (excitation between 350 and 400) and chlorophylls (excitation between 431 and 450). The presence of pheophytin *b* cannot be confirmed because its fluorescence is overlapped by the presence of chlorophyll *a* (excitation wavelength of 431 nm). A broad signal is detected at the emission wavelength of 650 nm. Ziegman *et al.* [73] described this region as composed by chlorophyll *a* and phycocyanin. The presence of phycocyanin (excitation 510–710 nm and emission 550–680 nm) is well reported in cyanobacteria *Microcystis aeruginosa* [73] and some blue-green microalgae [75], such as *Spirulina platensis*. However, since phycocyanin is not present in *D. salina*, this signal may be due to chlorophyll *a*.

Fluorescence spectra present a clear difference between the spectra obtained in the beginning of the experiment and immediately before cell disruption (Figure II.2 (a) and (c), respectively). This difference is not clear between spectra obtained immediately before and after cell disruption (Figure II.2 (b) and (c), respectively). For this reason, PCA was used to compress EEM information into a lower number of new orthogonal variables, reducing redundancy and enabling the study of complex correlations with the off-line analyses performed.

Firstly, PCA analysis was performed for each harvesting experiment individually looking for direct correlations between PCs and cell concentration and viability of the microalgae. An example of this correlation is shown in Figure II.3. Although it was not uniform for all the experiments, for some of them it was possible to observe a trend when plotting PC 2 of concentrate versus the permeate weight removed (in kilograms). If the beginning of cell disruption is defined as a 10 % decrease of viability (corresponding to noticeable cell damage by microscope observation), it matches the transition to positive values of PC 2. This feature is coincident with a higher distance between number of total and viable cells, indicating an increased loss of viability.

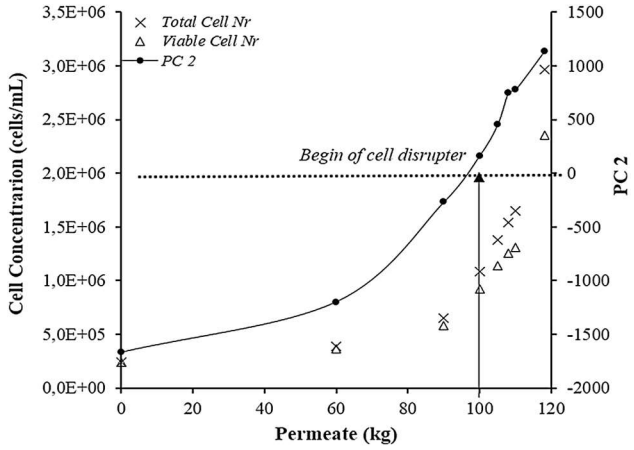


Figure II.3 Total (x) and viable cell number (Δ) during the membrane harvesting experiment (0.25 bar and 0.6 m/s). PC 2 (●–●) resulted from the PCA applied to the fluorescence EEMs acquired from the concentrate stream.

These observations show that 2D fluorescence spectroscopy can be a sensitive technique for monitoring biological systems, as described earlier by Galinha *et al.* (2011) [69].

In order to successfully study the complex correlations between the principal components (PCs) resulting from fluorescence spectra, acquired in the concentrate and in the permeate of all experiments performed and the analytical measurements of cell concentration and viability, a PLS modelling approach was used.

The first models were developed using seven PCs resulting from the PCA applied to the fluorescence EEMs acquired only on the concentrate. The training and validation results from the PLS models achieved for prediction of cell concentration and viability are shown in Figure II.4. Figure II.4 (a) and (c) show the prediction models for viability and cell number, respectively, where the observed values (y-axis) are plotted against the values predicted by the model (x-axis). In order to identify outlier data, a line corresponding to two times the standard deviation of the output experimental data was added. In these plots, a good model fitting should have slope near 1, meaning that the correlation between observed and predicted values follows the same trend.

As shown in Figure II.4 (a), the first model developed to predict the percentage of cell viability (using only fluorescence data from the concentrate) had low coefficient of determination (R2) for training and validation data (0.56 and 0.34, respectively). This model captured 56.4 % of variance using two of the seven initial inputs (the first and fourth PCs, Figure II.4 (b)). Although the training slope was 1.00, the validation slope was low (0.54), and the root mean square error of prediction (RMSEP) was 8.7 % viability, which means that the distance, on average, from a data point to the prediction value is around 9 %. This error is high, since a decrease of 10 % from the initial biomass viability corresponds to noticeable cell damage by observation at the microscope.

The prediction of cell concentration using only fluorescence data from concentrate was more successful. With a good fitting for both training and validation sets (0.86 and 0.83, respectively) and using only two PCs, the first and the second, it was possible to capture 86.3 % of the variance with a RMSEP of 1.27×10^5 cells/mL. The average of the standard deviations (calculated for the replicas of each sample assessed by the MUSE Cell Analyser) is 8.4×10^4 cells/mL, meaning that the prediction error cannot be lower than this value. The prediction is affected by the experimental

error of the cell concentration measurements (used to calibrate the model) and also of the fluorescence measurements, used to make the previsions. Therefore, a model with a RMSEP lower than two times the experimental error of the cell concentration ($1.7x10^5$) may be regarded as a good prediction.

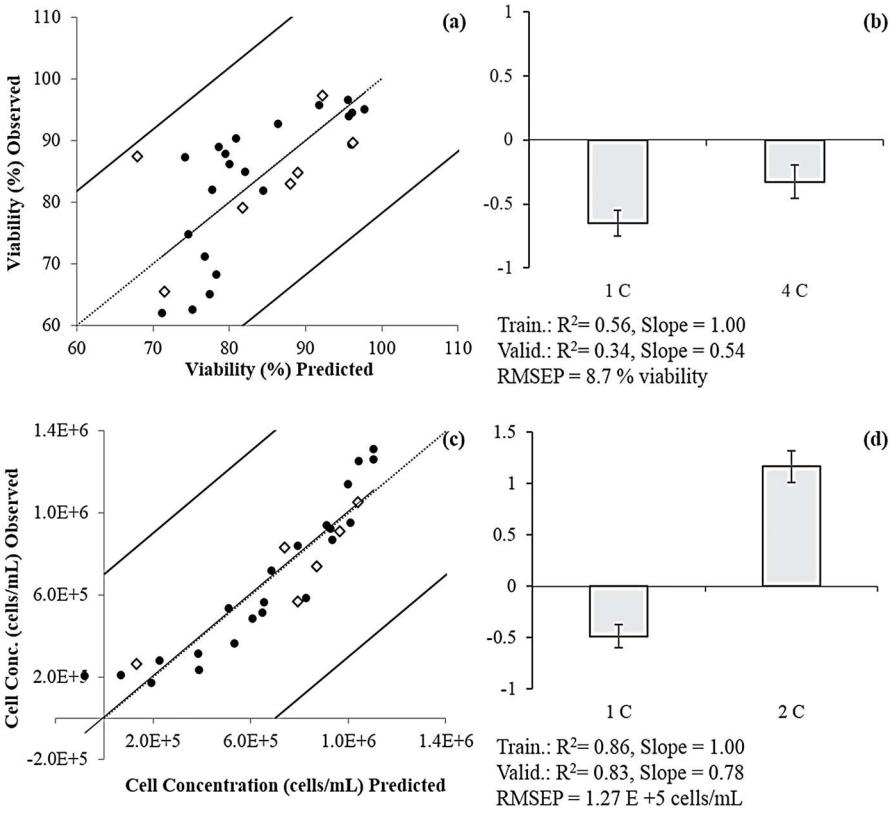


Figure II.4 Prediction models and normalized regression coefficients for viability percentage (a and b, respectively) and cell concentration (c and d, respectively) using PCs from the concentrate only. Training (n=21 for viability and n=22 for cell concentration) (●) and validation (n=7 for viability and n=6 for cell concentration) (◊) data are presented as percentage for viability, and cells/mL for cellular concentration. 1C, 2C and 4C are, respectively, PC1, PC2 and PC4 from the concentrate spectra.

Aiming at improving the model's prediction ability, seven PCs, resulting from PCA of fluorescence EEMs obtained with permeate samples, were added as inputs in PLS modelling. As shown on Figure II.5 (a), there was a significant improvement by adding the permeate PCs as model inputs to predict cell viability. The training R^2 increased to 0.91, and the validation R^2 increased to 0.82. The training slope maintained at 1.00, and validation slope also improved to 0.70. The RMSEP decreased to 5.3 % of viability and the variance captured increased to 90.6 % with the use of five PCs, three from the initial seven PCs from the concentrate (PC 1, PC 2 and PC 7) and two from the initial seven of the permeate (PC 1 and PC 2) (Figure II.5 (b)).

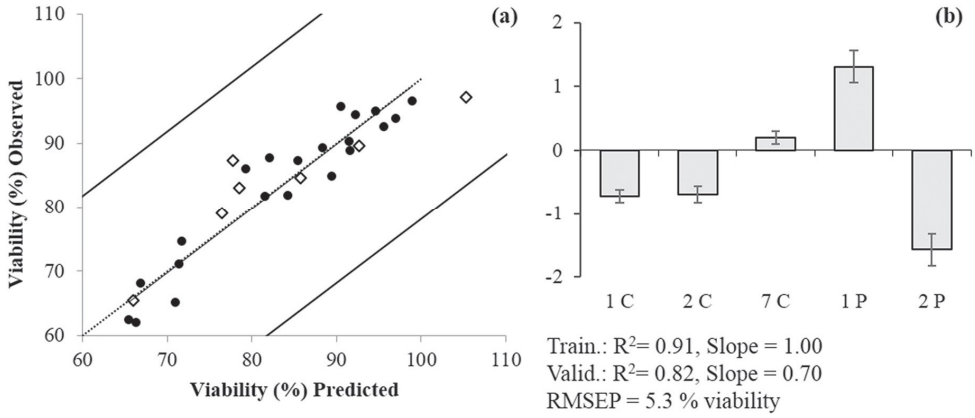


Figure II.5 Predicted vs observed values for viability percentage (a), with training (n=21) (●) and validation (n=7) (◊) values presented, using PCs from the concentrate and the permeate. Regression coefficients (b) are in normalised units. 1C, 2C, 7C, 1P and 2P are, respectively, PC 1, 2, 7 from the concentrate and PC 1 and 2 from the permeate spectra.

It is possible to see in Figure II.5 (b) that the presence of the permeate PCs is very important for a good prediction of the model developed, as they also have a great weight in the prediction (the regression coefficients of the two PCs of the permeate are higher than the other three of the concentrate). For the cell number prediction, no improvement was achieved with the addition of the seven PCs from the permeate. This trend was expected since no intact cells are expected to pass through the membrane to the permeate, only some small cell debris and intracellular compounds resulting from cell disruption.

The models obtained for "green" *D. salina* show that the loss of cell viability during the experiments is also correlated with the release of cellular material to the permeate. Therefore, the permeate stream comprises information useful for determination of cell disruption and loss of viability. However, for the prediction of total cell number such information is not needed and only the PCs from the concentrate are necessary. The ability of 2D fluorescence spectroscopy to detect simultaneously the fluorophores present in the concentrated cell medium, and in the permeate, enhances the potential of this technique as an on-line monitoring tool in harvesting procedure.

In order to test the presence of nonlinear correlations between fluorescence data and the predictive parameters, an attempt to improve the quality of the mathematical correlations was done. The fluorescence PCs defined previously as major contributors in the predictive models were, then, multiplied between each other (interaction terms) and by themselves (quadratic terms), and used as input parameters. Even though this approach was used previously [29], no significant improvement was achieved in the present work, and for that reason these results are not shown.

II.3.1.2. "Orange" culture

The pilot and lab-scale harvesting experiments for the carotene induced culture ("orange") of *D. salina* were monitored by microscope observation. The methodology used for "green" cells, with the Muse® Cell Analyser equipment, cannot be applied directly to assess cell number of

"orange" *D. salina*. Adaptation of this technique is required in the future. Therefore, cell disruption was defined as 0 for unbroken cells, and 1 from the moment that some cells are noticeably disrupted.

The PLS model was developed using twenty PCs resulting from the PCA applied to the fluorescence EEMs, ten obtained with the samples from the concentrate and another ten from the permeate. A higher initial number of PCs was used aiming at capturing more information from the inputs, since the output was based only on unbroken/broken (0/1) cells. The training and validation results of the PLS model for cell disruption are shown in Figure II.6. As shown in Figure II.6 (a), the prediction values ranged between −0.2 and 0.2 for unbroken cells, and between 0.7 to 1.1 to disrupted cells. The RMSEP of 0.19 is acceptable since is lower than 0.5, the division number between broken (1) and unbroken (0) cells. This model to predict cell disruption captured 94.5 % of variance using four PCs from initial twenty, three of the concentrate samples (PC 4, 5 and 6), and one of the permeate (PC 3). As observed with the model to predict cell disruption for "green" *D. salina*, Figure II.6 (b) shows again the importance of the PCs from the permeate to predict cell disruption. The regression coefficient relative to the permeate PC is once again higher than the three PCs from the concentrate.

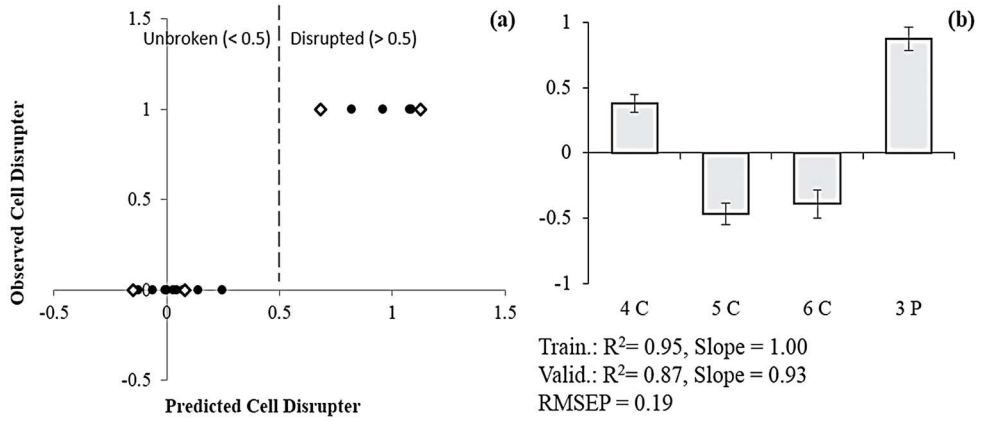


Figure II.6 Prediction model for cell disruption of "orange" culture during harvesting (a), using PCs from the concentrate and permeate streams. Training (n=13) (●) and validation (n=4) (◊) data are presented as unbroken (< 0.5) and disrupted (> 0.5) cells. Regression coefficients (b) are in normalized units. 4C, 5C, 6C and 3P are, respectively, PC 4, PC 5 and PC6 of the concentrate and PC 3 of the permeate spectra.

For PLS modelling of both "green" and "orange" culture, the fluorescence EEMs from the concentrate and the permeate were important, most probably due to the release of cellular compounds during cell rupture, more easily detectable in the cleaner permeate. The use of 2D fluorescence spectroscopy to monitor harvesting, at real-time, in a biorefinery context, can be achieved with the use of two optical probes with a switch box to acquire the fluorescence spectra simultaneously from both streams. The continuous flux of information that can be captured with the 2D fluorescence optical probe, without having any time delay, may contribute to a better control of harvesting operating conditions, yielding improved *D. salina* concentration for maximum product recovery.

II.3.2. Validation with bioreactors: experiments with green *Dunaliella salina* culture

Five lab-scale bioreactors were operated in order to follow the growth kinetics of "green" microalga *D. salina*. During cell growth, samples were taken to analyse cell concentration, using a Neubauer Modified chamber, and 2D fluorescence spectroscopy. These results were added to those obtained for the harvesting experiments (also with the "green" microalgae) in order to see if it was possible to predict the increase of cellular concentration that occurs in both processes, harvesting and cell growth. In the first approach, the harvesting data was used to develop the PLS model as training set, and the data from the bioreactors as validation set.

As it can be seen in Figure II.7 (a), the values of validation set are more disperse, corresponding to a R^2 of 0.58. The training R^2 of the fitted model for cell number prediction was 0.97, and the slopes 1.0 and 0.78 for training and validation, respectively. Despite the weak prediction for the bioreactor growth, it is possible to see a tendency in the validation set. All PCs provided to the model were necessary to capture 97.5 % of variance but, as represented on Figure II.7 (b), the standard deviations associated with the regression coefficients of PC 6 and 9 were high.

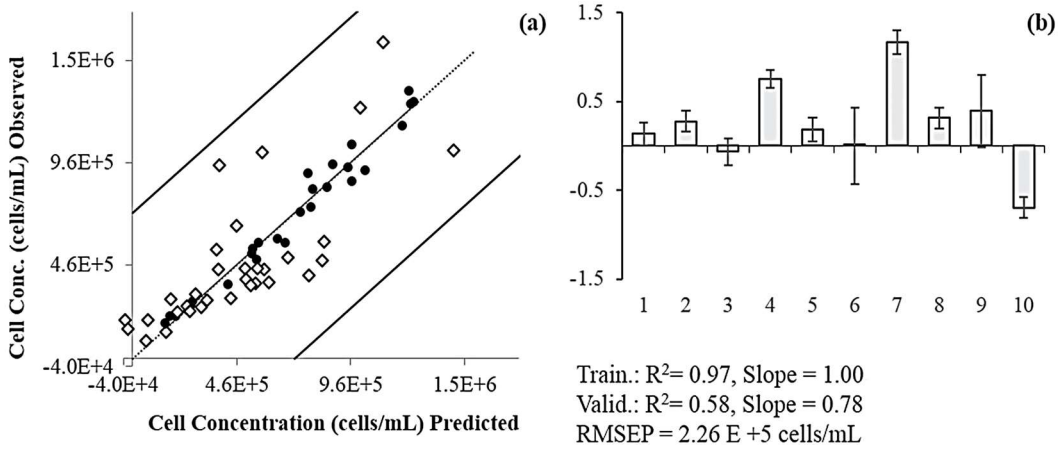


Figure II.7 Cell number prediction model (a) using harvesting experiments as training (n=28) (●) and bioreactors as validation (n=32) (◊), both in cells/mL. Regression coefficients (b) of model inputs are in normalised units, using PCs from 1 to 10.

It is described in literature that *D. salina* secretes metabolites to the surrounding medium during its growth while consuming nutrients, and this fact increases the differences between the cell culture medium along growth kinetics [76]. This major difference between growth and harvesting experiments leads to the need of recalibrating the model using data from both processes (growth and harvesting). Thus, considering a different and random validation data set, it was possible to find a correlation with cellular concentration as shown in Figure II.8 (a).

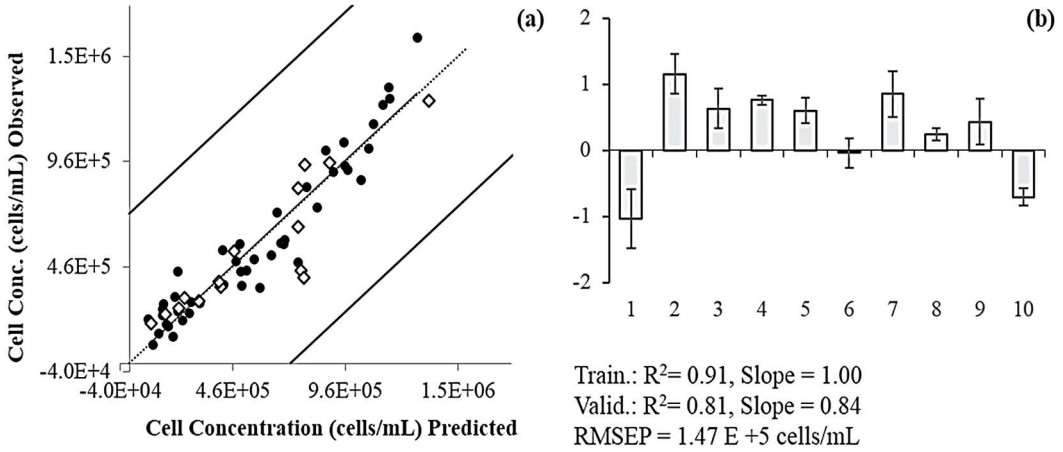


Figure II.8 Cell concentration prediction model (a) using random sets for training (n=45) (●) and validation (n=15) (◊), both in cells/mL. Regression coefficients (b) of the model inputs are in normalised units, using PC 1 to 10.

The validation R^2 and slope improved to 0.81 and 0.84, respectively, with a minor decreased of the training R^2 to 0.91. A slight decrease on the variance captured was noted, from 95 % to 91 %, but the number of PCs also decreased with the exclusion of PC 6, which had previously the highest standard deviation. The RMSEP decreased to 1.47×10^5 cells/mL (Figure II.8 (b)).

The present findings show that 2D fluorescence spectroscopy has the potential of monitoring green *D. salina* cell concentration *in-situ*, at real-time, during different processes within a biorefinery concept, such as cell culture and harvesting. 2D Fluorescence spectroscopy has already proved to be suitable for online monitoring of proteins, vitamins and several other compounds in bioprocess bioreactors [2,29]. Therefore, with the additional presence of several pigments in microalgae biomass, 2D fluorescence spectroscopy presents a high potential for monitoring also other parameters within a biorefinery concept, either during bioproduction or the downstream processes [68].

II.4. Conclusions

Cell integrity and concentration are two important parameters to evaluate during microalgae harvesting. Nowadays, several equipment and lab analyses can be performed to determine these parameters, but both require sampling the system and time-consuming procedures. Fluorescence spectroscopy can be used for the monitoring of two processes (cell growth and harvesting), directly at the bioreactor or at the feed tank of the membrane harvesting apparatus, even with different composition matrices. This non-invasive technique allows the acquisition of measurements online, without sampling.

The present work proved the ability of using 2D fluorescence spectroscopy for monitoring cell concentration and integrity of green *D. salina*, and the potential to monitor cell damage of “orange” culture, during the cell harvesting. Microalgae harvesting being an expensive process and with high impact on downstream processing, the methodology developed can be used to achieve maximum cell concentration without losing target compounds to the permeate stream.

Furthermore, this approach has the potential to monitor “green” *D. salina* cell concentration *in-situ*, at real-time, during different processes within a biorefinery concept, such as cell culture and harvesting. In sight of the present results, further work should be developed in order to find correlations for other parameters (key products), for both “green” and “orange” cultures and during different biorefinery processes, from the cultivation to the recovery and purification of added value compounds.

II.5. Acknowledgements

The authors would like to acknowledge the financial support provided by the European KBBE FP7 project “D-Factory”, King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under award no. OSR-2016-CPF-2907-05, and also the FCT, Portugal, for the Post-Doctoral fellows grants SFRH/ BPD/95864/2013 of Claudia Galinha (orcid="0000-0003-0045-2528") and SFRH/BPD/79533/2011 of Carla Brazinha, and PhD Fellow grant SFRH/BD/108894/2015 of Marta de Sá. A special thanks to the company A4F – Algae for future (Portugal), for sharing their extensive knowledge and supplying all the biomass needed to develop this work, as well as NBT Ltd (Israel) and The Marine Biological Association (Devon, UK).

Development of a monitoring tool based on fluorescence and climatic data for pigments profile estimation in *Dunaliella salina*

III

Abstract

When growing microalgae for biorefinery processes, a high product yield is desired. For that reason, monitoring the concentration of the desired products during growth and products induction procedure is of great interest. 2D Fluorescence spectroscopy is a fingerprinting technique, used *in situ* and at real-time, with a high potential for online monitoring of biological systems. In this work, *D. salina* pigments content were monitored using fluorescence data coupled with chemometric tools. Climatic parameters were also used as input variables due to their impact on the pigments profile in outdoor cultivations. Predictive models were developed for chlorophylls content (*a, b* and total) with variance captured between 50 and 90 %, and R^2 varying between 0.6 and 0.9 for both training and validation data sets. Total carotenoids models captured 70 % to 80 % of variance, and R^2 between 0.7 and 0.9, for training and validation. Models for specific carotenoids (zeaxanthin, α-carotene, all-trans-β-carotene and 9-cis-β-carotene) captured variance between 60 % and 90 %, with validation and training R^2 between 0.6 and 0.9. With this methodology it was possible to calibrate a monitoring tool for pigments quantification, as a bulk and as individual compounds, proving that 2D fluorescence spectroscopy and climatic data combined with chemometric tools can be used to assess simultaneously and at real time different pigments in *D. salina* biomass production.

Keywords: Monitoring; Spectroscopy; Mathematical Modelling; Industrial biotechnology; Bioprocesses

Published as: Sá, M., Ramos, A., Monte, J., Brazinha, C., Galinha, C. F., Crespo, J. G., 2019. Development of a monitoring tool based on fluorescence and climatic data for pigments profile estimation in *Dunaliella salina. Journal of Applied Phycology (DOI: 10.1007/s10811-019-01999-z).*

III.1. Introduction

In the current industrialization of compounds from biological sources, such as microalgae-based biorefinery, the control and monitoring of the cultivation and process parameters rely on physical, chemical and biological parameters. Physical and chemical parameters are currently monitored online using sensors like pH, temperature, dissolved O_2 or CO_2 [1,2]. However, biological parameters monitoring, such as cell concentration or product formation, are still mostly performed off-line, where a sample has to be withdraw from the cultivation/process and subjected to different laborious and time-consuming procedures, losing the window of opportunity to take decisions at real-time [3]. The development of an *in situ* and online sensor that could deliver a continuous stream of information would result in very short response time, enabling important control decisions in the spot. Additionally, the possibility of measuring several parameters simultaneously, without the need to sample, would be a great advantage for the microalgae biorefinery. Several spectroscopies have been studied for this purpose, namely fluorescence spectroscopy [2,4,5,8,11,14,22,29].

Fluorescence spectroscopy is able to measure several analytes simultaneously by scanning through a wide range of excitation-emission wavelengths (two dimensional (2D) scanning) [4,12,14]. This technique is non-invasive and highly sensitive to the presence of natural fluorophores, intra or extracellular. It is also reported that fluorescence spectroscopy can indirectly provide information about compounds that are non fluorophores but that interfere with the fluorophores in the sample [1]. Due to the complexity of fluorescence excitation-emission matrices (EEM), the use of chemometrics tools is advised to extract quantitative information from the fluorescence spectra and to resolve the occurrence of some limitations, such as inner filter effects or quenching, specially at high concentrations. The possibility of coupling an optical probe makes this technology suitable to be used in different processes within the biorefinery for an online and *in situ* monitoring [69,77–79].

In this study, microalga *Dunaliella salina* was selected for its current industrial production of biomass retailed as natural source of carotenoids. Its ability to grow in hypersaline lagoons, with low nitrogen and high solar light intensity, protects the culture from contaminants [80] and increases the carotenoid content, being β-carotene present at higher concentration (more than 12% of its dry weight) [81,82]. Photosynthetic organisms, like *D. salina*, harvest light energy due to their pigments, which can be chlorophylls, carotenoids and/or phycobilins. Of those, chlorophyll and carotenoid molecules are known for being used as natural colorants and antioxidants in different food products [82–84]. Specifically carotenoids are also important nutraceuticals, due to their anti-oxidant, anti-ageing, anti-inflammatory, anti-angiogenic, cardio and hepatoprotective properties [80]. There are several types of chlorophylls present in nature (named *a*, *b*, *c* and *d*), which differ in their side-group substitutes. Carotenoids are lipophilic compounds able to dissipate excess light as heat, protecting the cells from reactive oxygen species (ROS) generated during photosynthesis and high light intensity [80,84–86]. Carotenoids are accessory pigments transferring the absorbed energy to chlorophylls, expanding the light absorbing spectrum of the algae [87].

