Analysis of Ecological Change in Marine Ecosystems via Novel Imaging and Spectroscopic Techniques on Microalgae Cells

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Morgan Breonne McConico-Lewis

May 2013

Dedication

To my Lord and Savior, Jesus Christ, who has given me the strength to pursue this degree. To my mom, I am forever grateful for your sacrifice and support.

Proverbs 16:3
“Commit thy works unto the Lord, and thy thoughts shall be established.”

Acknowledgements

This work was supported by the National Science Foundation under CHE-1058695. In addition, I would like to acknowledge the University Of Tennessee Chemistry Department for awarding me with the DuPont Diversity Fellowship as well as the Gordon Research Conference for awarding me with the Carol Storm Underrepresented Minority Fellowship which allowed me to travel and present my research. I would like to thank them all for their support it is greatly appreciated.

Abstract

With the increase in anthropogenic pollutants such as: atmospheric CO_2 [carbon dioxide], fossil fuel combustion and fertilizer run off, the fate of the environment is of great importance. Through industrialization, the greenhouse gas, CO_2, is a major contributor to global warming; while nitrogen based fertilizer, causes the production of toxins due to harmful algal blooms. Both chemical impacts dissolve into marine ecosystems effecting the chemical composition of organisms such as microalgae. Microalgae remediate these changes by sequestering carbon and nitrogen; thus there is a need to understand how these impacts affect the ecosystem by detecting the change and relating it to its chemical source. It was demonstrated that microalgae respond sensitively to changes in their environment, the present dissertation expands the utilization of phytoplankton as an *in-situ* biological probe.

The *goal* of this dissertation is to develop analytical methodologies for investigations of chemical impacts on marine ecosystems via microalgae for examining the relationship between ambient chemical changes and microalgae remediation processes. Thus, experiments performed simulated the chemical shifts of selected chemical parameters. Impacts are measured by means of microscopic imaging and FT-IR spectroscopy for determining physical and chemical responses of microalgae. In order to quantitate these processes, novel chemometric algorithms were developed for determining ambient conditions based on imaging and spectroscopic data.

The first step to understanding these effects was developing an innovative image analysis technique to measure cell size and shape as a function of species and ambient conditions. This technique discriminated microalgae cells based on their physical parameters. For chemical analyses, FT-IR spectroscopy was used for detecting species chemically induced shifts in microalgae's chemical composition. Within mixture species spectrum new chemical information was found and the suppression of chemical features was also revealed. Results demonstrated chemical change is present in the biomass and are nutrient source and concentration dependent. These novel analytical methodologies developed have demonstrated the ability to detect chemical changes under varying environmental conditions.

Table of Contents

1. Link Between Microalgae and Analytical Chemistry: Microalgae as an *in situ* Sensor 1
   1.1. Why Chemical Analyses of Microalgae? 1
   1.2. Utilization of Microalgae as a Biological Probe 2
   1.3. Microalgae Characteristics 7
   1.4. Dissertation Overview 11
2. Experimental Techniques for Assessing Impacts of Environmental Change of Microalgae 13
   2.1. Introduction 13
   2.2. Microalgae Culturing Methods 13
   2.3. Nutrient Induced Changes Imposed on Microalgae's Environment 25
   2.4. Cellular Image Acquisition 26
   2.5. Harvesting Microalgae Cells for Spectroscopic Analyses 28
   2.6. Summary 30
3. Investigation of Microalgae's Physical Parameters 31
   3.1. Exploration of Chemical Impacts on Cell Cultures 31
   3.1.1. Extracting Cellular Size and Shape Information 33
   3.1.2. Qualitative Assessments 36
   3.1.3. Quantitative Predictions of Ambient Chemical Parameters 38
   3.1.4. Modeling Environmental Effects on Cells' Physical Parameters 41
   3.2. Results of Monitoring Cell Size and Shape 43
   3.2.1. Assessing the Fit Quality 44
   3.2.2. Discrimination of Cell Species and Chemical Environments 48
   3.2.3. Concentration Prediction via Cell Size and Shape 55
   3.2.4. Analysis of Concentration-dependent Fitting 59
   3.3. Modeling Nutrient Impacts of Cell Growth and Size 68
   3.3.1. Determining the Concentration Dependency of Growth Curves 70
   3.3.2. Depicting Cell Size Concentration Dependency 72

3.3.3. Polynomial Order Selection 73
3.4. Effects of Nutrients on Growth Rate vs. Cell Size 74
4. Chemical Impacts of Microalgae Mixtures Studied by FT-IR Analysis 87
4.1. Assessment of Chemical Composition of Microalgae Cell Mixtures 89
4.2. Results 101
4.3. Conclusions 105
5. Summary and Conclusions 106
List of References 108
Appendices 115
Appendix 1 116
Appendix 2 118
Appendix 3 123
Appendix 4: 125
Vita 127

List of Figures

Figure 1. Inoculations of each species in the growth cabinet. From left to right: *Dunaliella salina, Nannochloropsis oculata* and *Dunaliella parva*. 20

Figure 2. Each microalgae species on the hemocytometer in 0.25 nL cells at their exponential phase: (A) *Dunaliella salina,* (B) *Nannochloropsis oculata,* (C) *Dunaliella parva.* 22

Figure 3. *Dunaliella salina* (A) 0.35 mM NO_3^-, (B) 0.873 mM NH_4^+, *Nannochloropsis oculata* (C) 0.873 mM NH_4^+, (D) 0.873 mM NO_3^-, *Dunaliella parva*(A) 0.16 mM NO_3^-, (B) 0.549 mM NH_4^+. .24

Figure 4. Microscope images (10X magnification, 380 x 300 pixels) of cultures of three different algae species; the cells were still dispensed in ESAW - (left) *Dunaliella parva*, (center) *Nannochloropsis oculata*, and (right) *Dunaliella salina*; the visible illumination inhomogeneity can be assumed to be homogeneous across 20 x 20 pixel subimages (cp. Figure 6 (A) - (C)). 27

Figure 5. (A) dried algae sample, (B), algae and KBr mixture at a 0.6 weight percent and (C) pelleted algae and KBR samples. 29

Figure 6. (A) – (C) gray-scale subimages (20 x 20 pixels, raw data) from top-to-bottom *Dunaliella parva*, *Nannochloropsis oculata*, *Dunaliella salina* cultured under 0.549 mM nitrate as N-source and 2.071 mM bicarbonate as C-source; (D) – (F) same as left column but in 3D-representation; (G) – (I) Gaussian fits (6) of the middle column's raw images. 45

Figure 7. The overall fit quality was assessed via correlation coefficients between subimages (Figure 6 (D) – (F)) and their fits (panels (G) – (I)).....47

Figure 8. (A) Species discrimination based on cell size distributions – different species feature histogram maxima at different size. Panel (B), (C), (D) comparing species pairwise by means of t-tests applied to the individual histogram bins; see text – the nutrient situation chosen for this and the following graph are the standard conditions [88]. 49

Figure 9. same analyses as in Figure 2 but applied to shape histograms for assessing the species' different cell shapes (4); 0.549 mM NO_3^- and 2.071 mM HCO_3^- have been used to culture these cells. 51

Figure 10. Analyses of nutrient concentrations c on *D. salina*'s shape histogram. For different nutrients and their concentrations c, the shape histograms have been fitted by (17) (cf. panel (A) in Figure 9); the resulting fit parameters U and k have been plotted over c of the respective nutrient. Not shown are $x0$ and $a0$ as they only shift the histograms but have no influence on the histograms' decay properties. 54

Figure 11. Predictor Surfaces for size (3) histograms obtained from microalgae grown under varying nutrient concentrations; these Predictor Surfaces depict how the histograms change as a function of nutrient availability. *Dunaliella salina*: (A) carbon, (B) ammonium, (C) nitrate; *Nannochloropsis oculata*: (D) carbon, (E) ammonium, (F) nitrate; *Dunaliella parva*: (G) carbon, (H) ammonium, (I) nitrate; the Size-axes had to be given here in 'bin #' but they cover the same 'pixel-units' as panel (A) in Figure 8. 56

Figure 12. Predicted versus true nutrient concentration utilizing Predictor Surfaces (Figure 11) of polynomial order $P = 2$ (9) compared to PLS predictions; the stated error percentages are averaged absolute values of relative errors. 58

Figure 13. Cell Size fit results obtained with $P = 2, QS = 2$; **(i) – (III)** $S0c1, c2$, $Ac1, c2$, and $\sigma c1, c2$ (14), (15) for *Dunaliella salina* grown under varying [C] and [NH_4^+]; **(IV)** $S0c1, c2$, for *Dunaliella*

salina grown under varying [C] and [NO_3^-]; **(V)** same as (IV) but for $P = 1, QS = 1$; **(VI)** and **(IX)** $S0c1, c2$ for *Nannochloropsis oculata* grown under varying [C] and [NH_4^+] as well as under varying [C] and [NO_3^-]; **(VII)** and **(VIII)** $S0c1, c2$ for *Dunaliella parva* grown under varying [C] and [NH_4^+] as well as under varying [C] and [NO_3^-]. 61

Figure 14. (top row) Correlation coefficients obtained for *Dunaliella salina* for different combinations of P and QS; (middle row) *Nannochloropsis oculata*; (bottom row) *Dunaliella parva*. 62

Figure 15. Comparing the fitted model $Y\mathrm{size}S, \mathbf{c}$ (14) to the measured size S histograms obtained from *Dunaliella salina*, nutrients: C and NH_4^+ with $P = 2, QS = 2$; as described in [40], the sizes of the cell were determined from imaging data and thus are measured in pixels. The error bars have been obtained from images of independently prepared replicate samples – their values are due to the strong but natural in-class variability of these biological samples. 64

Figure 16. Same as Figure 15 but for nutrients C and NO_3^-; the left panel shows two fits with lower but acceptable quality. The fit quality of the examples in the right panel is below a desirable level. 65

Figure 17. Surfaces $S0c1, c2$ (15) for *Dunaliella salina* grown under varying concentrations of C and NH_4^+ as obtained with different polynomial orders; top to bottom $P = 0, \ldots, 3$; left to right $V = 0, \ldots, 2$; the center right graph equals Figure 13 panel (I). 67

Figure 18. (top row) *Dunaliella salina* grown under varying carbon concentrations - (left column) Growth rate data, (right column) Cell size data [40]; (A) growth rate surface $Y\mathrm{count}t, c$ (18), with $P = 1$; (B) size surface $Y\mathrm{size}S, c$ (20) with $P = 1$ and $Q = QS = 0$ after subtracting the polynomial background surface which is here simply $\alpha 0$ because $Q = QS = 0$; (C) 'predicted', i.e. computed from $Y\mathrm{count}$, versus measured, i.e. 'true', cell concentration - see text; (D) predicted, normalized cell counts versus true counts - see text. 75

Figure 19. Mean 'prediction errors' for growth curves, i.e. $yt, c - Y\mathrm{count}t, c$, obtained for the three species cultured under varying concentrations of three nutrients; the most accurate model (polynomial order P) has been indicated by a (*); models not represented here did not result contain acceptable model parameters (22). 77

Figure 20. Mean 'prediction errors' for size histograms, i.e. $yS, c - Y\mathrm{size}S, c$, obtained for the three species cultured under varying concentrations of three nutrients; the most accurate model (polynomial order combination P, QS) has been indicated by a (*) 78

Figure 21. Growth surfaces $Y\mathrm{count}t, c$ which describe the experimental data best; see models marked by (*) in Figure 19. 80

Figure 22. Size histogram surfaces $Y\mathrm{size}S, c$ after subtracting the background polynomial surface which describe the experimental data best; see models marked by (*) Figure 20. 81

Figure 23. Growth rates sc as computed from models $Y\mathrm{count}t, c$ (18) such as shown in Figure 18 (A); the legends for the most accurate models (cp. Figure 19) were underlined and printed in bold face. 83

Figure 24. Cell sizes $S0c$ as computed from models $Y\mathrm{size}S, c$ (14) such as shown in Figure 18 (B); the legends for the most accurate models (cp. Figure 20) were underlined and printed in bold face. 84

Figure 25. Cell size $S0c$ versus growth rate sc for the best models (Figure 20 and Figure 19); the direction of increasing nutrient concentration c is indicated by arrows. 86

Figure 26. Replicate spectra obtained from 6.72 mM HCO_3^- are averaged to obtain mean spectrum with error bars within the wavenumber region of 3500-2800 cm^{-1} and within the fingerprint region from 1800-950 cm^{-1} [100]. The region of 2800-2200 cm^{-1} has been cut out due to atmospheric CO_2 as well as that region did not contain any pertinent chemical information. (A) and (B) mean spectra and standard deviations. (C) Result of t-test of Species #1 vs. Mixture #2 (D) Result of t-test between Species #3 vs. Mixture #2. 92

Figure 27. (A) Averaged spectra from Species #1, Species #3 and Mixture #2. Resultant comparison test depicts the chemical contributions contained within Mixture #2 at each wavenumber position: (B) Zoomed in from 3500 – 2800 cm^{-1} (C) Zoomed in from 1800 – 950 cm^{-1} 95

Figure 28. 2.07 mM HCO_3^- analysis of: (A) Species #2 and Mixture #3 mean spectra (B) Species #3 and Mixture #3 mean spectra (C) Results of comparison test 97

Figure 29. All comparison test results depicted for Mixture #2 displaying all Situations at varying carbon concentrations from 0.16 – 8.26 mM HCO_3^-. The data represents the chemical information present in Mixture#2 between 3500 – 950 cm^{-1}. 100

Figure 30. Concentration differences depicted in comparison test results of all three nutrient sources. 104

List of Tables

Table 1. Reagents used to make Salt Solution I for ESAW medium [9]. 16

Table 2. Reagents used to make Salt Solution II in ESAW [9]. 17

Table 3. Nutrient stock solutions for ESAW medium [9]. 18

Table 4. Species reference numbers 91

Table 5. Nine situation outcomes for Mixture spectrum depicted in Figure 27. This analysis is performed at each wavenumber position. 94

Table 6. FT-IR spectroscopy band assignments [53], [59], [95], [143] and [144]..... 98

1. Link Between Microalgae and Analytical Chemistry: Microalgae as an *in situ* Sensor

1.1. Why Chemical Analyses of Microalgae?

Marine ecosystems are experiencing rapid changes in comparison to its geological history [1]. Some of these changes are attributed to anthropogenic chemical impacts; thus there is a need for detailed assessment of the chemical and physical implications on the biological environment.

Due to the increase in industrialization, production of anthropogenic CO_2 is increasing [2] and the impact of this greenhouse gas is becoming a serious concern [3], [4].

The majority of CO_2 entering the atmosphere dissolves in oceans, increasing carbon dioxide and bicarbonate concentration levels [5] and lowering the pH [6]; thus altering the chemical composition of organisms within the ecosystem [1].

Phytoplankton, microalgae, are one of the major consumers of carbon dioxide and producers of oxygen, with a habitat covering more than 71% of the Earth's surface [7], [8]. Microalgae contribute about 40 - 50% of the oxygen in the atmosphere [1], [7], [9], by photosynthesis [10] and absorbs over one hundred million tons of CO_2 per day [11]. More oxygen is produced by algae compared to all forests combined. Conversely, microalgae also utilize atmospheric CO_2 for biomass production and thus their carbon sequestering counter-balances CO_2 emissions [11] - [17]. Understanding phytoplankton's part in sequestering anthropogenic CO_2 is particularly important because of the indications that the ocean's carbon storage capacity has begun to diminish [18].

Human activities involving nitrogen such as, fossil fuel combustion and fertilizer production, cause drastic changes in marine ecosystems. For example, mineral fertilization has a negative effect on the environment by killing off species or diminishing habitat quality [19]. Harmful algal blooms are also of great concern due to their potential for the production of toxins [20], [21]. The largest single source of nitrogen pollution in some regions is due to atmospheric deposition from fossil fuel combustion [22]. Nitrogen pollution affects cells by reducing the

protein content and enhancing lipid and carbohydrate storage. This demonstrates how the biochemical composition of microalgae responds to environmental conditions [23].

To study such anthropogenic impacts, intracellular biochemical processes were investigated [24] - [30] to understand phytoplankton's sequestering of carbon and other inorganic compounds. Prior studies demonstrated [30] - [35] that microalgae respond to changing environmental conditions [13], [15], [36] - [43] by altering chemical composition. Because these adaptations impact the chemical composition of phytoplankton, a more detailed understanding of these changes is necessary; which requires advancement in chemical sensing methodologies.

The *goals* of this research are to utilize physical and chemical changes in microalgae cells as an innovative indicator for assessing the chemical state of marine ecosystems. These studies will give insight into how the ecological community is affected by nutrient induced changes. Novel models for the analysis of chemical effects on marine ecosystems have been developed. This combination of analytical and statistical analyses facilitates a deeper understanding of how chemical changes affect the ecological system.

1.2. Utilization of Microalgae as a Biological Probe

An introduction of microalgae use as a biological probe is presented. One analytical tool applied was Laser Scanning Cytometry (LSC) as an image analysis method to measure algae cells based on cell size and shape. Vardon et al. determined that LSC was able to measure the shape and size of the *Spirulina platensis* and *Heterocapsa pygmaea* cell over a large concentration range [44].

Biological effects are multifaceted and require analytical techniques to provide information at cellular and sub-cellular levels. Nano-scale secondary-ion mass spectrometry (Nano-SIMS) imaging detects the copper-ion distribution within microalgae cells (*Chlorella* kessleri). Nano-SIMS ion microprobe determines ion distribution, detects trace elements and yields information regarding the dynamics of different biological processes of single cells. These cells were

exposed to micromolar and nanomolar concentrations of copper, nitrogen and phosphorus. $^{63}Cu^-$, $^{12}C^{14}N^-$, and $^{31}P^-$ ions maps were captured and analyzed to establish a correlation of $^{63}Cu^-$ ion maps to maps of $^{12}C^{14}N^-$, and $^{31}P^-$. The experiment provided imaging analysis of single, microalgae cells and their copper distribution [45].

X-ray techniques measure bioaccumulation of nutrients in microalgae. For example, particle-induced X-ray emission (PIXE) was used for studying the bioaccumulation of Zn^{2+} within microalgae. Iwata cultured *Nannochloropsis sp.* with known amounts of Zn^{2+}, to determine the bioaccumulation of Zn^{2+}. Zn^{2+} concentrations increased within the *N. sp.* for six hours until stabilizing. The concentration of nutrient uptake within microalgae cells was determined [46].

Cesareo et al analyzed microalgae to investigate their usefulness in dietary supplements and vitamins. Energy-dispersive x-ray fluorescence was used for the analysis of *Arthospira maxima* and *Chlorella,* and the identification of several trace elements such as: Zn, Ca, Sr, K, Ni, Cu, Mn, and Fe. Although the trace elements were able to be determined, there were several sources of error, such as sample preparation as well as the instability of the x-ray source [47].

Stokes Raman scattering spectra of nitrogen starved and healthy algae cells of *Chlorella sorokiniana* and *Neochloris oleoabundans* were analyzed. Chemical composition analysis of algae was performed by mapping dense locations of the algal sample over small designated areas. Key peaks were identified from the Raman image scans of single algae cells or a group of algal cells. Chemical composition analysis was performed in order to locate triglyceride in nitrogen starved *N. oleoabundans* cells along with identifying healthy *N. oleoabundans* cells. Stokes Raman scattering enabled fast, non-intrusive qualitative analysis of the algal composition. However, the strong fluorescence overpowers peaks of interest, there by limiting Raman usefulness [48].

For nutrient prediction of *Dunaliella tertiolecta* cells, *in-vivo* analysis was performed using Raman microscopy. A 782 nm excitation laser was used and the resulting microalgae spectra contained characteristic signatures of β-carotene and chlorophyll *a*. *Dunaliella tertiolecta* showed broad features such as sloping baseline from fluorescence, a range of spectral intensity and a baseline offset, which was due to the light illumination from the laser power of the instrument. The relative abundances of the β-carotene and chlorophyll differed under limiting the nitrogen source; the band of the β-carotene became more intense and the chlorophyll band became weaker. Spectra from nitrogen starved and nitrogen replete changed. Distributions were significant enough that partial least squares (PLS) [49], [50] predicted the change of

nutrient abundances from Raman spectra. The benefit of this analysis is that the population variability is able to be analyzed along with assessing the nutrient response that occurs within the microalgae [51].

A complementary technique to Raman spectroscopy is Fourier Transform Infrared (FT-IR) spectroscopy. It has been shown that FT-IR spectroscopy is a promising tool for chemical investigations of microalgae [52], [53] - [56] because most macromolecules within the microalgae (i.e.: lipids, nucleic acids, carbohydrates and proteins) are infrared active. When FT-IR analysis is performed on microalgae cells, the spectra reflect the biochemical composition of the microalgae sample [57]. Although different species have similar chemical compositions, small differences are reflected in the IR signatures which facilitate their discrimination [56]. Microalgae cells have been observed to adapt their chemical composition to environmental changes. For instance, algae chemical composition is influenced by the nitrogen source (nutrient), light, and atmospheric CO_2 levels. Such changes were seen in the mid-IR spectra of three independent replicates of *Chlorella marina* grown under two different nitrogen conditions. Cell cultures grown under the same conditions showed much smaller sample-to-sample variations in comparison to changing the nitrogen source from NH_4^+ to NO_3^- or vice versa [53], [56].

The nutrient status of four microalgae species, *Sphaerocystis schroeteri, Phormidium luridum var olivace, Scenedesmus quadricaud* and *Nitzchia* sp.*,* was determined using FT-IR and nutrient induced fluorescence transient (NIFT). The four species were grown under P-replete or P-limited conditions. The FT-IR was combined with a microspectrometer (FT-IR MS) in order to obtain spectral information of single algae cells. The FT-IR MS was then coupled with NIFT for detecting the changes within microalgae due to phosphate (PO_4^{3-}) exposure. The results for the NIFT response were not as advantageous as the FT-IR. Starting cultures with limited nutrients produced very low responses with NIFT. The FT-IR MS showed that differences within the spectra of the four microalgae cells [57].