The development of a monitoring tool able to detect all these pigments at real-time will enable a better understanding of the pigment's formation through the life cycle of *D. salina*, and at

industrial scale it is useful to increase the biomass potential in a biorefinery context. Therefore, in this study, fluorescence EEMs were acquired during several pilot scale cultivation experiments of *D. salina*. Predictive models using Principal Component Analysis (PCA) and Projection to Latent Structures (PLS) were developed in order to correlate the EEMs with the pigments profile using two different off-line calibration tools: spectrophotometric and HPLC methodologies. The importance of environmental conditions during the carotenogenesis and their impact on the pigments profile was also studied through the use of climatic parameters as input variables.

III.2. Material and Methods

III.2.1. *Dunaliella salina* growth and carotene induction conditions

D. salina DF40 was collected from Monzón Biotech (Spain), isolated by Marine Biological Association (United Kingdom) and scaled-up and produced at pilot scale by A4F - Algae for Future (Portugal).

Batch cultivation was carried out in artificial salt water (ASW) under continuous aeration and mixed by bubbling using 0.2 µm-filtered atmospheric air enriched with 2 % of CO_2. The temperature was set at 25 °C and continuous illumination was provided by fluorescent tubes (150 µmol/(m^2.s)).

For carotenoid induction experiments, the culture was scaled-up to pilot scale flat-panel photobioreactors (Green Wall™, GW) and the same ASW medium was used with depletion of nitrogen and increased salinity. All batches were performed between January and October of 2017, exposed to outdoor weather conditions, with control of maximum temperature, but no control of the minimum temperature. In total, six batches were performed and monitored from the inoculation, with non-stressed "green" cells, until reaching a highly stressed "orange" culture.

III.2.2. Sampling procedure and analysis

Samples were taken every other day until a stable carotenoid/chlorophyll's ratio was reached. Pigments quantification and fluorescence spectroscopy assessment were performed for each sample.

III.2.2.1. Pigments analysis

Pigments quantification was performed by two methodologies, spectrophotometry and HPLC.

Briefly, *D. salina* cells (2 mL) were collected by centrifugation (5000 g, 5 minutes incubated at 60 °C for 40 minutes, followed by cooling on ice for 15 minutes. After centrifugation, the supernatant was collected, and the extraction was repeated until a white pellet was achieved. Quantification was performed in a UV/Vis spectrophotometer [88]. The modified's Arnon's equations were used to calculate chlorophyll and carotenoid contents [89]:

$$Chl_a = (16.72 \times A_{665} - 9.16 \times A_{652}) \times dilution\ factor\ [\text{mg/L}]$$

$$Chl_b = (34.9 \times A_{652} - 15.28 \times A_{665}) \times dilution\ factor\ [\text{mg/L}]$$

$Chl_{tot} = Chl_a + Chl_b$ [mg.L^{-1}]

$Car_{tot} = (dilution\ factor \times 1000 \times A_{470} - 1.63 \times Chl_a - 104.96 \times Chl_b) / 221$ [mg/L]

Two different methods were used to quantify pigments by HPLC, one for chlorophylls and another for carotenoids. The HPLC system consisted of a Waters Alliance Separations Module e2695 (Waters, Dublin, Ireland) coupled to a Photodiode Array Detector Module e2998 (HPLC-PDA).

For chlorophyll analysis, the extraction was performed in fresh biomass pellet using 100% methanol at 4 °C, repeatedly until a white pellet was reached. Separation of chlorophyll *a* and *b* was achieved using a reverse phase C18 (Vydac 201TP^{TM}, 250 x 4.6 mm, 5 µm, Hichrom, Berkshire, United Kingdom) and an isocratic elution with 100 % methanol, at a constant flow rate of 1 mL/min. The column oven temperature was set at 25 °C and the injection volume was 20 µL.

For carotenoids analysis, the extraction was performed in fresh biomass pellet with 100 % methyl tert-butyl ether (MTBE). After centrifugation, the supernatant was recovered, and the extraction was repeated until a white pellet was obtained. Separation of carotenes was achieved using a reverse phase C30 (YMC Carotenoid, 250 x 4.6 mm, 5 µm, YMC Europe GmbH, Dislaken, Germany) and an isocratic elution with 90 % methanol and 10 % MTBE, at a constant flow rate of 1 mL/min. The column oven temperature was set at 25 °C and the injection volume was 20 µL.

Standard stock solutions of chlorophylls (0.1 mg/L) and carotenes (0.2 mg/L for all-trans-β-carotene, 9-cis-β-carotene and α-carotene; 0.1 mg/L for lutein and zeaxanthin) were prepared with 100 % methanol and 100 % MTBE, respectively. For each compound, standards were diluted from the stock solutions to obtain a concentration range between 0.0025 mg/L and 0.1 mg/L. Independent replicates were prepared from the stock solutions in the beginning of each run and were used to assess the precision of the method. All stock solutions and calibration standards were stored at 4 °C.

III.2.2.2. 2D Fluorescence spectroscopy

All culture samples collected for pigments analysis were also analyzed by 2D fluorescence spectroscopy directly through the immersion of an optical fiber probe in a stirred sampling tube, preventing cell sedimentation. Fluorescence EEMs were acquired using a fluorescence spectrophotometer Varian Cary Eclipse, equipped with excitation and emission monochromators, and with a fluorescence optical fiber probe. As described in previous work[78], fluorescence data was collected in an excitation wavelength between 250 and 690 nm, with an excitation slit of 10 nm and increments of 5 nm, and emission wavelength between 260 and 700 nm, with an excitation slit of 20 nm and increments of 5 nm.

III.2.3. Climatic data

Carotenoid induction experiments were performed in A4F - Algae 4 Future outdoor facilities, thus climatic conditions at which the experiments were exposed were used as input in the development of the PLS models. Two strategies were attempted using the climatic data: i) based on the values assessed on the sampling day; ii) using a cumulative approach, where the values

used for each parameter are the sum of the values assessed from time zero until the sampling day. The climatic parameters used are shown in Table III.1.

Table III.1 Climatic parameters used in the development of the PLS models

Input nr and parameter		Units	Data Source	Frequency
11	Temperature	°C	ECWMF	daily
12	Total Precipitation	mm	ECWMF	daily
13	Sunlight	min	timeanddate.com	daily
14	Cloud	dimensionless	MODIS/MOD06	daily
15	Fpar	dimensionless	MODIS/MCD15A3H.006	4-days
16	Irradiance	$MJ/m^2.day$	measured locally	daily

III.2.3.1. Temperature range, precipitation and sunlight

Temperature (°C) and precipitation (mm) data were taken from the ERA Interim global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) for the region of Lisbon [90]. Maximum and minimum temperature products, taken at 2 m from the surface at every 6 hours, were used to calculate the temperature range. The daily total precipitation input was the sum between the convective and stratiform precipitation, and used as accumulated precipitation of the day. Sunlight data (in minutes) was obtained from the website timeanddate.com, with measurement location at Lisbon Portela Airport, 3.8 km from A4F – Algae for Future experimental unit. Through the hour of sunrise and sunset was determine the day length in hours per day.

III.2.3.2. FPAR and clouds fraction

The fraction of photosynthetically active radiation (FPAR) is defined as the fraction between the wavelengths 400 and 700 nm of incident photosynthetically active radiation that is absorbed by the green elements of vegetation. This parameter is important when measuring biomass production because vegetation development is related to the rate at which radiant energy is absorbed. FPAR data was acquired by the MCD15A3H version 6 MODIS (Moderate Resolution Imaging Spectroradiometer) Level 4, as a 4-day composite data set with a 0.5-km-pixel resolution for the Lisbon area [91].

Cloud fraction is defined by the Earth fraction which is covered by clouds relative to the fraction that is not covered. Clouds play an important role in regulating the amount of energy that reaches the Earth from the Sun and also the energy that the Earth reflects back into space. This parameter is a product of radiance and reflectance measurements acquire by Cloud Mask (MOD 35) of MODIS, as a daily data set with a 1-km-pixel resolution. The data set was acquired through NEO website (NASA Earth Observations).

III.2.3.3. Irradiance

Irradiance was recorded by WatchDog® Weather Station (Spectrum Technologies, Inc., Illinois, USA) at A4F-Algae 4 Future outdoor facilities, every 15 minutes. The value used in the PLS modelling corresponds to the mean value of the day.

III.2.4. Development of multivariate models

The modelling methodology followed in this work is represented in Figure III.1.

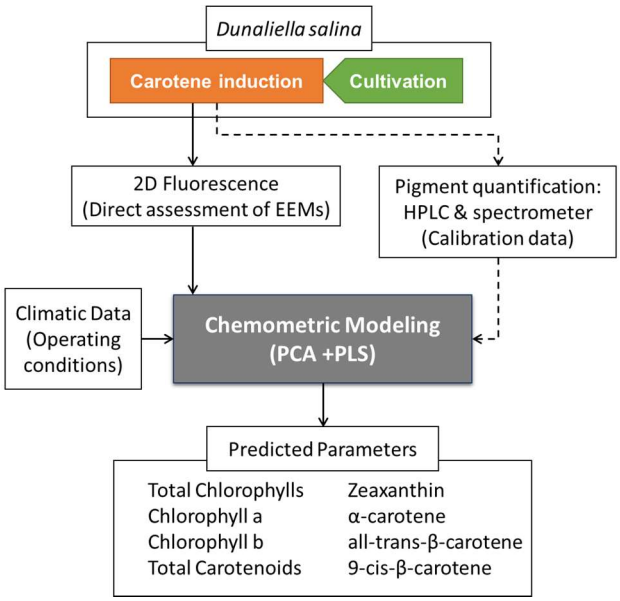


Figure III.1 Schematic representation of the methodology followed. Dashed arrows represent the off-line measurements required for model calibration.

Firstly, Principal Component Analysis (PCA) was used to compress and deconvolute the fluorescence excitation-emission matrices (EEMs) into principal components (PCs), by extracting the most relevant information. Shortly, the first PC was selected for explaining the higher variance possible; then a second PC was selected for explaining the higher remaining variance possible with the constrain of being orthogonal to the first one; and so on. The calculated PCs are uncorrelated and ordered according to the variance explained. PCA was applied to all spectra acquired and the first ten PCs were selected to be used as input in the Projection to Latent Structure (PLS) modelling, since they captured more than 99% of the variance.

Projection to Latent Structures (PLS) modelling was used to stablish multilinear correlations between the PCs of fluorescence and climatic conditions (inputs) and the pigments concentration, chlorophylls and carotenoids (outputs). The models were developed using 41 fluorescence spectra, corresponding to 41 samples. From this initial data set, 75 % was randomly select to train the model, i.e., to find the best model that explains the higher variability found in the data with the lower errors. Then, the remaining 25% of the data were used to validate the model developed previously and assessed the prediction quality. All data were normalized before being used in the PLS models.

For each pigment, three strategies were attempted based on the initial inputs used: first using only ten PCs from the fluorescence; second, using PCs from the fluorescence and climatic data of the day of the sampling; third, using PCs from the fluorescence and climatic data with cumulative effect towards the sampling day, which means, for each sampling day was considered not only the values of that day, but also the values of the previous sampling days (one cumulative parameter is the sum of all measures of that parameter until the day of sampling).

The selection of the useful PCs and climatic inputs for the models was performed by iterative stepwise elimination (ISE) [42]. The quality of each model was evaluated by the percentage of the variance captured, the slope and coefficients of determination (R^2) of the validation and the training data sets, as well as the root mean square errors of cross validation (RMSECV) and prediction (RMSEP). All algorithms were implemented in Matlab, using the n-way tool box for PCA and PLS through PARAFAC and nPLS functions, respectively [72].

III.3. Results and Discussion

Six pilot-scale flat-panel experiments were performed to induce carotenoids formation in *D. salina*, and two main approaches were evaluated, the importance of using climatic inputs and the importance of choosing the adequate methodology to calibrate the models. Fluorescence EEMs were obtained for each sample and the pigments profile was quantified. The content of two chlorophylls (*a* and *b)* and five carotenes (lutein, zeaxanthin, α-carotene, β-carotene and 9-cis-β-carotene) were measured using two different methodologies: spectrophotometric and by HPLC.

Figure III.2 shows the fluorescence spectra acquired from *D. salina* culture during carotenogenesis in the sixth batch for days one, six and fourteen after the inoculation. It is possible to distinguish 2 fluorescence regions, one at excitation-emission wavelengths of 275 nm and emission between 300 and 350 nm, and another at higher emission wavelengths, above 650 nm. The first region is defined as protein-like fluorescence, mainly because of aromatic aminoacid tryptophan fluorescence [69]. The second region is defined as the pigment band [75]. The differences between the spectra from the inoculation day (day one, Figure III.2a) and six days after (Figure III.2b) is quite visible, mainly due to the pigments band fluorescence intensity. In these samples the ratio carotenoid/chlorophyll increased from 0.21, on day one, to 3.04, on day six. However, when analyzing the fluorescence spectra of day fourteen (Figure III.2c) some differences can be also spotted, although the ratio carotenoid/chlorophyll was 3.01, similar to day six. During carotenogenesis experiments, the biomass concentration also increased through time, resulting in the intensification of the emission and excitation light scatter. It is noteworthy that fluorescence was assessed directly on the samples, without any dilution or sample preparation, as it would be if it was measured using an optical probe directly coupled to the bioreactor. Therefore, the effect of increased turbidity and color were also a source of information assessed by fluorescence spectroscopy. For these reasons, the direct interpretation of fluorescence spectra is too complex, and so chemometric tools were used.

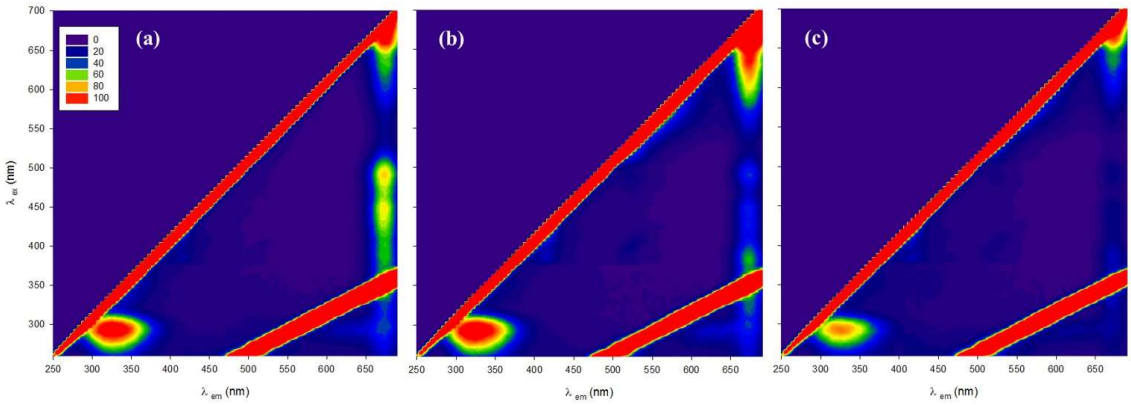


Figure III.2 2D Fluorescence spectra of D.salina during the sixth batch; (a) day one (inoculation), (b) after six days, and (c) after fourteen days. X-axis displays the wavelengths of emission, Y-axis the wavelengths of excitation and the intensity of the fluorescence is represented through color gradient. Two distinct fluorescence regions can be noticed, a protein-like region (excitation wavelength of 275 nm and emission wavelengths between 300 and 350 nm) and a pigment band (emission wavelengths above 650 nm).

Three input strategies were studied to develop the PLS models for prediction of each compound individually: without climatic inputs; and with daily or cumulative climatic inputs. However, all models needed the input of the climatic parameters to successfully estimate the pigments content, which validates the importance of climatic conditions on microalgae cultivation in outdoors facilities. Therefore, the models developed without climatic conditions inputs are not shown in this work. The use of daily data as input can give extra information about the impact of climatic conditions on the culture status of the sampling day, although the cumulative effect can be important when studying the induction of carotenoids.

The most relevant models achieved for chlorophylls and carotenoids are shown in Table III.2 and III.3, respectively.

III.3.1. Chlorophylls

III.3.1.1. Chlorophyll a

For chlorophyll *a* modelling, and using spectrophotometric methodology as calibration method, two models are presented, one with daily climatic inputs and another with cumulative climatic inputs. In this case, the use of cumulative climatic inputs did not improve greatly the quality of the model, although a slight increase was observed. The variance explained increased from 83.3 %, for daily climatic inputs, to 89.2 %, for cumulative. An important difference between the models is the selection of the inputs used by each model. When calibrating the model with daily climatic inputs, less inputs were selected in general, either from fluorescence PCs or climatic. Only one climatic input was selected, the fraction of photosynthetically active radiation (FPAR), an important parameter in this study, and particularly for this model, because of the direct correlation between photosynthetic activity and content of green pigments. When calibrating with cumulative climatic data, more fluorescence PCs were selected, as well as two different climatic inputs, the presence of clouds and irradiance. The irradiance is a parameter that indicates the amount of light that reaches the Earth surface and was measured daily at the A4F - Algae for Future facilities. Unlike FPAR measurements, all light flux is considered in irradiance measurements. The presence of clouds can also give an indication of the amount of light that is being blocked from

the Earth surface. Although different light measurements are used to explain the variability observed in these experiments, it is safe to enhance the importance of light settings when working in outdoor systems such as in this study.

Table III.2 Statistical parameters of the selected models for chlorophylls prediction. The climatic inputs code is shown in Table III.1

	AM	Input	Var (%)	RMSE P	RMSE CV	Validation		Training		Selected Inputs	
						R^2	Sl	R^2	Sl	Fluor.PCs	Climatic
Chl *a*	*Spect*	daily	83.3	$4.42x10^{-10}$	$7.17x10^{-10}$	0.85	0.99	0.83	1.00	1 2 4 5 6 8 9 10	15
		cumul.	89.2	$5.97x10^{-10}$	$6.26x10^{-10}$	0.78	0.85	0.89	1.01	1 2 3 4 5 6 7 8 9 10	14 16
	HPLC	cumul.	69.7	$2.04x10^{-09}$	$3.17x10^{-09}$	0.71	1.25	0.70	1.00	1 2 3 6 8 10	15 16
Chl *b*	*Spect*	daily	47.5	$2.99x10^{-10}$	$5.76x10^{-10}$	0.75	1.17	0.47	1.00	3 4 5 6 9 10	-
	HPLC	cumul.	67.8	$3.77x10^{-10}$	$9.27x10^{-10}$	0.79	0.86	0.68	1.00	1 2 3 4 5 6 7 8 9 10	12 14 15 16
Total Chl	*HPLC*	cumul.	74.2	$8.20x10^{+04}$	$1.25x10^{+05}$	0.79	0.88	0.74	1.00	1 2 4 5 6 7 8 9 10	11 12 13 14 16

Chl a – Chlorophyll a; Chl b – Chlorophyll b; Total Chl – Total Chlorophyll; AM – Analytical Method; Spect – Spectrophotometric; Input – Type of climatic input; cumul. – cumulative; Sl – slope; Fluor.PCs – Fluorescence PCs.

(*) RMSEP and RMSECV values are presented as mg/cell, except for the models of total Chlorophylls, which are in area units.

Using a different methodology to calibrate the model, like HPLC, several differences are noteworthy, such as less fluorescence PCs inputs were selected. However, less variance was explained (69.7 %) and the prediction error (RMSEP) was higher. This tendency was also observed in the experimental error of the methodologies, $3.99x10^{-10}$ mg/cell for HPLC and $1.16x10^{-10}$ mg/cell for spectrophotometer, which can be also related with differences between extraction methodologies (sonication and 60 °C were used for the spectrophotometric analysis, and 4 °C for HPLC analysis). However, the average chlorophyll *a* concentration was higher when measured with HPLC methodology, thus, the error related to the average for both methods was similar (around 5 %).

Different methods for pigments quantification and extraction have been discussed in the literature in the past years. For different microalgae or biological state ("green" or "orange"), the selection of the extraction methodology and the analytical method should be carefully chosen. In this study, when aiming to monitor chlorophyll *a* in *D. salina*, similar models were achieved using HPLC or spectroscopic data. However, the spectrophotometer method presented lower error values being a more accurate method. Since this methodology is less laborious and less time consuming, using it for the model calibration can be an advantage.

III.3.1.2. Chlorophyll b

When aiming to monitor chlorophyll *b*, differences in the methodology to calibrate the model are also noticeable. Although daily climatic inputs were provided in the model calibrated with spectrophotometer quantification, none of the parameters were selected. Together with a low number of fluorescence PCs selected by the model, a low variance was explained (47.5 %). When using HPLC to calibrate the model, a higher R^2 for both training and validation sets was observed. However, the difference between RMSEP and RMSECV shows that the model is not robust to predict chlorophyll *b* content. Furthermore, high RMSECV may indicate some degree of overfitting (most probably due to the high number of inputs required). Like the tendency observed for chlorophyll *a*, chlorophyll *b* concentration models have a higher RMSEP when using HPLC results than spectrophotometric. However, this tendency was not observed in the analytical error of the methodologies, 7.73×10^{-11} mg/cell for HPLC and 1.52×10^{-10} mg/cell for spectrophotometry.

The RMSEP values for monitoring chlorophyll *b* content were higher than desired. The high standard deviation found on the analytical measurements interferes greatly with the model calibration, leading to strong limitations. These findings confirm the importance of carefully choosing the analytical methodology for both extraction and quantification. Furthermore, monitoring chlorophyll *b* content by 2D fluorescence spectroscopy could be enhanced with the addition of more experimental points through a longer calibration time. Nevertheless, even with the low accuracy due to the results of the analytical method for the model calibration, the possibility of using this technology online is a great advantage since none of the current methodologies enables it.

III.3.1.3. Total chlorophylls

After analyzing the individual content of each chlorophyll, and due to the high variability observed, a different approach was attempted to monitor the chlorophyll content as total amount, by the sum of chlorophyll *a* and *b* peak areas obtained with HPLC analysis. A variance of 74.2 % was explained using nine of the ten PCs of fluorescence and several cumulative climatic parameters. Similar R^2 for validation and training data sets (0.79 and 0.74, respectively) and a small difference between RMSEP and RMSECV were observed.

When comparing all the models for chlorophyll *a*, *b* and total content, less climatic inputs are needed to predict chlorophyll *a* content, with more variability explained. Regarding the two analytical methodologies, spectrophotometric and HPLC analysis, due to a different extraction procedure, it is difficult to assert which one is more suitable. Slightly better results were obtained for models calibrated with spectrophotometric analysis, but that can be related with the extraction procedure. To confirm this approach, the impact of the heat versus room temperature extraction using the same apparatus should be further tested.

Simultaneous quantification of chlorophylls and the interferences between them has been a topic of study by several authors. Moberg *et al*. [75] used 2D fluorescence spectroscopy to analyze a mixture of six pigments standards, chlorophylls *a*, *b* and *c*, and their respective degradation products, pheophytins *a*, *b* and *c*. The authors observed that the signals from the different pigments overlap, and for that reason, their direct quantification using only a specific region of the spectra can be misleading. When validating 2D fluorescence spectroscopy using

HPLC they found that the RMSEP for chlorophyll *a* was six times higher than with fluorescence spectroscopy. Also, chlorophyll *b* quantification was not always possible due to limitations in the quantification limit of the HPLC methodology [75].

The models presented were selected for their better results on the prediction of chlorophylls content (*a*, *b* or total) when using 2D fluorescence spectroscopy as input. It is noteworthy that most of these models also used climatic data as input, either as daily or cumulative. To our knowledge, this approach of using climatic inputs to help monitor and predict pigment induction online was not used before. The results obtained enhance the importance of using the operating conditions (climatic data), coupled with the 2D fluorescence spectroscopy, to monitor chlorophylls profile when cultivating *D. salina* in outdoor facilities.

III.3.2. Carotenoids

III.3.2.1. Total carotenoids

When aiming to predict total carotenoids content in *D. salina* biomass, better result was achieved when using spectrophotometric quantification to calibrate the model, and also using cumulative effect of the climatic data. Lower RMSEP (2.51×10^{-9} mg/cell) and a lower RMSECV (3.85×10^{-9} mg/cell) were observed, indicating the robustness of the model, and higher R^2 for validation and training data sets (0.89 and 0.82, respectively). For both models, all climatic inputs were selected, once again revealing the importance of the climatic conditions in the outdoor cultivation of microalgae.

It is well-known that there is no harmonized protocol for carotenoids extraction, with many options available in literature depending on physical characteristics, composition and amount of water in the samples, although a chromatographic separation seems to be unanimous. Nevertheless, optimization of the chromatographic separation is required to obtain a good resolution of peaks, which is not always achieved, due to the complexity of these extracts and presence of compounds with similar structures. Petri *et al.* study on orange fruit extract showed that, when analyzing 52 carotenoids by LC-DAD-MS/MS, most of them co-elute, revealing the challenge of separating and quantifying every single carotenoid [92]. Through the HPLC methodology used it was possible to discriminate several different pigments like zeaxanthin, α-carotene, all-trans-β-carotene and 9-cis-β-carotene. Therefore, different models were developed to assess each compound individually (as shown in Table III.3 and in Figure III.3).

Table III.3 Statistical parameters of the selected models for carotenoids prediction. The climatic inputs code is shown in Table III.1

	AM	Input	Var (%)	RMSE P	RMSE CV	Validation		Training		Selected Inputs	
						R^2	Sl	R^2	Sl	Fluor.PCs	Climatic
Total Carot	*Spect*	daily	79.6	$2.75x10^{-09}$	$4.54x10^{-09}$	0.79	1.04	0.80	1.00	1 2 3 4 5 6 7 8 9 10	11 12 13 14 15 16
		cumul.	81.7	$2.51x10^{-09}$	$3.85x10^{-09}$	0.89	0.96	0.82	1.00	1 2 3 4 6 7 8 9 10	11 12 13 14 15 16
	*HPLC**	cumul.	64.9	$6.30x10^{+06}$	$7.35x10^{+06}$	0.72	0.91	0.65	1.00	2 3 7	12
Zeaxanthin		daily	74.4	$1.94x10^{-10}$	$1.99x10^{-10}$	0.69	0.86	0.74	1.00	2 3	13
		cumul.◄	79.3	$2.52x10^{-10}$	$1.78x10^{-10}$	0.62	1.11	0.79	1.01	3 8	13 16
α-Carotene		daily◄	87.8	$4.90x10^{-10}$	$3.21x10^{-10}$	0.63	1.27	0.88	1.00	1 2 3 4 5 6 7 8 10	11 12 13 15 16
all-trans-β-Carotene		daily	70.4	$4.97x10^{-09}$	$4.61x10^{-09}$	0.79	1.47	0.70	0.99	1 2 3 6 8	12 13
		cumul.◄	79.1	$3.31x10^{-09}$	$4.05x10^{-09}$	0.80	1.18	0.79	1.00	2 3 4 10	13 16
9-cis-β-Carotene		daily	64.2	$2.17x10^{-09}$	$3.80x10^{-09}$	0.73	0.83	0.64	1.00	2 3 6	13
		cumul.◄	65.7	$2.65x10^{-09}$	$3.73x10^{-09}$	0.88	1.29	0.66	1.00	2 3 4 5 6 7 8 10	11 12 13 14 15 16

Total Carot – Total Carotenoids; AM – Analytical Method; Spect – Spectrophotometric; Input – Type of climatic input; cumul. – cumulative; Sl – slope; Fluor.PCs – Fluorescence PCs.