Hirschmugl et al used FT-IR micro-spectroscopy in conjunction with a high-intensity synchrotron IR source for the analysis of all major cell pools (lipids, sugars and amides) in *Euglena gracilis*. A synchrotron is a very useful tool for such analyses, because it is a very intense, non-destructive IR light source that results in exceptional signal-to-noise. A synchrotron enhances small spectroscopic details for studying the effects of nutrient induced changes in algae cells. Algae cells were analyzed under nutrient replete and deplete conditions. The results

demonstrated that nutrient replete cells show a larger protein to carbohydrate ratio in comparison to nutrient deplete cells. This technique gave way to quick studies of microalgae cells' chemical composition in their growth environment [52].

Multivariate data analysis (principal component analysis, PCA) was used to predict individual independent colonies of fixed *Scenedesmus quadricauda* cells' nutrient status via FT-IR microscope data. P-replete species had a higher protein and lower carbohydrate level when compared to P-depleted cells. FT-IR spectromicroscopy measured the variability of the changes of nutrients in microalgae cells under different phosphorus conditions. PCA enabled the prediction of the nutrient status of the colonies based off of the FT-IR spectra [58].

Murdock et al measured single cell nutrient supplies of *Achnanthes affinis* and *Fragilaria virescens* using Infrared microspectroscopy (IMS) with chemical imaging. A synchrotron was combined with IMS to study ecological interactions on the microscale. This dual technique measured nutrient distributions of macromolecular variation in individual cells as a function of varying chemical distributions [59].

The microalgae species *Chaetoceros muellerii* was analyzed using Fourier transform infrared spectroscopy to determine responses in the presence of changes in nitrogen (nitrate or ammonium) sources. When grown under either depleted NH_4^+ or NO_3^- spectra were similar except for variances in the amide I and II peaks. When algal species were grown in the presence of ammonium, there was a 20% - 30% increase in the amide I and amide II peaks compared to nitrate grown cells. When cells were grown under a nitrogen enriched environment, there was an increase in the levels of intensity for both amide bands. The lipid band under ammonium decreased by 50% and under nitrate it decreased by 65%. Through the utilization of FT-IR, determination of relative levels of biomolecules were obtained in microalgae and their responses to changing environmental conditions [53].

Microalgae were studied via FT-IR spectroscopy combined with a novel non-linear chemometric method called, Predictor Surfaces, to obtain a better understanding of non-linear systems such as marine ecosystems through the use of microalgae cells.

Predictor Surfaces were developed because linear multivariate systems are inadequate for extraction of qualitative and quantitative information in biologically complex systems. Predictor Surfaces stem from multivariate Taylor expansions for approximation of nonlinear model functions. For this study, three microalgae species were cultured with ammonium, nitrate and

inorganic carbon under varying concentrations. A non-linear dependency between environmental conditions and the microalgae's spectroscopic signatures was demonstrated. Also, quantitative predictions were made using principle component regression (PCR) and Predictor Surfaces. The Predictor Surfaces were determined to quantitatively predict concentrations more accurately than PCR. The Predictor Surface method was found to be a novel tool for quantitative analyses of complicated non-linear biological systems. The combination of FT-IR spectroscopy with Predictor Surfaces was able to prove that microalgae are capable of being utilized as an innovative biological tool for ecological investigations [39].

Changes in the chemical composition of selected microalgae cells have been monitored by means of a novel image analysis technique coupled with vibrational spectroscopy [60], [61]. However, there is a lack of methodologies for determining how microalgae respond physically to changing environmental conditions as well as how cells respond chemically when grown within mixtures. Hence, there is a need for a technique that is simple and does not require extensive sample/ sensor preparations for extracting physical parameters. Imaging is the technique of choice due to its ability to analyze numerous cells and thus gains a culture representative insight rather than a potentially nonrepresentative single-cell response. A novel image analysis method has been developed to measure cell size and shape of a large number of microalgae cells as a function of species and ambient conditions. The hypothesis for this study is that cells' physical appearance [5], [14], [62] - [64] reflects cell type and/or chemical impacts imposed onto them by the environment. Image analysis provides a simple straightforward study of: (1) Cell discrimination based on physical appearance such as size and shape and (2) Building predictive models that relate the measured cell appearance to chemical parameters in the environment.

While the aforementioned investigations addressed the responses of different species to ambient conditions or gained an overview of microalgal biodiversity, it has not yet been investigated how multiple, coexisting species chemically interact for nutrients. Although different species have similar chemical compositions, small differences (i.e. species or nutrients) have been reflected in the IR-spectra of single species which facilitated the discrimination [65]. An analytical methodology has been developed via FT-IR spectroscopy for detecting chemical interactions in microalgae's chemical composition. The hypothesis is that different microalgae species within the ecosystem chemically interact for nutrients and these chemical processes effect the chemical composition of the biomass. Thus, microalgae's chemical response to changing conditions has been studied.

1.3. Microalgae Characteristics

Algal belong to a diverse group of eukaryotic and prokaryotic organisms from the Eucarya genus [7], [66]. Eukaryotic cells contain organelles (nuclei, flagella, mitochondria, Golgi bodies, and, plastids) that control the functionality of the cell. Prokaryotic cells, unlike eukaryotic cells, lack organelles and are more similar to bacteria than to algae [5]. Species are either multicellular or single celled and microscopic or macroscopic in size [7], [8], [66]. One difference between the two algal strains is that macroscopic algae are multicellular and usually attach to rocks. Whereas microscopic algae can be singled celled organisms, that either float freely or move by flagella. Since some microalgae are a single celled species, they often reproduce asexually by binary fission. Binary fission is the process of the cell duplicating its deoxyribonucleic acid (DNA) and then splitting into two cells. On other occasions, sexual reproduction can occur by a process called meiosis [7] occuring when the chromosomes are reduced from two to one. Algae species also differ in shapes and sizes ranging from 0.5 μm to over 50 meters long [5], [7], [63], [66].

While algae have several differences there are several similarities such as they are categorized by their ability to acclimate to a large range of environmental conditions [66]. Another similarity is that algae have rigid cell walls that are composed of cellulose and contain a membrane bound nucleus. They also contain organelles in their cytoplasm such as mitochondria and chloroplast. All species have characteristic pigments which are due to chloroplast containing chlorophyll.

Specifically all microalgae contain chlorophyll *a*, which is necessary for photosynthesis [7], [67] and crucial to their continued existence [5]. Those pigments absorb light at different wavelengths and the color is based off of the reflected wavelength. The chlorophyll absorbs the photons and emits a pigment. For example, green algae, absorbs blue light (430nm) and red light (655nm) which in turn transmits the green color [68]. Photosynthesis utilizes sunlight which provides the energy to convert CO_2 and H_2O into oxygen and glucose [7], [69]. The photosynthetic process involves two steps, 1) capturing light and 2) converting the light into O_2 and adenosine triphosphate (ATP), chemical energy. ATP is a compound used by cells for the

storage of energy. This energy is used in respiration in order to aid in growth [5] allowing microalgae to increase their mass [66].

The term algae are not a strict term of classification, but organisms under algae are grouped in order to identify the different properties. As stated before, all algal species contain green chlorophyll, chlorophyll *a*, but some are hidden due to the photosynthetic pigments which is what breaks them up into their separate divisions [7], [8]. Microalgae are classified in six different Divisions composed of several thousands of species [70]. Each Division (phyla) is classified based off its physiological characteristics. The Blue Green algae also known as, Division Cyanopyta, Cyanobacteria differs from the other divisions due to its prokaryotic nature. This species is microscopic and are characterized by their lack of: a defined nucleus, sexual reproduction and chloroplasts [71]. The second group is Division Chrysophyta, called yellow-green algae. This group also can be found in both marine and fresh waters. Their cell wall is composed of pectose with silicates utilizing both asexual and sexual reproduction. Division Rhodopyta, red Algae, is the third division that mostly contains algae found in marine waters including approximately twelve other species that have habitats in brackish and fresh waters. These plants vary in size from the microscopic to macroscopic level; where some have the ability to reach up to a meter in length. Their cells are composed of cellulose and pectin and reproduction is asexually and sexually. Brown algae, Division Phaeophyta, are vary in size. Their cell walls are mainly composed of cellulose and pectin and reproduction is either sexual or asexual [7], [71]. The fifth division is Division Euglenophyta or Euglenoids. This group is unicellular and exists mainly in fresh waters and contains chlorophyll *a* and chlorophyll *b*. Euglenoids are the only group that are absent of a cell wall not containing any cellulose or pectin. They only reproduce asexually and by binary fission. The next group is Division Pyrrhophyta known as Dinoflagellates which contain many species that are very diverse in their color. They are found in salt water as well as fresh water and are unicellular organisms. Dinoflagellates mainly reproduce asexually and have a cell wall that is composed of cellulose and pectin. The final division is Chlorophyta also known as Green Algae. This group includes plants that contain both fresh and marine water species, but can also be found in a terrestrial environment. This group's cell wall is also composed of cellulose and pectin. In addition to chlorophyll *a*, it also contains chlorophyll *b* for its pigment; which is similar to the Euglenophyta division. This group reproduces asexually and by multiple fission. The species, *Dunaliella salina*, *Nannochloropsis oculata* and *Dunaliella salina*, which are used for analysis within this dissertation, are found in this division.

Each species was carefully chosen for analysis due to its ease of culturing [72]. The *Dunaliella* genus is characterized by being unicellular [73], [74] and its ability to grow in high salinity levels [75]. The cells vary in shape from spherical to ellipsoid. *Dunaliella parva* have green or orange cells that range from cylindrical, spherical and elliptical in shape and 4 - 10 µm wide and 9 - 16 µm long; while, *Dunaliella salina* have either green or red cells and are spherical, oval or ellipsoidal in shape with a cell width of 3.8 - 20.3 µm and a cell length of 5 - 29 µm. *Dunaliella parva* is discriminated from *Dunaliella salina* by their size with *D.* salina being the larger of the two species [72]. *Nannochloropsis oculata* is a unicellular green algal strain [46] that are characterized by their small size of 2 - 5 µm and its spherical or oval shape and can be found in brackish, fresh and marine bodies of waters [76], [77]. Similar to the two *Dunaliella* species, *N. oculata* is able to grow in a wide range of salinities.

In their natural environment, microalgae are characteristically found close to the surface of fresh or salt water. But, in order to analyze microalgae for manufacturing purposes, it must be grown commercially [78]. There are two examples of commercial cultivation: open air ponds and photobioreactors (PBR) where both are maintained under close observation. Open ponds have been used since the 1950's [5] and are categorized as ponds, lakes, natural waters and containers [79]. The advantages to utilizing open pond systems are that the cost is minimal as well as being relatively easy to construct and maintain. The drawbacks are the lack of a controlled environment which can cause growth rates to fluctuate [5], [66], [80], the loss of CO_2 into the atmosphere, evaporation along with the potential of pollution and contamination [79]. These systems are not as resourceful for biomass production [66] in comparison to photobioreactors [5]. PBRs differ by utilizing a controlled, closed system that creates an optimal environment of temperatures, carbon dioxide, pH levels and light for the mass production of algae. This system is costly, but important in order to be able to control the temperature of the system for the algae to grow because the optimal growth temperature for microalgae is between 20 − 30 °C. If temperatures increase above 35 °C, this is generally detrimental to the species; whereas if temperatures are lower than 16 °C microalgae's growth rate will diminish. The consistency of the algae growth cycle is essential for manufacturing [8]. For this purpose, PBRs are more common in both the commercial and scientific field due to their ability to provide a controlled environment and high productivity [81]. Open pond cultivation has been utilized as the main cultivation method for years. Nevertheless, photobioreactors are now becoming more popular [79]. Microalgae have been proven to respond quickly to environmental change [39] by adapting their metabolism to various environments [78]. The

ability to survive under extreme conditions combined with their production of secondary metabolites makes microalgae very valuable to the food and health industries. Because microalgae is quite promising for the production of human food, cosmetics, medicines and biofuels [80], it is vital to obtain a consistent growth rates to ensure microalgae have the ability to thrive.

Operating under these characteristics, microalgae have become an area of focus for many uses. Algae contain important chemical constituents that are used in pharmaceuticals and cosmetics [66], [79], [80]. Due to pharmacologically active compounds found in microalgae, over the past decade there have been several leads to their use as antibiotics. Specifically marine microalgae are great sources of proteins and polyunsaturated fatty acids which are important for the prevention and treatment of diseases such as coronary heart disease and cancer [82], [83]. Many studies have proven that the growth rate of microalgae is essential for the optimization of the use of fatty acid. *Dunaliella salina* (which is studied within this dissertation) is of interest for medicinal purposes due to the high beta-carotene content as well as its anti-cancer drug containing properties. Within the human body, beta-carotene is converted into vitamin A which assists in the functionality of the immune system. Another algal component that is affected by algal growth and environmental conditions are proteins that are used in health foods. Studies have proven algae have a relatively high nutritional value in comparison to proteins that are found in conventional vegetables. The ease of culturing microalgae combined with their biochemical properties makes them a valuable resource for pharmaceuticals [82] and food consumption [8], [80].

Although microalgae are used for the production of pharmaceuticals and food, possibly the most sought after area of interest is in the study of biofuel production. The ability for microalgae to produce biofuel is based on their fast growth rate [66], [84], their ability to double their biomass with one days turnaround and that their oil content greatly exceeds other crops [8], [81], [85]. Algae production is 6 - 12 times higher than the first generation biofuels such as switchgrass, sugarcane and corn, because it is more effective in converting sunlight energy into organic compounds [8], [66], [80], [85]. In 2009, 20,000 gallons of algal fuel was sold to the U.S. Navy for $8.5 million. The following year they made another purchase of 150,000 gallons. Not only will algae based biofuel have a direct effect on transportation fuels, but it will also have a great impact on economic and environmental resources [8]. These are few reasons why algae are among the most potentially significant sources [81] utilized for human usage.

1.4. Dissertation Overview

The overall objective for this project is monitoring changes that occur within an ecosystem by detecting how these changes impact microalgae species. Microalgae are at the bottom of the food chain within trophic systems, therefore can yield a view into how the chemical changes affect a system. The following studies investigated these changes through the use of novel sample preparation, imaging analyses and spectroscopic studies, all supported by chemometric techniques. These tools demonstrated the applicability of microalgae as an *in-situ* biological sensor. The combination of these novel analytical and chemometric tools provides the necessary means for an assessment of microalgae's marine ecosystems and the changes that occur.

The first section of this dissertation focuses on the development of an image analysis method that quickly detects and fits the shape and size of cells to relate to nutrient induced changes. Due to the variability of cells, a large number of cells were utilized for analysis to obtain species specific features. For testing this hypothesis, microalgae were cultured under well-defined but varying nitrogen and inorganic carbon concentrations for extracting chemical specific size and shape information. The analysis was unable to be performed utilizing a linear model due to the variability of samples. Thus a nonlinear response model, coined Predictor Surfaces, was used for predicting concentrations of the growing medium from the measured shapes and sizes of the cell. A quantitative analysis technique was employed to determine if the shape and size of the cells have a measureable impact on the growth conditions that the cells were cultured in. Imaging histograms were averaged and t-tested to determine statistical relevance. The analysis method demonstrated that as the growth condition and concentration of the nutrient source changed the shape and size varied.

The next section utilized mathematical techniques to express the relationship between microalgae cell size and growth rate under various nutrient induced environments. Two nonlinear regression models were utilized to correlate nutrient source types and concentrations to growth rates through the use of sigmoidal functions of time as well as cell size distributions

through a modified Gaussian curve. The analyses determined that the relationship between cell size and growth rate was species, nutrient source and concentration specific.

Single microalgae species were proven to have distinct different chemical signatures when grown under varying environmental conditions [39]. In chapter 4, this analysis moved closer to a real-world environment to study how microalgae species chemically interact utilizing Fourier Transform Infrared Spectroscopy. Determining the role that microalgae play in the ecological community partly depends on understanding the chemical changes that occur to microalgae as they undergo disturbances.

The focus of this dissertation is to develop analytical methodologies for detecting chemical changes in marine ecosystems. Therefore, novel pathways have been investigated by way of microalgae's use as an *in-situ* probe which have measureable physical and chemical adaptations to environmental changes.

2. Experimental Techniques for Assessing Impacts of Environmental Change of Microalgae

2.1. Introduction

To better understand how ambient chemical conditions affect microalgae species physically and chemically, three species were cultured under two different sources, carbon and nitrogen, and under varying concentrations. The physical parameters of microalgae were studied utilizing imaging analyses to determine the size and shape of microalgae cells. Chemical parameters were analyzed by FT-IR spectroscopy because several biological components are found within the mid-infrared region.

2.2. Microalgae Culturing Methods

Stock bottles used for growth medium, flasks, test tubes and utensils were washed using deionized water and powdered Contrex AP labware detergent (Fisher Scientific). Cotton plugs were made for all flasks. Next sterilization is needed and very important in phycological research in order to maintain sterile cultures for living organisms. The combination of sterile cultures and sterile equipment is necessary in order to reduce the possibility of contamination [9]. To ensure that all items had reached the sterilization temperature, autoclave tape was placed onto each item.

Medium, flasks, test tubes, beakers and pipet tips were all placed into an autoclave (Getinge Vacuum/Gravity Steam Sterilizer, 733LS) which reached 121°C at 2.0 atm steam pressure for 20 minutes. Autoclaving is the most popular technique to sterilize liquids and heat-resistant materials [9].

Starting cultures of *Dunaliella parva* (#LB1983)*, Nannochloropsis oculata* (#LB2164) and *Dunaliella salina* (#LB200) were obtained from UTEX at the University of Texas, Austin. For each species, about 15 mL of aliquots were suspended in Erdschreiber's [9], [86] medium. Once obtained, the caps were unscrewed and air exchange occurred for two hours before inoculation. All three species were transferred to a different growth medium called Enriched Artificial Seawater (ESAW) medium [9], [87], [88]. The ESAW medium was made from reference [89]. All medium was prepared utilizing the same steps. First each major nutrient source (i.e. ammonium, carbon and nitrate) was prepared in stock solutions under varying concentrations to later be added to the medium. The growth medium was prepared in two flasks and later combined before autoclaving. Salt Solution I was prepared with 600 mL of deionized water and the reagents in Table 1 and salt solution II was prepared with 300 mL of deionized water and with the reagents listed in Table 2. After each salt solution was made, they were combined in a one liter PYREX stock bottle and one mL of each nutrient listed in Table 3 except vitamin stock was added. The vitamin stock is added after autoclaving because the high temperatures during autoclaving will cause the reagents to precipitate [87]. To maintain a pH of about 8.2, 10 mL of tris hydrochloride solution (1M, Fisher Scientific) was added. Next 83 mL of deionized water was added in order to bring the final volume to 999 mL before autoclaving. Once all reagents are added, the lid on the stock bottles must be loosened prior to being placed in the autoclave in order to prevent pressure build up inside the bottle. After the ESAW has been autoclaved, the medium must cool for a minimum of 24 hours. Once it is cooled, the vitamin stock is filtered by using a mixed cellulose filter (Fisher Scientific) with a pore size of 0.22 μM into a sterilized flask under an Air Science Purair VLF laminar flow cabinet/ PCR (polymerase chain reaction) cabinet. The flow hood uses an ultraviolet light to sterilize all materials and surfaces that will come in contact with the microalgae. Also, constant air flow inside of the hood is utilized to suppress airborne matter from coming in contact with any materials by using an ultra-low penetration air (ULPA) filter. Before handling equipment of cultures, 70% ethanol must be used to wipe down gloves and surfaces as a part of the sterilization procedure [9]. Once the medium is cooled, sterilization is complete and the vitamin stock is filtered, one mL of vitamin stock is added. After all reagents have been added, each stock bottle is tested using a pH meter to ensure that the final pH of each growth medium is approximately 8.2. After sterilization and pH testing of growth medium, microalgae were inoculated into the growth medium listed in Table 1 - Table 3 inside of the flow hood to further avoid contamination. The test tubes containing the microalgae sample were removed within the hood and the top of the vial was flamed with a Bunsen burner in order to burn off anything that

could possibly cause contamination. Next, beakers were flamed and the algae stock cultures were poured into these sterilized beakers. In the case they do not grow upon the first transfer, the cultures obtained from UTEX can be kept for a few weeks after their arrival and another transfer needs to occur.

Table 1. Reagents used to make Salt Solution I for ESAW medium [9].

Reagent	Mass (g)	Final Concentration in ESAW
NaCl (Fisher Scientific)	21.19	363 mM
Na_2SO_4 (Fisher Scientific)	3.55	25 mM
KCl (Fisher Scientific)	0.599	8.04 mM
$NaHCO_3$ (Fisher Scientific)	0.174	2.07 mM
KBr (Fisher Scientific)	0.0863	725 μM
H_3BO_3 (Fisher Scientific)	0.023	372 μM
NaF (Fisher Scientific)	0.0028	65.7 μM

Table 2. Reagents used to make Salt Solution II in ESAW [9].

Reagent	Mass (g)	Final Concentration in ESAW
$MgCl_2 \cdot 6H_2O$ (Fisher Scientific)	9.592	41.3 mM
$CaCl_2 \cdot 2H_2O$ (Fisher Scientific)	1.344	9.14 mM
$SrCl_2 \cdot 6H_2O$ (Acros Organics)	0.0218	82 μM

Table 3. Nutrient stock solutions for ESAW medium [9].

Nutrient	Reagent	Mass (g)	Final Concentration in ESAW
Nitrate	$NaNO_3$ (Fisher Scientific)	46.7	549 µM
Phosphate	$NaH_2PO_4 \cdot H_2O$ (Fisher Scientific)	3.09	21 µM
Silicate	$Na_2SiO_3 \cdot 9H_2O$ (Sigma-Aldrich)	15	105 µM
Metal Stock I	$Na_2EDTA \cdot 2H_2O$ (Sigma-Aldrich)	3.09	6.56 µM
	$FeCl_3 \cdot 6H_2O$ (Fisher Scientific)	1.77	6.56 µM
Metal Stock II	$ZnSO_4 \cdot 7H_2O$ (Fisher Scientific)	0.073	254 nM
	$CoSO_4 \cdot 7H_2O$ (Fisher Scientific)	0.016	5.16 nM
	$MnSO_4 \cdot 4H_2O$ (Alfa Aesar)	0.54	2.42 µM
	$Na_2MoO_4 \cdot 2H_2O$ (Acros Organics)	1.48×10^{-3}	6.1 nM
	Na_2SeO_3 (Sigma-Aldrich)	1.73×10^{-4}	1 nM
	$NiCl_2 \cdot 6H_2O$ (Sigma-Aldrich)	1.49×10^{-3}	6.3 nM
	$Na_2EDTA \cdot 2H_2O$ (Sigma-Aldrich)	2.44	8.29 µM
Vitamin Stock	thiamine-HCl (Fisher Scientific)	0.1	297 nM
	Biotin (Fisher Scientific)	0.002	4.09 nM
	Vitamin B_{12} (Fisher Scientific)	0.001	1.47 nM

For the inoculation, a 250 mL Erlenmeyer flask was flamed and filled with 100 mL of ESAW. After sterilization of each flask, two mL of microalgae was inoculated into the medium using a Finnpipette adjustable volume pipette (Fisher Scientific) with a sterile pipette tip. The flask was then flamed again and plugged with a cotton plug in order to allow for air exchange. ESAW contains a number of additional, nutrients at low concentrations, which were all composed from the same stocks in order to minimize fluctuations in the chemical composition of the growing medium. To incorporate the naturally occurring, often considerable replicate-to-replicate variability into these analyses, five replicate cultures were grown for each species and nutrient situation. Each flask was placed into the environmental growth chamber (Precision 818, Thermo Scientific) at 20°C under 24 hours of illumination for a controlled environment until each species had reached exponential growth phase [9], [89].



Figure 1. Inoculations of each species in the growth cabinet. From left to right: *Dunaliella salina, Nannochloropsis oculata* and *Dunaliella parva*.

After each species has been inoculated into a new medium, a growth curve was obtained for each species under each growth parameter. The growth curves determine when each species reaches its exponential phase for harvest [90]. Also, growth curves are able to provide information on the cell concentration of the cultured population which is able to determine the population size and the biomass [9].

To begin growth curves acquisition, each flask was swirled to ensure a homogeneous distribution of cells [91]. 200 µL of culture was pipetted into a sterilized test tube and mixed with 20 µL of Lugol's iodine solution (Sigma-Aldrich) using a Finnpipette with a sterilized tip. The Lugol's solution is used as a cell stain and preserves the algae in order to prevent the algae from moving to obtain an accurate cell count [92]. One drop of the mixture was placed on a hemocytometer (Cole-Palmer) 0.1 mm chamber depth slide and covered with a coverslip. Once each side of the hemocytometer was filled, the cells were allowed to settle for three to five minutes before obtaining a count. The slide was then placed under a 10x microscope objective and examined to see if any contamination may have taken place [9]. Ten counts from ten independent squares were taken and averaged in order to obtain a mean value of cells for each species at a specific day. For each species and each nutrient situation, several cell counts were taken at the same time of day and averaged.

Because each species differs in size, there are different grids used for counting. For smaller cells, the volume calculations (number of cells/ mL of solution) used the smaller squares (0.25 nL) and for larger cells, larger squares (6.25 nL) were used [92].

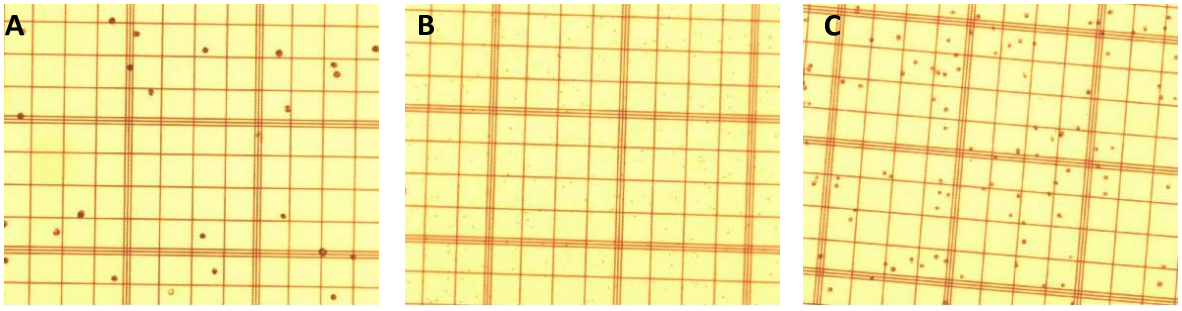


Figure 2. Each microalgae species on the hemocytometer in 0.25 nL cells at their exponential phase: (A) *Dunaliella salina,* (B) *Nannochloropsis oculata,* (C) *Dunaliella parva.*

The following equation was used to obtain the concentration for each species at a given day. The volume calculations (number of cells/mL of solution) were dependent on the size of the grid (0.0025 mm^2 for smaller squares which were used for these growth curves) [92].

$$\left(\frac{\text{Average Number of Cells}}{(\text{Square Volume } (\text{mm}^2) \times .1 \text{ mm depth}) \times (\text{number of squares counted})}\right) \times \left(\frac{1\ mm^3}{.001\ mL}\right) \quad (1)$$

From which data, growth curves were obtained for the concentration of each species at a specific nutrient concentration and condition for a given day.

Replicate samples and nutrient source were averaged and growth curves were obtained. These growth curves were used throughout this entire dissertation for all experiments.

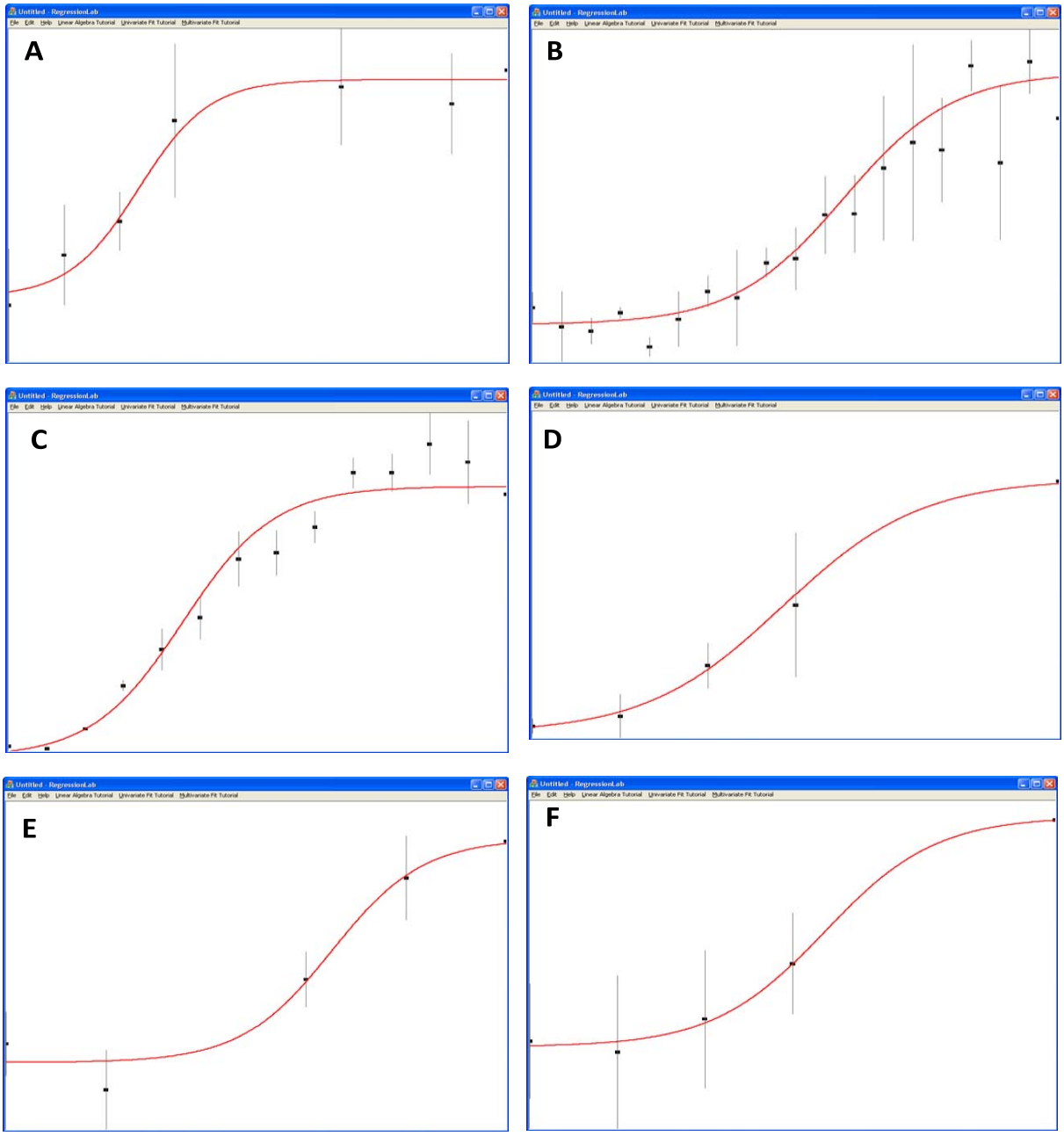


Figure 3. *Dunaliella salina* (A) 0.35 mM NO_3^-, (B) 0.873 mM NH_4^+, *Nannochloropsis oculata* (C) 0.873 mM NH_4^+, (D) 0.873 mM NO_3^-, *Dunaliella parva*(A) 0.16 mM NO_3^-, (B) 0.549 mM NH_4^+.

2.3. Nutrient Induced Changes Imposed on Microalgae's Environment

It has been proven that microalgae respond sensitively and are indicators of changes that occur from their environmental conditions [1]. Because microalgae spectra have revealed a non-linear response to nutrient induced change, their ability to redistribute organisms makes it an indicator for change[6], [39].

Nutrient concentration fluctuations greatly influence the growth of microalgae by altering their growth [54]. Determining how the microalgae's physical and chemical parameters are affected due to nutrient variability is of great importance in further understanding human induced chemical impacts on the marine ecosystems. There are two nutrients in particular that contribute to these changes; carbon and nitrogen which are necessary for algae growth [8], [91] - [93]. In response to these nutrient stressors, microalgae cells undergo changes in their chemical composition [23], [39].

One way nitrogen affects microalgae cells is by the reduction of protein content and the enhancement of lipids and carbohydrates thus altering the chemical composition. These changes have occurred due to such human activities as nitrogen fossil fuel combustion and nitrogen fertilizer production [19], [22]. This in turn demonstrates how biochemical composition of microalgae can be altered based off of environmental conditions [23].

Another source that affects microalgae composition is carbon. Most of the CO_2 that enters the atmosphere caused by humans dissolves in the oceans and increases the dissolved CO_2 and bicarbonate ion HCO_3^- [5] concentration levels while lowering the carbonate ion concentration and pH [6] and effects the distribution of algae. Each species has a different reaction to the raised levels of CO_2 concentration. Because of this increase, the growth rate of some species will increase, while others will not be favored. This in turn will also change and alter their cellular composition [1]. It has been shown that algae can use both HCO_3^- and CO_2 as a carbon uptake method. In some cases, specific species will only take up CO_2 and other species will only take up HCO_3^- [31].

For this dissertation, carbon and nitrogen were varied and measured using imaging analysis along with spectroscopy in order to obtain a deeper understanding of what occurs within the chemical composition as well as the physical parameters of the cells. These impacts were studies by utilizing two different nitrogen sources, sodium nitrate and ammonium chloride. Each

nitrogen source was grown under varying concentrations: 0.16 mM, 0.35 mM, 0.549 mM (normal nitrate concentration), 0.873 mM, 1.28 mM, 1.47 mM, 1.65 Mm. The carbon source was introduced into ESAW utilizing sodium bicarbonate under the following concentrations: 0.16 mM, 1.11 mM, 2.07 mM (normal condition), 3.63 mM, 5.18 mM, 6.72 mM and 8.26 mM.

2.4. Cellular Image Acquisition

Once each species has reached its exponential phase, it is time for the cells to be harvested. Before harvesting, two mL of culture are transferred to a new medium in order continue the growth of each species. About 40 mL of microalgae culture is poured into a flamed, sterilized beaker. For algal fixation, 10 mL of the microalgae culture and 10 µL of Lugol's solution are pipetted from the beaker into borosilicate glass centrifuge tubes (Fisher Scientific) 15 mL in diameter with a 100 mm height. The centrifuge tubes are then placed in an Eppendorf 5702 centrifuge and spun at 4400 rotations per minute for five minutes. Upon completion, the supernatant is discarded leaving the microalgae pellet at the bottom of the tube. These steps are repeated at least two additional times and each time adding 10 mL of culture mixed with 10 µL of Lugol's solution to the centrifuge tube with the algae pellet. Once there is enough pellet for spectroscopic analysis, two mL of 0.5 M isosmotic ammonium formate solution (Alfa Aesar) is used to wash the pellet. The ammonium formate is used to remove any excess medium that may have carried over into the pellet [94]. The salt crystals are IR transparent, but could cause light scattering inference which could affect the spectroscopic signatures [95]. The ammonium formate was added to the pellet and centrifuged and then repeated. After the two mL of ammonium formate is added the second time, 100 µL of the microalgae ammonium formate mixture is then pipetted into another test tube for obtaining microscope images of cells. With an inoculating loop a few drops of washed algae suspension were then positioned onto a calcium fluoride (CaF_2) microscope slide. From each sample multiple images (Figure 4) were acquired as 24-bit color bitmaps using a 10X microscope in transmission mode.

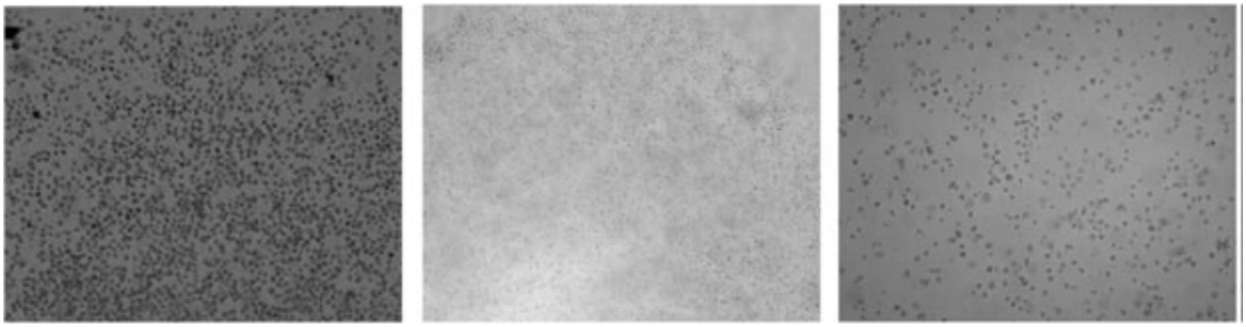


Figure 4. Microscope images (10X magnification, 380 x 300 pixels) of cultures of three different algae species; the cells were still dispensed in ESAW - (left) *Dunaliella parva*, (center) *Nannochloropsis oculata*, and (right) *Dunaliella salina*; the visible illumination inhomogeneity can be assumed to be homogeneous across 20 x 20 pixel subimages (cp. Figure 6 (A) - (C)).

2.5. Harvesting Microalgae Cells for Spectroscopic Analyses

After harvesting, the pelleted cells are then dried for 4-5 days at 60°C. Dried algae samples tend to clump and lead to inhomogeneity's which cause IR light scattering and baseline drifts [95]. To account for this, a mechanical mixer (WIG-L-BUG, Rinn Crescent) was utilized in order to produce a homogenous mixture. Once microalgae are dry, microalgae samples were ground and mixed with potassium bromide (KBr) for two minutes. The hygroscopic KBr was not ground to prevent increasing surface area exposing it to ambient moisture which would affect the FT-IR spectra; thus it was stored in a desiccator under a low vacuum. For all species and conditions, algae in KBr were prepared in concentrations of 0.6 weight percent. The mixture was then pressed into a 3 mm inner diameter washer to form an IR-transparent pellet containing approximately 5 mg of powder mixture taken from a larger sample. For a reference spectrum, KBr was pelleted and analysis was performed using a FT-IR spectrometer to acquire transmission spectra using a Bruker Vertex 70 FT-IR spectrometer, equipped with a deuterated triglycine sulfate (DTGS) detector and controlled by OPUS 6.5 software package. Sixty-four scans were acquired at a 4 cm^{-1} resolution co-added over the range of 7500-370 cm^{-1} utilizing a 5 mm aperture setting. The aperture setting was selected by recording spectra at each decreasing aperture size until the size became too small and the signal level decreased from the infrared beam not illuminating the sample. Information between 3500-2800 cm^{-1} and 1800-950 cm^{-1} were utilized for analysis. The region from 2800-1800 cm^{-1} was omitted in order to exclude atmospheric CO_2 fluctuations from the spectra that are not reproducible.

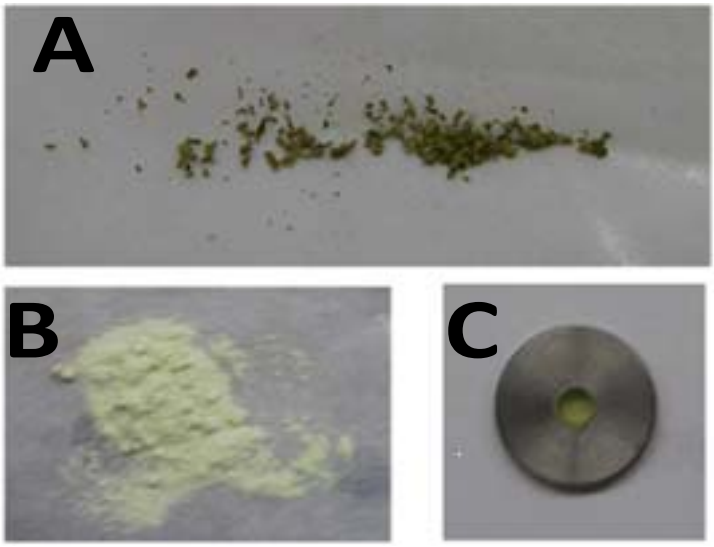


Figure 5. (A) dried algae sample, (B), algae and KBr mixture at a 0.6 weight percent and (C) pelleted algae and KBR samples.

2.6. Summary

Three microalgae species, *Dunaliella salina*, *Nannochloropsis oculata*, *Dunaliella parva*, are chosen and cultured for imaging and spectroscopic analysis due to their ability to adapt to their environment. Each species is grown under well-defined, sterile conditions and exposed to two known environmental nutrient sources as well as being grown under various concentrations of these two nutrient stressors. These nutrient sources are studied to determine if their physical and chemical parameters will be able to reflect back to the environment they were exposed. After culturing and harvesting, the next step is to obtain microscope images in order to study how the physical parameters of a cell will change when exposed to different environmental conditions and how well the cells will be able to relate back to those conditions.

3. Investigation of Microalgae's Physical Parameters

3.1. Exploration of Chemical Impacts on Cell Cultures

Chemical analyses of biological materials often encounter challenges imposed by the chemical complexity, strong replicate-to-replicate variations, minute species-dependent signal features, and overshadowing influences of the cells' chemical environment. Several analytical techniques such as high-performance liquid chromatography [96] - [99] and Fourier transform infrared spectroscopy [52] - [54], [56], [100] have been applied to analyses of microalgal samples, which, however, require destroying the sample or are not applicable to live and thus water-containing samples. Furthermore, the chemical similarity of many cell species superimposed by shifts of their chemical signature induced by the cells' changing environment [39] makes cell classification challenging if not too demanding for a single sensor. Image analyses, on the other hand, can be performed in situ on large numbers of living cells and thus can augment other measurement techniques to enhance their species and/or analyte selectivity. Hence, combining a sensing technique with imaging sensors can considerably enhance the discrimination power. It was hypothesized for this study that the cells' physical appearance [5], [14], [62] - [64] reflects cell type and/or chemical impacts imposed onto them by their environment. To utilize this information, a novel image analysis method has been developed to measure cell size and ellipticity of a large number of, here, microalgae cells (Figure 4) as a function of species and ambient conditions. Compiling size and ellipticity results from numerous cells into histograms and basing subsequent analyses on them incorporates the natural and often considerable cell-to-cell fluctuations. These histograms can then be investigated with respect to cell discrimination and/or environmental impacts. Since such histograms are multivariate data sets, linear or nonlinear multivariate least-squares regression (MLR) [101] - [104] can be utilized for quantitative investigations relating the cells' physical appearance to selected ambient parameters. In the remainder, "cell class" or simply "class" will refer to a cell type cultured under a certain set of growing conditions. Understanding the various responses on a cellular level

affords the option for gaining more in depth knowledge of microalgae's response to nutrients. The ability for microalgae to detect these changes can be based on cell size [51].

When measurement conditions are well-defined, imaging sensors with their aforementioned advantages can be sufficient for analyses of an unknown sample's cell. Often however, the sheer number of possible cell types multiplied by numerous ambient influences can be too demanding for a single sensor. Then combining a sensing technique with imaging sensors can considerably enhance the discrimination power. If chemical information and the cells' visual appearance are consistent, the identification of chemically similar cells and/or their environmental parameters is re-enforced. If however an unknown cell is chemically similar to a reference cell but their physical appearances are inconsistent, the unknown and reference cannot be positively identified as belonging to the same class.