(*) RMSEP and RMSECV values are presented as mg/cell, except for the models of Total Carotenoids, which are in area units.

(◄) Models represented in Figure III.3.

III.3.2.2. Zeaxanthin

Zeaxanthin is widely present in photosynthetic multicellular organisms, including microalgae, and is always found together with lutein and β-carotene [93]. Two models were selected to monitor zeaxanthin content in *D. salina* cells, using daily and cumulative data. For both models, two fluorescence PCs were selected, and the number of hours of sunlight was a common climatic input. However, the use of cumulative climatic data and the inclusion of the irradiance information, led to a model able to explain more data variability (79.3 %), with a slight increase on RMSEP. It is noteworthy the low number of inputs required to predict zeaxanthin, and the similarity between RMSEP and RMSECV, meaning that robust models were achieved based on fluorescence and sunlight.

III.3.2.3. α-Carotene

The model developed has the highest variance explained of all carotenoids in this study (87.8 %). The climatic parameters with impact for the determination of α-carotene were the thermal amplitude, precipitation and parameters related with light, such as number of hours of sunlight,

FPAR and irradiance. When studying the influence of cumulative effect of climatic parameters, the models obtained did not revealed an improvement (data not show).

III.3.2.4. β-Carotene

β-Carotene is the carotenoid present in higher concentration in *D. salina*. Analysis of the oily globules' structures where β-carotene is stored showed mainly two stereoisomers: all-trans and 9-cis. It is reported in the literature that the light intensity at which the cultures are exposed influences greatly the final ratio between all-trans and 9-cis, and also in the total amount of β-carotene accumulated within each cell [81].

For both forms of β-carotene, the models where climatic parameters were used as cumulative effect led to higher variance explained when compared with the ones when using daily climatic inputs. For all-trans-β-carotene, the use of cumulative climatic inputs also decreased both errors, RMSEP and RMSECV. Between the climatic parameters provided, the exposure to sun light was important to explain the development of carotenoids during nitrogen starvation, especially when referring to pilot-scale outdoor *D. salina* cultivation. For 9-cis-β-carotene, slightly higher errors (RMSEP and RMSECV) were observed when using the climatic input as cumulative, however the difference between them was lower, revealing a more robust prediction model.

In general, all four carotenes models presented similar results. When plotting the models (Figure III.3) it is possible to observe some dispersion of the values, resulting in lower R^2, but no outliers were found, which lead us to conclude that performing more assays would benefit the prediction capability. This study proves that 2D fluorescence spectroscopy is a powerful tool that can simultaneously detect and quantify different carotenoids, with the advantageous of giving an estimation at real-time.

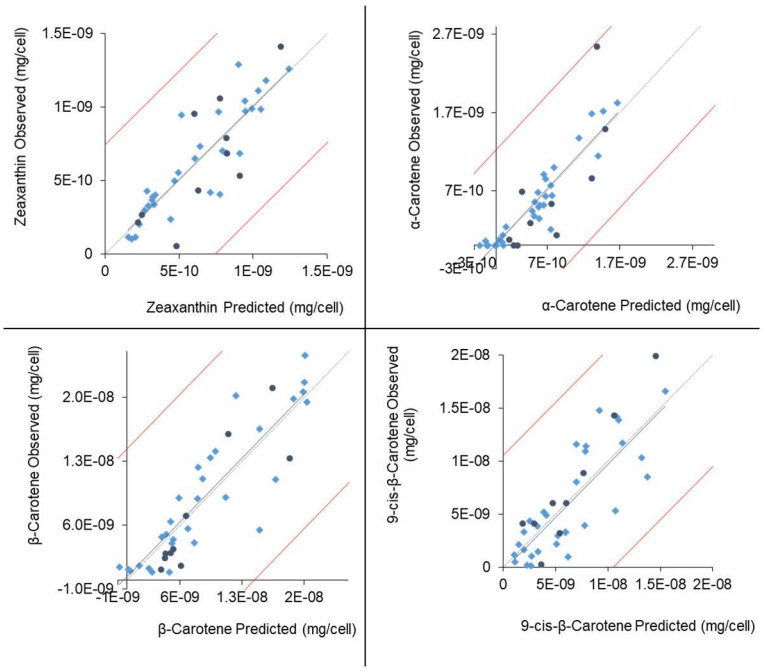


Figure III.3 Carotenoids concentration prediction models, from left to right and top to bottom: zeaxanthin, α-carotene, β-carotene and 9-cis-β-carotene. Training (◊) and validation (●) data are presented as mg/cell. Statistical parameters of the models represented are displayed in Table III.3.

III.3.3. Application perspectives

It is known today that, for the microalgae industry to be as competitive as other plant-based refineries, the biorefinery needs to be improved and the tendency is to produce and recover high value compounds. The industrialization of microalgae increased the need for better methodologies to monitor (and control) these complex biological systems. Among several advantages, 2D fluorescence spectroscopy can be coupled with an optical probe directly immersed in the bioreactor/process under study, and enables the collection of information frequently and at real-time, without the need for sampling.

When aiming for pigment content monitoring, there are several methodologies in the literature, but they require a sampling step and extraction procedure, which normally are time consuming analysis. Other spectroscopic and fluorometric methods studied so far also require sampling, either for extraction of pigments, for dilution or for the use of other methods able to remove interferences that prevent a direct correlation [94–96]. Furthermore, such tests aim at measuring total pigments or a specific compound but are not able to discriminate simultaneously between different pigments. Therefore, even if the accuracy of the prediction model can be lower than the analytical quantification, the possibility of having several and repeated fluorescence measurements at real time is a motivating advantage in industrial production.

For an industrial application, the prediction accuracy of such models as developed in the present work can be improved with the increase of the calibration period, allowing the collection of more data points to attenuate the prediction error and to account for higher variability of the operating conditions. Also, the incorporation of extra information about the cultivation process, such as parameters that are known to influence the final biomass composition (like media composition) can possibly lead to an improve in the model's prediction capability.

III.4. Conclusions

The approach presented in this work validates the use of 2D fluorescence spectroscopy and climatic data combined, coupled with chemometric modelling, as monitoring tool for pigments as a bulk, total chlorophylls and total carotenoids, but also as individual compounds, in *D. salina* biomass.

Mathematical models were successfully developed to assess chlorophylls content as chlorophyll *a, b* and total chlorophyll, and carotenoids content as total content, and specifically, as zeaxanthin, α-carotene, all-trans-β-carotene and 9-cis-β-carotene. To validate the models, a careful choice of the analytical method is required, since experimental errors and accuracy of the analytical methods will highly influence the model prediction quality. Furthermore, the possibility to include climatic data, related with the environmental conditions during microalgae production in outdoor facilities, enhances the prediction of pigments content. Additionally, the possibility of using a fluorescence probe directly in the production system avoids the need of sampling and of an extraction step.

Although the analytical methodologies commonly used (such as HPLC) are accurate and able to assess different compounds simultaneously, they are expensive, time consuming and require a laborious extraction step, while 2D fluorescence spectroscopy (coupled with appropriate

mathematical tools) proved to be an effective solution for the industrial application to address the monitoring of these high value and in high demand compounds.

III.5. 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”, under the topic “The CO2 Microalgae Biorefinery”, and by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2016-CPF-2907-05. FCT/MCTES is also acknowledged for the Post-Doctoral Fellows grants SFRH/BPD/95864/2013 and SFRH/BPD/79533/2011, and PhD Fellow grant SFRH/BD/108894/2015. The authors would like to thank the company A4F-Algae for future (Portugal), who performed all the pilot scale cultivation trials and provided the microalgae used in this work, The Marine Biological Association (Devon, UK) and NBT Ltd (Israel).

IV

Fluorescence coupled with chemometrics for simultaneous monitoring of cell concentration, cell viability and media nitrogen during production of carotenoid-rich *Dunaliella salina*

Abstract

Two-dimensional (2D) fluorescence spectroscopy was investigated as a monitoring tool for cultivation, harvesting and effluent treatment of carotenoid rich *Dunaliella salina* biomass, 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, based on fluorescence excitation-emission matrices (EEMs) acquired directly from media. Cell concentration during cultivation and harvesting can be predicted by a single model capturing 92.0% of 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 of 0.73 for validation. Nitrate concentration is 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 of 0.80 for validation. Therefore, this work proves 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.

Published as: Sá, M., Monte, J., Brazinha, C., Galinha, C. F., Crespo, J. G., 2019. Fluorescence coupled with chemometrics for simultaneous monitoring of cell concentration, cell viability and medium nitrogen during production of carotenoid-rich *Dunaliella salina. Algal Research 44:101720..*

IV.1. Introduction

Dunaliella salina, a “green” unicellular microalga, is remarkably halo-tolerant due to its capability of *de novo* synthesis of glycerol with the increase of salt stress. Also, under salt and light stress conditions, nutrient depletion or extreme temperatures, *D. salina* accumulates a high concentration of carotenoids, being nowadays considered as a food source [54,55,97]. Nevertheless, the industrial production of *D. salina* is still done for niche markets, like high-value products such as pigments. To increase the economic interest of microalgae cultivation, in general, the biomass cost needs to be reduced while the production scale increases [47].

One of the problematics of microalgae cultivation is the dilute suspension that characterises these cultures [52]. In fact, increasing the concentration of biomass during the cultivation can lead to a decrease of 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 [52]. In that sense, membrane-based harvesting systems offers a potential solution to concentrate biomass since it is a highly reliable and direct scalable technology, enables a hybrid harvesting process, due to the possibility of coupling to other technologies like centrifugation, and it is low energy and low cost [51,52,98–101]. As described by Monte *et al.* [53], including a membrane pre-concentration step before the usual centrifugation procedure, reduces the number of centrifuges needed, reducing the energy consumption (up to 63 %) and capital cost (up to 60 %) of the procedure.

Furthermore, the use of a membrane dewatering system enables the recovery of the permeate, suitable for microalgae cultivation, a critical requirement to minimise the use of new fresh media [52]. The costs associated with the cultivation media make the reuse and recycling of this permeate a key parameter to reduce water footprint [102–104]. Although this permeate is usually clear, free of cells and of high molecular weight contaminants, it can contain some remaining nutrients, cell debris and organic compounds [102]. Therefore, before reusing the permeate, a further intermediate treatment step, like ozone or UV light, needs to be reassessed [105].

As it has been occurring in other complex biological systems, the microalgae biorefinery can enhance its potential 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 with resort to an optical density (OD) probe at several wavelengths. However, it is known that pigments and organelles affect the light absorbance significantly and the microalgae cellular composition may change along the cultivation time. Thus, this methodology requires a careful selection of the right wavelengths [106,107]. 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 [108].

Being able to monitor cell concentration and integrity online is of extreme importance when thinking of a membrane pre-concentration step within *D. salina* biorefinery. The absence of rigid cell wall coupled with the shear stress inherent to a membrane filtration process may lead to severe cell damage and, consequently, to the loss of highly valuable compounds to the medium.

As mentioned by Lamers *et al.* [109], nitrogen depletion, usually in the form of nitrate or ammonium, is a frequently practiced strategy to enhance carotenoids concentration in *D. salina*. The possibility of monitoring nitrogen concentration during cell growth and induction of carotenoids, and also during the treatment of the permeate recovered from the membrane harvesting step, will make it possible to improve the design and operation of these processes.

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), being suitable for applications such as online monitoring of complex samples [29,77]. 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, about the interactions between them and the interactions with the media, scattering also being a source of information about the suspended material present.

The use of *in situ* fluorometry technique was reported by Shin *et al.* [110] 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 an online 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 aminoacids) [78].

In order 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).

IV.2. Material and Methods

IV.2.1. Growth conditions of Dunaliella salina and carotene induction

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

The inoculum was maintained and scaled-up in A4F – Algae for Future (Lisbon, Portugal) facilities. The culture was maintained in artificial salt water (ASW), an A4F Industrial Medium [105] and successively scaled-up volumetrically in round flasks up to 5 L, at 25 °C and under continuous illumination (150 µmol/(m^2.s)) provided by fluorescent tubes. Continuous aeration (0.1-0.3 % vvm) was provided through 0.2 µm filtered air supplemented with 2 % of CO_2, 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), with an optical

pathlength between 7 and 10 cm, and adjustable volume through adjustable length. The culture was exposed to outdoors 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 9.3x105 cells/mL and Car/Chl (carotenoids/chlorophylls) ratio between 3 to 14 (Table IV.1).

Table IV.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.h))	MWCO (kDa)	Memb. relaxation (y/N)	Consecutive biomass increment (Y/n)	Car/Chl*	Initial cell concentration (cells/mL)
30	100	y	n	14.2	1.6E+05
25	100	N	n	12.5	7.9E+05
25	500	y	n	7.7	7.2E+05
25	100	y	n	4.6	9.3E+05
25	100	y	n	3.0	1.6E+05
25	100	y	n	2.2	5.5E+05
15	100	y	n	5.1	5.9E+05
10	100	y	Y	5.5	6.1E+05
10	100	y	Y	4.5	3.3E+05
10	100	y	n	5.6	6.3E+05
10	100	y	n	7.3	4.7E+05

**Carotenoids to chlorophylls ratio*

IV.2.2. Operation conditions for membrane harvesting

A small pilot filtration unit composed by hollow fibres of polyethersulfone (GE Healthcare, UK) was used for the harvesting experiments (63.5 cm of length and 0.48 m^2 of membrane area). 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 100 kDa and with a MWCO of 500 kDa; imposed permeate volumetric flux – 10, 15 and 25 L/(m^2.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 IV.1. Three replicates were performed at a permeate flux of 25 L/(m^2.h), using the 100 kDa membrane and applying relaxation. For an imposed permeate flux of 10 L/(m^2.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^2.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 IV.1. The experiment started with an initial biomass load of 20 kg (concentration factor (CF) of 1) and was filtered until reaching 10 kg (CF 2), when a load of 10 kg of new fresh biomass was added to the concentrate (CF 1.5). This second load of 20 kg was filtered until reaching 10 kg (CF 3), and a third load of fresh biomass (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.* [105].

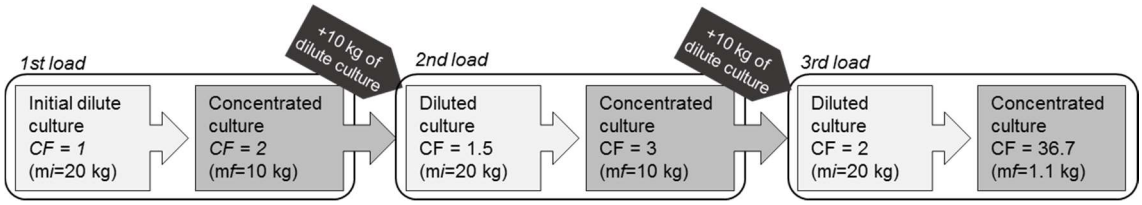


*Figure IV.1 Schematic representation of the procedure used during the consecutive membrane harvesting experiments. CF represent the concentration factors during the experiment, calculated in function of mass of biomass, in the beginning of the load (*m*i) and in the end of the load (*m*f).*

IV.2.3. Permeate treatment: oxidation, photodegradation and membrane purification

The permeate obtained during harvesting of *D. salina* by ultrafiltration was pre-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^2.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_2O_2), and with pH correction to 9. Both pH correction and H_2O_2 addition are known to accelerate the ozone decomposition and the oxidation process [104]. Further details about the treatment of the permeate by oxidation and photodegradation is provided in Monte *et al.* [105].

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

IV.2.4. Sampling procedure and analysis

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 IV.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 analysis were performed for the samples collected as described below.

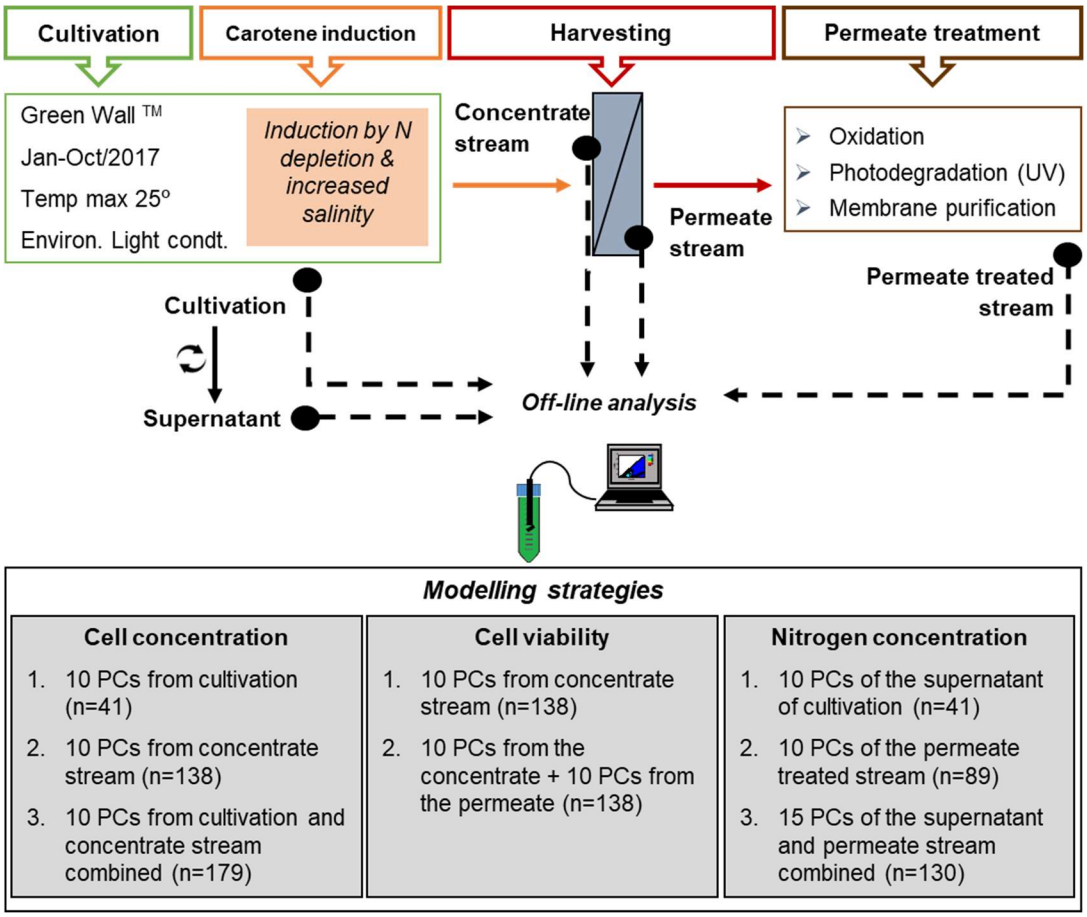


Figure IV.2 Schematic representation of the overall process and indication of the sampling points. Samples were withdrawn from cultivation broth and its supernatant, from the concentrate and permeate streams during membrane filtration, and from during permeate treatment.

IV.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 maximum 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, viability percentage was used as cell integrity indicator.

IV.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 colorimetric method implemented in a segmented flow analyser (Skalar San++, Skalar Analytical, The Netherlands).

IV.2.5. Fluorescence spectroscopy

All the samples in this study were analysed by 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 immersion of an optical fibre probe directly in the samples without 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.* [78]: 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).

IV.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 *n* number of initial variables, *n* linear combinations will be obtained, the PCs, characterised for being uncorrelated, orthogonal and ordered according the variance explained (the first PC explains the higher part of the variance, and so on) [16,35].

Fluorescence spectra were acquired in 5 streams: cultivation broth and its supernatant from carotenogenesis, concentrate and permeate from harvesting, and treated permeate samples (Figure IV.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 variance. For models where the fluorescence of two streams were 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 nr of PCs used for each model is shown in Table IV.2 and Figure IV.2.

Projection to Latent Structures (PLS) modelling was used to describe multilinear correlations between the fluorescence PCs (inputs) and the off-line measures of cell concentration, cell viability and nitrate concentration (outputs). More detailed information about the PLS modelling can be found here [29,111]. Table IV.2 summarise 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 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.

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

	Inputs			Prediction model
Output	**Operational Process**	***n***	**Strategy PCs**	
Cellular conc.	Carotenogenesis	41	10 PCs	Fig IV.4 (a)
	Harvesting	138	10 concentrate PCs	Fig IV.4 (c)
	Carotenogenesis+ Harvesting	179	10 PCs	Fig IV.4 (e)
Cellular viability	Harvesting	138	10 concentrate PCs	Fig IV.5 (a)
			10 concentrate + 10 permeate PCs	Fig IV.5 (c)
Nitrate conc.	Carotenogenesis	41	10 PCs	.
	Permeate treatment	89	10 PCs	-
	Carotenogenesis + Permeate treatment	130	15 PCs	-
		58	15 PCs	-

To develop the PLS models, 75% of the experimental data was randomly attributed for training purposes and the remaining 25% for model validation. 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 [42–44]. 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 tool box of Matlab was used to implement PARAFAC and nPLS [112].

IV.3. Results and discussion

In Figure IV.3 is shown an emission-excitation fluorescence spectrum acquired during a carotenogenesis cultivation experiment. Is it noteworthy 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). A more detail description about the fluorescence signal acquired for *D. salina* microalgae is provided in Sá *et al.* [78].

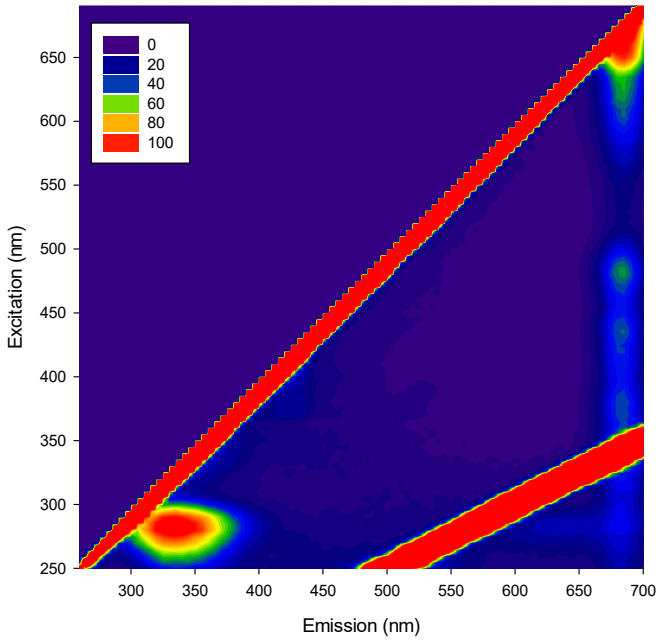


Figure IV.3 Fluorescence spectrum of Dunaliella salina biomass, acquired during carotenogenesis experiments. The emission wavelengths is in the x-axis, the excitation on the y-axis, and the fluorescence intensity in colour-grade scale. Two distinct areas can be identified: protein-like region for excitation 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.

IV.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 (Figure IV.4).

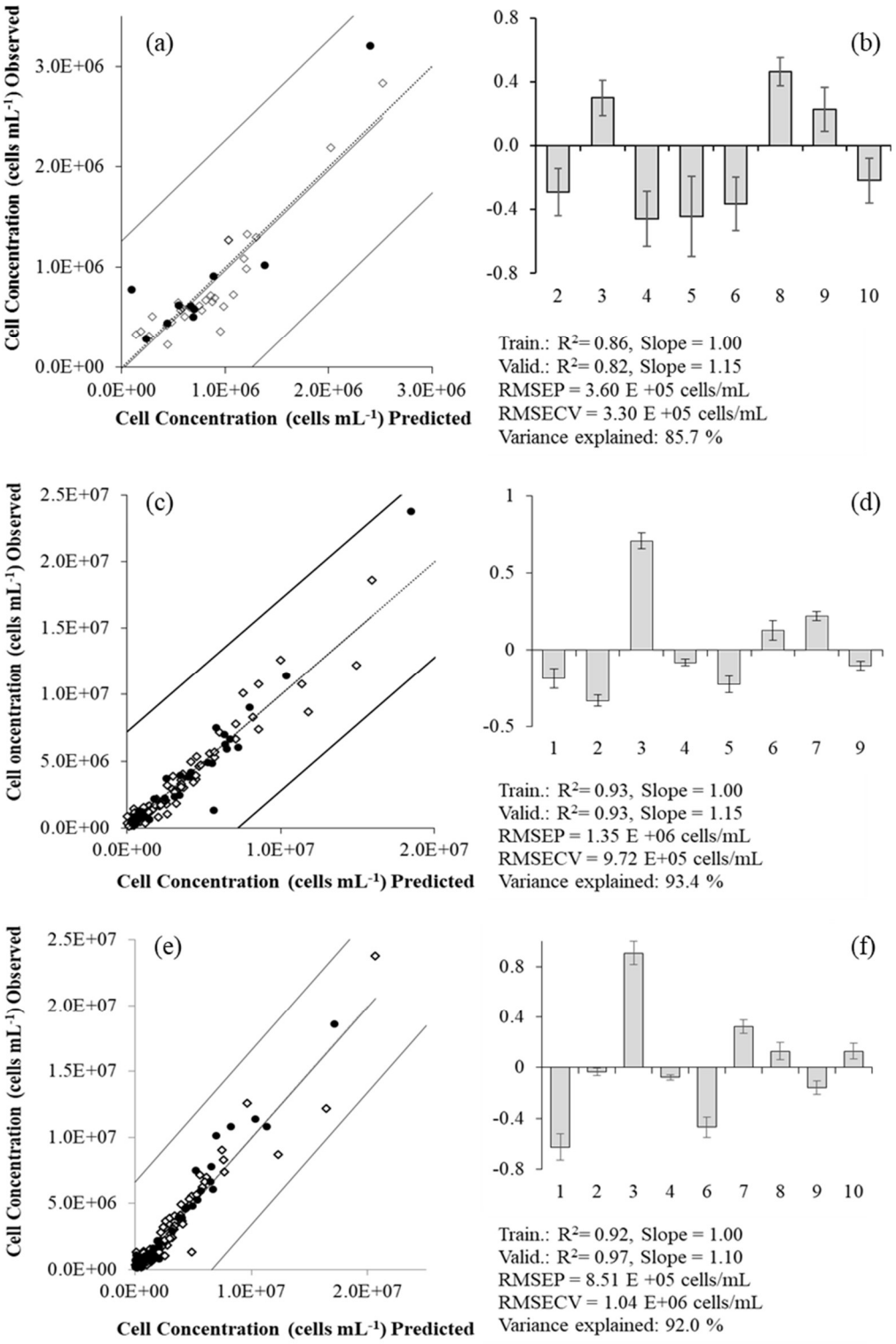


Figure IV.4 Cell concentration prediction models (a) for cultivation experiments samples; (c) for membrane harvesting experiments, using only PCs acquired in the concentrate stream; and (e) for two distinct processes, cultivation and harvesting, using fluorescence EEMs from both processes compiled together. Respective normalized regression coefficients are shown in (b), (d) and (f), in the same order. Training (◊) and validation (●) 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.