From an environmental analytical chemistry perspective, microalgae have a real-world relevance as they have been reported to adapt their chemical compositions to changes in aqueous ecosystems [12] - [14], [36]. Furthermore, microalgal biodiversity has been proposed as an system-wide response to chemically changing environments [37]. In recent investigations, it could be shown that microalgae can be discriminated based on the FT-IR signature [53], [56], [65], [105], [106] and that their chemical composition is nonlinearly but reproducibly related to their nutrients' availability; this has enabled indirect predictions of these selected ambient parameters [39]. Microalgae cells have the potential as innovative biological *in-situ* sensors for monitoring chemical changes in large ecosystems. In this context, the identification of cell species is of importance because only those cells for which calibration models are available should be included into an environmental assessment. To support advances in utilizing microalgae cells as biosensors, they have been chosen as a proof-of-principle application to assess the applicability of image analyses developed in this study. Three microalgae species were cultured under numerous, well-defined ambient conditions namely the concentration of inorganic carbon (source: bicarbonate) and nitrogen (sources: ammonium and nitrate) [12] - [14], [36], [37], [39], [53], [56], [65], [105], [106]. Two types of experiments have been performed, i.e.: (i) species discrimination based on cell size and ellipticity when grown under the same environmental conditions and (ii) assessment of changes in a cells' physical appearances as a function of its nutrient induced change.

The automated image analysis developed in this study fits a 2D function (surface) to each cell's 3D image (Figure 6) and derives the wanted cell properties from the resulting fit parameters

(section 3.1.1). In a second step (section 3.1.2), statistical tests were employed to assess differences among cell types as well as the cells' changes in their physical appearance to changing ambient conditions. Finally, quantitative models were developed (section 3.1.3) for predicting chemical parameters of the cells' environment based on the cells' adapting size and shape. Due to the unavoidable cell-to-cell variations within a cell class, a large number of cells need to be analyzed to derive class-representative histograms. Thus, multiple images have been acquired per cell culture and in each image; several hundreds of cells were found and a fit was acquired for all samples (see e.g. Figure 4).

3.1.1. Extracting Cellular Size and Shape Information

A digital image of $H \times M$ pixels can be depicted as a 3D plot in which x and y represent the spatial dimensions and $z(x_m, y_h)$ the light intensity at a given pixel (x_m, y_h) (Figure 6). These cell images are fitted to an application-dependent model function $Z(x, y)$ from whose fit parameters the cells' size and shape are then derived. In a transmission microscopy setup as used for these investigations, the cells cast shadows (Figure 4) which carry information about the wanted cell features. In a reflection or fluorescence measurement, cells would act like a small light source and the presented data analyses remain valid after multiplying the raw data by a factor -1. In the given proof-of-principle application, the cell shadows could be well described by down-pointing 2D Gaussians ((2), [107], Figure 6) [40]. To keep this discussion straightforward, it is assumed that the illumination of the sample is homogeneous and can be described by a constant background light level a_0. Such a homogeneous illumination superimposed by a cell shadow of 2D Gaussian shape is modeled by the model function $Z(x, y)$ (equation (2)) with $A < 0$:

$$Z(x,y) = a_0 + A \cdot \exp\left\{-\frac{1}{2} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}^{\mathbf{T}} \cdot \boldsymbol{\Sigma}^{-\mathbf{1}} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}\right\} \qquad (2)$$

If homogeneous illuminations are not given, a polynomial surface [104] might be more appropriate such as a plane $a_0 + a_1 \cdot x + a_2 \cdot y$ or a paraboloid which would then replace the

term a_0 in (2). The matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho_{xy} \cdot \sigma_x \cdot \sigma_y \\ \rho_{xy} \cdot \sigma_y \cdot \sigma_x & \sigma_y^2 \end{pmatrix}$ whose *inverse* is contained in equation (2) is a 'variance-covariance matrix' with σ_x and σ_y describing the widths of the 2D Gaussian along its two principal axes. A 2D Gaussian whose principal axes are not aligned along the x and y-axes introduces a correlation $(\rho_{xy} \neq 0)$ between the spatial dimensions x and y and hence nonzero off-diagonal elements in $\mathbf{\Sigma}$. After deriving $\mathbf{\Sigma}$ in the procedure described in the following section, the cells' elliptical footprint area (= size) is calculated as below:

$$\text{size} = \pi \cdot \sigma_x \cdot \sigma_y \qquad (3)$$

As a measure of a cell's 'shape', or more precisely its deviation from circularity, a height-to-width ratio has been defined as: [1]

$$\text{shape} = \frac{\max(\sigma_x,\ \sigma_y)}{\min(\sigma_x,\ \sigma_y)} \qquad (4)$$

For the size (3) and shape (4) measures, σ_x and σ_y need to be derived by fitting the measured light intensities $z(x_m, y_h) = Z(x_m, y_h) + \varepsilon_{m,h}$ to the model function $Z(x, y)$ (equation (2)). For this purpose, measured at pixel (m,n) the exponent in (2) is re-written taking advantage of the symmetry of $\mathbf{\Sigma}$ and thus $\mathbf{\Sigma^{-1}}$:

$$\underbrace{((x-x_0)\quad (y-y_0))}_{=\binom{x-x_0}{y-y_0}^{\mathbf{T}}} \cdot \underbrace{\begin{pmatrix} B & C \\ C & D \end{pmatrix}}_{=\mathbf{\Sigma^{-1}}} \cdot \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} = B \cdot (x-x_0)^2 + 2 \cdot C \cdot (x-x_0) \cdot (y-y_0) + D \cdot (y-y_0)^2 \qquad (5)$$

In this revised notation [40], the model function reads:

$$Z(x, y) = a_0 + A \cdot \exp\left\{-\frac{1}{2} \cdot [B \cdot (x-x_0)^2 + 2 \cdot C \cdot (x-x_0) \cdot (y-y_0) + D \cdot (y-y_0)^2]\right\}$$

[1] Utilizing the maximum among σ_x and σ_y in the nominator of (4) and the minimum in the denominator avoids any ambiguity of the ratio of the two widths.

(6)

Once $\hat{B}$, $\hat{C}$, and $\hat{D}$ have been estimated[2] from fitting $Z(x,y)$ to the measured $z\left(x_{m=1,\dots,M}, y_{h=1,\dots,H}\right)$, they are written into a 2×2 matrix $\mathbf{\Sigma^{-1}}$ whose inversion then results in $\mathbf{\Sigma}$, and thus, the required $\hat{\sigma}_x$ and $\hat{\sigma}_y$; $\hat{\rho}_{xy}$ will not be needed here. During this least-squares fitting, the sum of squared errors, $\mathrm{SSE}\left(\hat{a}_0, \hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{x}_0, \hat{y}_0\right) = \sum_{m=1}^{M}\sum_{h=1}^{H} \varepsilon_{m,h}^2 = \sum_{m=1}^{M}\sum_{h=1}^{H}[z(x_m, y_h) - Z(x_m, y_h)]^2$, is minimized. The necessary condition for a minimum of the SSE is:

$$\begin{aligned}\boldsymbol{\nabla}\mathrm{SSE}\left(\hat{a}_0, \hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{x}_0, \hat{y}_0\right) &= 2 \cdot \sum_{m=1}^{M}\sum_{k=1}^{K}[z(x_m, y_k) - Z(x_m, y_k)] \cdot \boldsymbol{\nabla}\left(z(x_m, y_k) - Z(x_m, y_k)\right) \\ &= -2 \cdot \sum_{m=1}^{M}\sum_{k=1}^{K}[z(x_m, y_k) - Z(x_m, y_k)] \cdot \boldsymbol{\nabla} Z(x_m, y_k) = \mathbf{0}\end{aligned}$$

(7)

The second equal sign in (7) holds because the $z(x_m, y_k)$ are measured numbers and not functions of the fit parameters; thus $\boldsymbol{\nabla} z(x_m, y_k) = \mathbf{0}$. The gradient operator $\boldsymbol{\nabla} = \left(\frac{\partial}{\partial \hat{a}_0} \quad \frac{\partial}{\partial \hat{A}} \quad \frac{\partial}{\partial \hat{B}} \quad \frac{\partial}{\partial \hat{C}} \quad \frac{\partial}{\partial \hat{D}} \quad \frac{\partial}{\partial \hat{x}_0} \quad \frac{\partial}{\partial \hat{y}_0}\right)^{\mathbf{T}}$ in (7) leads to a system of here seven nonlinear equations. This nonlinearity prohibits the use of standard linear least-squares regression [101] - [103] and hence, iterative solvers [108], [109], were employed instead.

The biological nature of the samples introduces a considerable but unavoidable variability among cells even in replicate experiments. Hence, a large number of cells need to be analyzed to extract a cell type's characteristics and influences of the chemical state of the cells' environment. It was deemed to be more straightforward to split a larger image (cp. Figure 4) into multiple subimages (Figure 6 (A) – (F)), because then the iterative algorithm solving (7) has to incorporate fewer unknowns and is thus less likely to being caught in a local minimum of the SSE. The algorithm has to first determine the number *J* of cells in each subimage (or the full image for that matter); it then incorporates a superposition of *J* cells into equation (8) rather than isolating each individual cell and use equation (6). Since it is not known *a priori* how many cells are contained in an image nor is it reasonable to manually count the cells for each of the 285 subimages, J must be derived automatically for each subimage. For this purpose, an automated search for a light intensity minimum, that is, a cell shadow, starts at a randomly

[2] In the remainder, a hat ˆ indicates a least-squares estimate of the unhated item.

chosen pixel and then keeps stepwise, moving the intermediate position of the minimum to that one of its eight neighbor pixel that has the lowest signal intensity. This procedure is continued until the current pixel position has a lower signal than all eight neighbors. This final position is marked as a local minimum. This procedure is repeated a few thousand times from randomly selected starting pixels and accumulates a considerable number of (hits) endpoints in the true local minima. In order to prevent that noise influences this minima search too much, a 2D-fast Fourier transform low-pass filter [110] has been applied to remove noise spikes.

$$Z(x,y) = a_0 + \sum_{j=1}^{J} A_j \cdot \exp\left\{-\frac{1}{2} \cdot \left[B_j \cdot \left(x - x_{0,j}\right)^2 + 2 \cdot C_j \cdot \left(x - x_{0,j}\right) \cdot \left(y - y_{0,j}\right) + D_j \cdot \left(y - y_{0,j}\right)^2\right]\right\} \quad (8)$$

From fitting (8) to a subimage, J sets of fit parameters $\left(\hat{A}_j, \hat{B}_j, \hat{C}_j, \hat{D}_j, \hat{x}_{0,j}, \hat{y}_{0,j}\right)$ are obtained from which J values for size (3) and for shape (4) were derived (Figure 6 (G) – (I)). If cells of different shapes [5], [63] are to be analyzed, the model function $Z(x,y)$ needs to be adapted [40].

3.1.2. Qualitative Assessments

Due to the inherent variability of biological samples, a single number for size or shape is unlikely to be representative for a cell class - a class will rather be described by *distributions* of size (3) and shape (4) values. From experimental data, approximations for these distributions are derived in form of histograms which will be the foundation of subsequent analyses of species discrimination and investigations of environmental impacts on cells. The narrower a histogram's bins are chosen, the more precisely are underlying distributions approximated; however, using a large number of narrow bins runs the risk of spreading information thinly and burying it among sample-to-sample fluctuations. In order to balance close approximation and spreading out information, the bin widths were not set up to be equal but zoomed into certain areas of interest while remaining wide in others. For a given application, such areas of interest for size histograms might be determined in preliminary tests were adapted to the microalgae cells under

investigations. If, for instance, cells appear to be fairly circular, the bin resolution for shape histograms should be chosen high for $\mathrm{shape} = 1 \ldots 2$ and low for more elliptic cells. Of course, all size and all shape histograms, respectively, were set up with the same bin widths so that they are compatible in subsequent analyses. When incorporating bins of different widths, wide bins get over-emphasized. To assign an equal weight to each bin and to ensure that the histograms do not look 'distorted', the bins' counts were divided by that bin's width. On a similar note, the different cell counts found within different images prevent a direct comparison of the histograms extracted from different images. Thus, histograms were also normalized by the number of cells they had been compiled from. In the remainder, only *normalized histograms* will be utilized; in figures, their y-axis will be labeled 'normalized counts' or 'relative counts'.

For generating a measure for the natural replicate-to-replicate variability of these normalized histograms, multiple images were acquired per class. But instead of composing fit results from replicate images into one histogram, the histograms of each image were averaged bin-wise and corresponding standard deviations for each bin were computed. With such mean normalized counts and the corresponding standard deviation, normalized histograms of two classes can be compared bin-by-bin by means of t-tests to determine whether the counts in a certain bin are significantly different or not. Testing individual bins has the advantage of identifying at which size and shape values significant changes were found. If for instance the size distributions of two cell classes are to be compared and cell type A is generally larger than cell type B, such a bin-by-bin t-test will determine that type A has significantly fewer counts than type B in bins representing small cell sizes. On the other hand, type A will have significantly more counts than cell type B in bins representing large cell areas. In other words, the two histograms' maxima will be shifted against each other and the t-test will determine in these histogram regions. Obviously, for such analyses, the two histogram normalization steps of bin widths and overall cell count are crucial. Throughout this study, a confidence level of 95% has been employed. For visualization, the outcome of such t-tests can be composed into 't-histograms': a negative value, say -0.1, is chosen to indicate that in a certain bin cell type A has a significantly *lower* count than cell type B; a positive value, e.g. $+0.1$, specifies that the first cell type's count (i.e. A's) is significantly *higher* than the second cell type's count (i.e. B's). Obviously, it matters which cell type (i.e. A vs. B) is listed first in a comparison and that order cannot change throughout a comparison or the plus and minus signs in a t-histogram become incorrect [40].

3.1.3. Quantitative Predictions of Ambient Chemical Parameters

In the next step, a quantitative analysis has been developed to investigate whether changes in cell size and shape are quantitatively related to chemical parameters in the microalgae cells' growing environment. As proof-of-principle application (see section 2.3), the nutrient concentration c has been chosen as ambient chemical parameter which is to be predicted. For building quantitative calibration models, two chemometric methods have been employed, i.e. standard Partial Least-Squares (PLS) [111], [112] and Predictor Surfaces [39]: PLS is a linear method which has been frequently found to be superior to conventional Multivariate Least-Squares Regression (MLR) [101] - [103] and serves as comparison to the novel Predictor Surfaces that is based on nonlinear, multivariate polynomial models. In this study, both algorithms consider the concentration c of a selected nutrient as the sole predictor variable x for the Predictor Surfaces; the N bins of a normalized histogram $\mathbf{h}_{(N\times 1)}$ (i.e., distributions of cell sizes) play the role of the response variables $\mathbf{y}_{(N\times 1)} = (y_1 \quad \cdots \quad y_N)^{\mathbf{T}}$. In this section, x and y refer to predictor and response variables and not to spatial dimensions as in Figure 6 for instance.

Since only one predictor variable is involved here and thus no coupling between multiple predictors has to be incorporated, Predictor Surfaces $Y_{\boldsymbol{\beta}}(x)$ [39] simplify to N univariate polynomials $Y_{\boldsymbol{\beta}_{n=1,\dots,N}}(x) = \beta_{n,0} + \beta_{n,1}\cdot x + \cdots + \beta_{n,P}\cdot x^P$ of user-selected order P:

$$Y_{\boldsymbol{\beta}}(x) = \begin{pmatrix} \beta_{1,0} & \beta_{1,1} & \cdots & \beta_{1,P} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{N,0} & \beta_{N,1} & \cdots & \beta_{N,P} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ x \\ \vdots \\ x^P \end{pmatrix} \quad \text{with:} \quad \mathbf{y} = Y_{\boldsymbol{\beta}}(x) + \boldsymbol{\delta} \tag{9}$$

The calibration model $[\beta_{n,p}] = \boldsymbol{\beta}$ in (9) is linear in the model parameters $\beta_{n,p}$ which hence can be estimated by a standard MLR: $[\hat{\beta}_{n,p}]_{(N\times P+1)} = [y_{n,k}^{\mathrm{cal}}]_{(N\times K)} \cdot [x_{n,k}^{\mathrm{cal}}]^{\mathbf{T}} \cdot \left([x_{n,k}^{\mathrm{cal}}] \cdot [x_{n,k}^{\mathrm{cal}}]^{\mathbf{T}}\right)^{-\mathbf{1}}$ with

$$[x_{n,k}^{\mathrm{cal}}] = \begin{pmatrix} 1 & \cdots & 1 \\ x_1^{\mathrm{cal}} & \cdots & x_K^{\mathrm{cal}} \\ \vdots & \cdots & \vdots \\ {x_1^{\mathrm{cal}}}^P & \cdots & {x_K^{\mathrm{cal}}}^P \end{pmatrix} \tag{10}$$

Such a calibration requires $K \geq P+1$ histograms obtained from cell cultures grown under different known ambient conditions, i.e. nutrient concentrations $c^{\mathrm{cal}}_{k=1,\dots,K} = x^{\mathrm{cal}}_{k=1,\dots,K}$. Utilizing $\widehat{\boldsymbol{\beta}}$ (10), a Predictor Surface can be visualized by calculating $\mathbf{Y} = \widehat{\boldsymbol{\beta}} \cdot \begin{pmatrix} 1 \\ c_{\mathrm{visual}} \\ \vdots \\ c^{P}_{\mathrm{visual}} \end{pmatrix}$ with simulated values covering the calibration range $c_{\mathrm{visual}} = \min(c_{\mathrm{cal}}), \dots, \max(c_{\mathrm{cal}})$ in sufficiently fine steps. Such surfaces themselves are not involved in subsequent prediction steps though and only serve for visual inspections of how histograms and thus cells change as a function of an environmental parameter (nutrient concentration).

The prediction step should *not* be based on another MLR estimating an unknown $\mathbf{x}_{\mathrm{unk}} = (1 \quad x \quad \cdots \quad x^P)^{\mathbf{T}}_{\mathrm{unk}}$ from a measured histogram $\mathbf{y}_{\mathrm{unk}}$ via $\hat{\mathbf{x}}_{\mathrm{unk}} = \left(\widehat{\boldsymbol{\beta}}^{\mathbf{T}} \cdot \widehat{\boldsymbol{\beta}}\right)^{-\mathbf{1}} \cdot \widehat{\boldsymbol{\beta}}^{\mathbf{T}} \cdot \mathbf{y}_{\mathrm{unk}}$. This MLR fit would estimate every element of the vector $(1 \quad x \quad \cdots \quad x^P)^{\mathbf{T}}_{\mathrm{unk}}$ independently which is obviously not the case. In reference [39], this has led to low precision and accuracy. Nonetheless, the least-square principle is applied to predict $\hat{x}^{\mathrm{unk}}$ via minimizing

$$\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right) = \|\boldsymbol{\delta}\|_2^2 = \left\|\mathbf{y}^{\mathrm{unk}} - \mathbf{Y}_{\widehat{\boldsymbol{\beta}}}\left(\hat{x}^{\mathrm{unk}}\right)\right\|_2^2 = \sum_{n=1}^{N} \frac{\left[y_n^{\mathrm{unk}} - Y_{\widehat{\boldsymbol{\beta}}_n}\left(\hat{x}^{\mathrm{unk}}\right)\right]^2}{\sigma_n^2} \qquad (11)$$

The necessary condition for a minimum is $\frac{\partial \mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)}{\partial \hat{x}^{\mathrm{unk}}} = 0$ from which a univariate polynomial in $\hat{x}^{\mathrm{unk}}$ of order $2 \cdot P - 1$ is derived. In subsequent steps, the definition $b_{n,0} = \frac{1}{\sigma_n^2} \cdot \left(y_n^{\mathrm{unk}} - \hat{\beta}_{n,0}\right)$ and $b_{n,p} = -\frac{1}{\sigma_n^2} \cdot \hat{\beta}_{n,p}$ for $p \geq 1$ will make the notations more concise. Applying the chain rule of differentiation and making the summation over n the inner-most loop in the second step of equation (12) results in the following:

$$\begin{aligned} \frac{d\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)}{d\hat{x}^{\mathrm{unk}}} = 0 &= -2 \cdot \sum_{n=1}^{N} \left[\sum_{p_1=0}^{P} b_{n,p_1} \cdot \left(\hat{x}^{\mathrm{unk}}\right)^{p_1}\right] \cdot \left[\sum_{p_2=1}^{P} p_2 \cdot b_{n,p_2} \cdot \left(\hat{x}^{\mathrm{unk}}\right)^{p_2-1}\right] \\ &= \sum_{p_1=0}^{P} \sum_{p_2=1}^{P} \underbrace{\sum_{n=1}^{N} b_{n,p_1} \cdot p_2 \cdot b_{n,p_2}}_{=B_{p_1,p_2}} \cdot \left(\hat{x}^{\mathrm{unk}}\right)^{p_1+p_2-1} = \sum_{p_1=0}^{P} \sum_{p_2=1}^{P} B_{p_1,p_2} \cdot \left(\hat{x}^{\mathrm{unk}}\right)^{p_1+p_2-1} \end{aligned} \qquad (12)$$

The $2P-1$ roots [108], [109] of the polynomial (12) minimize (or maximize) $\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)$ and are thus candidates for the wanted $\hat{x}^{\mathrm{unk}}$. Obviously, complex roots have no chemical meaning and can be discarded. In some applications when $\hat{x}^{\mathrm{unk}}$ represents a concentration for example, negative roots can be omitted as well. Furthermore, only those root for which

$$\frac{d^2\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)}{d\hat{x}^{\mathrm{unk}\,2}} = 2 \cdot \sum_{p_1=0}^{P} \sum_{p_2=1}^{P} (p_1 + p_2 - 1) \cdot B_{p_1,p_2} \cdot \left(\hat{x}^{\mathrm{unk}}\right)^{p_1+p_2-2} > 0 \qquad (13)$$

correspond to a minimum of SSE; the other roots are maxima or saddle points and have thus to be excluded from the list of potential solutions. To pick the best among the remaining roots, the unknown's response vector $\mathbf{y}^{\mathrm{unk}}_{(N\times 1)}$ is 'reconstructed' from root $\#p$ denoted by $\hat{x}^{\mathrm{unk}}_{\#p}$ resulting in a vector $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}\left(\hat{x}^{\mathrm{unk}}_{\#p}\right)_{(N\times 1)} = \begin{pmatrix} Y_{\boldsymbol{\beta}_{n=1}}\left(\hat{x}^{\mathrm{unk}}_{\#p}\right) \\ \vdots \\ Y_{\boldsymbol{\beta}_{n=N}}\left(\hat{x}^{\mathrm{unk}}_{\#p}\right) \end{pmatrix} = \hat{\boldsymbol{\beta}} \cdot \begin{pmatrix} 1 \\ \hat{x}^{\mathrm{unk}}_{\#p} \\ \vdots \\ \hat{x}^{\mathrm{unk}\,P}_{\#p} \end{pmatrix}$. That root $\hat{x}^{\mathrm{unk}}_{\#p}$ which results in the highest correlation coefficient $\mathrm{corr}\left(\mathbf{Y}_{\hat{\boldsymbol{\beta}}}\left(\hat{x}^{\mathrm{unk}}_{\#p}\right), \mathbf{y}^{\mathrm{unk}}\right)$ between reconstructed and originally measured response vector is used as the prediction's outcome [40].