Figure IV.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 outlier data. 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 prediction 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 $3.60x10^5$ 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 to explained 85.7 % of the variance captured (Figure IV.4(b)).

Cell concentration is a vital factor for the economic feasibility of microalgae biorefinery. A similar study was performed by Sá *et al.* [78] 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 “orange” *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 (Figure IV.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 (Figure IV.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 variance (Figure IV.4(d)) and a RMSEP of $1.35x10^6$ 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. 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 10 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 % (Figure IV.4(e)). From the ten new PCs given to the model, nine were selected (Figure IV.4(f)). The difference between RMSEP ($8.51x10^5$ cells/mL) and RMSECV ($1.04x10^6$ cells/mL) was still higher than desired. As it is possible to observed from the data dispersion in Figure IV.4(e), more experimental data at the high cell concentration range could be useful to increase the robustness of the model, since only few points were observed for concentrations above $1.50x10^7$ cells/mL. Nevertheless, the addition of cultivation data to the harvesting model increased the prediction quality of the model. However, since the use of an

extended range of concentrations (namely in harvesting) results in higher errors, 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 [69]. The use of this technique in a microalgae biorefinery was first described by Sá *et al.* [78], 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 now developed for cell concentration can be used despite cellular variations in composition along the cultivation time.

IV.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 loss of valuable intracellular compounds to the medium. Thus, monitoring cell viability during harvesting is of extreme significance 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 Figure IV.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 prediction value. According to Sá *et al.* [78], values of loss of viability above 10% are considered high when aiming to harvest *D. salina*, since it corresponds to a noticeable cell damage.

To improve the predictability of the cell viability model using 2D fluorescence spectroscopy, a second input strategy was attempted 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 (Figure IV.5(d)). This phenomenon was previously reported by Sá *et al.* [78] 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) (Figure IV.5(c)). Both slopes were close to 1 and the RMSEP decrease to 9.5 %. Using five PCs from the fluorescence EEMs acquired in concentrate and three PCs acquired in permeate, the model can explain 76.3 % of the variance captured 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.

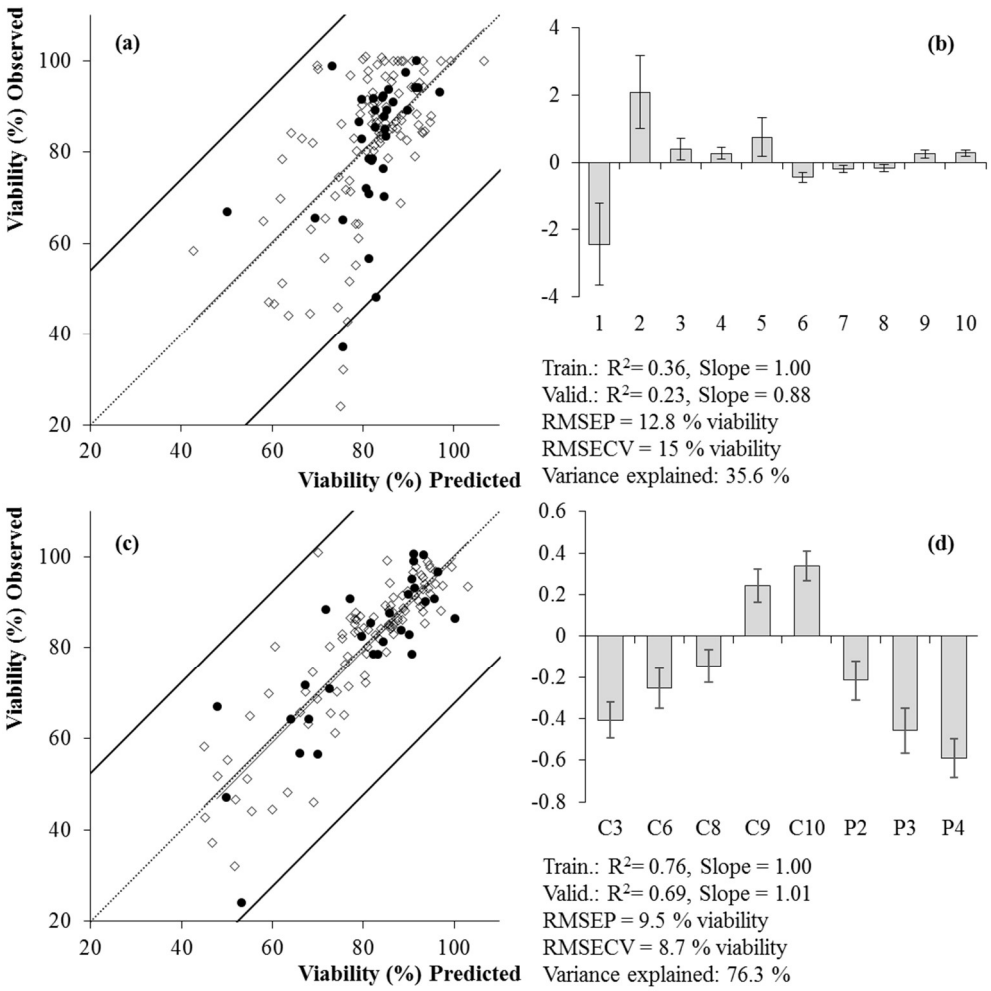


Figure IV.5 Prediction models for viability percentage during membrane harvesting experiments: using ten PCs of the fluorescence acquired in the concentrate stream (a) and adding ten PCs of the fluorescence acquired in the permeate stream to the ten PCs of the concentrate stream (c). The respective regression coefficients (b and d) are in normalised units. The letters "C" or "P" before the PC number indicates its origin, from concentrate or permeate streams, respectively. Training (◊) (n=104) and validation (●) (n=34) data are presented as percentage of viability.

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, a high cell integrity is desired to guarantee that there is no loss of intracellular compounds to the concentrate and permeate streams. In fact, the model obtained predicts better (with less dispersion) at higher viability values than at lower values. However, the accuracy of such models may possibly be improved, if required, by the used of physical parameters, such as pH or salinity, that can be related with these compounds. Nevertheless, the ability of fluorescence spectroscopy to sense such small variabilities in matrices so different as cell broth and clean permeate, boosts this technique potential to be used as a monitoring tool.

IV.3.3. Nitrate concentration modelling

D. salina capability for high production of carotenoids is achieved by imposing stress factors such as depletion of nitrogen. Monitoring of nitrogen concentration in pilot scale cultivations, being nitrate one of the most common forms of nitrogen used, improves the operator knowledge about the biological state of the culture and may help taking decisions at real-time. 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 before recycling it for cultivation (oxidation, photodegradation and nanofiltration), and nitrate concentration was monitored as a potential oxidation by-product formed during these treatments.

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 processes (Table IV.3).

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

	Nr Inp.	*n*	Var (%)	RMSE CV	RMSE P	Validation		Training		Selected Fluor. PCs
						R^2	Sl	R^2	Sl	
Carot.	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
P.treat.	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
Carot. + P. treat.	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

Carot. – Carotenogenese; P.treat. – Permeate treatment; Nr.Inp. – Number of inputs used; n – number of samples; Sl – slope; Selected Fluor.PCs – Selected Fluorescence PCs.
RMSEP and RMSECV values are presented as ppm N.

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 justified 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 membrane 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 (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 comparing 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 a 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. Nitrogen can be present in these samples as a combination of peptides and proteins (as released products of cellular lost integrity) and nitrate (as a nitrogen source for culture growth). The interference between all nitrogen forms is complex and probably not all caught by fluorescence in the first PCs. Thus, the models developed with cultivation and permeate treatment data combined, used fifteen PCs acquired from PCA of the fluorescence EEMs.

Since the nitrate concentration in most of the culture experiment samples is lower than 20 ppm, and in treated permeate there are more samples with higher concentrations (higher than 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 variance captured. Furthermore, the difference between RMSEP and RMSECV is lower in this model than in the other (Table IV.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 prediction 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 models (even if they are negative, due to the calibration of the analytical equipment). This is an advantage from an industrial point of view, 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 nitrogen prediction models require a pre-treatment before being analysed by fluorescence spectroscopy, since the presence of cells need to be removed by centrifugation. Nevertheless, the possibility of predicting nitrogen concentration without the need of an elaborate protocol is a great advantage in the daily routine of microalgae production.

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

IV.3.4. Application perspectives

The use of microalgae as source of several high value compounds has been studied for several authors and several constrains 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, as fluorescence spectroscopy. The direct use of fluorescence is already a reality but, unfortunately, to measure only chlorophyll content through a specific pair of excitation-emission wavelengths [113]. As shown in this work, the potential of this technology is wider than that. The possibility of obtaining a matrix increases the amount of information acquired from the processes, and since each analysis can take up to 5 minutes, it is possible to have a real-time control of the processes. It is possible to couple an optical probe directly immersed in the system under monitoring, like a bioreactor or a harvesting vessel, by a port just like for 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 a software that allows the operator to choose the desired sampling times. This is already an important advance in the use of fluorescence spectroscopy as monitoring tool, but from the authors point of view, this technology would benefit from a user-friendly software where the result showed to the operator was already processed through the model.

IV.4. Conclusions

This work proves 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) is advised to have 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 decision about the optimal point to start the induction of carotenoids or to harvest.

Cell viability monitoring can lead to higher product recovery yields, so higher percentage is desirable at the end of the biomass recovery. The cell viability model based on fluorescence data showed 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 a better control of the operating conditions throughout the entire production chain, including the induction and recovery of biomass and of added-value compounds.

IV.5. 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.

Monitoring of eicosapentaenoic acid (EPA) production in the microalgae *Nannochloropsis oceanica*

V

Abstract

With the increase awareness for a healthier food regime and greener environmental processes, microalgae are being looked as a solution for a sustainable production of polyunsaturated fatty acids, such as omega-3 eicosapentaenoic acid (EPA). *Nannochloropsis oceanica* is an oleaginous microalga, well-known for the ability of EPA accumulation, although higher lipid productivities are still required to make the process competitive. Therefore, three cultivation parameters were tested in the present work (temperature, light cycles and nitrogen supply) in order to study the EPA profile in the polar (PL) and neutral (TAG) fractions of the cells. In addition, an online monitoring tool based on a fluorescence spectroscopy technique was developed with the aim of increasing process knowledge at real-time. The results of this work show that nitrogen depletion induces the highest variability in EPA accumulation in TAG fraction. However, to increase the EPA content in the PL fraction a different strategy needs to be implemented, such as decreasing the cultivation temperature or the light available per cell. Chemometric models were developed through PCA (Principal Component Analysis) and PLS (Projection to Latent Structures), using only fluorescence spectra as inputs, enabling the monitoring of EPA in both fractions separately. High explained variance was observed (above 85%) in both fractions, with R^2 above 0.81 and slopes above 0.93 for both validation and training data sets. Lower values of cross-validation and prediction errors were observed (between 0.29 and 0.49 % g/g_{DW}). The results obtained show that fluorescence spectroscopy is a powerful technique for online monitoring of non-fluorophore molecules, such as EPA, in complex process like microalgae cultivation.

Keywords: EPA, omega-3 fatty acid, *Nannochloropsis oceanica*, Fluorescence spectroscopy, Chemometric modelling

Published as: Sá, M., Ferrer-Ledo, N., Wijfells, R., Crespo, J. G., Barbosa, M., Galinha, C., 2019. Monitoring of eicosapentaenoic acid (EPA) production in the microalgae Nannochloropsis oceanica. Algal Research 45:101766..

V.1. Introduction

The importance of long-chain polyunsaturated fatty acids (PUFA), such as omega-3 (ω-3), has been extensively studied in the past years with the increased concern in the western world for a better and more equilibrated food regime. This class of lipids proved to have several pharmaceutical and nutraceuticals applications [114–116] and since they are essential for humans and most animals, and neither have the capacity to produce them, food and feed are considered the main vehicles for their supply [117]. Among the ω-3 fatty acids present in Nature, eicosapentaenoic acid (EPA, 20:5 ω-3), plays an essential role in long term health benefits of cardio and immune system [59,114–116].

The main source of EPA is fish and krill oil. To satisfy the human requirements, the World Health Organisation (WHO) advices a dietary intake of fish oil once or twice per week [118,119]. However, this solution is not sustainable in a long term since the supply of fish and krill oil is limited. As an alternative, marine photosynthetic organisms, like microalgae, are being regarded as a solution in aquaculture and terrestrial livestock feed as well as in human supplements, since they are the primary producers of PUFAs like EPA [115–117,120,121]. Microalgae can accumulate lipids in two distinct fractions, the polar lipids fraction (PL), mainly glycolipids and phospholipids; and the neutral lipids fraction (NL), in the form of triacylglycerols (TAG) [114,116,117]. There is some ongoing discussion about which of these two fractions is the best carrier of the EPA in food and feed, with some authors defending the TAG fraction [115] and others the PL fraction [122].

Several strategies are used to increase the lipid content in microalgae cells, namely nutrient limitation (nitrogen or phosphorous deprivation), high salinity, temperature, and high light intensity [114,116,117,123,124]. It is well known that under nitrogen limitation, TAG concentration increases [116]. For example, *Nannochloropsis* sp., a well-known oleaginous microalgae specie, can accumulate between 25 to 45 % of total fatty acids under photoautotrophic conditions, among them the EPA [59,116,117]. It is also known that under different environmental conditions, the content and composition of fatty acids in PL and TAG fractions can vary substantially [59,114].

Using microalgae biomass has great advantages, such as the possibility to grow in non-arable land, using sea water and residual nutrients [115–117]. However, the production of lipids from microalgae biomass still faces some challenges. Although *Nannochloropsis* genus is considered to be a model organism for lipid production [124], EPA content is still low, up to 4.3 % on dry weight basis in *N. gaditana* [114,116,121] and between 2.7 and 5.2 % in *N. oceanica* [116]. To make the process economically competitive, higher lipid productivities are required to decrease production costs [121] associated with high energy requirements for water management, and for lipid extraction of the biomass [115,117].

When aiming for the production of fatty acids, specially EPA, a real-time and online monitoring of the culture would bring great advantages. Fatty acid analysis is known for being a laborious method, involving several steps: extraction, separation into different fractions, methylation and quantification. Different organic solvents are needed in this method, making the process non-green, and most of these steps are time consuming. Thus, the development of an online monitoring tool will enable to understand the effect of the process parameters in the product accumulation at real-time, allowing the possibility to take important decisions in the moment, such

as harvesting the culture when the maximum product content is achieved. To achieve this goal the development of a sensitive probe is of great importance for the overall economic efficiency of the microalgae production and biorefinery. Several spectroscopic techniques have been reported in the literature for the online monitoring and control of bioreactors, such as fluorescence spectroscopy, due to the possibility of tracking different metabolites simultaneously (substrates and products), and being a non-invasive and non-destructive tool. Fluorescence spectroscopy detects the presence of natural fluorophores in the media (extrinsic) and in the cells (intrinsic). The interaction between the two is rather complex, and for this reason, chemometric methods are often needed to deconvolute and find correlations between the concentration of substrates and products and the fluorescence captured [3,4,12].

In the present study, *N. oceanica* was cultivated in highly controlled lab-scale photobioreactors under different environmental conditions in order to better understand the effect of process conditions on EPA accumulation and to develop a monitoring tool specific for this fatty acid. Different temperatures (15, 20, 25 and 30 °C), light cycles (24 hours of light or a day/night cycle of 16 hours: 8 hours) and nitrogen supply (with or without) were tested. The EPA content was analysed in the PL and TAG fractions. To explore the possibility of using fluorescence spectroscopy as a monitoring tool for EPA production, excitation-emission matrices (EEMs) of the fluorescence spectra were acquired during *N. oceanica* cultivation, and chemometric tools such as PCA (Principal Component Analysis) and PLS (Projection to Latent Structures) were used to develop prediction models.

V.2. Material and Methods

V.2.1. Strain, cultivation and pre-culture conditions

Nannochloropsis oceanica was provided by NECTON, S.A. (Olhão, Portugal). Pre-cultures of *N. oceanica* were incubated in 250 mL Erlenmeyer flasks with 100 mL of culture, at 25 °C, in an orbital shaker (90 rpm) with an incident light of 100 µmol/(m^2.s) and 0.2 % CO_2. The cultures were kept under a day/night cycle light regime (16:8 hours) until further use. The cultivation media is constituted by natural sea water (Eastern Scheldt, the Netherlands) and filtered (0.2 µm) before use, supplemented with 10.7 mM of $NaNO_3$ and 0.535 mM of KH_2PO_4. A nutrient solution was added, Nutribloom from Phytobloom®, containing: 4 µM $ZnSO_4.7H_2O$, 2 µM $MnCl_2.4H_2O$, 0.2 µM $Na_2MoO_4.2H_2O$, 0.2 µM $CoCl_2.6H_2O$, 0.2 µM $CuSO_4.5H_2O$, 52.8 µM $Na_2EDTA.2H_2O$, 4 µM $MgSO_4.7H_2O$ and 40 µM $FeCl_3.6H_2O$. HEPES buffer (20 mM) was added and the pH was set to 7.8 with NaOH 5 M. The medium was directly filtered sterilized by Sartobran® Capsule 0.2 µm into the Erlenmeyer flasks or the bioreactor vessel. Prior to the photobioreactor inoculation, inoculums were placed under day/night cycle or continuous light, according to the light regime of the experiment.

V.2.2. Photobioreactor setup

All experiments were performed in a flat-panel airlift-loop photobioreactor, heat-sterilized, with a light path of 20.7 mm and working volume of 1.8 L (Labfors 5 Lux, Infors HT, Switzerland, 2010). The pH was controlled at 7.8 by CO_2 injection, and the culture is homogenised by filter sterilized

air with a flow rate of 1 L/min. The temperature was controlled by a water-jacket in direct contact with the cultivation vessel. The incident light was provided by 260 LED lamps (28 V, 600 Watt) on the culture side of the photobioreactor, with warm white spectrum (450 - 620 nm), and the back side was covered to prevent interference of the ambient light. The incident light started at 200 µmol/(m^2.s) and was increased to 636 µmol/(m^2.s) when the back light reached 50 µmol/(m^2.s).

V.2.3. Photobioreactor operation conditions

In this study eight experiments were performed, in batch mode, following a two-step approach, except the control experiment.

In the first step, bioreactors were inoculated at a cell concentration between 1.0 and 1.5 x10^7 cells/mL and run under replete nitrogen conditions. This was followed by a second step with nitrogen starvation. Briefly, the bioreactors were emptied, the biomass was centrifuged (2500 rpm for 15 minutes) and washed with free-nitrogen medium, and the bioreactor was then refilled with the culture and free-nitrogen medium to prevent limitation of other nutrients, until a specific light supply rate of 1x10^{-13} µmol/(cells.s) (called N-starvation experiments).

The experiments are divided in two categories according to the light cycle – continuous light (24 hours) or day/night (d/n) cycle (16:8 hours) (Figure V.1). In the continuous light experiments, four temperatures were studied – 15, 20, 25 and 30 °C, for a period of four days, starting at N-starvation. The temperature was set in the beginning of the experiment and maintained during nitrogen depletion. In the d/n cycle experiments, three batches were performed to study the effect of N-starvation and sudden temperature decrease, versus a control batch, and followed for ten days. The first step of the cultivation was performed at 25 °C, and in the second step the culture was either submitted to nitrogen-deplete medium or the temperature was decrease to 15 °C. In the control experiment, the bioreactor was inoculated and followed for ten days.

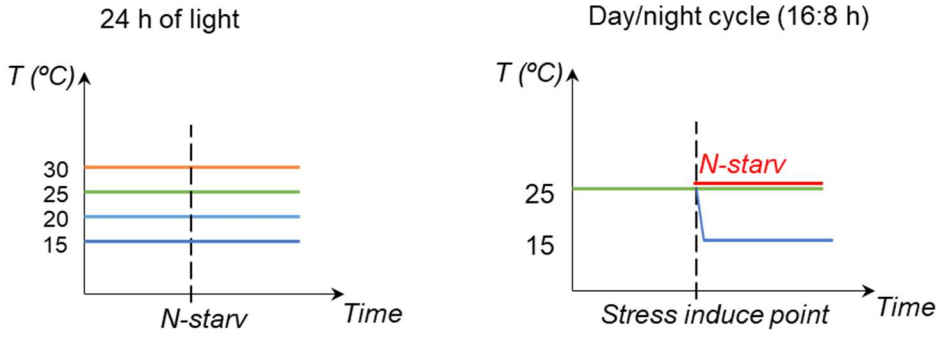


Figure V.1 Photobioreactor operating conditions; environmental conditions tested

V.2.4. Off-line measurements

Biomass concentration was assessed by measuring dry weight (DW) and cell concentration. Dry weight was measured in triplicates, as described by Kliphuis *et al.* [125], and 0.5 M ammonium formate was used to remove the salts from the culture. Cell concentration and cell size distribution was measured in duplicates, using Isotone II diluent to dilute the samples, in a Multisizer II (Beckman Counter) using a 50 µm aperture tube.

Biomass volumetric production rate was calculated according to the following equation (Equation 1), where *DW(t)* and *DW(0)* correspond to the dry weight measured for day *t* and day *0*, respectively, and *t* is the time of the experiment (4 or 10 days):

$$Biomass\ volumetric\ porduction\ rate\ (g/(L.d)) = \frac{DW(t) - DW(0)}{t}$$

Lipid composition of *N. oceanica* was measured during the entire period of the cultivation, before and after the point of stress induction by nitrogen starvation or temperature decrease. Biomass samples were centrifuged, washed with 0.5 M ammonium formate, and stored at -20 °C until lyophilisation. Lipids were extracted, separated into triacylglycerol (TAG) and polar (PL) fractions, and quantified as reported by Breuer *et al.* [126] and Leon-Saiki *et al.* [127]. Briefly, 10 mg of lyophilized biomass was disrupted with a beat beater and lipids were extracted with chloroform:methanol (1:1.25, v:v) containing the internal standards for TAG and PL fractions, 170 µg/mL of tripentadecanoin (9:0) and 170 µg/mL of 1,2-dipentadecanoyl-sn-glycero-3-[phosphor-rac-(1-glycerol)] (sodium salt) (15:0) respectively. TAG and PL fatty acids were separated by different elution solvents, hexane:diethylether (7:1, v:v) and methanol:acetone:hexane (2:2:1, v:v) respectively, in a SPE silica gel column (Sep-Pak Vac 6cc, Waters). Both fractions were methylated and quantified by gas chromatography (GC-FID). The results are expressed in g_{EPA}/g_{DW}.

Fluorescence spectroscopy excitation-emission matrices (EEMs) were measured in a Shimadzu RF-6000 spectrofluorophotometer with a cuvette. Each analysis takes around 5 minutes, where no cell sedimentation was observed. The spectra were acquired through an excitation wavelength range between 250 and 790 nm, with 5 nm steps, and emission wavelength range between 260 and 800 nm, also in 5 nm steps. Excitation and emission monochromator slit widths were 3 nm, with a scan speed of 12000 nm/min.

V.2.5. Chemometric models development

The development of the chemometric models included three steps:

1) Spectra pre-treatment: water scatter peaks, like Rayleigh scatter of first and second order, can be a source of interference when aiming for a quantitative analysis using 2D fluorescence spectra. This wavelength-dependent scatter (peak emission ± 10 nm at each excitation wavelength) was removed with the use of an algorithm developed by Bahram *et al.* [20], and replaced by interpolation of the surrounding data points. The program is available at www.models.kvl.dk.

2) Principal Component Analysis: through PCA is possible to compress and reduce the information in the EEMs, with minimal loss of information, by dividing the initial data into *n* linear combinations that have to follow two main rules, be uncorrelated and be ordered according to the explained variance they captured. The first principal component (PC1) will capture maximum variance in a certain direction (axis); then PC2 will be orthogonal to PC1 and will capture less variance; and so on. In total, 73 fluorescence spectra were compressed into 20 PCs, capturing more than 99% of the variance.

3) Projection to Latent Structures modelling: in order to correlate the fluorescence PCs (inputs) with the EPA concentration in TAG and PL fractions of *N. oceanica* (outputs), multivariate statistical modelling was used, namely PLS. In PLS modelling two sub data sets were created, one for training (calibration) the model, i.e. to build the function that better correlates the outputs with the inputs; and another for validation (prediction), to test the quality of the model to predict a new data set. Two validation approaches were studied, first using each batch at a time (batch-by-batch), and secondly using a random data set corresponding to 25 % of the total data. The cross-validation (CV) of the models was performed with the remaining data, used for calibration (seven batches in the case of batch-by-batch validation, or random 75 % of all data). Shortly, several models were created using an independent set of data, selected by leave one out strategy (LOO) and repeatedly evaluating the errors of the models (RMSECV). Not all the PCs provided as inputs are required. Thus, to select the useful PCs for each model, an iterative stepwise elimination (ISE) was used [42]. To assess the quality of the models several parameters are evaluated, such as the variance captured (%), the root mean square error of cross-validation (RMSECV) and prediction (RMSEP), and the coefficients of determination (R^2) and slopes of the validation and training sets.

All multivariate statistical analysis were performed using n-way toolbox for MATLAB - MathWorks® [72].

V.3. Results and Discussion

In this work, the effect of several environmental conditions on the accumulation of EPA was studied in apolar (TAG) and polar (PL) fractions. The environmental conditions studied were light – continuous light (24 h) or day/night cycles (16 h : 8 h); temperature – low temperatures (15 and 20 °C) and high temperatures (25 and 30 °C); and nitrogen starvation. Samples were taken along the entire length of the experiments to measure fatty acid profile and obtain the EEMs.

To evaluate the effect of different stress factors on EPA accumulation, the fatty acid profile of the culture was compared in the beginning of the “stress phase”, by nitrogen depletion or decrease of temperature, until the end of the batch, in a total of 4 days for 24 h light experiments, and 10 days for d/n cycle experiments. To monitor the evolution of the EPA during the experiments, models were performed with all the samples acquired through each batch, before and after the “stress phase”. This strategy enables a higher data set to be used to calibrate and validate the models and also captures the variability expected in these experiments, from inoculation until EPA production.

V.3.1. Biomass concentration and cell size

In the experiments where light was provided during 24 h (Figure V.2A), no differences were found in the biomass concentration between cultivation at 20, 25 and 30 °C after 4 days of nitrogen depletion. The biomass volumetric productivities were on average 0.9 g_{DW}/(L.d) for 20

and 30 °C, and 1.0 g_{DW}/(L.d) for 25 °C. As expected, the biomass volumetric production rate was lower in the 15 °C batch, 0.3 g_{DW}/(L.d).

In the d/n cycle experiments (Figure V.2B), the biomass concentration shows a similar profile in all three experiments for the first 4 days. However, nitrogen depletion seems to have a negative effect on the final biomass concentration than decreasing the temperature from 25 to 15 °C. The biomass volumetric production rates were on average 0.5, 0.8 and 0.9 g_{DW}/(L.d) for nitrogen starvation, decrease in temperature and control experiment, respectively. In the decreased temperature and in the control experiments, no differences were found in the final biomass concentration.