In order to introduce a slight noise reduction, the images' three color channels were averaged. The multi-2D Gaussian model function (8) that is fitted to images of J cell shadows requires solving nonlinear equation systems (7) with $J \cdot 6 + 1$ unknowns. In the given application, images (Figure 4) typically comprised $J \gg 100$ cells and it was found that using equation (8) in equation (7) takes a fairly long computation time per image and more importantly that the iterative solver [109] of such large nonlinear equation systems frequently derived an obviously suboptimum solution as it got stuck in local minima of the very high-dimensional function $\mathrm{SSE}\left(\hat{a}_0, \hat{A}_{j=1,\ldots,J}, \hat{B}_{j=1,\ldots,J}, \hat{C}_{j=1,\ldots,J}, \hat{D}_{j=1,\ldots,J}, \hat{x}_{0\,j=1,\ldots,J}, \hat{y}_{0\,j=1,\ldots,J}\right)$. For improving computation speed and fit quality, each full image of 380 x 300 pixels was segmented into 285 subimages of 20 x 20 pixels each of which contained a few tens of cells at most. The information about cell sizes (3) and shapes (4) extracted from these subimages was then composed to a final result of the original full image. Thus, all cells were fit and analyzed. It was found that analyzing a large number of subimages is considerably faster than one global fit of the full image and that the fit quality measured by correlation coefficients improved. Furthermore, a few conditions for accepting a fit result were established: In some instances, the area between two closely spaced

cells was modeled as an up-pointing Gaussian $(A > 0)$ that was removed from the list of fits as an obvious artifact. Similarly, the individual cells in cell clusters were sometimes not discriminable and thus modeled as one exceptionally large or elongated cell; such cases were rejected as well. On a few occasions, a dead pixel or a noise spike were picked up as 2D Gaussian with unreasonably small footprint (3) which is not included into the histograms.

The fitting procedures, i.e. (8) incorporated into (7), histogram evaluations as well as the Predictor Surfaces calibration (10) and predictions (12) were programmed in C/C++ using the gcc 4.7.0 compiler [113] under Linux CentOS 5.8 and run on a 3.47GHz Intel Xeon computer. For solving systems of nonlinear equations (7), MLR calibration (10), and root finding (12), the GNU Scientific Library (version 1.15, [109], [114]) was employed which in turn was built on LAPACK (version 3.4.1, [115]) and the ATLAS library (version 3.8.4, [116]) for linear algebra computations [40].

3.1.4. Modeling Environmental Effects on Cells' Physical Parameters

Microalgae cells are major component in the oceanic food web and while their chemical composition determines their nutritional value, their size distribution determines the amount of available biomass. Due to microalgae's ecological relevance and due to their susceptibility to environmental change, they have the potential to serve as an indicator for chemical shifts in marine ecosystems. Utilizing this potential requires new analytical methods for measuring their chemical composition as well as their size distributions. To this end, reference [39] utilized FT-IR spectroscopy to acquire the nutrient-dependent chemical signatures of three microalgae species. For nonlinear multivariate modeling of the relation between IR spectra and ambient chemical parameters (here: nutrient concentrations), Predictor Surfaces were introduced. The cells' physical appearance, i.e. their size and shape, has been measured by means of a novel image analysis technique [117]. It was found that cell size and shape distributions reflect environmental parameters as well. After adapting the aforementioned Predictor Surface algorithm for modeling spectroscopic information to modeling cell size and shape histograms, quantitation of ambient parameters based on these distributions became feasible. However, Predictor Surfaces are a purely empirical approach as they fit a polynomial to the concentration-

dependent values of each histogram bin individually and hence (i) they cannot derive a mathematical model describing the relation between the cells' physical appearance and ambient parameters such as nutrient concentrations. (ii) Furthermore the interrelation of $V > 1$ ambient parameters had not been incorporated into Predictor Surfaces [117]. Information on (i) and (ii) however is key for predicting how changing environments drive the amount of microalgal biomass, an important contributor to the trophic chain. Thus, the present study [41] builds on image analyses as presented in [117] but utilizes the concentration-dependent size and shape distributions for mathematical modeling of coupled environmental parameters.

In [40], it was found that the size distributions of microalgal cells can be described well by a normal distribution which is superimposed on a polynomial of order Q_S which accounts for artifacts. In order to derive a mathematical model Y_size for the cell counts y_size as a function of cell size S and concentrations $\mathbf{c}_{(V\times 1)}$ of V nutrients, the model parameters A, S_0, σ, and $\alpha_{0,\dots,M_\mathrm{S}}$ in (14) are not introduced as constants but as $\mathbf{c}$ –dependent:

$$Y_{size}(S,c) = A(c)\cdot\exp\left\{-\frac{\left(S-S_0(c)\right)^2}{2\cdot\sigma(c)^2}\right\} \; + \; \alpha_0(c)+\alpha_1(c)\cdot S+\cdots+\alpha_{Q_S}(c)\cdot S^{Q_S}$$

(14)

In the absence of additional information about the model parameters' concentration dependency, they can be approximated by polynomial (hyper) surfaces [109] of user-selected order P. (This polynomial order P is not to be confused with the background's polynomial order Q_S.) In the experimental section, $V = 2$ will be considered and thus the polynomial surface

$$S_0(c_1,c_2) = \gamma_{S_{0,0}}+\gamma_{S_{0,1}}\cdot c_1+\gamma_{S_{0,2}}\cdot c_2+\gamma_{S_{0,3}}\cdot c_1^2+\gamma_{S_{0,4}}\cdot c_1\cdot c_2+\gamma_{S_{0,5}}\cdot c_2^2+\cdots$$

(15)

describes the concentration dependency of the most common cell size. The constant fit parameters $\gamma_{S_0,w}$ need to be derived via nonlinear least-squares regression. Similarly, $\sigma(\mathbf{c}) = \gamma_{\sigma,0}+\gamma_{\sigma,1}\cdot c_1+\gamma_{\sigma,2}\cdot c_2+\cdots$ describes the concentration dependency of the size distribution's width and $A(\mathbf{c}) = \gamma_{A,0}+\gamma_{A,1}\cdot c_1+\gamma_{A,2}\cdot c_2+\cdots$ the concentration dependency of the normal distributions' maximum height. Altogether, the model function (14) comprises $[3+(1+Q_\mathrm{S})]\cdot W$ model parameters γ with $W = W(Q_\mathrm{S},P)$ [118].

Equation (14) is fundamentally different from the Predictor Surfaces [40] as it models the entire size distribution rather than one histogram bin at a time. Furthermore, $Y_{\mathrm{size}}(S, \mathbf{c})$ considers cell size and concentration information in one fit rather than in two subsequent fits.

In [40], the cell shape was introduced as a cell's height-to-width ratio R whose distribution Y_{shape} could be well modeled by an exponentially decaying function. A species specific decay parameter $\mu(\mathbf{c})$ (16) describes how strongly a given species' shape deviates from being circular (i.e. $R = 1$). Again, artifacts in the images from which these histograms had been derived required an additional background polynomial of order Q_{R}. Equation (16) has been introduced here to concurrently model cell shape and concentration information and thus to gain insights into the concentration dependency of $\mu(\mathbf{c})$ [41].

$$Y_{shape}(R, c) = E(c) \cdot \exp\{-\mu(\mathbf{c}) \cdot (R - R_0(\mathbf{c}))\} + \beta_0(c) + \beta_1(c) \cdot R + \cdots + \beta_{Q_R}(c) \cdot R^{Q_R} \quad (16)$$

3.2. Results of Monitoring Cell Size and Shape

These image analysis tools can then either be used as a standalone technique in well-defined measurement situations or to generate complementary information for enhancing the specificity of chemical sensors. Microalgae cells have been chosen as proof-of-principle system and the nutrient sources and concentrations in the culturing medium as ambient chemical parameters.

This results section covers three main aspects: First, the novel method's applicability is demonstrated. Then, qualitative analyses of cell size and shape are presented to investigate these physical cell parameters' feasibility to discriminate cell species and ambient chemical conditions. In a third step, the cells' adaptations were utilized to quantify nutrient concentrations in the cells' environment.

3.2.1. Assessing the Fit Quality

First, the appropriateness of a chosen model function $Z(x, y)$ (8) for fitting the given image data has to be tested because the size (3) and ellipticity (4) of cells are derived from the resulting fit parameters. Selected examples of raw imaging data and the corresponding fits are shown in (Figure 6).

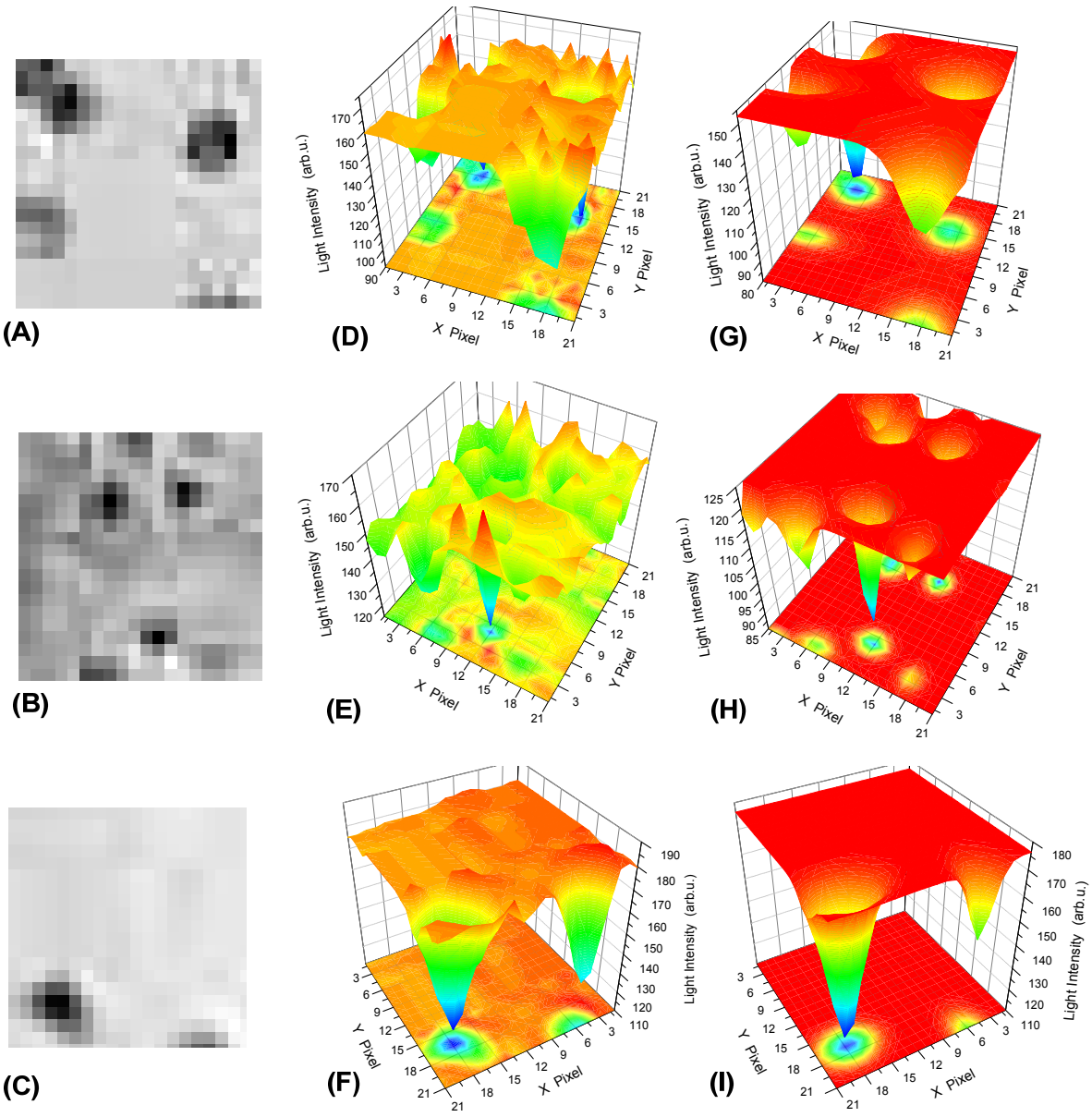


Figure 6. (A) – (C) gray-scale subimages (20 x 20 pixels, raw data) from top-to-bottom *Dunaliella parva*, *Nannochloropsis oculata*, *Dunaliella salina* cultured under 0.549 mM nitrate as N-source and 2.071 mM bicarbonate as C-source; (D) – (F) same as left column but in 3D-representation; (G) – (I) Gaussian fits (6) of the middle column's raw images.

Despite the considerable noise level, most cells were found and fitted to a visually good agreement with the raw data. What appears to be a misfit (panel (A) top left corner) is a small cluster of three cells from which only one cell fit was returned but with reasonable fit parameters (cf. panel (G)). In order to gain an overview of the fit quality in general, a large number of fits need to be tested objectively for the different species and nutrient situations. For this purpose, correlation coefficients have been employed which have been obtained after unfolding the raw image, a matrix, and the fit image, another matrix, into two vectors of 20 x 20 = 400 elements each. Fits that resulted in a correlation coefficient <0.7 were assumed to be impacted by artifacts in the images such as cell clusters, non-cell objects and were thus discarded. Furthermore, for the reason that the iterative algorithm [109] for solving nonlinear equation systems (7) may get stuck in a suboptimum solution, it is restarted with different initialization values until a minimum correlation (here: 0.9) was achieved or the preset number of repeats (here: 50) was reached. This minimum correlation was chosen empirically and reflects the rather low signal-to-noise in the given application. Since each full image was split into 285 subimages and despite rejecting a few fits, a considerable number of correlation coefficients were obtained per full image. These correlation coefficients were compiled into one histogram per full image, averaged over multiple images taken from replicate samples, and depicted in (Figure 7) for different species, nutrient sources, and concentrations.

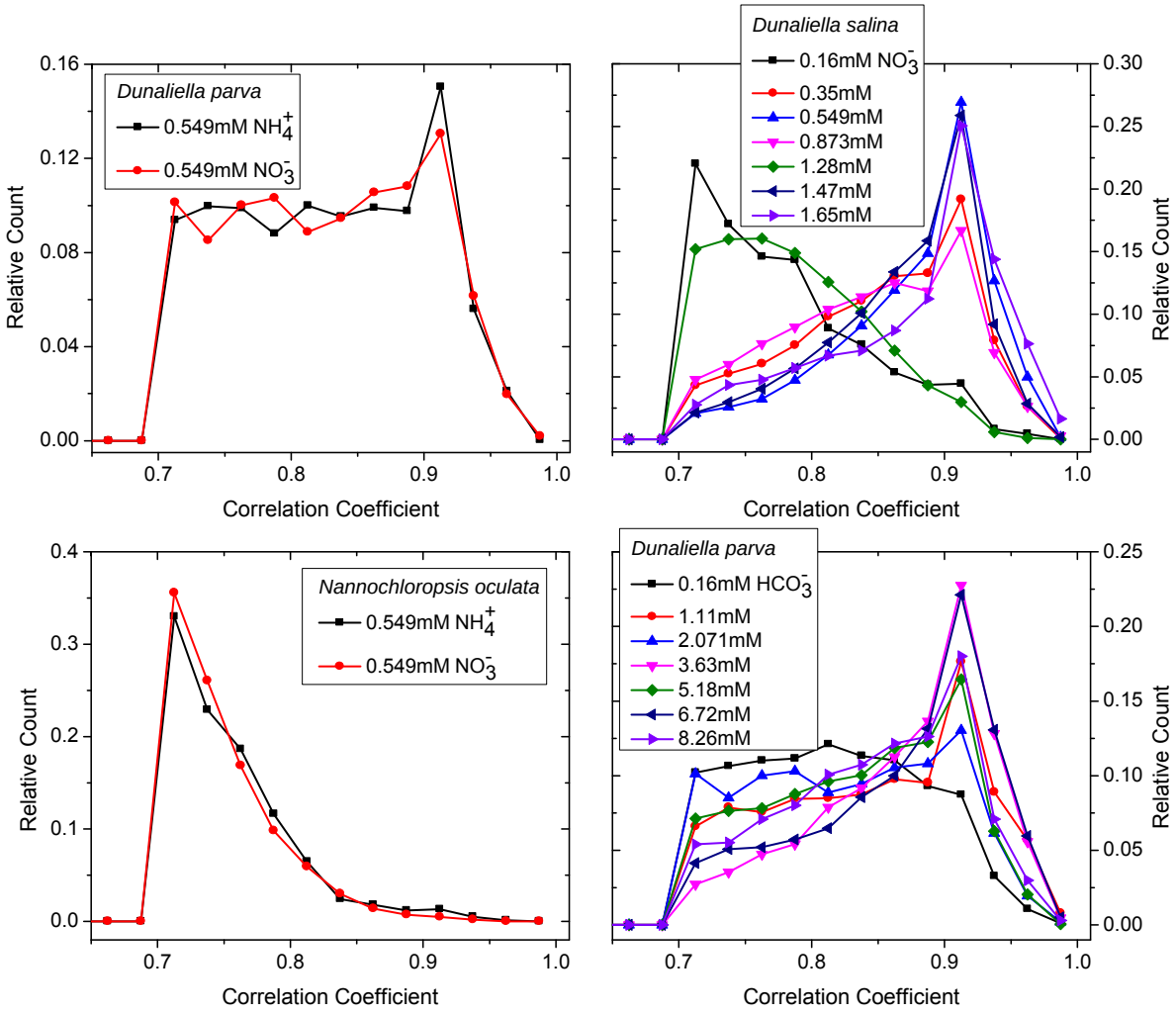


Figure 7. The overall fit quality was assessed via correlation coefficients between subimages (Figure 6 (D) – (F)) and their fits (panels (G) – (I)).

Most of the correlation coefficient histograms (Figure 7) derived from the two *Dunaliella*'s peak near a correlation coefficient of 0.9 and drop towards lower correlations. This indicates a high overall fit quality. Only for two of *Dunaliella salina*'s correlation histograms, the majority of correlation coefficients were obtained for considerable lower but still acceptable values. One source of low correlation might have been a rather high cell density in the images causing cells to cluster. Due to *Nannochloropsis oculata* cells being rather small (Figure 4), their area is comparable to the length scale of noise fluctuation which then has a stronger influence on the fits (cp. Figure 6 (E) vs. (D) and (F)). This is one explanation for this species' generally lower correlation coefficients (Figure 7, top left). Overall, however, the correlation coefficients and visual inspections (Figure 6) indicate that the obtained fit parameters and in particular σ_x and σ_y reflect the cell size and shapes well.

3.2.2. Discrimination of Cell Species and Chemical Environments

After computing cell sizes (3) and shapes (4) for each of the accepted cell fits, these cell parameters have been compiled into histograms as outlined in section 3.1.2. Multiple full images (380 x 300 pixels) have been acquired for each cell class and from each of these full images, one histogram has been prepared. After averaging all class-specific histograms bin-wise and calculating the corresponding standard deviations at each bin (cf. e.g. Figure 8 (A)), histograms can be compared by means of t-histograms in order to determine size and shape differences among cell types and/or nutrient situations.