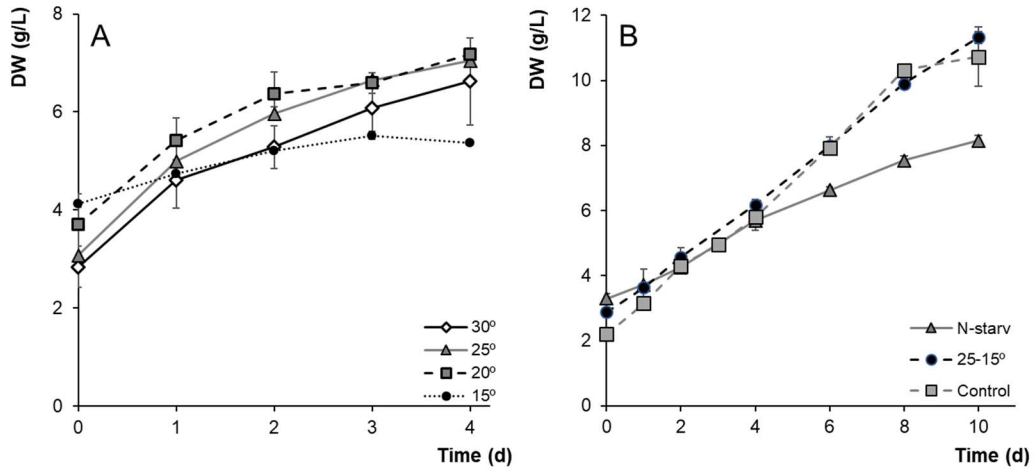


Figure V.2 Average biomass concentration and respective standard deviation bars, expressed as dry weight (g/L), measured over time, from the moment of stress induction. A) For 24h of light and nitrogen starvation, comparison between different temperatures: 15 (circles), 20 (squares), 25 (triangles) and 30 °C (diamonds); B) For day/night cycle (16:8), comparison between nitrogen starvation (N-starv, triangles), decrease of temperature from 25 to 15 °C (25-15°, circles) and a control experiment (no starvation and no decrease of temperature, squares).

The environmental conditions imposed to the culture also had an influence on the cell size (Table V.1). For all experiments performed with 24 h of light and nitrogen depletion, it is possible to notice an increase in cell size. Nitrogen starvation is one of the strategies often used to increase the TAG concentration in microalgae cells, which is done in lipid bodies leading to an increased volume. Only in the experiment performed at 15 °C the final cell size was slightly lower than in the beginning of the starvation phase, indicating that the culture was probably already accumulating TAG when the nitrogen depletion was performed. This result can also explain why the biomass dry weight was higher in the beginning of the starvation phase of this experiment (d0) when compared with the other temperatures tested.

In the d/n cycle experiments, when the temperature was decreased (from 25 to 15 °C), the cell size increased in the first two days (to 3.5 µm), but immediately after that it decreased to 3.0 and remained stable until the end of the batch. Cell concentration was constant on the first four days, only increasing after that (data not shown). This increase in the cell size can be explained by the accumulation of fatty acids inside lipid bodies, as an acclimation to the decrease in

temperature. The cells of the control batch without starvation or any other stress factor, increased size gradually throughout the experiment, accompanied by an increase in cell concentration. The process of increasing the amount of membrane surface was reported before as an adaption to low light conditions, as the one observed in the control batch, in order to increase the photosynthetic capacity to capture more light [121].

When comparing the two experiments with nitrogen depletion at 25 °C and different light regimes, the same final cell size was achieved in the end of the experiments (d4 and d10), although the cell size increased faster for the culture exposed at 24 h of light as expected.

Table V.1 Cell size (µm) of N. oceanica at the first day of the "stress phase", and after 4 and 10 days. Experimental error inferior than 2.32 % for all measurements.

	Day 0	Day 4	Day 10
15	3.08	2.98	-
20	2.92	3.14	-
25	2.88	3.08	-
30	2.98	3.28	-
N-starv	2.85	2.85	3.04
25-15	3.10	3.30	3.00
Control	2.79	3.05	3.34

V.3.2. EPA production

Nitrogen starvation, a well-known methodology to induce lipid accumulation in microalgae lipid bodies, was performed at four different temperatures, 15, 20, 25 and 30 °C (Figure V.3A).

For lower temperatures (15 and 20 °C) it is possible to notice a higher content of EPA in the TAG fraction in the beginning of the starvation (d0) than for higher temperatures (25 and 30 °C). In fact, the EPA content in TAG fraction is inversely proportional to the temperature, i.e., higher contents were observed in lower temperatures, and vice-versa. This might be due to the combined effect of temperature and light per cell, since at lower temperatures the biomass productivity decreases, which means that the culture will take longer to multiply, so more light will be available per cell. As a result, TAG content was already higher before starting the nitrogen starvation phase. Yet, when comparing the EPA accumulation in the TAG fraction during the four days of nitrogen starvation (final content of EPA compared to the initial), higher temperatures led to higher TAG productivities. For 30 °C, the EPA accumulated in the TAG fraction was 0.017 g_{EPA}/g_{DW}, 0.012 g_{EPA}/g_{DW} for 25 °C, 0.007 g_{EPA}/g_{DW} for 20 °C and 0.005 g_{EPA}/g_{DW} for 15 °C. The increase in long chain FA accumulation with increasing temperatures was also reported by Ördög *et al.* in three *Chlorella* strains [128].

As studied before [121], nitrogen depletion leads to an increase in TAG content associated with a decrease of the PL fraction. One of the mechanisms proposed is the *de novo* synthesis of TAG by conversion of membrane lipids, which consists mainly of PL. In this study it is possible to confirm that the increase observed in the TAG fraction was accompanied by a decrease in the PL fraction.

A decrease in temperature has been reported to increase the EPA content in the cell membranes, especially during growing conditions [114,129]. Lipid accumulation occurs during the day period, while cell division occurs at night [130,131]. Aiming to increase the EPA content in the PL fraction of *N. oceanica*, a second set of experiments were performed, keeping the culture under d/n cycle: nitrogen depletion (N-starv), temperature decrease (from 25 to 15 °C) and a control batch, where the algae were allowed to grow without any imposed stress (Figure V.3B).

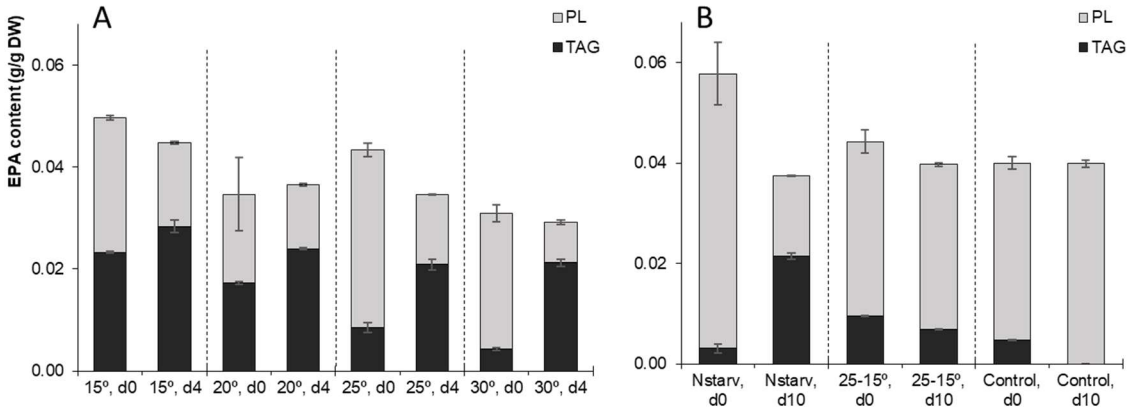


Figure V.3 EPA content expressed per dry weight of biomass (g EPA/g DW) in the TAG (dark grey) and PL (light grey) fractions. A) For 24h of light and nitrogen starvation, comparison between different temperatures: 15, 20, 25 and 30 °C. B) For day/night cycle (16:8), comparison between nitrogen starvation (N-starv), decrease in temperature from 25 to 15 °C (25-15) and a control experiment (no starvation and no decrease in temperature);

In the beginning of the experiments, N-starv and control batches started with a similar EPA content in the TAG fraction. Nevertheless, after ten days of nitrogen depletion media, the culture accumulated a large amount of EPA in the TAG fraction at the expense of the content in the PL fraction (Figure V.3B, N-starv d0 and d10). These results confirm what was stated by other authors regarding the effect of nitrogen starvation [121,124,132]. The effect of lowering the temperature resulted in a slight increase of EPA in the PL fraction, accompanied with a decrease in the TAG fraction (Figure V.3B, 25-15° d0 and d10). According to previous studies in *Nannochloropsis salina*, *Phaeodactylum tricornutum* and *Chlorella sp.* [116,133,134], temperature reduction increased the content of EPA and PUFA's, due to the need of increasing membrane fluidity. However, these studies point to previous research [135] where the FA were analysed in total lipid profile of the whole cell, and no distinction was done between TAG or PL fractions. The highest content of EPA in the microalgae membranes (PL fraction) was achieved in the control batch, where the temperature was maintained at 25 °C with d/n cycle (Figure V.3B, d0 and d10). In this experiment, the EPA content in TAG fraction decreased after ten days from 0.005 g/g_{DW} to 0 g/g_{DW}.

Although there is no difference on the final dry weight between the control experiment and the decreased temperature (Figure V.2), the cell size may be the contribution for the higher content of the EPA in PL fraction in the control batch. A higher cell concentration was achieved (data not shown), meaning that less light was available per cell. To adapt to this environmental condition, the microalgae increase the plastid membrane intending to increase the photosynthetic

apparatus, and consequently the EPA content in this fraction increases [121]. And, as mentioned before, a cell size increase was noticed in this experiment (Table V.1).

The experiments performed resulted in several scenarios in the accumulation of EPA by *N. oceanica*. Nitrogen depletion enables the accumulation of EPA in lipid bodies (TAG fraction) independently of the temperature, although higher temperatures led to higher accumulation. For nitrogen replete cultivation conditions, light played an essential role. Although the incident light was the same in all experiments, the biomass concentration of the control batch was higher, so the cells perceive lower light, and this led to an increase of EPA in PL fraction. Also, temperature decrease might slightly increase the EPA content in the PL fraction. When performing nitrogen depletion at 25 °C, a similar EPA final content in the biomass was reached for both 24h and d/n cycle batches, with the main difference being the time needed.

V.3.3. EPA monitoring

The experiments in this study led to several responses in *N. oceanica* biomass regarding cell concentration, size and physiological state (from non-stressed green cells, to yellowish when TAG synthesis is induced). This variability in the cells originates fluorescence EEMs with different characteristics. Changes in the media composition, such as nitrate concentration, were also reported in other microalgae cultivations to originate a different fluorescence profile (Chapter 4). Also, the variability of EPA content that can be found in the two fractions of the cell, apolar (TAG) and polar (PL), increases the complexity of monitoring such product. Depending on the commercial destination of the lipid enriched biomass, having information about the content and location of the EPA in the cell can be extremely useful namely for its recovery. Thus, a tool that can distinguish the content of EPA in both fractions of *N. oceanica* cells, simultaneously, was developed.

When aiming for industrialisation, a batch-by-batch validation approach seems logical, since the final model acquired with these experiments will then be validated with new data acquired from a new batch. However, the experiments of this work were thought to lead to a wider range of EPA concentrations, to be able to have broader range of scenarios. For that reason, some of the batches are not representative of a cultivation to produce EPA but were important to acquire concentration points in a lower or higher data ranges. Thus, a random data set was also used to create a general model, that would be more suitable for future use.

The prediction models are represented in a graphic (Figures V.4 and V.5 for the TAG fraction; Figures V.6 and V.7, for the PL fraction) where the observed values (y-axis) are plotted against the predicted values by the model (x-axis). Two parallel lines were added to each graph corresponding to two times the standard deviation of all the experimental data acquired. Points represented outside these lines may be considered outlier observations.

Model parameters including variance explained (%), root mean square error of cross-validation (RMSECV) and prediction (RMSEP), validation and training R^2 and slopes, and the number of inputs selected, are shown in Tables 3 and 4 for TAG and PL fractions, respectively. The quality of a model can be described by a high explained variance, with coefficient of determination (R^2) and slope near to 1. Lower values of root mean square errors are preferable,

and the values between the cross-validation and prediction errors should be close, which means that both data sets (validation and training) are representative of all data variability.

V.3.3.1. Prediction models for EPA in TAG fraction

As mentioned before, the batches performed aimed to acquire a wide range of EPA concentration values. In Figure V.4 it is possible to see that the experiments of d/n cycle (the validation data in the upper three graphics in Figure V.4) show different data distribution. For instance, the decrease in temperature did not increase the EPA content in the TAG fraction. A similar distribution was observed in the control batch. For that reason, the EPA concentration range observed in these two batches is limited, meaning that when these batches are used for validation, the combination of R^2 and slope of the validation set is low (Table V.2). Also, for both models, a high number of inputs is necessary (16 for both models) to explain the variance captured (around 91 % for both models). For the experiment N-starv, the validation set has four outliers (of a total of ten data points), experimental points represented above the line of two times the standard deviation of all data. This means that, without the data points of this batch to train the model, the prediction ability decreases (52.6 % of variance explained), together with a low R^2 (0.26) for the validation set and with the highest root mean square errors (0.58 and 0.69 % g/g_{DW} for RMSECV and RMSEP, respectively).

The models obtained using the experiments performed at 24 h of light and nitrogen starvation as validation data set were more consistent (Figure V.4, four lower graphics). Nitrogen depletion is known to induce the TAG accumulation in the cells, giving a broad range of concentrations through each experiment. The variance captured by these models was above 80 %, with RMSECV ranging from 0.32 to 0.41 % g/g_{DW} and RMSEP between 0.26 and 0.31 % g/g_{DW} (Table 2). The fact that the values of RMSECV and RMSEP are close reveals that the validation data set is representative of values of the training set.

For the model using as validation set the experiment performed at 20 °C, a lower R^2 was found for both validation and training set (0.64 and 0.80, respectively). When using the batch performed at 30 °C as validation set, the training R^2 was also lower (0.81). This can be explained by the fact that in both models only seven PCs were selected to build the model, and some outliers can be found in the data distribution. Nevertheless, the remaining models, the R^2 of the validation were above 0.81, and above 0.89 for the training set, revealing the robustness of the fluorescence spectroscopy as a technique to monitor EPA concentrations in TAG fractions of *N. oceanica*.

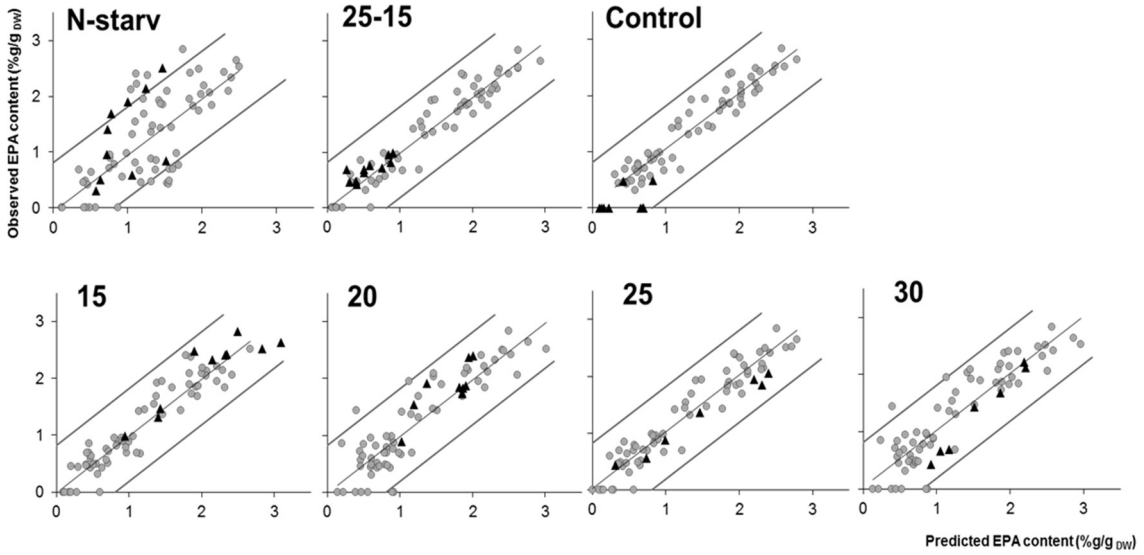
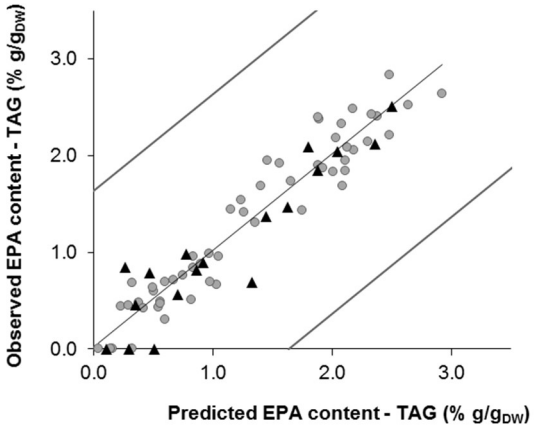


Figure V.4 Prediction models of EPA content content (% g/g$_{DW}$) in TAG fraction of N. oceanica. In Figure 4, predicted values (x-axis) are plotted against observed values (y-axis). Training (●) and validation (▲) data are represented in percentage of grams of EPA per grams of dry weight.

Table V.2 Statistical parameters of the prediction models of EPA content (% g/g$_{DW}$) in TAG fraction of N. oceanica.

	Var	RMSECV	RMSEP	Validation		Training		Nr of inputs
	(%)	*%g/g$_{DW}$*		R^2	Slope	R^2	Slope	
N-starv	93.1	0.29	1.55	0.01	0.21	0.93	1.00	16
25-15	89.5	0.42	0.21	0.78	0.53	0.89	1.00	14
Control	89.5	0.46	0.21	0.49	0.69	0.89	1.00	17
15	85.9	0.46	0.31	0.68	1.23	0.86	1.00	11
20	67.4	0.62	0.29	0.79	0.92	0.67	1.00	5
25	89.9	0.42	0.47	0.74	0.99	0.90	1.00	14
30	89.6	0.41	0.31	0.93	0.77	0.90	1.00	15

A general model was built using 25 % of the total data as a validation set, randomly chosen, and the remaining 75 % as training set. As it can be seen in Figure V.5, the validation set is widely spread through the entire concentration range of EPA found in the TAG fraction. Thirteen PCs, of twenty inputs given, were selected to explain 92.1 % of the variance captured by the model, which means that the fluorescence spectroscopy is a robust method for capturing the variability found in these experiments. The RMSECV and RMSEP were close (0.30 and 0.29, respectively) and both validation and training R^2 and slopes were high (all above 0.87).



Variance	*(%)*	92.1
RMSECV	*% g/g$_{DW}$*	0.30
RMSEP		0.29
Validation	*R^2*	0.87
	Slope	0.95
Training	*R^2*	0.92
	Slope	1.00
# Inputs		13

Figure V.5 Prediction model of EPA content in TAG fraction using 75 % of the data for training (●) and 25 % for validation (▲), represented in percentage of grams of EPA per grams of dry weight. Normalised regression coefficient of the represented prediction model.

The variability found in the quality of the models shows the importance of the calibration data set used. Not only it is important to provide as much data as possible, but also that data should be representative of the several scenarios that can be found. Overall, fluorescence spectroscopy proved to be a strong tool to monitor EPA concentration in the TAG fraction of *N. oceanica*. EPA is a fatty acid and therefore it is known for not being a natural fluorophore. Fluorescence spectroscopy was reported to be a scanning technique able to detect natural fluorophores, but also the interaction of those with the medium and other non-fluorophores components can provide information about the matrix constitution. For this reason, the monitoring of a non-fluorophore molecule like EPA, reveals the potential of this technique to monitor other lipid components.

V.3.3.2. Prediction models for EPA in PL fraction

In Figure V.6 it is possible to see that the experiments of d/n cycle (the validation data in the upper three graphics) show different patterns. In the control batch, where the culture was allowed to grow at optimal conditions (25 °C, d/n cycle), at a certain point the light became a limiting factor. It was shown in literature that decreasing light per cell increases EPA content, due to the increase of cell membranes [121]. This justifies the fact that EPA concentration was the highest in this batch. Decreasing the cultivation temperature was also reported to increase the unsaturated fatty acids in the PL fraction of the algae, and it is possible to observe a high content of EPA in this experiment. However, and as reported for the models of EPA in TAG fraction, this means that using these data as validation set (meaning they are not used for calibration) is not a correct approach. Both models reach 89.5 % of variance explained, but there is a wide gap between RMSECV and RMSEP, and the combination between R^2 and slope of the validation set was not ideal (Table V.3). When using the N-starv batch as validation, six of the ten points used for validation are considered outliers. Also, while observed values of EPA content varied between 1 and 6 % g/ g_{DW}, no prediction was performed above 4 % g/ g_{DW}. As described previously in the model of the EPA in the TAG fraction, the quality of the model decreases when these data points are not used to build the prediction model. Although the variance explained was 93.1 %, the

difference between RMSECV and RMSEP is the highest observed in these models, with the lowest validation R^2 (0.01) and slope (0.21) (Table V.3).

In the models obtained for the batches performed at 24h light and nitrogen starvation, a wider range of EPA concentration was observed (Figure V.6, four lower graphics). Except for the batch at 20 °C, the variance captured was above 85 %, with RMSECV ranging from 0.41 to 0.46 % g/g_{DW} and RMSEP between 0.31 and 0.47 % g/g_{DW} (Table V.3). Using as validation set the data of the batch performed at 20 °C, more outlier values were observed, and lower number of PCs were selected as inputs, resulting in a model with lower variance explained (67.4%), higher difference between RMSECV and RMSEP (0.6 and 0.29 % g/g_{DW}, respectively) and the lowest R^2 of the training data set (0.67) (Table V.3).

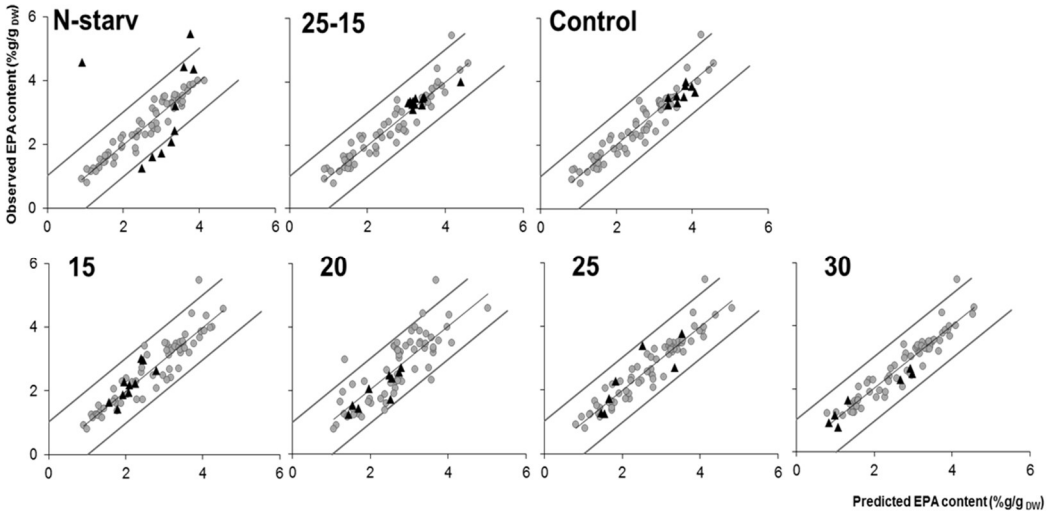
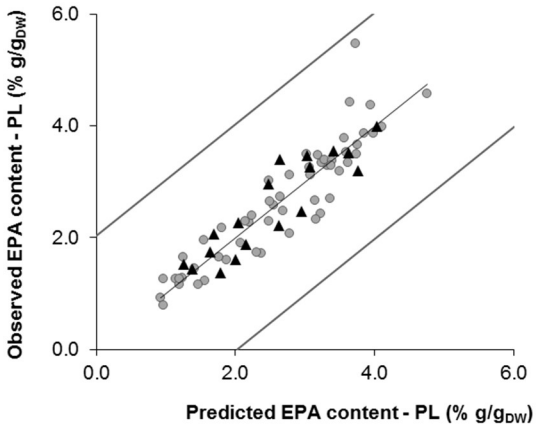


Figure V.6 Prediction models of EPA content content (% g/g_{DW}) in PL fraction of N. oceanica. In Figure 4, predicted values (x-axis) are plotted against observed values (y-axis). Training (●) and validation (▲) data are represented in percentage of grams of EPA per grams of dry weight.

Table V.3 Statistical parameters of the prediction models of EPA content (% g/g_{DW}) in PL fraction of N. oceanica.

	Var	RMSECV	RMSEP	Validation		Training		# Inputs
	(%)	% g/g_{DW}		R^2	Slope	R^2	Slope	
N-starv	93.1	0.29	1.55	0.01	0.21	0.93	1.00	16
25-15	89.5	0.42	0.21	0.78	0.53	0.89	1.00	14
Control	89.5	0.46	0.21	0.49	0.69	0.89	1.00	17
15	85.9	0.46	0.31	0.68	1.23	0.86	1.00	11
20	67.4	0.62	0.29	0.79	0.92	0.67	1.00	5
25	89.9	0.42	0.47	0.74	0.99	0.90	1.00	14
30	89.6	0.41	0.31	0.93	0.77	0.90	1.00	15

When using 25 % of random data as validation set (Figure V.7), eleven of the twenty PCs were selected to explain 84.8 % of the variance and no outliers were observed. The quality of the model was high, with errors ranging between 0.37 and 0.49 % g/g_{DW} (RMSECV and RMSEP, respectively). Both R^2 of validation and training set were higher than 0.80 and slopes near 1.00.



Variance	*(%)*	84.8
RMSECV	*% g/g$_{DW}$*	0.49
RMSEP		0.37
Validation	*R^2*	0.81
	Slope	0.93
Training	*R^2*	0.85
	Slope	1.00
# Inputs		11

Figure V.7 Prediction model of EPA content in PL fraction using 75 % of the data for training (●) and 25 % for validation (▲), represented in percentage of grams of EPA per grams of dry weight. Normalised regression coefficient of the represented prediction model.

As mentioned previously, a suitable calibration data set should be selected in order to capture the entire concentration range of EPA in the PL fraction, and thus, build a model representative of the different scenarios that can be encountered. The ability of the fluorescence spectroscopy to monitor EPA concentration in the cell membrane of *N. oceanica* is of high importance for the industrial application of this microalgae in the food and feed industry. It was described that the fatty acids present in the microalgae membrane are more bioavailable for fish metabolism [122]. Having a tool able to determine EPA concentration at real-time, with a fast and non-invasive methodology, allow the monitoring of the process as well as stirring it in order to have the desire amount of EPA in the biomass, improving the cultivation efficiency and thus the economic gain of the overall process.

V.4. Conclusions

For the first time, fluorescence spectroscopy coupled with chemometric tools was used to monitor EPA in microalgae biomass.

To enable the development of PLS prediction models, different environmental cultivation conditions were tested to induce different biological responses in the biomass profile and EPA accumulation. Stress factors, like nitrogen depletion and low light per cell, led to an increase in the cell size of *N. oceanica*. Nitrogen depletion led to an increase of EPA in the TAG fraction of the cells, while low light or decreased temperature lead to an increase of EPA in the PL fraction.

Statistical modelling combined with fluorescence spectroscopy is a step forward in the real monitoring of complex biological systems like microalgae cultivation. In this work, prediction

models were developed that enable the monitoring of EPA content in both TAG and PL fractions of the cell. Two validation strategies were studied, batch-by-batch and random 25% of the total data, to access the importance of the batch conditions in the prediction capability of the method. The models developed demonstrate the potential of using fluorescence spectroscopy as a monitor tool of non-fluorophore molecules, such as EPA, revealing the potential of this technique to monitor lipid components. The use of such technology enables not only the possibility to monitor the cultivation system, but also of taking decisions at real-time, empowering the optimisation of the cultivation of *N. oceanica* when aiming for EPA production.

V.5. Acknowledgments

The authors would like to acknowledge the Analytical Chemistry & Chemometrics group of the Institute for Molecules and Materials (IMM) at Radboud University (Nijmegen, The Netherlands), lead by professor Jeroen Janssen, for their help in the spectra pre-treatment used in this work. This work was supported by the project titled “Microalgae As a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by cost effective New Technologies” [MAGNIFICENT], funded by Bio-based Industries Joint Technology Initiative under European Commission [project ID: 745754], by the Associate Laboratory for Green Chemistry- LAQV which is financed by national funds from FCT/MCTES (UID/QUI/50006/2019), and by the follow Fellow grants SFRH/BPD/95864/2013 and SFRH/BD/108894/2015 of FCT/MCTES.

Fluorescence spectroscopy coupled with chemometric modelling for the simultaneous monitoring of cell concentration, chlorophyll and fatty acids

VI

Abstract

Microalgae industrial production is nowadays viewed as a solution for environmental and sustainable alternative products in the fields of fuel, feed and food. Online monitoring of algal biotechnological processes however still requires development to support economic sustainability. In this work, fluorescence spectroscopy coupled with chemometric modelling is studied to monitor simultaneously several compounds of interest, such as chlorophyll and fatty acids, and also the biomass as a whole (cell concentration). Fluorescence excitation-emission matrices (EEM) were acquired in experiments where different environmental growing parameters were tested, namely light regime, temperature and nitrogen (replete or deplete medium). The prediction models developed have a high R^2 for the validation data set for all five parameters monitored, specifically cell concentration (0.66), chlorophyll (0.78), and fatty acid as total (0.78), saturated (0.81) and unsaturated (0.74). Regression coefficient maps of the models show the importance of the pigment region for all outputs studied, and the protein-like fluorescence region for the cell concentration. These results demonstrate the potential of fluorescence spectroscopy to monitor these parameters simultaneously in *Nannochloropsis oceanica* production.

Keywords: Fluorescence spectroscopy, PLS modelling, *Nannochloropsis oceanica*, Cell concentration, Chlorophyll, Fatty acids

Submitted as: Sá, M., Bertinetto, C., Ferrer-Ledo, N., Janssen, J., Wijfells, R., Crespo, J. G., Barbosa, M., Galinha, C., 2019. Fluorescence spectroscopy coupled with chemometric modelling for the simultaneous monitoring of cell concentration, chlorophyll and fatty acids. Scientific Reports.

VI.1. Introduction

Microalgae industrial production is still a niche industry, although efforts are being done to improve the economic viability of the overall process. Online monitoring through spectroscopic techniques is already a reality in several other bio-based industries [1,15,23,28]. Development of an appropriate tool able to monitor several metabolites simultaneously would be a great advantage. Nowadays, biological parameters are monitored off-line, where a sample is withdrawn, from the cultivation or from process streams in the biorefinery, to be analysed. These analyses can take some minutes or a couple of hours, or days, depending on the parameter to be measured and techniques involved. For example, cell concentration is a rather fast and simple measurement, while assessing metabolites' content is much slower, because extraction steps and chromatographic techniques are often needed. In this sense, fluorescence spectroscopy represents a viable solution since it is a non-invasive and non-destructive technique that already proved its value in other industries, like the food industry [1,15,23], or even in wastewater treatment plants that are characterized for being complex biological systems [28,29,69]. This technique is able to detect natural fluorophores but is also sensitive to interactions between fluorophores and non-fluorescent compounds, increasing the range of compounds that can be monitored by it [12]. Also, with the improvement of chemometric tools, it is possible to extract more information from such signals.

Nannochloropsis oceanica is a promising microalga due to its ability to produce high amount of lipids. This microalgae is cultivated in sea water, which has been pointed as the most sustainable solution for microalgae production [59,116,117]. The most common product of a microalgae industry is still the whole biomass, being of extreme importance to monitor cell concentration. It is also known that chlorophyll is an important pigment in several industries [136,137]. Being a oleaginous microalgae, there is an increasing interest in industrializing *N. oceanica* production due to its ability to produce high quantities of lipids, and also high quality oils with nutritional value [59,116].

Previous work demonstrated the advantage of using fluorescence spectroscopy in a microalgae biorefinery context [78] (and Chapters 2 and 3). *Dunaliella salina*, also a saline microalga, is capable of producing high contents of carotenoids, being already produced at industrial scale for this purpose. In these works, fluorescence spectroscopy coupled with chemometric tools was used to develop predictive models for several parameters, such as cell concentration and viability, carotenoids content and nitrate concentration in the medium. Cell concentration of *Chlorella vulgaris* and *Spirulina* sp. was also determined by Shin *et al.* using an *in situ* fluorometry technique [113].

In previous work [78] (and Chapters 2 and 3), the parameters under study were natural fluorophores, such as pigments, or parameters strictly correlated with them, like cell concentration, which is correlated with pigment and protein concentrations. To our knowledge, the use of this technique was never used to assess the content of non-fluorophores molecules, such as fatty acids. The spectral regions with higher relevance for the prediction models of all parameters is also studied, providing a better understanding of how fluorescence can be useful to monitor fluorophores and non-fluorophores' molecules. The goal of this work is to develop

models to predict simultaneously cell concentration, chlorophyll content and fatty acid composition in *N. oceanica*.

VI.2. Material and Methods

VI.2.1. *Nannochloropsis oceanica* pre-culture and cultivation experiments

Nannochloropsis oceanica was provided by NECTON, S.A. (Olhão, Portugal). The inoculum was kept in 250 mL Erlenmeyer flasks, under the follow conditions: 25 °C, 90 rpm in an orbital shaker, 100 µmol/(m^2.s) of incident light, day/night cycle (16/8 hours) light regime, and 0.2% CO_2. The cultivation media is composed by filtered (0.2 µm) natural sea water (Eastern Scheldt, the Netherlands), supplemented with 10.7 mM of $NaNO_3$, 0.535 mM of KH_2PO_4 and Nutribloom from Phytobloom®. HEPES buffer (20 mM) was added and the pH was set to 7.8. Sterilization of the medium was performed by filtration (Sartobran® Capsule, 0.2 µm).

Experiments were performed in batch mode, in a heat-sterilized flat-panel, with a 1.8 L of working volume and a light path of 20.7 mm (Labfors 5 Lux, Infors HT, Switzerland, 2010). The light was provided by LED lamps (28V, 600 Watt) with warm spectrum (450 – 620 nm). In the beginning of the experiment light was set at 200 µmol/(m^2.s) and was increased to 636 µmol/(m^2.s) when the back light reached 50 µmol/m^2. Filter sterilized air in an airlift-loop homogenizes the culture at a flow rate of 1L/min, and pH was controlled by CO_2 injection. A water-jacket was used to control the bioreactor temperature.

Three cultivation parameters were tested in eight experiments (Table VI.1): light regime, temperature, and nitrogen supply – with (√) or without (X) nitrogen. The light regime was set in the beginning of the experiment and two approaches were tested: 24 hours of light or 16 h of light and 8 h of dark (d/n cycle). Temperature was set in the beginning (15, 20, 25 and 30 °C) and kept through the experiment, except in one batch where the temperature was decreased from 25 to 15 °C when a specific light supply rate of $1x10^{-13}$ µmol/(cell.s) was reached. All batches started with a replete nitrogen medium to enable biomass growth. For six of the eight batches (Table VI.1, Nitrogen supply "X"), a second step, the nitrogen depletion phase, was performed. Briefly, the biomass was collected, centrifuged (2500 rpm, 15 minutes) and washed with nitrogen deplete medium, and the bioreactor was then refilled with culture and nitrogen deplete medium until reaching a specific light supply rate of $1x10^{-13}$ µmol/(cell.s).

Table VI.1 Experimental conditions of the eight batch experiments performed

Temp (°C)	Nitrogen supply[(a)]	Light cycle (hours)
15	X	d (24)
20	X	d (24)
25 [(b)]	X	d (24)
25	X	d/n (16/8)
25	√	d/n (16/8)
25 → 15	√	d/n (16/8)
30	X	d (24)

[(a)] X = absent; √ = present; [(b)] this batch was performed twice;

VI.2.2. Off-line measurements

Samples were taken every day to measure cell concentration, chlorophyll content, fatty acid composition and spectrofluorescence.

Cell concentration was measured in a Multisizer II (Beckman Counter), in duplicates, using a 50 µm aperture tube and Isotone II diluent to dilute the samples.

Chlorophyll content was assessed by a spectrophotometric method as described by Leu & Hsu (1987) [88]. An aliquot of 2 mL was centrifuged (5000 g, 5 min) and stored at -80 °C until further analysis. Extraction was performed with 2 mL of methanol, samples were sonicated for 5 min and incubated for 40 min at 60 °C, following by cooling for 15 min in ice. Extraction steps were repeated until a white pellet was recovered. The modified Arnaud equation was used to calculate the chlorophyll content:

$$Chla = (16.72 \times A_{665} - 9.16 \times A_{652}) \times \ dilution\ factor\ (\mathrm{mg/L})$$

Lipid composition was measured in lyophilized biomass samples, previously washed with 0.5 M ammonium formate, as described by Breuer et al [126] and Leon-Saiki [127]. Briefly, 10 mg of sample were disrupted by beat beater and an extraction was performed with chloroform:methanol (1:1.25, v:v), containing the internal standards for triacylglycerol (TAG) and polar lipid (PL) fractions, 170 µg/mL of tripentadecanoin (9:0) and 170 µg/mL of 1,2-dipentadecanoyl-sn-glycero-3-[phosphor-rac-(1-glycerol)] (sodium salt) (15:0) respectively. TAG and PL were then separated in a SPE silica gel column (Sep-Pak Vac 6cc, Waters) using hexane:diethylether (7:1, v:v) and methanol:acetone:hexane (2:2:1, v:v) respectively. Methylation was performed in both fractions prior to quantification by gas chromatography (GC-FID). The results are expressed in percentage of total, saturated and unsaturated fatty acids in a dry weight basis.

Fluorescence spectra were acquired with a Shimadzu RF-6000 spectrofluorophotometer. Each sample took 5 minutes to analyse, in a cuvette, and no sedimentation was observed. The excitation-emission matrices (EEMs) from the spectra ranged from 250 and 790 nm for excitation wavelength, and between 260 and 800 nm for emission wavelength, in steps of 5 nm. Excitation and emission monochromator slit widths were 3 nm, with a scan speed of 12000 nm/min.

VI.2.3. Chemometric models development

The EEMs were pre-processed to prepare them for further analysis. Rayleigh scatter of first order was removed and replaced by empty values; the second order was replaced with an interpolation of surrounding data points [20]. Any fluorescence signal corresponding to emission wavelengths (y-axis) shorter than the excitation wavelengths (x-axis) was replaced by zeros. Inner filter effects, caused by excessive concentration in some samples, were also corrected by the algorithm whenever present.

The pre-processed EEMs were correlated to five biological parameters – cell concentration, chlorophyll and fatty acids content as total, saturated and unsaturated – using Projection to Latent Structures (PLS). This method finds linear combinations of the observed variables (i.e. the excitation-emission wavelengths, the inputs) that yield the best linear regression to the predicted variables (i.e. the biological parameters, the outputs). These linear combinations are known as Latent Variables (LVs), and can be seen as underlying structures or patterns that are correlated

to the predicted parameters more directly than the original spectral variables measured by the experiment. More extensive descriptions of PLS can be found elsewhere [111]. In the current work, the multiway version of PLS, known as n-PLS, was used [138], which allows for fully exploiting the mathematical relationships among the three modes of the EEM data (i.e. sample, excitation, emission).

To facilitate the modelling task, the predicted variables were converted into their logarithm with base 10 to normalize their distribution. The predictive models were validated by a 4-fold double cross-validation [139]. In short, the data set was split randomly into a training and a validation set, consisting of 75 % and 25 % of the total data, respectively. The training set was used to calibrate the model and optimize the number of LVs, using a leave-one-out cross-validation (LOOCV), in which one sample from the training set is removed, a PLS model is built on the remaining training samples and assessed on the left-out one; this procedure is then repeated with a different leave-out sample until all samples have been rotated. The external validation set is held out of this loop and is used to validate the model with the optimal number of LVs as determined by the LOOCV. The whole procedure (LOOCV + external validation) is repeated three more times, using the external validation data that was previously used for training, until every sample has appeared in the external validation set once.

To assess the quality of the models several parameters were evaluated, such as the variance explained in the biological parameters (%), the root mean square error of cross-validation (RMSECV) and prediction (RMSEP), and the R^2 and slopes of the training and external validation (from now on mentioned as validation) sets.

After evaluating the quality of the models, a final model was built for each predicted parameter using the whole data set, using the optimal number of LVs defined by the previous validation, in order to determine and examine the regression coefficient values.

All multivariate statistical analysis were performed in MATLAB (The MathWorks®) using the drEEM toolbox (http://www.models.life.ku.dk/dreem) and n-way toolbox [72].

VI.3. Results and Discussion

Nannochloropsis oceanica production can aim at different final products. The whole biomass of microalgae is rich in pigments and fatty acids, but contains also high valuable proteins and carbohydrates. Depending on the end-product desired, the production of the biomass can be tuned to reach higher yields. For that reason, the experiments of this work were designed to increase the concentration range of three different products, biomass (as cell concentration), chlorophyll and fatty acids. Having a wider range of scenarios increases the range of the outputs, which, in the end, increases the strength of the prediction models.

Microalgae cultivation is characterized by low biomass concentrations, mainly to avoid dark zones in the bioreactors that could lead to lower photosynthetic efficiencies [50]. For this reason, microalgae samples have high water content, that results in the presence of high intensity Rayleigh scatter (Figure VI.1*a*). The presence of water scatter will have a high impact in the estimation of the regression coefficients of the final models (presented in section VI.3.4) and, since its signal is not proportional to the water content in the sample, it was removed before the

PLS modelling, as described in Section VI.2.3 [19]. An example of the pre-processed spectra is shown in Figure VI.1*b*.

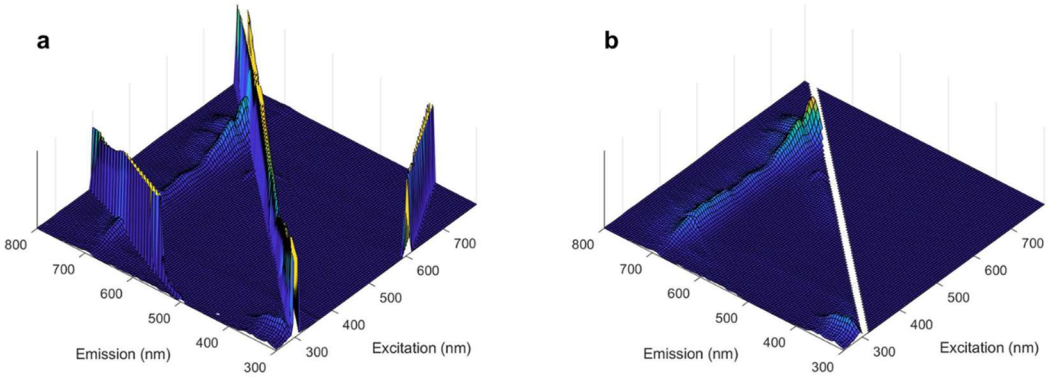
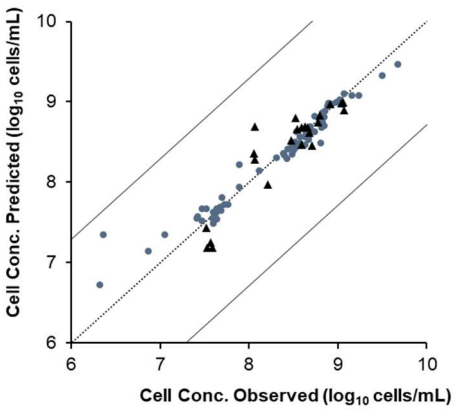


Figure VI.1 Fluorescence spectra of a Nannochloropsis oceanica sample: a) original spectra; b) final spectra used as inputs in the PLS models. Rayleigh scatter of first order was removed and replaced by empty values; the second order was replaced with an interpolation of surrounding data points. Fluorescence signal corresponding to emission wavelengths (y-axis) shorter than the excitation wavelengths (x-axis) was replaced by zeros. Inner filter effects were also corrected whenever present.

VI.3.1. Cell concentration

The model obtained to monitor cell concentration of *N. oceanica* can explain 84 % of the variability captured by the fluorescence spectroscopy (Figure VI.2) with five LVs. A low root mean square error of *N. oceanica* concentration prediction (RMSEP) was observed (0.27 log_{10} cells/mL), which represents the average distance between the observed values and the ones predicted by the model. The relative error (in percentage), calculated as a quotient between the prediction error (RMSEP) and the observed cell concentration average value, is 3.18 %. The root mean square error of cross-validation (RMSECV) of 0.30 log_{10} cells/mL indicates absence of model overfit. However, the R^2 of the validation data set was lower than the one of the training set, mainly because of the poor predictions for samples with lowest cellular concentrations.



Variance	*(%)*	83.69
RMSECV	*(log_{10} cells/mL)*	0.30
RMSEP		0.27
Training	R^2	0.88
	Slope	0.85
Validation	R^2	0.66
	Slope	0.88
Number of LVs		5

Figure VI.2 Cell concentration prediction model (one of the four partitions of training/validation data sets). Training (●) (n=69) and validation (▲) (n=23) data are represented in log_{10} cells/mL.

In the current industrial scenario, most microalgae products are sold as whole biomass powder, making total biomass a key parameter to control process efficiency. During cultivation at industrial scale, too high or too low biomass concentration can have an influence in several biological parameters. Low concentrations can result in an inefficient light absorption or photo inhibition, while high concentrations result in dark regions in the bioreactor that provoke endogenous respiration [50].

Biomass concentration can be monitored using different parameters, such as optical density (OD), dry weight (g/L) and cell concentration (cells/L). In the present work, biomass concentration was measured as cell concentration. The main reason was because the experiments were designed to give a wide range of cell concentration, but some experiments induce also other biological changes, such as coloration or accumulation of fatty acids. For example, as mentioned by Janssen *et al.* (2018) [140], accumulation of fatty acids in lipid bodies, due to nitrogen depletion medium, leads to an increase of the dry weight while the cell concentration reaches a plateau. This phenomenon was also observed with the experiments of this work (data not shown). Experiments using a day/night cycle lead the biomass to follow circadian rhythms, which means that the cell size increases during the day while the cell concentration increases during the night (due to cell division) [130,141].

Previous studies reported the use of fluorescence spectroscopy to monitor cell concentration of a different microalgae, *Dunaliella salina*, in different biological states, as non-stressed ("green") and stressed ("orange") cultures, where carotenoids were induced due to nitrogen depletion or increased salinity [78](Chapter IV). A slightly different modelling approach was used then, since PCA (principal component analysis) was performed on the EEMs, without the need to remove the scatter, prior to PLS modelling. Nevertheless, a similar explained variance was observed (between 85.7 and 86.3%), with similar values of R^2 for training and validation sets (between 082 and 0.86). These results, together with the results of this work, demonstrate the potential of using fluorescence spectroscopy as a monitoring tool for cell concentration with different microalgae biomass.

VI.3.2. Chlorophyll

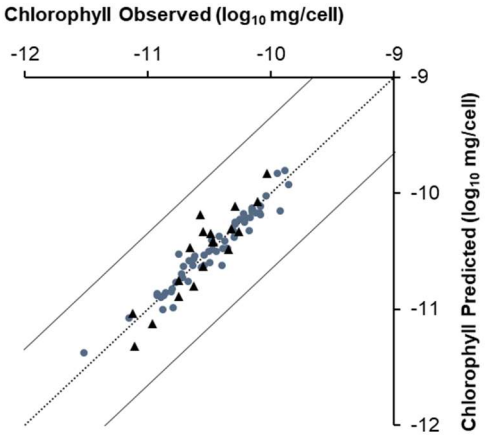
Chlorophyll is the most abundant light harvesting pigment in nature, enabling the photosynthesis, and is a molecule well-studied for its potential use in several fields. In the feed and food supplements industry, chlorophyll is relevant due to its anti-oxidant properties. Because of its bright green colour, it is also an appealing dye for the food and paint industries [136,137].

Chlorophyll content in the microalgae is tightly correlated with the light intensity and circadian rhythms. It was reported that chlorophyll content increases during the light period, and starts to decrease with the beginning of a dark period [130,141,142]. This phenomenon is explained by the fact that the cell division mechanism is favourable in the dark period, and since *Nannochloropsis* genus divide by binary fission, the chlorophyll content of the "adult" cell is divided by its new cells [130].

The experimental conditions induced a considerable variability in the chlorophyll content. As mentioned previously, day/night cycles induce the chlorophyll to oscillate during the experiments. Moreover, microalgae are known for their ability to adapt their photosynthetic apparatus to

different light conditions, a process called photoacclimation. High light intensities reduce chlorophyll content to protect the cell, while low intensities induced the photosynthetic apparatus to synthesise chlorophyll, to provide the cell with a highest light harvesting capacity [140,143]. Nitrogen starvation, however decreases chlorophyll content and increases carotenoid concentrations to equip the cell with stress defence mechanisms [130,140].

The chlorophyll content studied in these experiments enabled the development of a accurate model (Figure VI.3), with a relative error of 1.31 % using five LVs. The validation and training sets show high R^2 and low errors, both RMSECV and RMSEP.



Variance	*(%)*	86.47
RMSECV	*(log_{10} mg/cell)*	0.39
RMSEP		0.32
Training	R^2	0.88
	Slope	0.97
Validation	R^2	0.78
	Slope	0.97
Number of LVs		5

Figure VI.3 Chlorophyll content prediction model (one of the four partitions of training/validation data sets). Training (●) (n=57) and validation (▲) (n=19) data are represented in log_{10} mg/cell.

Fluorescence spectroscopy was previously used to monitor and model the content of chlorophyll *a* and *b* in *Dunaliella salina* (Chapter III). In that work, cultivation experiments were performed in outdoor pilot scale facilities, and so the importance of climatic conditions was studied by its incorporation in the PLS modelling as input, additionally to the fluorescence EEMs. Although extra input information had to be used to successfully model the chlorophylls content, a high explained variance (ranging from 83.3 to 89.2 %) and a high R^2 for validation and training data sets (between 0.75 and 0.89) were also obtained.

The possibility of monitoring chlorophyll content online provides information about the physiological state of the microalgae at real time. With this knowledge it is possible to take decisions during the cultivation process, without the need to perform the time-consuming lab analysis that are usually required. In this work, fluorescence was acquired off-line, however it would be rather simple to perform the acquisition online, by coupling an optical probe to the cultivation system.

VI.3.3. Lipids

Several microalgae are being studied for their potential to produce high content and/or high-quality lipids. Having in mind the different opportunities for an enriched-lipid biomass, the *N. oceanica* lipid profile was assessed through a set of different experiments. Nitrogen depletion is

a well-documented strategy to increase lipid content in the TAG fraction [116,121] while low light conditions were documented to lead to an increase of lipids in the cellular membranes [121]. Also, low temperatures were described to increase the content of unsaturated fatty acids [116,133,134], whereas high temperature favour the saturated [128].

Models were obtained to monitor fatty acids as total content (Figure VI.4*a*), as well as saturated (Figure VI.4*b*) and unsaturated (Figure VI.4*c*) content only. Values of explained variance ranged between 87 and 92 %. When compared with the previous models of this work, a higher number of LVs is needed to explain the variability of the fluorescence (between 9 and 10). Relative error of prediction for total and unsaturated fatty acids was 5.89 % and 5.62 %, respectively, lower than the value found for saturated fatty acids (9.54 %). All R^2 of validation and training sets were above 0.74 and slopes close to 1.

When cultivated under optimal growing conditions, most of the lipid fraction in microalgae is present on the cellular membrane [117]. However, under stress growing conditions some microalgae, like *Nannochloropsis*, can accumulate up to 45% of their dry weight in triacylglycerol (TAG) [116,121]. According to the region of the cell where the lipid is accumulated and its profile, the final destination of the lipid-enriched biomass can vary from feed or food supplements to biodiesel production [115,117,122].

Fatty acids can be classified into saturated and unsaturated, the latter into mono- or poly-unsaturated according to the number of chemical double bonds. Biomass produced with the aim of supplying feed or food supplements industries is desired to be rich in unsaturated fatty acids, preferably omega-3 fatty acids such as EPA (eicosapentaenoic acid) or DHA (docosahexaenoic acid) [117,121]. Biomass produced for biofuel applications needs to fulfil quality parameters such as ignition and combustion performance, and those are directly correlated with saturated and unsaturated content [116,123].

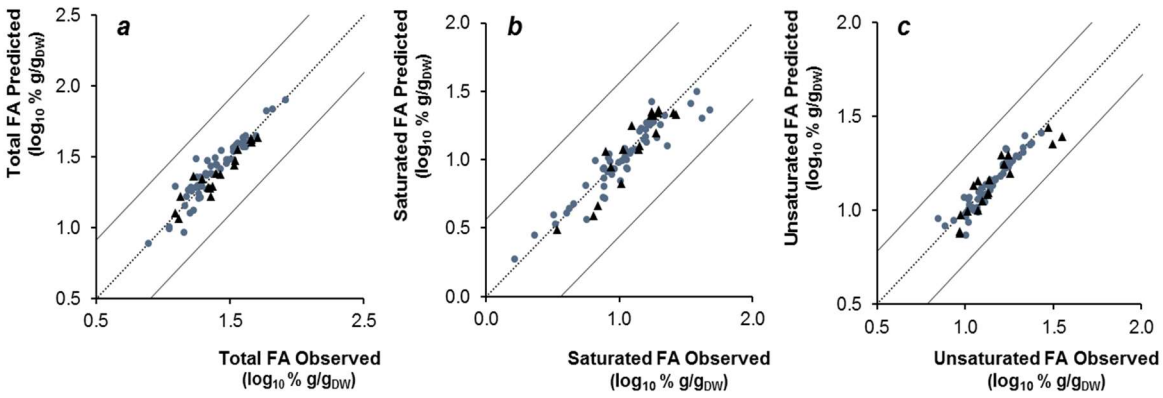


Figure VI.4 Fatty acids (FA) prediction models for total (a)*, saturated* (b) *and unsaturated* (c) *FA (one of the four partitions of training/validation data sets). Training (●) (n=54) and validation (▲) (n=18) data are represented in* log_{10} *% g/g* $_{DW}$.

Table VI.2 Prediction models parameters for total, saturated and unsaturated fatty acids.