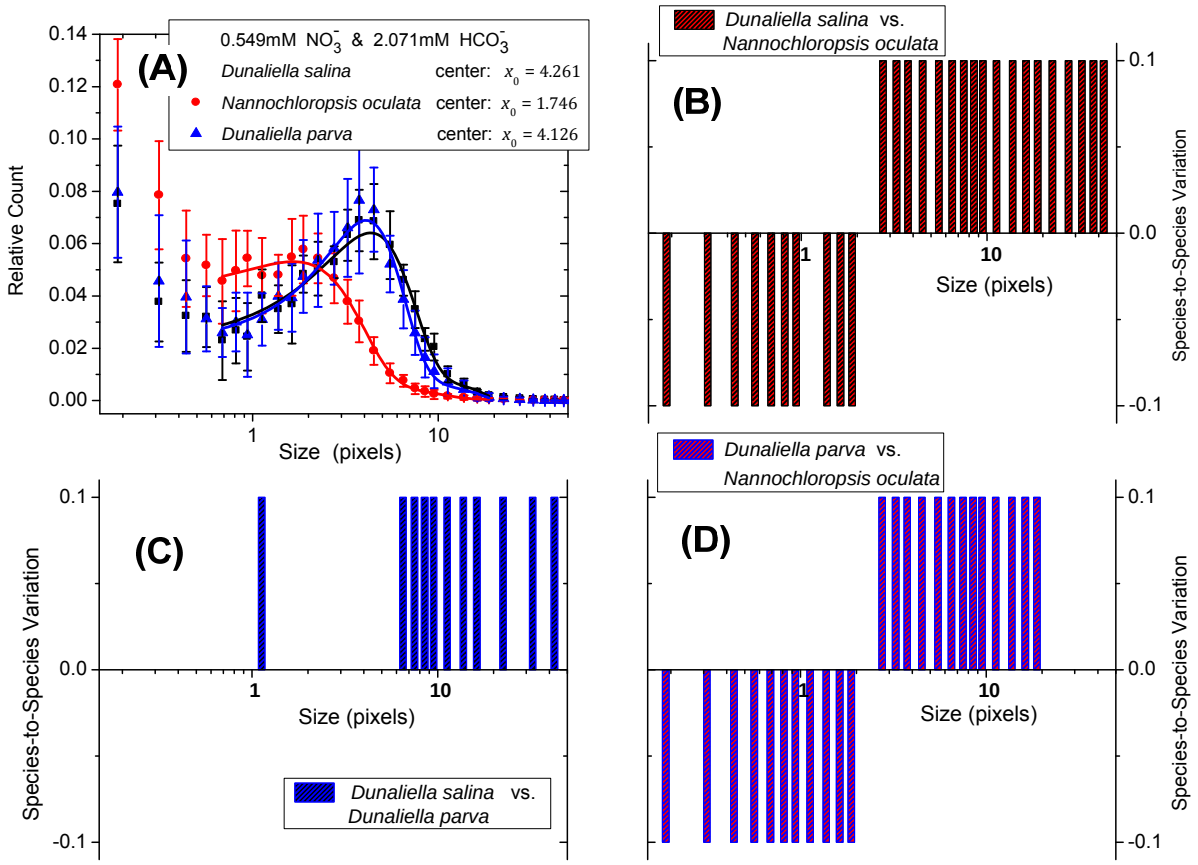


Figure 8. (A) Species discrimination based on cell size distributions – different species feature histogram maxima at different size. Panel (B), (C), (D) comparing species pairwise by means of t-tests applied to the individual histogram bins; see text – the nutrient situation chosen for this and the following graph are the standard conditions [88].

According to Figure 4 and Figure 6, *N. oculata* cells are considerably smaller than both *D. parva* and *D. salina* which in return are of similar size. These results from these visual inspections must be reflected in the three species' size histograms. Figure 8 (A) gives a first impression of the fit algorithm's performance: The maxima's positions stated in panel (A) were obtained by fitting a Normal distribution superimposed by a second order polynomial describing the 'background counts' to the histograms. This normal distribution's center x_0 are the most common size for a certain species and are consistent with the expectations, i.e. *N. oculata* is considerably smaller than the two *Dunaliellas* which in return are very similar in size with a slightly large cell size for *D. salina*. At first glance, it may be surprising that these cell sizes only amount to a few pixels but one has to consider that the σ_x and σ_y in equation (8) measure the 2D Gaussian's width at a height of $\sim 37\% \cdot A$ where it already has narrowed. One also has to keep in mind though that for this application a representative measure for a cell's size is sufficient and that its absolute size is not needed.

Given the considerable error bars Figure 8, panel (A)) which result from the considerable cell-to-cell size fluctuations, statistical testing was deemed to be required and t-histograms have been introduced for this purpose in section 3.1.2 . The panels (B) - (D) in Figure 8 pairwise compare the species' size histograms for differences. As expected, the t-histograms of either Dunaliellas versus N. oculata state that there are fewer counts in the Dunaliella histograms for cell sizes below ~2 pixels and more counts above sizes of ~3 pixels. Within this window of ~2-3 pixels, the histograms of the species cross each other and thus do not contain significantly different cell counts. For larger cells (>20 pixels) only D. salina was found to have a higher count than *N. oculata* this was confirmed in panel (B); in panel (C) which compares the two *Dunaliellas* to each other and for cells that large, *D. salina* has a significantly higher count than *parva*. Thus, all panels in Figure 8 are consistent in that *D. salina* is larger in cell size closely followed by *D. parva*; *N. oculata* was found to be less than half the size of the *Dunaliellas* at 0.549 mM NO_3^- and 2.071 mM HCO_3^-.

A similar study was performed on shape histograms which measure the overall ellipticity of the cells. No pronounced maxima were found in shape histograms (Figure 9), panel (A)) - all showed a featureless decay of the 'relative counts' with increasing height-to-width ratio. In the given application, a steeper decline means a more homogeneous distribution of the cell shapes or more precisely a larger percentage of circular cells. It has been hypothesized that these distributions' decay is characteristic and contains species-specific information. It was also investigated whether this 'decay characteristic' reflects changes in nutrient availability.

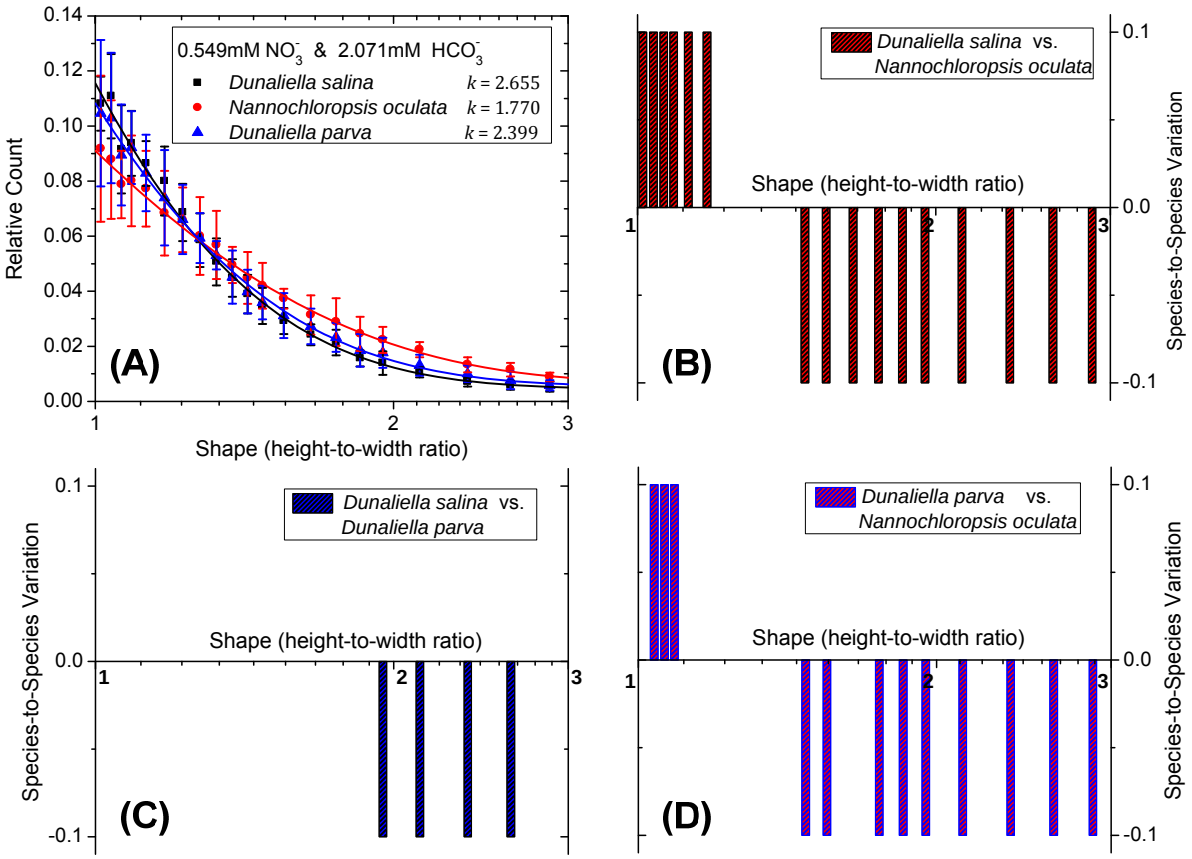


Figure 9. same analyses as in Figure 2 but applied to shape histograms for assessing the species' different cell shapes (4); 0.549 mM NO_3^- and 2.071 mM HCO_3^- have been used to culture these cells.

For testing these hypotheses, these shape histograms were fitted to an empirically chosen function $y(x)$ (17) in which x represents the height-to-width ratio (4) and y the 'relative count' in a certain bin [40].

$$y(x) = U \cdot \exp\{-k \cdot (x - x_0)\} + a_0$$

(17)

In equation (17), the parameter U is a measure for the counts at the histograms' left end and thus reflects the percentage of cells with a circular shape. The 'decay constant' k describes how quickly the counts drop with increasing height-to-width ratio and serves here as a measure for the shape homogeneity of a species. Furthermore, x_0 and a_0 are additional fit parameters to shift the curve in x and y direction; a constant offset a_0 is of importance in order to incorporate a certain level of background counts into the fit.

In (Figure 9 (A)), the counts in *N. oculata*'s shape histogram decay slowest with increasing shape ratio (4) and thus this species has the least homogeneous shape distribution. From a different perspective, both *Dunaliellas*' cell counts are higher for low shape values and thus these species are overall more circular compared to *N. oculata*'s cells. The t-histograms (panels (B) and (D)) comparing either one of the *Dunaliellas* to *N. oculata* depict that the *Dunaliellas* have a significantly higher count for height-to-width ratios $\lesssim 1.2$ and a significantly lower cell count $\gtrsim 1.4$. This means that the *Dunaliella* cells are in general more circular than the cells of *N. oculata* which are thus more elliptical. In the 'gap' between $\mathrm{shape} \approx 1.2$ and ≈ 1.4 in which the t-histogram is zero, both species have an equivalent percentage of slightly elliptical cells. However, for *D. parva* this gap is wider than for *D. salina* and hence it can be concluded that *D. parva* and *N. oculata* are more similar than *D. salina* and *N. oculata*. Since both *Dunaliellas* have fewer counts in bins with large height-to-width ratios and *D. salina* has fewer counts than *D. parva* for elliptic cells (panel(C)) it can be deduced that *D. salina* is most homogeneous in shape and most circular followed in this order by *D. parva* and *N. oculata*. Note that these statements are shape independent. In conclusion, the results depicted in (Figure 9) demonstrate that the cell shape can be used for species discrimination despite the fact that the differences in panel (A) appear to be minute and covered by error bars.

Next, it was investigated by means of *Dunaliella salina* whether the fit parameters in equation (17) are concentration c dependent (Figure 10) depicts these fit parameters $\hat{U}(c)$, $\hat{k}(c)$, $\hat{x}_0(c)$, and $\hat{a}_0(c)$ as a function of nutrient concentration (c =[HCO_3^-], [NH_4^+], and [NO_3^-], respectively)

along with linear fit polynomials. These linear fits were tested for their significance by means of ANOVAs [101] at 95% significance level. It was found that none of the four fit parameters was linearly related to one of the three nutrients' concentrations and it has to be concluded that these nutrients have no impact on the shape of *D. salina*. Whether the nutrient concentrations determine the cell size is investigated in a quantitative approach in the next section.

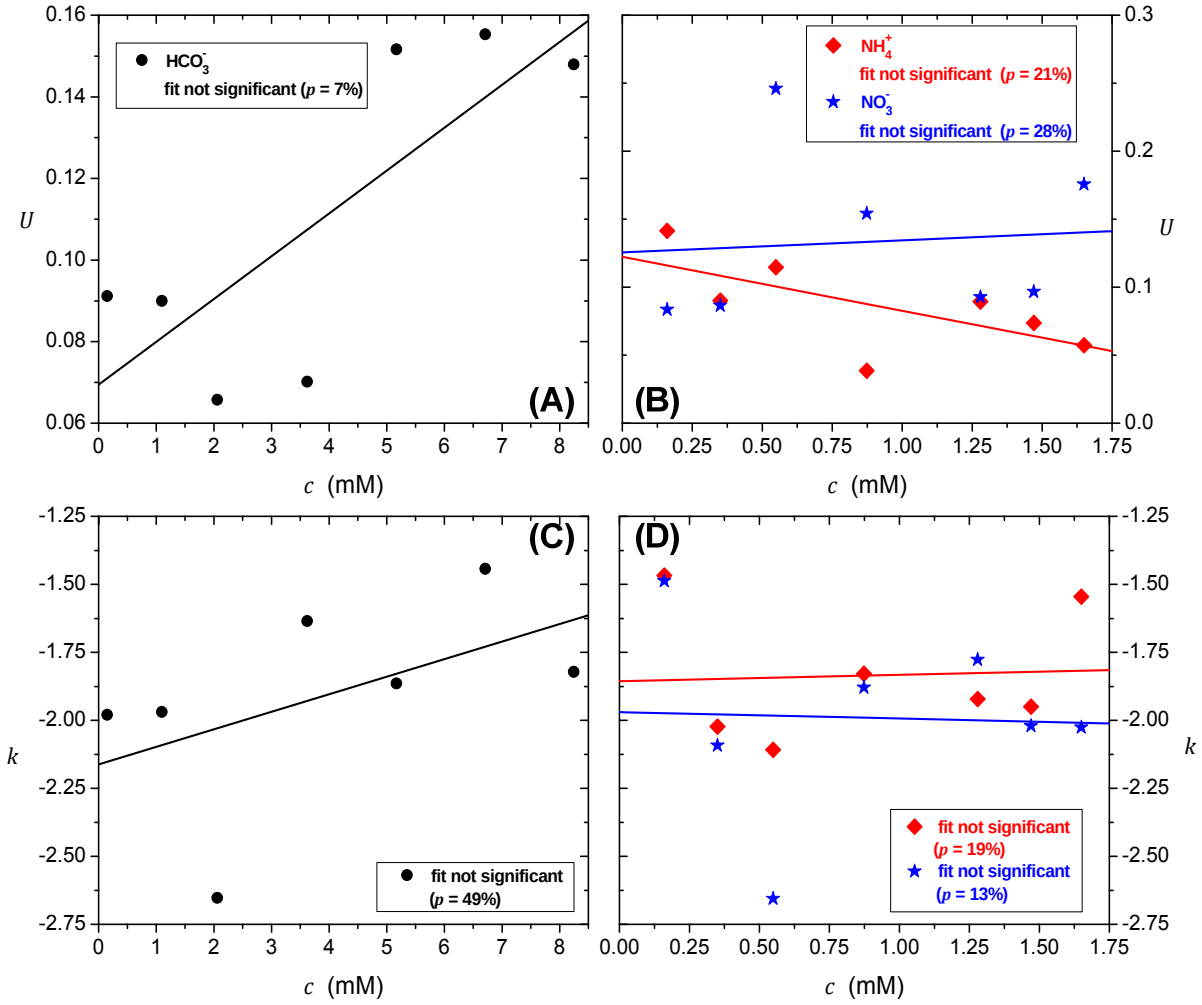


Figure 10. Analyses of nutrient concentrations c on *D. salina*'s shape histogram. For different nutrients and their concentrations c, the shape histograms have been fitted by (17) (cf. panel (A) in Figure 9); the resulting fit parameters $\hat{U}$ and $\hat{k}$ have been plotted over c of the respective nutrient. Not shown are $\hat{x}_0$ and $\hat{a}_0$ as they only shift the histograms but have no influence on the histograms' decay properties.

3.2.3. Concentration Prediction via Cell Size and Shape

If the prediction of nutrient concentrations in the culturing media based on size histograms is precise, it can then be concluded that the nutrient concentrations have a distinct and reproducible impact on the cell size. Building quantitative calibration models based on Predictor Surfaces (section 3.1.3) required that the calibration concentrations listed in section 2.3 were split into a calibration set and an independent test set. For all species, the bicarbonate Predictor Surface was built using replicate samples cultured under 0.16mM, 2.071mM, 5.18mM, and 8.26mM leaving the samples grown with 1.11mM, 3.63mM, and 6.72mM bicarbonate for prediction. Similarly, the ammonium and nitrate, respectively, calibration models were built utilizing the samples grown with 0.16mM (*Dunaliellas* only), 0.35mM (*N. oculata* only), 0.549mM, 1.28mM and 1.65mM; the remaining cultures were utilized to test the prediction capabilities of the respective models, i.e.: 0.35mM (*Dunaliellas* only), 0.873mM, and 1.47mM.

For visual inspection, the Predictor Surfaces with $P = 2$ in equation (9) were computed (cf. below equation (10) and displayed in (Figure 11). For all three selected species and all three investigated nutrients there are strong nonlinear responses in the size distributions to the nutrient concentrations. Such information by itself can be useful for biologists to gain a better understanding of the cells' size adaptation mechanism to their chemical environment. Furthermore, Partial Least-Squares (PLS) [111] has been utilized as chemometric reference methods to put the precision of Predictor Surfaces in perspective. For determining the numbers of latent variables boot-strapping has been employed which resulted in one to three latent variables depending on the species and nutrient. In cases when PLS was utilized more than one latent variables, PLS was able to approximate the observed nonlinearities in the data sets.

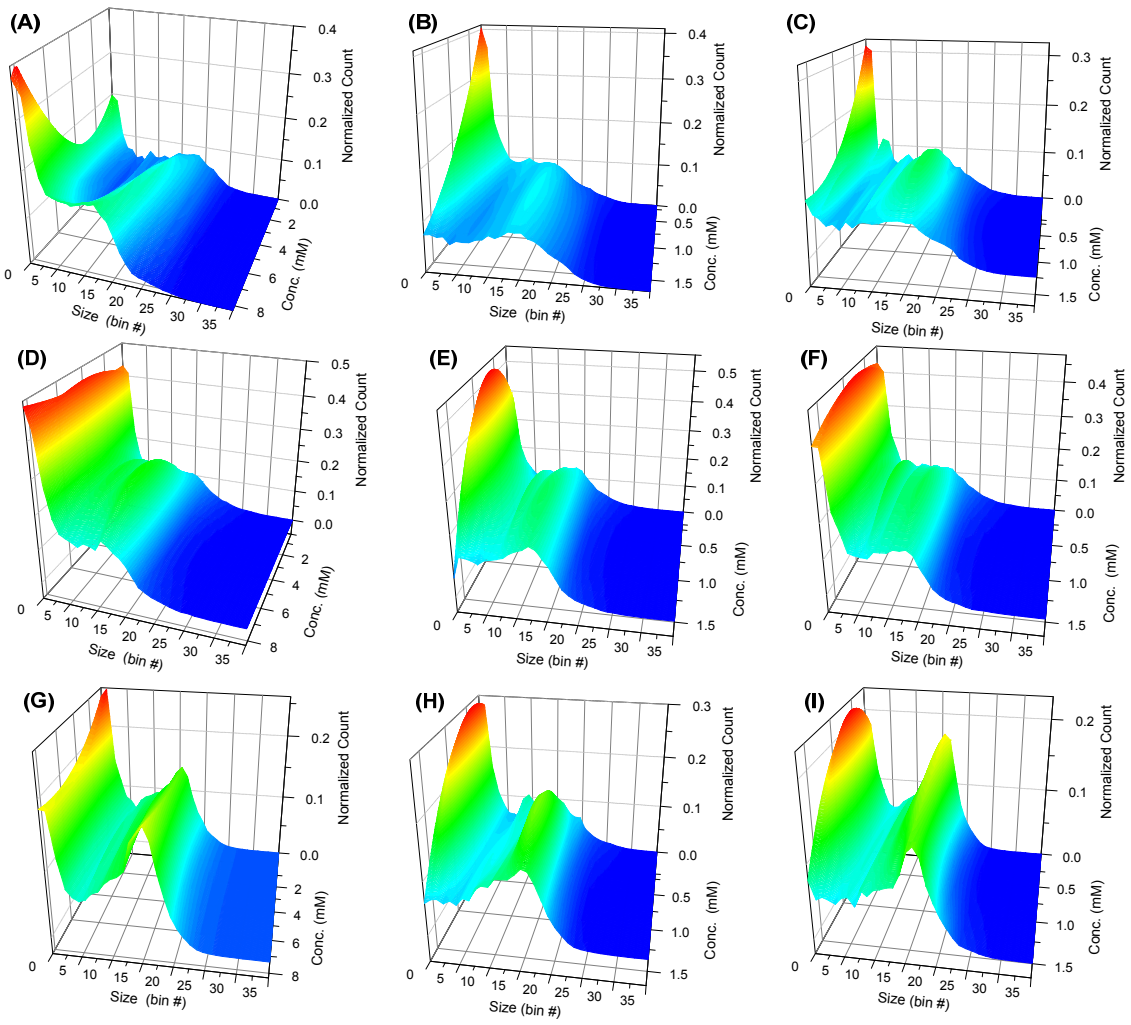


Figure 11. Predictor Surfaces for size (3) histograms obtained from microalgae grown under varying nutrient concentrations; these Predictor Surfaces depict how the histograms change as a function of nutrient availability. *Dunaliella salina*: (A) carbon, (B) ammonium, (C) nitrate; *Nannochloropsis oculata*: (D) carbon, (E) ammonium, (F) nitrate; *Dunaliella parva*: (G) carbon, (H) ammonium, (I) nitrate; the Size-axes had to be given here in 'bin #' but they cover the same 'pixel-units' as panel (A) in Figure 8.

The predicted nutrient concentrations as obtained by the Predictor Surfaces and PLS are plotted in (Figure 12) versus the true concentrations which were known due to the preparation of the culturing media. It is important to recall that the nutrient concentrations as the chemical environmental parameter, were measured indirectly, i.e. via their impact on the cells' physical size distributions. The bicarbonate concentrations could only be predicted with an acceptable precision via the size histograms of *N. oculata* and only by means of Predictor Surfaces; bicarbonate prediction for both *Dunaliellas* failed. The ammonium and nitrate predictions were found to be comparable regarding precision for *N. oculata* and *D. parva*, for *D. salina* the Predictor Surfaces were considerably more precise than PLS. One reason for this might be that the size distributions for *D. salina* show more and stronger nonlinear features with increasing concentrations (top row in Figure 11) than for the other species. Predictor Surfaces appears to be more capable of modeling these nonlinearities than PLS which has to linearly interpolate such nonlinear responses by means of additional latent variables.