		Total	Saturated	Unsaturated
Variance	*(%)*	90.88	87.86	85.33
RMSECV	*(log_{10} % g/g_{DW})*	0.25	0.31	0.17
RMSEP		0.17	0.25	0.12
Training	R^2	0.89	0.87	0.90
	Slope	0.98	0.84	1.01
Validation	R^2	0.89	0.84	0.83
	Slope	0.85	1.12	0.85
Number of LVs		9	10	9

To our knowledge, fluorescence spectroscopy was not previously reported as an monitoring tool for lipid content. It is known that fluorescence spectroscopy is highly sensitive to detect the presence of natural fluorophores and that lipid molecules are not natural fluorophores. In the presence of a complex matrix, like microalgae cultivation broth, fluorescence spectroscopy is able not only to identify the natural fluorophores (intra or extracellular) but also the relations between these compounds and the non-fluorophores [12]. For that reason, fluorescence spectroscopy spectra cannot be directly used for quantification, but is a powerful tool revealing the correlations between compounds that emit natural fluorescence, the ones that capture it, and the ones that somehow mask the fluorescence signal.

VI.3.4. Regression coefficients of the final models for cell concentration, chlorophyll and fatty acids

The experiments performed enabled to acquire a wide range of scenarios of a *N. oceanica* cultivation for different end products. After calibrating the models for the parameters cell concentration, chlorophyll and fatty acids, it is possible to confirm that fluorescence spectroscopy has a great potential for online monitoring of all three parameters simultaneously.

Aiming at the application of fluorescence spectroscopy, a final model was created for each parameter studied, using 100% of the data as training set. Figure VI.5 shows the regression coefficients obtained for each output, where the excitation and emission wavelengths (in nm) are in the x-axis and y-axis, respectively, and each square is an excitation/emission pair. The weight of each regression coefficient is represented in colour-scale.

For the cell concentration model, two main regions can be distinguished as having relevant regression coefficient weight (positive or negative): a band at emission wavelengths higher than 600 nm (whole excitation range), and a region for excitation and emission wavelengths lower than 400 nm. This reveals a similarity with the overall fluorescence signal of a sample, where these two regions have high fluorescence intensities. As described previously, these two regions of the spectra correspond to the pigments fluorescence band and the protein-like region (aromatic aminoacids), respectively [69,73,75,78]. Several differences can be noticed between the regression coefficients map of the cell concentration model and the remaining outputs. The protein-like region does not have the same weight as for cell concentration, and different weights are attributed to the pigments band. As expected, a high correlation is shown between the

pigments band and the regression coefficient map for chlorophyll content prediction. Also, specific areas of this pigment band are used for fatty acids content prediction (as total, saturated or unsaturated). These results confirm the relationship previously described between chlorophyll and fatty acids content [130] and that, although fatty acids do not emit fluorescence, they interfere with the signal of natural fluorophores like chlorophyll. The development of a simpler spectrofluorometric technique, able to acquire signal in these two regions of the spectra, instead of the entire range, would possibly simplify the analysis by decreasing the acquisition time of each data point.

Using fluorescence spectroscopy to simultaneously monitor several biological parameters has been described as one of the powerful characteristics of this technique [12]. By coupling the fluorescence EEMs and the regression coefficients developed by chemometric models, more knowledge about the cultivation process can be acquired, enabling important decisions at real time, like the optimum harvesting time.

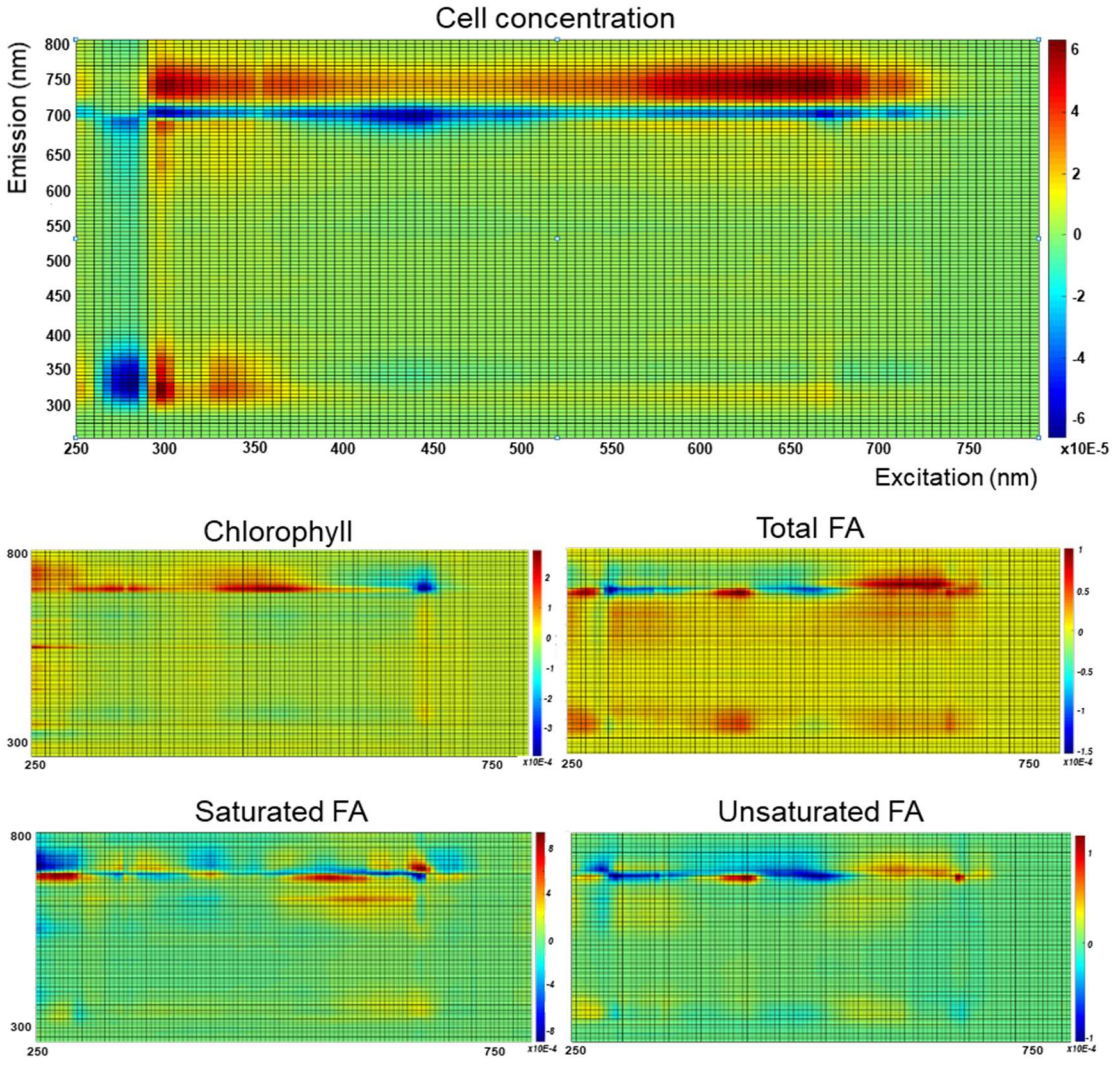


Figure VI.5 Regression coefficients of the prediction models for cell concentration, chlorophyll, and fatty acids (FA) as total, saturated and unsaturated; 100% of the data set was used as training set. Excitation wavelengths are represented in the x-axis, emission wavelengths in the y-axis, and intensity is represented in the colour bar on the right side.

VI.4. Conclusions

The present work shows the considerable potential of fluorescence spectroscopy as a monitoring tool for *N. oceanica* production. Not only it was possible to monitor natural fluorophores (chlorophyll), but also more complex and non fluorophorous' molecules (fatty acids) or even the whole cell (cell concentration). Different environmental conditions were tested in order to increase the response range of the parameters under study. That allowed the development of accurate models for all the parameters (explained variances between 81 and 91 %), with low errors (REMSECV and RMSEP between 0.12 and 0.40) and high R^2 (between 0.65 and 0.93). The regression coefficient maps highlight the importance of the pigment and the protein regions for the development of these models.

VI.5. Acknowledgements

This work was supported by the project "Microalgae As a Green source from Nutritional Ingredients for Food/Feed and Ingredients for Cosmetics by cost effective New Technologies" [MAGNIFICENT], funded by Bio-based Industries Joint Technology Initiative under European Commission [project ID: 745754], by the Associate Laboratory for Green Chemistry – LAQV which is financed by national funds from FCT/MCTES (UID/QUI/50006/2019), and by the PhD grant SFRH/BD/108894/2015 of FCT/MCTES.

Thesis overview and general conclusions

This thesis addresses several scientific questions related to the ability of fluorescence spectroscopy to monitor several biological parameters simultaneously, such as cell concentration and viability, pigments or fatty acids content, and nitrate concentration in the cultivation medium. Additionally, the possibility of using this tool in different process steps was tested, such as cultivation, harvesting and permeate treatment. For this work, two marine microalgae with industrial relevance were selected as case-studies, *Dunaliella salina* and *Nannochloropsis oceanica*.

In the second chapter of this thesis, cell concentration and cell viability were assessed during cultivation and harvesting (by membrane filtration) of "green" *D. salina*, non-stressed stage. Fluorescence spectra were acquired in the culture broth and in both streams of the filtration process, concentrate and permeate. Chemometric tools, namely principal component analysis (PCA) and projection to latent structures (PLS), were used to process the fluorescence information and to develop predictive models. PCA was used to compress the information in the fluorescence spectra and reduce the noise and collinearity. The PCs resulting from the PCA were used as inputs in the PLS modelling. For cell concentration, a model was developed using cultivation data and combining data from cultivation with harvesting, to increase the cell concentration prediction range. This work demonstrated that fluorescence spectroscopy is robust to predict cell concentration in a wide range, giving the opportunity to use the same model in two different processes (cultivation and harvesting). For prediction of cell viability in membrane filtration harvesting experiments, it was proved that fluorescence spectra from both streams (concentrate and permeate) are important to be used as inputs, since both contain information about molecular compounds that can be possibly released during cell disruption and retained differently by the filtration membrane.

In the third and fourth chapter of this thesis the study was focused in enriched carotenoid *D. salina* biomass, its cultivation, carotenoid induction, membrane harvesting and treatment of the respective permeate. The experiments for this work were performed outdoors, exposed to climatic conditions of light and temperature, where only the maximal temperature was maintained at 25 C, to avoid a culture crash. The same model strategy was applied, using PCA to compress the fluorescence information and PLS to develop the prediction models.

Carotenogenesis was monitored through pigment quantification during the transition of "green", low carotenoid content, to "orange" culture, carotenoid-enriched biomass. Two chlorophylls (*a* and *b*) and five carotenoids (zeaxanthin, *α*-carotene, all-trans-β-carotene and 9-cis-β-carotene) were studied, as well as the total content in chlorophylls and carotenoids. The models were developed using PCs of the fluorescence as input, but also information about the climatic conditions. The use of climatic inputs in these prediction models was described for the first time. Two effects were evaluated, using climatic data of the sampling day or its cumulative effect throughout the carotenogenesis. From the results obtained, it is possible to conclude that

providing information on light conditions in outdoor experiments, in the form of FPAR (fraction of photosynthetic active radiation), irradiance, number of hours of light or presence of clouds, leads to an improvement of the prediction models. Acquiring information about light is common practice in microalgae production facilities, so the inclusion of this input is easy to implement.

To calibrate the prediction models, two analytical methods of pigments' quantification were studied, spectrophotometry and HPLC, each with a specific extraction procedure. The main conclusions point out to the importance of carefully choosing the analytical method for calibration, since it highly influences the prediction quality of the models. It is difficult to conclude which methodology is more adequate for pigment quantification, because it depends on the main purpose of a specific microalgae cultivation. If it is necessary to monitor several individual molecules simultaneously it is possible, by HPLC, to obtain a pigment profile of the samples. However, if the goal is to monitor total chlorophylls or total carotenoids, the spectrophotometric quantification might be enough. Nevertheless, it is advised to optimise the extraction methodology for the microalgae in study.

As studied before, for the "green" *D. salina* biomass, a prediction model was developed for the cell concentration using data collected during the carotenogenesis, data from harvesting, and combining data from both processes. Once again, it was proven that fluorescence spectroscopy is a powerful and advantageous method prediction of the same biological parameters in different steps of microalgae production process. Furthermore, nitrate concentration was successfully predicted in two distinct processes, using combined data of the carotenogenesis and of the permeate treatment. As previously proven in the prediction models of cell viability in "green" *D. salina,* also for carotenoid-induced biomass models it was important to include as inputs the fluorescence acquired in the concentrate and permeate streams for carotenoid-enriched biomass models.

The work developed in this three chapters (2, 3 and 4) show the versatility of the fluorescence spectroscopy to monitor different products in different biological stages of the same microalga.

To further evaluate the potential of using fluorescence spectroscopy as a monitoring tool in microalgae processes, additional studies were performed using another marine microalga, *N. oceanica*, known for its ability to produce high concentrations of lipids (chapters 5 and 6). Although lipids are known for not having fluorescence properties, fluorescence spectroscopy features enable the use of this technique to monitor either fluorophores or non-fluorophore molecules. In the presence of a complex matrix, like microalgae cultivation broth, fluorescence spectroscopy can reveal the correlations between compounds that emit natural fluorescence, the ones that capture it, and the ones that somehow mask the fluorescence signal. Thus, several experiments were performed in order to increase the variability of biological scenarios, from low to high lipid content. The studies are divided in two, one focusing on the production of a specific polyunsaturated fatty acid, the omega-3 eicosapentaenoic (EPA), its accumulation in two different parts of the cell – cell membrane (PL) or lipid bodies (TAG) – and its monitoring by fluorescence spectroscopy (chapter 5); a second study was focused on the development of predictive models for cell concentration, chlorophyll and fatty acids (divided in total, saturated or unsaturated fatty acids) (chapter 6).

Different modelling approaches were studied in these two chapters. For both, fluorescence spectra were submitted to a standard pre-treatment before their use in the modelling, that

included scatter removal and inner-filter correction. A preliminary study was done to assess the quality of the prediction models for EPA using as input raw or pre-treated fluorescence spectra. The modelling strategy used was the same as for the previous studies, where the fluorescence spectra were converted into PCs, through PCA, before PLS modelling. The results show that no differences are found in the prediction quality of the models (Annex VII.1) and therefore, the pre-treatment was applied to all spectra prior to model development for the *N. oceanica* studies.

Two validation strategies were studied in the development of prediction models for EPA concentration, one using batch-by-batch and other using 25% of the total data randomly chosen. The results show that a suitable calibration data set should be used in order to capture a wider and representative range of EPA concentrations. Meaning, in this case, that a random 25 % of the total data is the most suitable approach. The experiments performed allow to conclude that nitrogen depletion favours EPA content to accumulate in the TAG fraction of the cell, while low light or low temperature favour its increase in the PL fraction.

In chapter 6, where models were developed to predict cell concentration, chlorophyll and fatty acids content, the pre-treated spectra were used as inputs in the PLS modelling, contrary to what was done before (where fluorescence spectra were converted to PCs though PCA, and then these PCs used as inputs in PLS). With this different input strategy, it was possible to determine the regression coefficients map, identifying the areas of the spectra with higher weight in the models. The regression coefficient maps show two main regions with relevant coefficient weight: the pigment region of the spectra, a band corresponding to emission wavelengths higher than 600 nm (whole excitation range); and the protein-like region, with excitation and emission wavelengths lower than 400 nm.

In these two last chapters (5 and 6) it is demonstrated the potential to monitor different molecules, fluorophores or not, in *N. oceanica* biomass using a single tool, fluorescence spectroscopy.

Overall, with the experiments developed through this thesis it was possible to evaluate the potential of fluorescence spectroscopy to monitor different microalgae, its physiological characteristics and their production of different value-added compounds. This work contributes to the validation of a new monitoring tool in microalgae production, based in a non-destructive and non-invasive technology, 2D fluorescence spectroscopy.

VII.1. Suggestions of future work and outlook

Fluorescence spectroscopy proved to be a powerful tool to monitor several biological parameters simultaneously. Nevertheless, some questions were raised where research effort should be devoted in the future. Three working guidelines for research are suggested:

1) Simplification of the spectrofluorometer

After identifying the regions of the spectra that are relevant to monitor the target biological parameters, it would be a great advantage to simplify the acquisition equipment by selecting and tuning only the wavelengths needed. Not only the time per analysis would decrease, but also the

cost of the equipment acquisition and maintenance. Nowadays some spectrofluorometers enable the possibility of coupling an optical fibre, but some work could be done to improve the connection of the probe to the bioreactors. With the use of a switch box, several probes can be connected to the same, enabling the acquisition of fluorescence signal from different processes simultaneously.

2) Simplification of the chemometric modelling & programming

Several chemometric strategies were studied in this thesis in order to increase the knowledge of the process. The use of raw spectra or the use of pre-treatment to remove water scatter and inner filter effects should be further evaluated aiming at improving prediction and decrease the computational effort. Similarly, the use of PCA to compress the fluorescence signal before the PLS modelling should be assessed thoroughly. An interface program should be designed to show the operator only the simultaneous measures that are acquired, running the code in background.

3) Calibrate the monitoring tool focusing on the product of interest instead of the microalgae of origin

The models developed in this thesis are specific per microalgae. For example, different prediction models for cell concentration were obtained for *D. salina* and *N. oceanica*. However, development of models able to predict the same product or physiological parameter, in different microalgae, would be of great interest for the industry. It is common practice in several microalgae producing companies to cultivate different microalgae according to the season of the year, known as “winter” and “summer” species. Nevertheless, the target compounds of the microalgae biorefinery not always change. For example, for fucoxanthin production, in the south of Portugal, *Isochrysis galbana* is cultivated during summer while *Phaeodactylum tricornutum* during winter. Therefore, the prediction tool could be developed having in consideration the final product, the value-added metabolite of interest, or a specific physiological parameter, such as cell concentration.

4) Application in industry

Fluorescence spectroscopy could be applied today in microalgae production if some improvements are made. The development of an interface program would be a key-factor. The adaptation of the fluorescence spectroscopy will depend on the main goal of the company. For example, if the goal is to monitor cell concentration and specific metabolites during one step of the process, such as the work developed in this thesis with *N. oceanica*, only one optical fibre per cultivation system would be needed. However, if we think about the work developed in this thesis for the *D. salina*, different processes were monitored, thus one optical fibre should be coupled to the cultivation system, one to the concentrate and one to the permeate tanks of the membrane harvesting, and one for the permeate treatment system. A switch-box needs to be developed and optimised to allow the acquisition of signals from different processes simultaneously.

The price of the spectrofluorometers and the optical fibre probes is highly dependent on the technology used to generate the fluorescence signal and the sensitivity of the signal acquisition. If the equipment works with a wide range of wavelengths, the complexity of the equipment

increases as the price. The same can be said for the optical fibre probes. A lab-bench size spectrofluorometer, with monochromator, the most complete and sensitive system possible, can cost between 15.000 € and 20.000 €. An optical fibre with 1.5 m can cost around 2.000 €, and the system to connect the fibre to the spectrofluorometer around 8.000 €. The price and sensitivity of any optical fibre is also dependent on the length of the fibre. If the goal is to monitor several processes, the overall cost increase per the price of each optical fibre.

There's spectrofluorometers that allow the possibility of installing a sipper unit, constituted by a peristaltic pump and an all-quartz flow-through cuvette, where a sample could be withdraw from the cultivation vessel (for example) and analysed in the spectrofluorometer. In this case, the cost of the optical fibre probe is not applicable, but a system of pumps and switching valves should be designed to adapt to each specific case.

There is advantages and disadvantages for each solution, and each microalgae production facility should carefully consider them according to their monitoring needs. Nevertheless, once the trigger monitoring points and main goals are stablished, a solution can be easily engineered.

Annex VII.1

Nannochloropsis oceanica was cultivated under different experimental conditions of light, temperature and nitrogen deplete/replete media. Eicosapentaenoic (EPA) accumulation was assessed in two different parts of the cell, cell membrane (PL) or lipid bodies (TAG). To evaluate the impact of the Rayleigh scatter of the fluorescence spectra in the final quality of the prediction models, the following study was performed. Two sets of inputs were generated, one using the raw data from the fluorescence, and a second one where the water scatter and the inner filter effect were corrected, as explained in Chapter 5. The two data sets were then submitted to the same chemometric treatment, first the fluorescence spectra were compressed to PCs (through PCA), and then 10 PCs (that explained 99.9% of the variance found in the fluorescence spectra) were used as input in the PLS modelling. The validation was performed four times, using a random data set, where each value was only used once. The results presented in Table A.1 are an average of the prediction model parameters obtained.

Table VII.1 Prediction model parameters parameters for EPA in TAG and in PL fractions of N. oceanica. The results showed are an average of 4 cycles of random validation data sets (25 % of the total data set). The same data sets where used in all prediction models, using raw or pre-treated fluorescence spectra.

		Var	RMSCEV	RMSEP	Validation		Training		N° inputs
		(%)	(g/g $_{DW}$)	(g/g $_{DW}$)	R^2	Slope	R^2	Slope	
EPA *in* TAG	*raw*	93.63 ± 1.31	0.26 ± 0.01	0.25 ± 0.02	0.89 ± 0.02	0.96 ± 0.04	0.94 ± 0.01	1.00 ± 0.00	12 ± 4
	pre-treat	92.10 ± 2.51	0.30 ± 0.01	0.29 ± 0.06	0.87 ± 0.08	0.95 ± 0.14	0.92 ± 0.03	1.00 ± 0.00	13 ± 3
EPA *in* PL	*raw*	89.82 ± 3.04	0.49 ± 0.03	0.50 ± 0.09	0.80 ± 0.06	1.00 ± 0.15	0.90 ± 0.03	1.00 ± 0.00	16 ± 2
	pre-treat	84.83 ± 3.37	0.49 ± 0.01	0.37 ± 0.02	0.81 ± 0.04	0.93 ± 0.04	0.85 ± 0.03	1.00 ± 0.00	11 ± 4

As it can be seen in the prediction model parameters, the differences observed using raw or pre-treated fluorescence spectra as input in the PLS modelling are minor. After this preliminary study, the spectra used in the prediction models of Chapter 5 and 6 used pre-treated fluorescence spectra.

References

[1] R. Ulber, J.G. Frerichs, S. Beutel, Optical sensor systems for bioprocess monitoring, Anal. Bioanal. Chem. 376 (2003) 342–348. doi:10.1007/s00216-003-1930-1.

[2] A. Glindkamp, D. Riechers, C. Rehbock, B. Hitzmann, T. Scheper, K.F. Reardon, Sensors in Disposable Bioreactors Status and Trends, Adv Biochem Eng Biotechnol. 115 (2009) 145–169. doi:10.1007/10_2009_10.

[3] J.G. Henriques, S. Buziol, E. Stocker, A. Voogd, J.C. Menezes, Monitoring Mammalian Cell Cultivations for Monoclonal Antibody Production Using Near-Infrared Spectroscopy, in: T. Scheper, G. Rao (Eds.), Opt. Sens. Syst. Biotechnol., Springer, 2010: p. 262. doi:10.1007/BFb0103032.

[4] S. Marose, C. Lindemann, R. Ulber, T. Scheper, Optical sensor systems for bioprocess monitoring, TIBTECH. 17 (1999) 30–34. doi:10.1002/elsc.201500014.

[5] K. Schügerl, Progress in monitoring, modeling and control of bioprocesses during the last 20 years., J. Biotechnol. 85 (2001) 149–173. doi:10.1016/S0168-1656(00)00361-8.

[6] L. Olsson, J. Nielsen, On-line and in situ monitoring of biomass in submerged cultivations, Trends Biotechnol. 15 (1997) 517–522. doi:10.1016/S0167-7799(97)01136-0.

[7] R.C. Dorresteijn, Software sensors as a tool for optimization of animal-cell cultures, Wageningen University and Research, 1997.

[8] K. Hantelmann, M. Kollecker, B. Hitzmann, T. Scheper, Two-dimensional fluorescence spectroscopy: A novel approach for controlling fed-batch cultivations, J. Biotechnol. 121 (2006) 410–417. doi:10.1016/j.jbiotec.2005.07.016.

[9] O.S. Wolfbeis, Materials for fluorescence-based optical chemical sensors, J. Mater. Chem. 15 (2005) 2657–2669. doi:10.1039/b501536g.

[10] H. Schulz, Spectroscopic Technique: Raman Spectroscopy, 2nd ed., Elsevier Inc., 2018. doi:10.1016/B978-0-12-814264-6.00005-0.

[11] C. Ranzan, L.F. Trierweiler, B. Hitzmann, J.O. Trierweiler, Fluorescence spectroscopy as a tool for ethanol fermentation on-line monitoring, IFAC Proc. Vol. 8 (2012) 940–945. doi:10.3182/20120710-4-SG-2026.00166.

[12] Joseph R. Lakowicz, Principles of fluorescence spectroscopy, 2006. doi:10.1007/978-0-387-46312-4.

[13] S. Marose, C. Lindemann, T. Scheper, Two-dimensional fluorescence spectroscopy: a new tool for on-line bioprocess monitoring., Biotechnol. Prog. 14 (1998) 63–74. doi:10.1021/bp970124o.

[14] B. Tartakovsky, M. Sheintuch, J.M. Hilmer, T. Scheper, Application of scanning fluorometry for monitoring of a fermentation process, Biotechnol. Prog. 12 (1996) 126–131. doi:10.1021/bp950045h.

[15] L. Lenhardt, R. Bro, I. Zeković, T. Dramićanin, M.D. Dramićanin, Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey., Food Chem. 175 (2015) 284–91. doi:10.1016/j.foodchem.2014.11.162.

[16] J. Sádecká, J. Tóthová, Fluorescence spectroscopy and chemometrics in the food classification - a review, Czech J. Food Sci. 25 (2007) 159–173. doi:10.1016/j.echo.2007.08.029.

[17] O. Podrazký, G. Kuncová, A. Krasowska, K. Sigler, Monitoring the growth and stress responses of yeast cells by two-dimensional fluorescence spectroscopy: first results., Folia Microbiol. (Praha). 48 (2003) 189–92. doi:10.1007/BF02930954.

[18] T. Larsson, M. Wedborg, D. Turner, Correction of inner-filter effect in fluorescence excitation-emission matrix spectrometry using Raman scatter, Anal. Chim. Acta. 583 (2007) 357–363. doi:10.1016/j.aca.2006.09.067.

[19] R.G. Zepp, W.M. Sheldon, M.A. Moran, Dissolved organic fluorophores in southeastern US coastal waters: Correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices, Mar. Chem. 89 (2004) 15–36. doi:10.1016/j.marchem.2004.02.006.

[20] M. Bahram, R. Bro, C. Stedmon, A. Afkhami, Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation, J. Chemom. (2006) 99–105. doi:10.1002/cem.978.

[21] M.-N. Pons, S. Le Bonté, O. Potier, Spectral analysis and fingerprinting for biomedia characterisation, J. Biotechnol. 113 (2004) 211–230. doi:10.1016/J.JBIOTEC.2004.03.028.

[22] A.D. Shaw, N. Kaderbhai, A. Jones, A.M. Woodward, R. Goodacre, J.J. Rowland, D.B. Kell, Noninvasive , On-Line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol Using Dispersive Raman Spectroscopy and Chemometrics, Appl. Spectrosc. 53 (1999) 1419–1428.