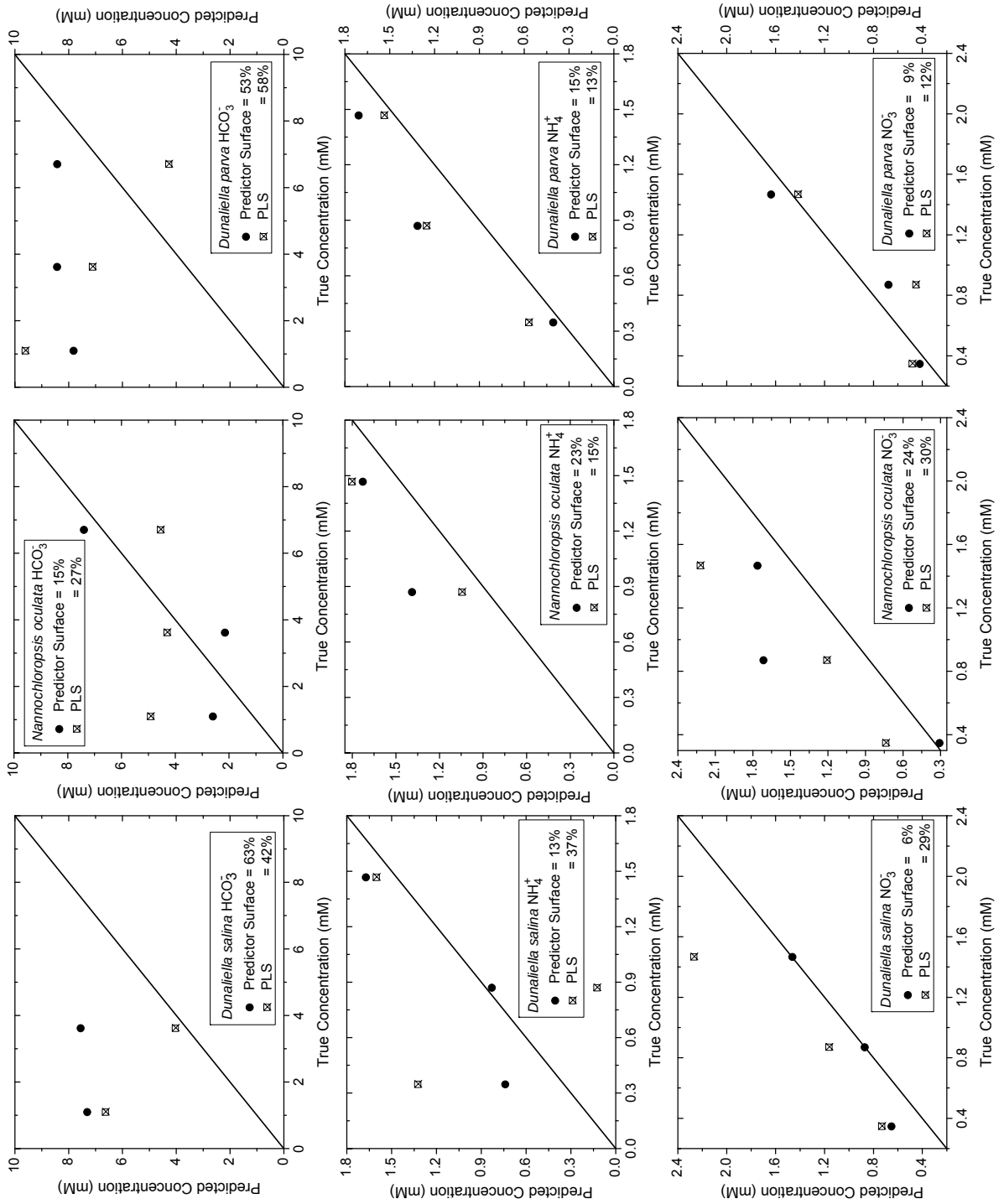


Figure 12. Predicted versus true nutrient concentration utilizing Predictor Surfaces (Figure 11) of polynomial order $P = 2$ (9) compared to PLS predictions; the stated error percentages are averaged absolute values of relative errors.

3.2.4. Analysis of Concentration-dependent Fitting

Size and shape distributions have been obtained as outlined in [40] from cultures of *Dunaliella salina, Dunaliella parva*, and *Nannochloropsis oculata* after being cultured under different $\mathbf{c} = (c_1, c_2)^{\mathbf{T}}$. Here, c_1 represents the concentration of inorganic carbon and c_2 the concentration of nitrogen dissolved in the growing medium; nitrogen has been provided via ammonium as well as via nitrate. By means of a Levenberg-Marquardt solver [109], the nonlinear, multivariate functions (14) and (16) with concentration dependent model parameters approximated by polynomial surfaces (15) were fitted to these experimental size and shape histograms, respectively. Results from preliminary and intermediate steps have been compiled.

The top row of Figure 13 depicts $S_0(c_1, c_2)$, $A(c_1, c_2)$, and $\sigma(c_1, c_2)$ obtained with $P = 2$ and $Q_S = 2$ for *Dunaliella salina* grown under different mixtures of C and NH_4^+. Panel (I) indicates that the most common cell size $S_0(c_1, c_2)$ (15) increases when either nutrient concentration increases. [C] obviously has a much stronger impact on the cell sizes than [NH_4^+] - but the ammonium impact becomes more pronounced with higher [C] as the slope in the [NH_4^+]-direction increases for higher [C] concentrations. Thus, the cells respond in a nonlinear way to the nutrient availability and both nutrients have a coupled impact. The corresponding normal distribution's maximum A decreases with increasing [C], while σ increases at the same time. Therefore, the cell size distribution becomes more spread out with increasing [C] and more shallow. For increasing ammonium concentration, the maximum and the distribution's spread slightly increases thus concluding that [C] has a stronger impact on this species' size than NH_4^+. Replacing *D. salina*'s nitrogen source with NO_3^- (panel (IV)) changed the concentration dependency of S_0 completely as the most common cell size now decreases with increasing [C] and [N]. To demonstrate that this trend is not dependent on the polynomial order, (V) shows the same trend for $(P = 1, Q_S = 1)$; however, the values for S_0 remain in a more realistic range. For *Dunaliella parva*, both nitrogen sources (panels (VII) and (VIII)) induce the same general trend in $S_0(c_1, c_2)$, i.e.: an increase with both [C] and [N] which is consistent with panel (I). The values for S_0 remain in the same range for both nitrogen sources but features a shallower concentration dependency for nitrate. The most common cell size of *Nannochloropsis oculata* also shows an impact of the nitrogen source (panel(VI)) where S_0 is more stable with [NH_4^+] than it is for [NO_3^-] which induces a considerable size increase for higher nitrate concentrations (panel (IX)).

In order to measure the fits' quality, Figure 14 depicts the correlation coefficients obtained for fitting (14) to the size distributions of all three microalgae species obtained for varying [C] and [NH_4^+] as well as [C] and [NO_3^-] combinations, respectively. These correlation coefficients were computed for different polynomial orders P (approximation level of the polynomial surfaces (15), [104]) as well as Q_S (modeling background counts (14)). First, it appears that the correlation coefficients in the left column are somewhat higher that those shown in the right column. Obviously, the cells cultured under [NH_4^+] rather than [NO_3^-] as the nitrogen source follows the chosen model function (14) more closely. Based on these correlation coefficients, $(P = 2, Q_S = 2)$ had been selected for the results displayed in Figure 13. The selection criterion was to increase P and Q_S until the correlation ceased to (considerably) improve – at that point, the model incorporated enough variance without introducing strong overfitting. Considering the level of natural sample-to-sample fluctuations (cp. Figure 15 and Figure 16), a correlation of mostly ≥0.9 is satisfactory and thus the conclusions drawn from Figure 13 are reliable [41].

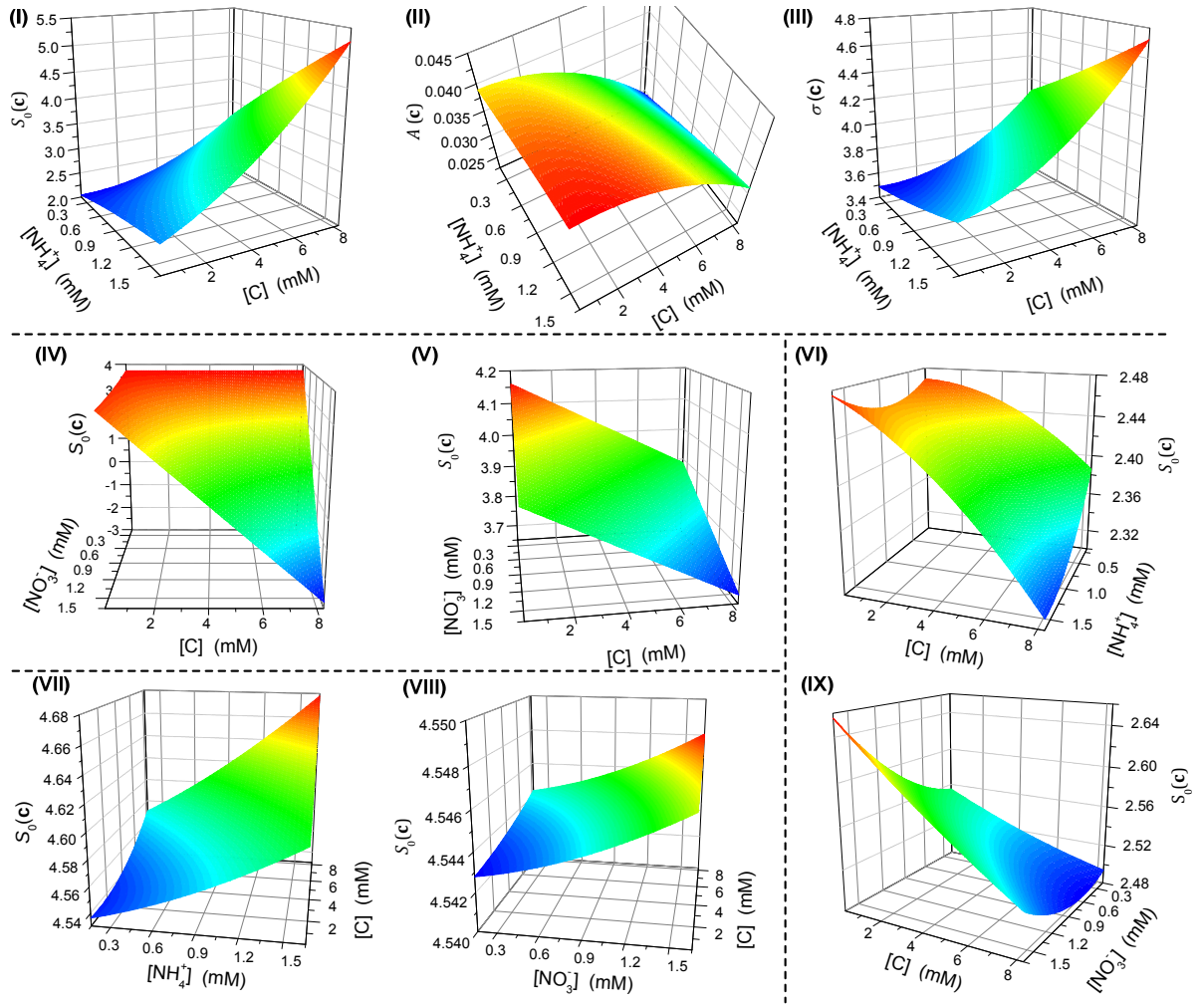


Figure 13. Cell Size fit results obtained with $P = 2, Q_S = 2$; **(i) – (III)** $S_0(c_1, c_2)$, $A(c_1, c_2)$, and $\sigma(c_1, c_2)$ (14), (15) for *Dunaliella salina* grown under varying [C] and [NH_4^+]; **(IV)** $S_0(c_1, c_2)$, for *Dunaliella salina* grown under varying [C] and [NO_3^-]; **(V)** same as (IV) but for $P = 1, Q_S = 1$; **(VI)** and **(IX)** $S_0(c_1, c_2)$ for *Nannochloropsis oculata* grown under varying [C] and [NH_4^+] as well as under varying [C] and [NO_3^-]; **(VII)** and **(VIII)** $S_0(c_1, c_2)$ for *Dunaliella parva* grown under varying [C] and [NH_4^+] as well as under varying [C] and [NO_3^-].

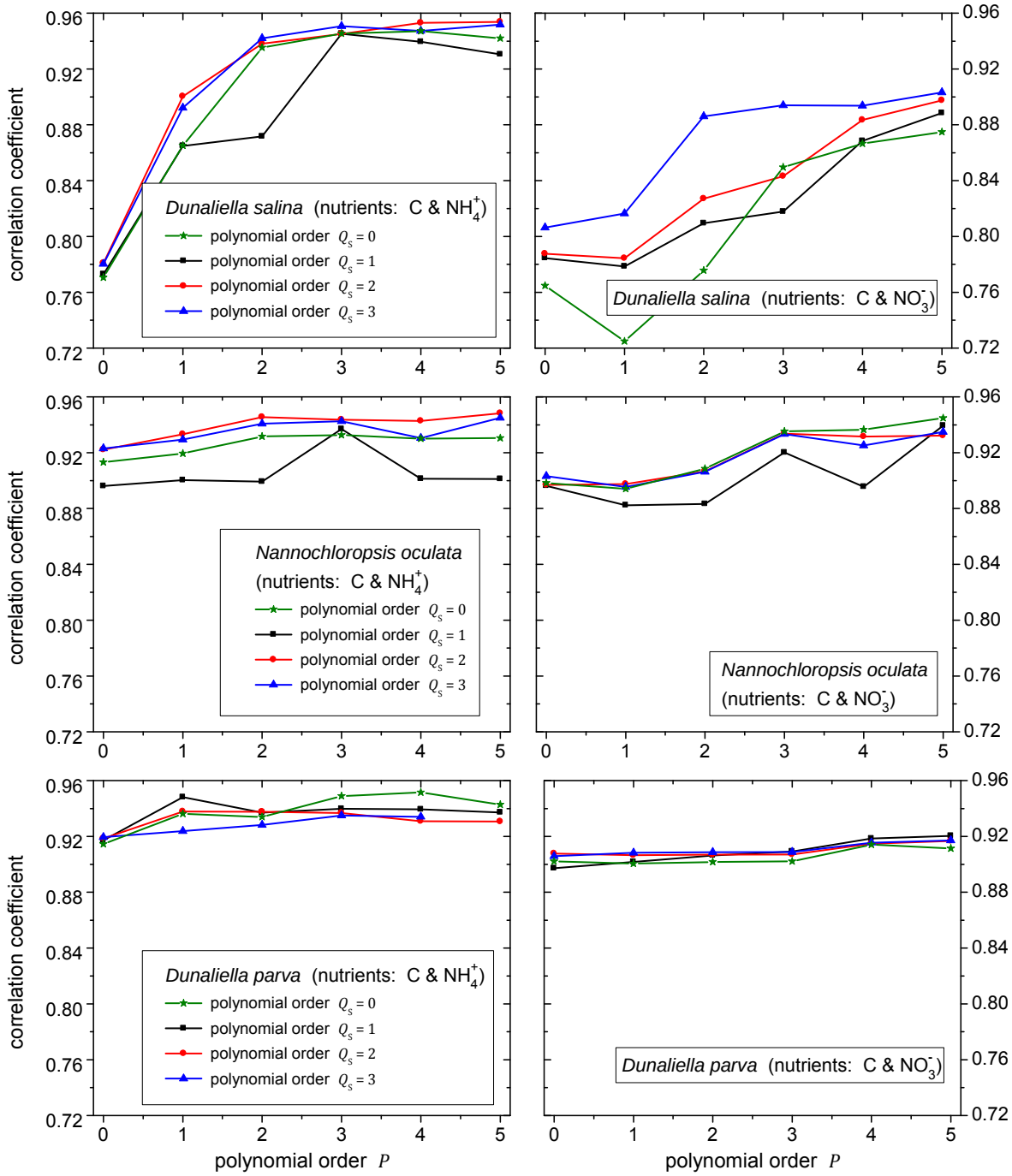


Figure 14. (top row) Correlation coefficients obtained for *Dunaliella salina* for different combinations of P and Q_S; (middle row) *Nannochloropsis oculata*; (bottom row) *Dunaliella parva*.

Figure 15 and Figure 16 compare selected fit results to the experimental data. Since the considered models are four-way data (S −axis, c_1 − axis, c_2 − axis, and y_{size} −axis, i.e. bins' counts), a graphical display of the model is not possible. In order to visualize the level of consistency between experimental data and the model function $Y_{\mathrm{size}}(S, \mathbf{c})$ (14), the latter has been computed for those $(c_1 \quad c_2)$ pairs for which experimental data were available. The results for *Dunaliella salina* grown under varying C and NH_4^+ concentrations (Figure 15) are reflecting the corresponding portion of the model function (14) well; especially in light of the replicate samples' considerable variability. As expected based on the correlation coefficients (Figure 14, top right), the model functions $Y_{\mathrm{size}}(S, \mathbf{c})$ and the size distributions for the same species cultured under NO_3^- (Figure 16) instead of NH_4^+ show a lower resemblance. While assessing these fits, one has to keep in mind, that all histograms for a species+nutrient series have to be fitted to *one* four-dimensional model function (14) rather than fitting one histogram at a time. In other words, each histogram corresponds to one $(c_1, c_2)-$ point in the (c_1, c_2) −plane and each histogram is a 'rod' along the S −direction within the (S, c_1, c_2) −cube.

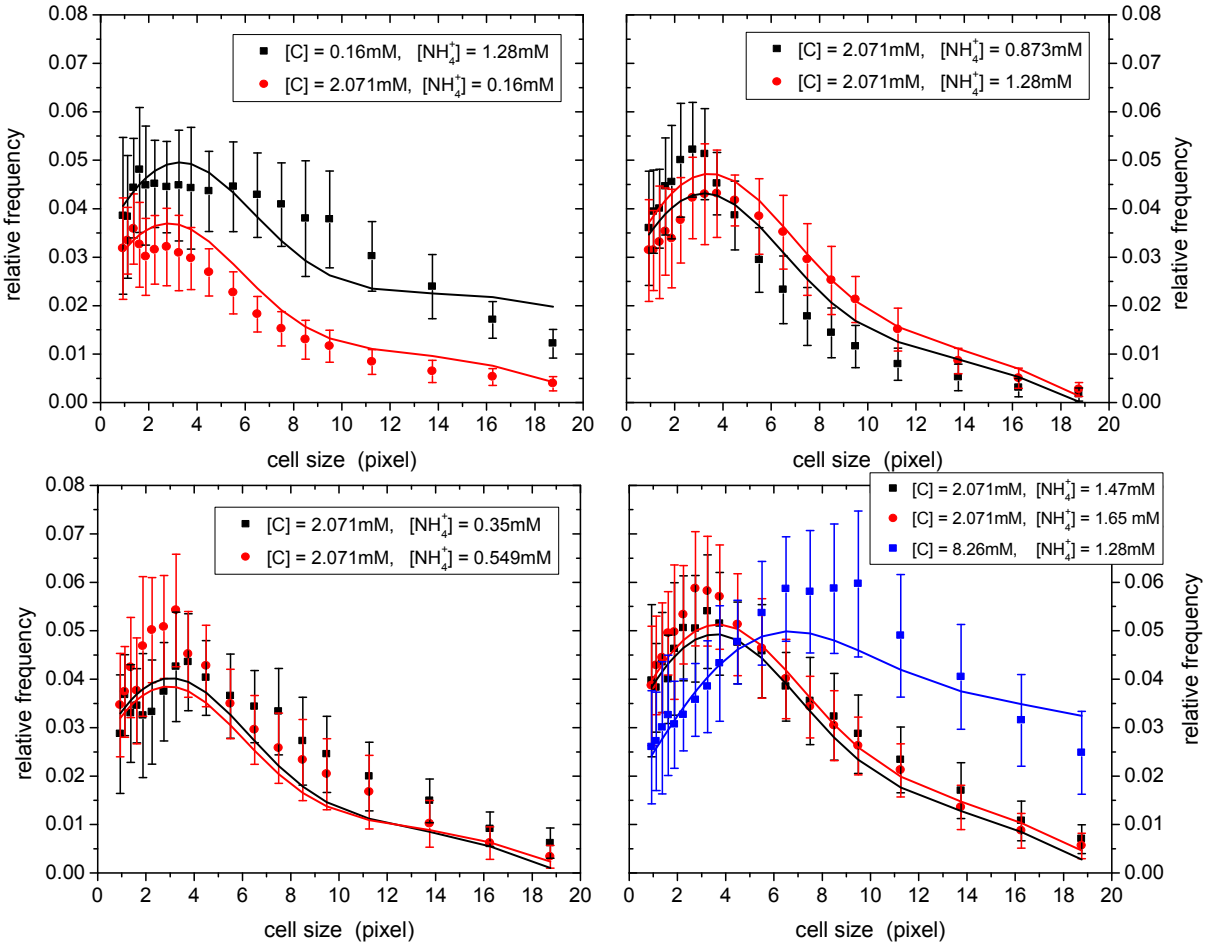


Figure 15. Comparing the fitted model $Y_{\text{size}}(S, \mathbf{c})$ (14) to the measured size S histograms obtained from *Dunaliella salina*, nutrients: C and NH_4^+ with $(P = 2, Q_S = 2)$; as described in [40], the sizes of the cell were determined from imaging data and thus are measured in pixels. The error bars have been obtained from images of independently prepared replicate samples – their values are due to the strong but natural in-class variability of these biological samples.