[23] F. Ammari, L. Redjdal, D.N. Rutledge, Detection of orange juice frauds using front-face fluorescence spectroscopy and Independent Components Analysis, Food Chem. 168 (2015) 211–217. doi:10.1016/j.foodchem.2014.06.110.

[24] T.J. Grundl, J.H. Aldstadt, J.G. Harb, R.W. St. Germain, R.C. Schweitzer, Demonstration of a method for the direct determination of polycyclic aromatic hydrocarbons in submerged sediments, Environ. Sci. Technol. 37 (2003) 1189–1197. doi:10.1021/es020940e.

[25] S. Babicheno, E. Erme, T. Ivkina, L. Poryvkina, V. Sominsky, A portable device and method for on-site detection and quantification of drugs, 2005111586, 2004.

[26] A.P. Teixeira, C.A.M. Portugal, N. Carinhas, J.M.L. Dias, J.P. Crespo, P.M. Alves, M.J.T. Carrondo, R. Oliveira, In situ 2D fluorometry and chemometric monitoring of mammalian cell cultures, Biotechnol. Bioeng. 102 (2009) 1098–1106. doi:10.1002/bit.22125.

[27] G. Wolf, J.S. Almeida, M.A.M. Reis, J.G. Crespo, Modelling of the extractive membrane bioreactor process based on natural fluorescence fingerprints and process operation history, Water Sci. Technol. 51 (2005) 51–58.

[28] E.M. Carstea, J. Bridgeman, A. Baker, D.M. Reynolds, Fluorescence spectroscopy for wastewater monitoring : A review, Water Res. 95 (2016) 205–219. doi:10.1016/j.watres.2016.03.021.

[29] 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.

[30] J.N. Louvet, B. Homeky, M. Casellas, M.N. Pons, C. Dagot, Monitoring of slaughterhouse wastewater biodegradation in a SBR using fluorescence and UV-Visible absorbance., Chemosphere. 91 (2013) 648–55. doi:10.1016/j.chemosphere.2013.01.011.

[31] A.C. Hambly, R.K. Henderson, M.V. Storey, A. Baker, R.M. Stuetz, S.J. Khan, Fluorescence monitoring at a recycled water treatment plant and associated dual distribution system - Implications for cross-connection detection, Water Res. 44 (2010) 5323–5333. doi:10.1016/j.watres.2010.06.003.

[32] G. Wolf, J.S. Almeida, C. Pinheiro, V. Correia, C. Rodrigues, M.A.M. Reis, J.G. Crespo, Two-dimensional fluorometry coupled with artificial neural networks: A novel method for on-line monitoring of complex biological processes, Biotechnol. Bioeng. 72 (2001) 297–306. doi:10.1002/1097-0290(20010205)72:3<297::AID-BIT6>3.0.CO;2-B.

[33] S. White, A. Anandraj, F. Bux, PAM fluorometry as a tool to assess microalgal nutrient

stress and monitor cellular neutral lipids, Bioresour. Technol. 102 (2011) 1675–1682. doi:10.1016/j.biortech.2010.09.097.

[34] J. Mandel, Statistical methods in analytical chemistry, J Chem Educ. 26 (1949) 534–539. doi:10.1002/0471728411.

[35] R. Leardi, Chemometric Methods in Food Authentication, 2nd ed., Elsevier Inc., 2008. doi:10.1016/B978-0-12-814264-6.00017-7.

[36] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J.M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools, Anal. Bioanal. Chem. 409 (2017) 5891–5899. doi:10.1007/s00216-017-0517-1.

[37] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J.M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part II: modeling, validation, and applications, Anal. Bioanal. Chem. (2018) 1–14. doi:10.1007/s00216-018-1283-4.

[38] R. Ulber, C. Protsch, D. Solle, B. Hitzmann, B. Willke, R. Faurie, T. Scheper, Use of bioanalytical systems for the improvement of industrial tryptophan production, Eng. Life Sci. 1 (2001) 15–17. doi:10.1002/1618-2863(200107)1:1<15::AID-ELSC15>3.0.CO;2-1.

[39] P. Yu, M.Y. Low, W. Zhou, Development of a partial least squares-artificial neural network (PLS-ANN) hybrid model for the prediction of consumer liking scores of ready-to-drink green tea beverages, Food Res. Int. 103 (2018) 68–75. doi:10.1016/j.foodres.2017.10.015.

[40] M.I.B. Lin, W.A. Groves, A. Freivalds, E.G. Lee, M. Harper, Comparison of artificial neural network (ANN) and partial least squares (PLS) regression models for predicting respiratory ventilation: An exploratory study, Eur. J. Appl. Physiol. 112 (2012) 1603–1611. doi:10.1007/s00421-011-2118-6.

[41] A.A. Oliveira, C.F. Lipinski, E.B. Pereira, K.M. Honorio, P.R. Oliveira, K.C. Weber, R.A.F. Romero, A.G. de Sousa, A.B.F. da Silva, New consensus multivariate models based on PLS and ANN studies of sigma-1 receptor antagonists, J. Mol. Model. 23 (2017). doi:10.1007/s00894-017-3444-3.

[42] R. Boggia, M. Forina, P. Fossa, L. Mosti, Chemometric study and validation strategies in the structure-activity relationships of new cardiotonic agents, Quant. Struct. Relationships. 16 (1997) 201–213. doi:10.1002/qsar.19970160303.

[43] T.P. Ryan, Modern Regression Methods, John Wiley & Sons, Inc., 1997.

[44] C. Duchesne, 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.

[45] J. Workman Jr, The state of multivariate thinking for scientists in industry: 1980 – 2000, Chemom. Intell. Lab. Syst. 60 (2002) 13–23.

[46] X. Zeng, M.K. Danquah, X.D. Chen, Y. Lu, Microalgae bioengineering: From CO2 fixation to biofuel production, Renew. Sustain. Energy Rev. 15 (2011) 3252–3260. doi:10.1016/j.rser.2011.04.014.

[47] 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.

[48] C. Posten, Design principles of photo-bioreactors for cultivation of microalgae, Eng. Life Sci. 9 (2009) 165–177. doi:10.1002/elsc.200900003.

[49] S. Esposito, A. Cafiero, F. Giannino, S. Mazzoleni, M.M. Diano, A Monitoring, Modeling and Decision Support System (DSS) for a Microalgae Production Plant based on Internet of Things Structure, Procedia Comput. Sci. 113 (2017) 519–524. doi:10.1016/j.procs.2017.08.316.

[50] J.H. de Vree, R. Bosma, R. Wieggers, S. Gegic, M. Janssen, M.J. Barbosa, R.H. Wijffels,

Turbidostat operation of outdoor pilot-scale photobioreactors, Algal Res. 18 (2016) 198–208. doi:10.1016/J.ALGAL.2016.06.006.

[51] 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.

[52] 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.

[53] 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:https://doi.org/10.1016/j.seppur.2017.08.019.

[54] A. Ben-Amotz, Accumulation of metabolites by halotolerant algae and its industrial potential, Annu. Rev. Microbiol. 37 (1983) 95–119.

[55] 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.

[56] P.P. Lamers, M. Janssen, R.C.H. De Vos, R.J. Bino, R.H. Wijffels, Carotenoid and fatty acid metabolism in nitrogen-starved Dunaliella salina, a unicellular green microalga., J. Biotechnol. 162 (2012) 21–7. doi:10.1016/j.jbiotec.2012.04.018.

[57] A. McWilliams, The Global Market for Carotenoids, 2018. https://cdn2.hubspot.net/hubfs/308401/FOD Report Overviews/FOD025F Report Overview.pdf?t=1542722219756&utm_campaign=FOD025F&utm_source=hs_automation&utm_medium=email&utm_content=62915556&_hsenc=p2ANqtz--6ThLx3pCwfLg8LrjVhMwuMISkQ4otxISEyiLdpb3vuPxOUiyfUjg.

[58] K. Wichuk, S. Brynjólfsson, W. Fu, Biotechnological production of value-added carotenoids from microalgae: Emerging technology and prospects, Bioengineered. 5 (2014) 204–208. doi:10.4161/bioe.28720.

[59] E.T. Chua, P.M. Schenk, A biorefinery for Nannochloropsis: Induction, harvesting, and extraction of EPA-rich oil and high-value protein, Bioresour. Technol. 244 (2017) 1416–1424. doi:10.1016/J.BIORTECH.2017.05.124.

[60] R. Radakovits, R.E. Jinkerson, S.I. Fuerstenberg, H. Tae, R.E. Settlage, J.L. Boore, M.C. Posewitz, Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana, Nat. Commun. 3 (2012) 610–686. doi:10.1038/ncomms1688.

[61] T. Tonon, D. Harvey, T.R. Larson, I.A. Graham, Long chain polyunsaturated fatty acid production and partitioning to triacylglycerols in four microalgae, Phytochemistry. 61 (2002) 15–24. doi:10.1016/S0031-9422(02)00201-7.

[62] M.A. Borowitzka, High-value products from microalgae-their development and commercialisation, J. Appl. Phycol. 25 (2013) 743–756. doi:10.1007/s10811-013-9983-9.

[63] M. Fachet, R.J. Flassig, L. Rihko-Struckmann, K. Sundmacher, A dynamic growth model of Dunaliella salina: parameter identification and profile likelihood analysis., Bioresour. Technol. 173 (2014) 21–31. doi:10.1016/j.biortech.2014.08.124.

[64] U.K. Bageshwar, L. Premkumar, I. Gokhman, T. Savchenko, J.L. Sussman, A. Zamir, Natural protein engineering: a uniquely salt-tolerant, but not halophilic, alpha-type carbonic anhydrase from algae proliferating in low- to hyper-saline environments., Protein Eng. Des. Sel. 17 (2004) 191–200. doi:10.1093/protein/gzh022.

[65] M. Plaza, M. Herrero, A. Alejandro Cifuentes, E. Ibáñez, Innovative natural functional ingredients from microalgae, J. Agric. Food Chem. 57 (2009) 7159–7170. doi:10.1021/jf901070g.

[66] A.A. Tammam, E.M. Fakhry, M. El-sheekh, Effect of salt stress on antioxidant system and the metabolism of the reactive oxygen species in Dunaliella salina and Dunaliella tertiolecta, J. Biotechnol. 10 (2011) 3795–3808. doi:10.5897/AJB10.2392.

[67] L. Christenson, R. Sims, Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts., Biotechnol. Adv. 29 (2011) 686–702. doi:10.1016/j.biotechadv.2011.05.015.

[68] I. Havlik, P. Lindner, T. Scheper, K.F. Reardon, On-line monitoring of large cultivations of microalgae and cyanobacteria., Trends Biotechnol. 31 (2013) 406–14. doi:10.1016/j.tibtech.2013.04.005.

[69] 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.

[70] C.F. Galinha, G. Guglielmi, G. Carvalho, C. a M. Portugal, J.G. Crespo, M. a M. Reis, Development of a hybrid model strategy for monitoring membrane bioreactors., J. Biotechnol. 164 (2013) 386–95. doi:10.1016/j.jbiotec.2012.06.026.

[71] Y. Collos, F. Mornet, A. Sciandra, N. Waser, A. Larson, P.J. Harrison, An optical method for the rapid measurment of micromolar concentrations of nitrate in marine phytoplankton cultures, J. Appl. Phycol. 11 (1999) 179–184. doi:10.1023/A:1008046023487.

[72] C.A. Andersson, R. Bro, The N-way Toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (2000) 1–4.

[73] M. Ziegmann, M. Abert, M. Müller, F.H. Frimmel, Use of fluorescence fingerprints for the estimation of bloom formation and toxin production of Microcystis aeruginosa., Water Res. 44 (2010) 195–204. doi:10.1016/j.watres.2009.09.035.

[74] N. Her, G. Amy, H.R. Park, M. Song, Characterizing algogenic organic matter (AOM) and evaluating associated NF membrane fouling, Water Res. 38 (2004) 1427–1438. doi:10.1016/j.watres.2003.12.008.

[75] L. Moberg, G. Robertsson, B. Karlberg, Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis, Talanta. 54 (2001) 161–170. doi:10.1016/S0039-9140(00)00650-0.

[76] A. Mishra, B. Jha, Isolation and characterization of extracellular polymeric substances from micro-algae Dunaliella salina under salt stress., Bioresour. Technol. 100 (2009) 3382–6. doi:10.1016/j.biortech.2009.02.006.

[77] 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.

[78] 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.

[79] S. Pawlowski, C.F. Galinha, J.G. Crespo, S. Velizarov, 2D fluorescence spectroscopy for monitoring ion-exchange membrane based technologies – Reverse electrodialysis (RED), Water Res. 88 (2016) 184–198. doi:10.1016/j.watres.2015.10.010.

[80] J. Hu, D. Nagarajan, Q. Zhang, J. Chang, D. Lee, Heterotrophic cultivation of microalgae for pigment production : A review, Biotechnol. Adv. 36 (2018) 54–67. doi:10.1016/j.biotechadv.2017.09.009.

[81] A. Ben-Amotz, Industrial Production of Microalgal Cell-mass and Secondary Products – Major Industrial Species: Dunaliella, in: A. Richmond (Ed.), Handb. Microalgal Cult. Biotechnol. Appl. Phycol., Blackwell Science, Oxford, 2004.

[82] E.B. Young, J. Beardall, Photosynthetic function in Dunaliella tertiolecta (Chlorophyta) during a nitrogen starvation and recovery cycle, J. Phycol. 39 (2003) 897–905.

[83] B. Schoefs, Chlorophyll and carotenoid analysis in food products. Properties of the pigments and methods of analysis, Trends Food Sci. Technol. 13 (2002) 361–371. doi:10.1016/S0924-2244(02)00182-6.

[84] R.K. Saini, Y.S. Keum, Microbial platforms to produce commercially vital carotenoids at industrial scale: an updated review of critical issues, J. Ind. Microbiol. Biotechnol. (2018). doi:10.1007/s10295-018-2104-7.

[85] P.P. Lamers, M. Janssen, R.C.H. De Vos, R.J. Bino, R.H. Wijffels, Exploring and exploiting carotenoid accumulation in Dunaliella salina for cell-factory applications., Trends Biotechnol. 26 (2008) 631–8. doi:10.1016/j.tibtech.2008.07.002.

[86] S. Orset, A.J. Young, Low-temperature-induced synthesis of a-carotene in the microalga Dunaliella salina (Chlorophyta), J. Phycol. 35 (1999) 520–527.

[87] M. Gong, A. Bassi, Carotenoids from microalgae: A review of recent developments, Biotechnol. Adv. 34 (2016) 1396–1412. doi:10.1016/j.biotechadv.2016.10.005.

[88] K.-L. Leu, B.-D. Hsu, A programmed cell disintegration of Chlorella after heat stress, Plant Sci. 168 (2005) 145–152. doi:10.1016/j.plantsci.2004.07.026.

[89] H. Liechtenthaler, Chlorophylls and carotenoids: pigments of photosynthetic biomembranes, Methods Enzym. 148 (1987) 350–382.

[90] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. Mcnally, B.M. Monge-Sanz, J.J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.N. Thépaut, F. Vitart, The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc. 137 (2011) 553–597. doi:10.1002/qj.828.

[91] R. Myneni, Y. Knyazikhin, T. Park, MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [Data set], NASA EOSDIS L. Process. DAAC. (2015). doi:10.5067/MODIS/MCD15A3H.006.

[92] F.C. Petry, A.Z. Mercadante, Composition by LC-MS/MS of New Carotenoid Esters in Mango and Citrus, J. Agric. Food Chem. 64 (2016) 8207–8224. doi:10.1021/acs.jafc.6b03226.

[93] Y. Zhang, Z. Liu, J. Sun, C. Xue, X. Mao, Biotechnological production of zeaxanthin by microorganisms, Trends Food Sci. Technol. 71 (2018) 225–234. doi:10.1016/j.tifs.2017.11.006.

[94] A. Solovchenko, C. Aflalo, A. Lukyanov, S. Boussiba, Nondestructive monitoring of carotenogenesis in Haematococcus pluvialis via whole-cell optical density spectra, Appl. Microbiol. Biotechnol. 97 (2013) 4533–4541. doi:10.1007/s00253-012-4677-9.

[95] J. Chen, D. Wei, G. Pohnert, Rapid estimation of astaxanthin and the carotenoid-to-chlorophyll ratio in the green microalga Chromochloris zofingiensis using flow cytometry, Mar. Drugs. 15 (2017) 1–23. doi:10.3390/md15070231.

[96] D.M.M. Kleinegris, M.A. van Es, M. Janssen, W.A. Brandenburg, R.H. Wijffels, Carotenoid fluorescence in Dunaliella salina, J. Appl. Phycol. 22 (2010) 645–649. doi:10.1007/s10811-010-9505-y.

[97] A. Ben-Amotz, M. Avron, The Potential Use of Dunaliella for the Production of Glycerol, β-Carotene and High-Protein Feed, in: Biosaline Res., Springer, 1982: pp. 207–214. doi:10.1007/978-1-4899-4998-1.

[98] J. Milledge, S. Heaven, A review of the harvesting of micro-algae for biofuel production, Rev Env. Sci Biotechnol. 12 (2013) 165–178. doi:http://dx.doi.org/10. 1007/s11157-012-9301-z.

[99] R. Baker, Membrane technology and applications, Wiley Online Library, 2000. doi:http://dx.doi.org/10.1002/0471238961.1305130202011105.a01.

[100] J. Trivedi, M. Aila, D. Bangwal, S. Kaul, M. Garg, Algae based biorefinery – how to make sense?, Renew Sustain Energy Rev. 47 (2015) 295–307. doi:http://dx.doi.org/10.

1016/j.rser.2015.03.052.

[101] R. Radmer, B. Parker, Commercial applications of algae: opportunities and con- straints, J Appl Phycol. 6 (1994) 93–98.

[102] 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.

[103] 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.

[104] 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.

[105] J. Monte, M. Sá, C. Parreira, J. Galante, A.R. Serra, C. Galinha, L. Costa, V. Pereira, C. Brazinha, J. Crespo, Recycling of Dunaliella salina cultivation medium by integrated membrane filtration and advanced oxidation, Algal Res. 39 (101460 (2019).

[106] R. Bosma, E. van Zessen, J. 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.

[107] 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.

[108] M. Córdoba-Matson, J. Gutiérrez, M. Porta-Gándara, No TitleEvaluation 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.

[109] 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.

[110] 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.

[111] S. Wold, M. Sjostrom, 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.

[112] 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.

[113] Y.-H. Shin, J.Z. Barnett, E. Song, M.T. Gutierrez-Wing, K.A. Rusch, J.-W. Choi, A portable fluorescent sensor for on-site detection of microalgae, Microelectron. Eng. 144 (2015) 6–11. doi:10.1016/j.mee.2015.01.005.

[114] M. Mitra, S.K. Patidar, S. Mishra, Integrated process of two stage cultivation of Nannochloropsis sp. for nutraceutically valuable eicosapentaenoic acid along with biodiesel, Bioresour. Technol. 193 (2015) 363–369. doi:10.1016/j.biortech.2015.06.033.

[115] R.B. Draaisma, R.H. Wijffels, P. (Ellen) Slegers, L.B. Brentner, A. Roy, M.J. Barbosa, Food commodities from microalgae, Curr. Opin. Biotechnol. 24 (2013) 169–177. doi:10.1016/J.COPBIO.2012.09.012.

[116] X.N. Ma, T.P. Chen, B. Yang, J. Liu, F. Chen, Lipid production from Nannochloropsis, Mar. Drugs. 14 (2016). doi:10.3390/md14040061.

[117] Y. Li-Beisson, J.J. Thelen, E. Fedosejevs, J.L. Harwood, The lipid biochemistry of eukaryotic algae, Prog. Lipid Res. 74 (2019) 31–68. doi:10.1016/J.PLIPRES.2019.01.003.

[118] E.J. Baker, E.A. Miles, G.C. Burdge, P. Yaqoob, P.C. Calder, Metabolism and functional effects of plant-derived omega-3 fatty acids in humans, Prog. Lipid Res. 64 (2016) 30–56.

doi:10.1016/j.plipres.2016.07.002.

[119] World Health Organisation, Population nutrient intake goals for preventing diet-related chronic diseases, (n.d.). http://www.who.int/nutrition/topics/5_population_nutrient/en/index13.html (accessed June 4, 2019).

[120] A.R. Ribeiro, A. Gonçalves, M. Barbeiro, N. Bandarra, M.L. Nunes, M.L. Carvalho, J. Silva, J. Navalho, M.T. Dinis, T. Silva, J. Dias, Phaeodactylum tricornutum in finishing diets for gilthead seabream: effects on skin pigmentation, sensory properties and nutritional value, J. Appl. Phycol. 29 (2017) 1945–1956. doi:10.1007/s10811-017-1125-3.

[121] J. Janssen, R. Wijffels, M. Barbosa, Lipid Production in Nannochloropsis gaditana during Nitrogen Starvation, Biology (Basel). 8 (2019) 5. doi:10.3390/biology8010005.

[122] M.J. Caballero, M.S. Izquierdo, E. Kjørsvik, D. Montero, J. Socorro, A.J. Fernández, G. Rosenlund, Morphological aspects of intestinal cells from gilthead seabream (Sparus aurata) fed diets containing different lipid sources, Aquaculture. 225 (2003) 325–340. doi:10.1016/S0044-8486(03)00299-0.

[123] I.A. Nascimento, S.S.I. Marques, I.T.D. Cabanelas, G.C. de Carvalho, M.A. Nascimento, C.O. de Souza, J.I. Druzian, J. Hussain, W. Liao, Microalgae Versus Land Crops as Feedstock for Biodiesel: Productivity, Quality, and Standard Compliance, Bioenergy Res. 7 (2014) 1002–1013. doi:10.1007/s12155-014-9440-x.

[124] D. Simionato, M.A. Block, N. La Rocca, J. Jouhet, E. Maréchal, G. Finazzi, T. Morosinotto, The Response of Nannochloropsis gaditana to Nitrogen Starvation Includes De Novo Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus, Eukaryot. Cell. 12 (2013) 665–676. doi:10.1128/ec.00363-12.

[125] A.M.J. Kliphuis, A.J. Klok, D.E. Martens, P.P. Lamers, M. Janssen, R.H. Wijffels, Metabolic modeling of Chlamydomonas reinhardtii: Energy requirements for photoautotrophic growth and maintenance, J. Appl. Phycol. 24 (2012) 253–266. doi:10.1007/s10811-011-9674-3.

[126] 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) 1–11. doi:10.1186/1754-6834-7-70.

[127] G.M. León-Saiki, I.M. Remmers, D.E. Martens, P.P. Lamers, R.H. Wijffels, D. van der Veen, The role of starch as transient energy buffer in synchronized microalgal growth in Acutodesmus obliquus, Algal Res. 25 (2017) 160–167. doi:10.1016/j.algal.2017.05.018.

[128] V. Ördög, W.A. Stirk, P. Bálint, A.O. Aremu, A. Okem, C. Lovász, Z. Molnár, J. van Staden, Effect of temperature and nitrogen concentration on lipid productivity and fatty acid composition in three Chlorella strains, Algal Res. 16 (2016) 141–149. doi:10.1016/j.algal.2016.03.001.

[129] M.H.A. Michels, J. Camacho-Rodríguez, M.H. Vermuë, R.H. Wijffels, Effect of cooling in the night on the productivity and biochemical composition of Tetraselmis suecica, Algal Res. 6 (2014) 145–151. doi:10.1016/j.algal.2014.11.002.

[130] J. Fabregas, A. Maseda, A. Dominguez, M. Ferreira, A. Otero, Changes in the cell composition of the marine microalga, Nannochloropsis gbitana, during a light:dark cycle, Biotechnol. Lett. 24 (2002) 1699–1703.

[131] S.S. Nikaido, C.H. Johnson, Daily and Circadian Variation in Survival From Ultraviolet Radiation in Chlamydomonas reinhardtii, Photochem. Photobiol. 71 (2004) 758. doi:10.1562/0031-8655(2000)071<0758:dacvis>2.0.co;2.

[132] G. Benvenuti, P.P. Lamers, G. Breuer, R. Bosma, A. Cerar, R.H. Wijffels, M.J. Barbosa, Microalgal TAG production strategies: Why batch beats repeated-batch, Biotechnol. Biofuels. 9 (2016) 1–17. doi:10.1186/s13068-016-0475-4.

[133] H. Jiang, K. Gao, Effects of lowering temperature during culture on the production of polyunsaturated fatty acids in the marine diatom Phaeodactylum tricornutum (Bacillariophyceae), J. Phycol. 40 (2004) 651–654. doi:10.1111/j.1529-8817.2004.03112.x.

[134] Y. Wang, B. He, Z. Sun, Y.F. Chen, Chemically enhanced lipid production from microalgae under low sub-optimal temperature, Algal Res. 16 (2016) 20–27. doi:10.1016/j.algal.2016.02.022.

[135] T. Sakamoto, S. Higashi, H. Wada, N. Murata, D.A. Bryant, Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium Synechococcus sp. PCC 7002, FEMS Microbiol. Lett. 152 (1997) 313–320.

[136] K.W. Chew, J. Yap, P. Show, N. Suan, J. Juan, T. Ling, D. Lee, J. Chang, Bioresource Technology Microalgae biorefinery : High value products perspectives, Bioresour. Technol. 229 (2017) 53–62. doi:10.1016/j.biortech.2017.01.006.

[137] D.B. Rodrigues, C.R. Menezes, A.Z. Mercadante, E. Jacob-lopes, L.Q. Zepka, Bioactive pigments from microalgae Phormidium autumnale, FRIN. 77 (2015) 273–279. doi:10.1016/j.foodres.2015.04.027.

[138] R. Bro, Multiway calibration. Multilinear PLS, J. Chemom. 10 (1996) 47–61. doi:10.1002/(SICI)1099-128X(199601)10:1<47::AID-CEM400>3.0.CO;2-C.

[139] P. Filzmoser, B. Liebmann, K. Varmuza, Repeated double cross validation, J. Chemom. 23 (2009) 160–171. doi:10.1002/cem.1225.

[140] J.H. Janssen, J. Kastenhofer, J.A. de Hoop, P.P. Lamers, R.H. Wijffels, M.J. Barbosa, Effect of nitrogen addition on lipid productivity of nitrogen starved Nannochloropsis gaditana, Algal Res. 33 (2018) 125–132. doi:10.1016/J.ALGAL.2018.05.009.

[141] L. de Winter, A.J. Klok, M. Cuaresma Franco, M.J. Barbosa, R.H. Wijffels, The synchronized cell cycle of Neochloris oleoabundans and its influence on biomass composition under constant light conditions, Algal Res. 2 (2013) 313–320. doi:10.1016/J.ALGAL.2013.09.001.

[142] R. Braun, E.M. Farré, U. Schurr, S. Matsubara, Effects of light and circadian clock on growth and chlorophyll accumulation of N annochloropsis gaditana , J. Phycol. 50 (2014) 515–525. doi:10.1111/jpy.12177.

[143] Z. Dubinsky, N. Stambler, Photoacclimation processes in phytoplankton: Mechanisms, consequences, and applications, Aquat. Microb. Ecol. 56 (2009) 163–176. doi:10.3354/ame01345.