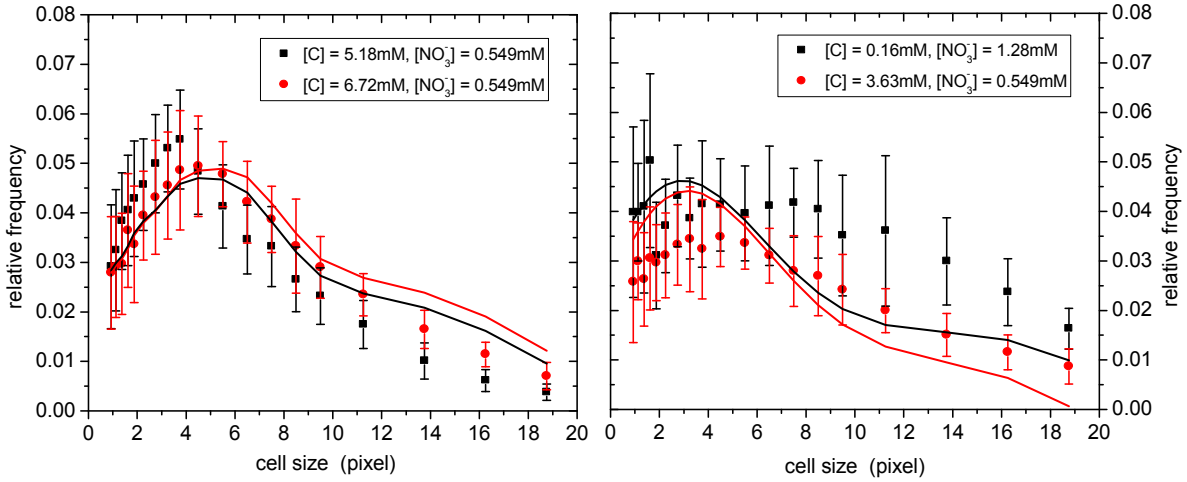


Figure 16. Same as Figure 15 but for nutrients C and NO_3^-; the left panel shows two fits with lower but acceptable quality. The fit quality of the examples in the right panel is below a desirable level.

Lastly, the dependency of the $S_0(c_1, c_2)$ −surfaces (15) on the polynomial orders P and Q_S have been analyzed. It has been expected that these surfaces do not change completely when utilizing slightly different polynomial orders. Figure 17 displays $S_0(c_1, c_2)$ results for *Dunaliella salina* grown under varying [C] and [NH_4^+] (Figure 14 top left, Figure 15). It must be noted that all (maybe with the exception of the $(P = 0, Q_S = 2)$-case) $S_0(c_1, c_2)$ values remain in the same order of magnitude. Furthermore, all surfaces with $P > 0$ show the same trend of increasing in both c_i −directions. Thus, the fits are not dominated by user-selected polynomial orders. Naturally though, the top row $(P = 0)$ cannot feature any concentration dependency because $S_0(c_1, c_2) = \gamma_{S_0,0}$ (15). In conclusion, the fit results are consistent and have an acceptable quality with a few exceptions (Figure 14 right). Thus, the conclusions to be drawn from the fits (14) and (16) and in particular the biological interpretations of the resulting surfaces such as $S_0(\mathbf{c})$ as well as $\mu(\mathbf{c})$ are meaningful.

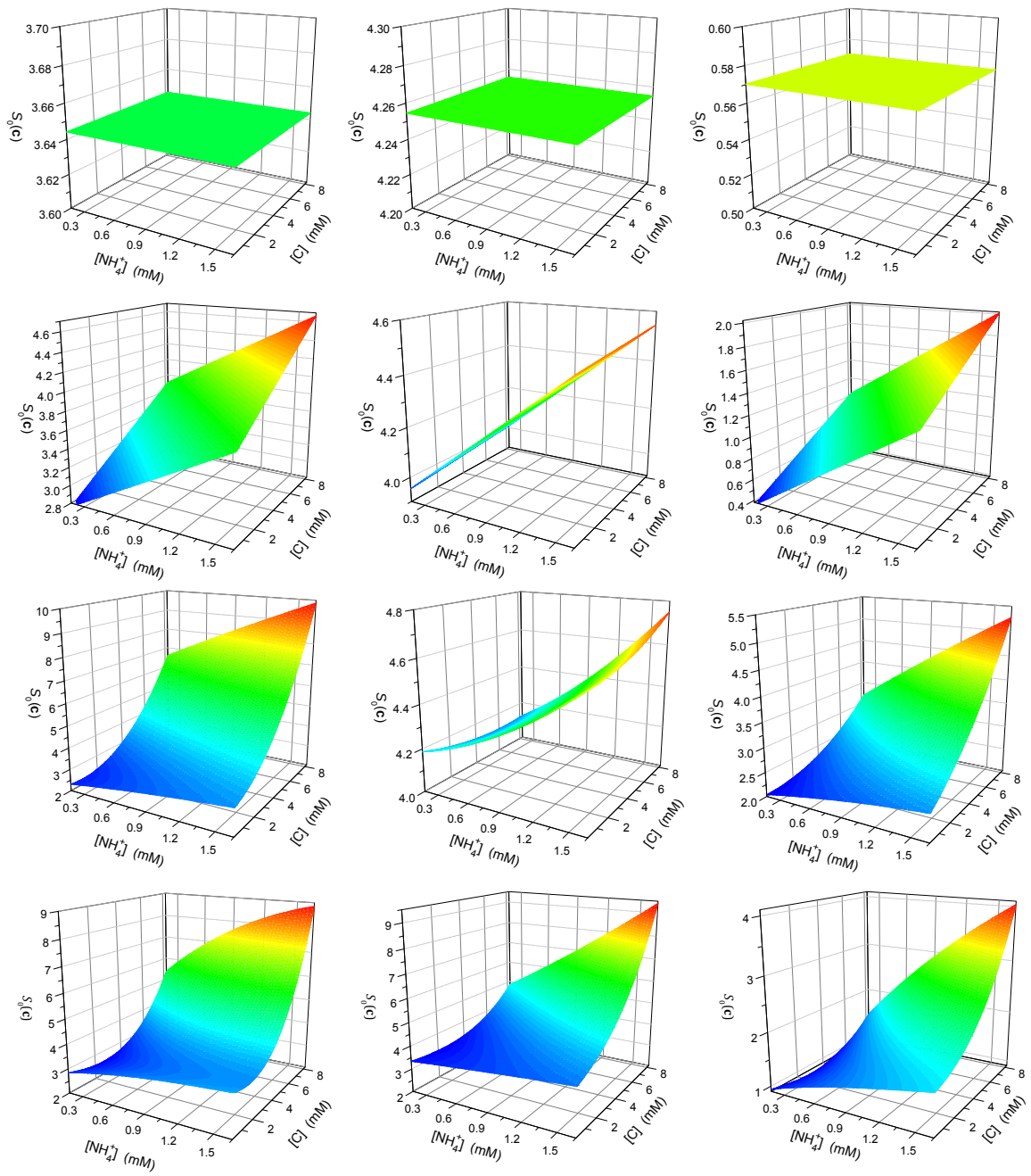


Figure 17. Surfaces $S_0(c_1, c_2)$ (15) for *Dunaliella salina* grown under varying concentrations of C and NH_4^+ as obtained with different polynomial orders; top to bottom $P = 0, \ldots, 3$; left to right $V = 0, \ldots, 2$; the center right graph equals Figure 13 panel (I).

For all investigated cases it was found that the cell shape distribution (16) does not change with nutrient concentrations, i.e. $\mu(\mathbf{c}) = \text{const}$. This type of information provides new insights for assessing the microalgal biomass production under given situations (e.g. algae blooms) and opens new perspectives in environmental monitoring [41].

3.3. Modeling Nutrient Impacts of Cell Growth and Size

The chemical composition and the amount of biomass in the form of phytoplankton is an important parameter for the food web embedded in aquatic ecosystems [12], [13], [36], [37]. The cells' chemical composition determines each cell's nutritional value which is amplified by the amount of their biomass. Furthermore, the cells' physical dimensions and buoyancy determines their spatial and temporal distribution. It has been hypothesized that the chemical make-up and size of microalgae cells are dependent on the microalgal nutrient availability [62], [5], [14], [40], [63], [64], [119], [120]. While microalgal chemical compositions have been studied for instance by means of FT-IR spectroscopy [52] - [54], [56], [100], the relation between cell size, their growth rate, and nutrient availability requires further investigations. It has been shown that growth rates of individual species are inversely related to size [121]. According to Kleiber's law [122], [123], smaller organisms have a more active metabolism than larger ones which in turn implies that photosynthetic rates and nutrient uptake will decrease as the size of the algae cell increases [123] - [125]. In reference [119], it has been shown that cell sizes increase as the growth rate decreases exponentially. Statistically significant differences in algal growth rates were found and it was shown that various algal species respond differently to nutrient sources. However, it is still under investigation how growth rate and cell size are related to nutrient sources and concentrations [91], [124] - [126]. Studies [121], [123], [127], [128] have been performed utilizing the allometric equation [129] in order to describe the growth rate and cell size relationship. Allometric modeling is dependent upon the empirical relationship between cell size and metabolic rates [128], [129]. An example of such a study is found in reference [126] which utilized the allometric equation to describe the metabolic rate of an organism's size to determine if this one equation was able to describe the overall relationship between metabolic rates and size. While

the allometric equation was able to estimate growth rates, it was not capable of deriving information on the relationship between cell size and growth rates within this particular study. These models are able to make good predictions of metabolic rates, but are not flexible and are dependent upon the size of the cell under various environmental conditions for different algal species [128]. Another study [121] determined the relationship between colonial organisms of the toxic cyanobacterium, *Microcystis aeruginosa*, using microscopy and digital image analysis. Growth rates were calculated for each colony by fitting the data utilizing linear regression. It was determined that there were significant variations within a single species between growth rate and colony size. In addition, the growth rates changed independent from the colony size within a genotype. Although the regression analysis in this study did provide more insight into the growth rate to size relationship, additional information is needed on how they relate under various nutrient sources and concentrations.

The goal of the present investigations was to derive a mathematical model linking cell size and growth rate as a function of nutrient sources and concentration. Based on such models, the impacts of an ecosystem's chemical parameters such as the nutrient situation on the microalgae community can be investigated. Such a mathematical description will then facilitate simulations of the potential impacts of chemical shifts on microalgae and their nutritional value for other organisms. Experimental data for proof-of-principle were acquired from three microalgae species (*Dunaliella salina*, *Dunaliella parva*, and *Nannochloropsis oculata*). Replicate cultures were grown under various, well-defined nutrient conditions namely the concentration of inorganic carbon (source: bicarbonate) and nitrogen (sources: ammonium and nitrate) [39], [40]. For the present study, growth rates have been determined via hemocytometer counting; cell and nutrient dependent size distributions have been obtained from microscope image analyses as described in reference [40].

In order to derive and investigate the relationship between nutrient availability, microalgal growth rates, and microalgae cell size, two types of mathematical models have been developed. The first model (section 3.3.1) is based on a sigmoidal function which expresses the cell concentration as a function of time and nutrient concentration. The second (section 3.3.2) assumes that the cell size is normally distributed and describes the absolute cell count as a function of cell size and nutrient concentration. Thus, these two least-squares models are 2D surfaces with two predictor variables [101], [104] with time and nutrient concentration for growth curves as well as cell size and nutrient concentration for size distributions. Both models comprise only one response variable, i.e. cell concentration in the case of growth curves and

cell count for the size distributions. In this study, one cell species and one nutrient at varying concentrations were considered at a time; all other nutrients were kept at standard conditions [9].

3.3.1. Determining the Concentration Dependency of Growth Curves

The number of cells per volume, Y_{count}, is a function of time and is described by a sigmoidal function [130] $Y_{\mathrm{count}}(t) = G \cdot \frac{1}{1+\exp\{-s\cdot(t-t_0)\}}$. The parameter G refers to the final or maximum cell concentration measured here in counts per milliliter, s states the growth rate and is measured here in days^{-1}; t_0 indicates the point of time at which $Y_{\mathrm{count}}(t_0) = \frac{1}{2} \cdot G$ and thus describes the center of the sigmoidal function. In order to introduce a dependency on the nutrient concentrations, the aforementioned model parameters G, s, t_0 are assumed to be functions of a nutrient's concentration c. Thus:

$$Y_{\mathrm{count}}(t, c) = G(c) \cdot \frac{1}{1 + \exp\{-s(c) \cdot (t - t_0(c))\}} \tag{18}$$

Since only one species and one nutrient have been changed in a measurement series, the species and nutrient type has not been included into this notation (18). The concentration dependency of the three model parameters $G(c)$, $s(c)$, and $t_0(c)$ were approximated by polynomials up to order P. The corresponding polynomial coefficients β (19) then carry information about the specific concentration dependency of these growth curves' parameters and will be determined from experimental data via nonlinear least-squares fits [117]:

$$\begin{aligned} G(c) &= \beta_{0,G} + \beta_{1,G} \cdot c + \cdots + \beta_{P,G} \cdot c^P \\ s(c) &= \beta_{0,s} + \beta_{1,s} \cdot c + \cdots + \beta_{P,s} \cdot c^P \\ t_0(c) &= \beta_{0,t_0} + \beta_{1,t_0} \cdot c + \cdots + \beta_{P,t_0} \cdot c^P \end{aligned} \tag{19}$$

The parameters β in (19) have the following meanings. $\beta_{0,G}$: cell concentration in absence of the nutrient under consideration. From a biological perspective, $c = 0$ may not be feasible but this parameter is required for modeling purposes. $\beta_{1,G}$ and $\beta_{2,G}$: linear and quadratic increase/decrease of the final cell concentration per nutrient concentration. $\beta_{0,s}$: growth rate in absence of nutrient $(c = 0)$. One can interpret this value as the 'base growth rate' which is adapted to concentration dependent effects by the following two parameters. $\beta_{1,s}$ and $\beta_{2,s}$: linear and quadratic change in growth rate with increasing nutrient concentration increase c; β_{0,t_0}: center in time of the exponential growth phase in absence of a nutrient concentration. This parameter to be interpreted as 'base time point' which is then modified by concentration dependent effects described by β_{1,t_0} and β_{2,t_0} which weigh linear and quadratic changes of center time point with concentration c.

In order to derive the parameters β, K cell concentrations $y_{t,c}$ were experimentally determined after growing cultures for t days under nutrient concentration c (see section 2.3). The model function $Y_{\text{count}}(t, c)$ was then fitted to these data points via minimizing the sum of squared errors[3] $\text{SSE}\left(\hat{\beta}_{0,G}, \hat{\beta}_{1,G}, \dots, \hat{\beta}_{0,s}, \hat{\beta}_{1,s}, \dots, \hat{\beta}_{0,t_0}, \hat{\beta}_{1,t_0}, \dots\right) = \sum_{k=1}^{K}\left[y_{t,c} - Y_{\text{count}}(t, c)\right]^2$. Again a necessary condition for a minimum of the SSE is:

$$\boldsymbol{\nabla}\text{SSE}\left(\hat{\beta}_{0,A}, \hat{\beta}_{1,A}, \dots, \hat{\beta}_{0,s}, \hat{\beta}_{1,s}, \dots, \hat{\beta}_{0,t_0}, \hat{\beta}_{1,t_0}, \dots\right) = -2 \cdot \sum_{k=1}^{K}\left[y_{t,c} - Y_{\text{count}}(t, c)\right] \cdot \boldsymbol{\nabla} Y_{\text{count}}(t, c) = \mathbf{0}$$

(20)

Since standard linear least-squares regression [101] - [103] cannot be applied to solve this system of $(P + 1) \cdot 3$ nonlinear equations, iterative solvers [108], [109] have to be employed to compute or more precisely estimate the parameters $\hat{\beta}$ [117].

[3] Parameters estimated by means of least-squares are denoted by a ^.

3.3.2. Depicting Cell Size Concentration Dependency

Distributions of cell size [4] S have been determined and compiled into species- and concentration dependent histograms as described in reference [40]. These histograms were utilized as input data for this study. Also, it had been shown in [40] that the histograms are centered at the most common cell size S_0 which was found to be species and nutrient dependent. These size distributions, i.e. of type $A \cdot \exp\left\{-\frac{(S-S_0(c))^2}{2\cdot\sigma(c)^2}\right\}$, are superimposed by a 'count background' which could be described by means of a polynomial of order Q_S, i.e. $\alpha_0(c) + \alpha_1(c)\cdot S + \cdots + \alpha_{Q_S}(c)\cdot S^{Q_S}$. Again (cp. (18)), the model functions parameters were assumed to be concentration *c* dependent as before seen in equation (14) and again shown below:

$$Y_{\text{size}}(S,c) = A(c)\cdot\exp\left\{-\frac{\left(S-S_0(c)\right)^2}{2\cdot\sigma(c)^2}\right\} \; + \; \alpha_0(c) + \alpha_1(c)\cdot S + \cdots + \alpha_{Q_S}(c)\cdot S^{Q_S}$$

Equivalent to (19), the nutrient concentration dependency of the model parameters $B(c)$, $S_0(c)$, and $\sigma(c)$ as well as $\alpha_{q=0,\ldots,Q}(c) = \beta_{0,\alpha_q} + \beta_{1,\alpha_q}\cdot c + \cdots + \beta_{P,\alpha_q}\cdot c^P$ were approximated by a polynomial of order P; if a parameter were concentration independent, the corresponding parameters $\beta_{p>0} = 0$ would be the result. Here it is important to note the different meanings of the two types of polynomials involved: The polynomial of order Q_S refers to the background approximation [5] of the entire histogram and the polynomial of order P indicates the approximation level of each model parameter. Following a procedure equivalent to (18) - (20) derives an equation system of $\underbrace{[(P+1)\cdot 3]}_{\text{Gaussian}} + \underbrace{[(P+1)\cdot(Q_S+1)]}_{\text{polynomial}}$ unknown model parameters β.

In the subsequent equation (21), $y_{S,c}$ refers to the experimental counts in the bin comprising cell size S of a culture that had been grown under nutrient concentration c [117].

---

[4] In (14), the capital letter S has been used to avoid confusion with the growth rate s in (18).

[5] This background term is a polynomial of order Q_S in S and in turn, each of this polynomial's parameters, $\alpha_{q=0,\cdots,Q_S}$, is a polynomial of order P in c. Hence, the term $\alpha_0(c) + \alpha_1(c)\cdot S + \cdots + \alpha_{Q_S}(c)\cdot S^{Q_S} = \beta_{0,\alpha_0} + \beta_{1,\alpha_0}\cdot c + \cdots + \beta_{P,\alpha_0}\cdot c^P + \beta_{0,\alpha_1}\cdot S + \beta_{1,\alpha_1}\cdot c\cdot S + \cdots + \beta_{P,\alpha_1}\cdot c^P\cdot S + \cdots + \beta_{0,\alpha_{Q_S}}\cdot S^{Q_S} + \beta_{1,\alpha_{Q_S}}\cdot c\cdot S^{Q_S} + \cdots + \beta_{P,\alpha_{Q_S}}\cdot c^P\cdot S^{Q_S}$ is a polynomial surface [104] in the (S,c)-plane.

$$\nabla \mathrm{SSE}\left(\hat{\beta}_{0,A}, \hat{\beta}_{1,A}, \hat{\beta}_{2,A}, \hat{\beta}_{0,S_0}, \hat{\beta}_{1,S_0}, \hat{\beta}_{2,S_0}, \hat{\beta}_{0,\sigma}, \hat{\beta}_{1,\sigma}, \hat{\beta}_{2,\sigma}, \underbrace{\cdots, \hat{\beta}_{0,\alpha_q}, \hat{\beta}_{1,\alpha_q}, \hat{\beta}_{2,\alpha_q}, \cdots}_{q=0,\ldots,Q}\right)$$
$$= -2 \cdot \sum_{k=1}^{K} \left[y_{S,c} - Y_{\mathrm{size}}(S,c)\right] \cdot \nabla Y_{\mathrm{size}}(S,c) = \mathbf{0}$$

(21)

3.3.3. Polynomial Order Selection

For subsequent investigations, the two polynomial orders P and Q_S introduced in sections 3.3.1 and 3.3.2 have been limited to the ranges $P = 0, \ldots, 5$ and $Q_S = 0, \ldots, 5$. For each species and each nutrient, this resulted in six different models for $Y_{\mathrm{count}}(t,c)$ and 6 x 6 = 36 different models for $Y_{\mathrm{size}}(S,c)$. For determining the appropriate models, correlation coefficients were not considered as useful since higher polynomial orders will describe minor artifacts and thus the correlation coefficient typically increases with increasing P and Q_S. Instead, the biological meaningfulness of the resulting fit parameters was considered. A model has been rejected, if the following requirements were not met for the relevant concentration ranges of the corresponding nutrients:

- $G(c) \geq 0$ and $A(c) \geq 0$
- $s(c) \geq 0$ and $\sigma(c) \geq 0$
- $t_0(c) \geq 0$ and $S_0(c) \geq 0$
- $\alpha_0(c) + \alpha_1(c) \cdot S + \cdots + \alpha_{Q_S}(c) \cdot S^{Q_S} \geq 0$ because the background counts (see footnote 5) are not supposed to compensate by means of negative values a model over-estimating the experimental data [117].

(22)

3.4. Effects of Nutrients on Growth Rate vs. Cell Size

Figure 18 presents the types of information to be gained from the nonlinear least squares introduced in sections 3.3.1 and 3.3.2 For the microalgae species *Dunaliella salina* cultured under varying carbon concentrations, panel (A) shows the concentration dependency of the growth curves, i.e. $Y_{\mathrm{count}}(t,c)$ (18) with $P=1$. The impact of the carbon concentration is reflected in the 'bending' of this surface with increasing concentration. Panel (B) depicts the concentration dependency of the background corrected[6] $Y_{\mathrm{size}}(S,c)$ (14), i.e. the peak of the normalized cell counts becomes narrower and higher with increasing carbon concentration. In order to assess the accuracy of these two models, the models values have been computed at the same time-concentration pairs (panel C) and size-concentration pairs (panel D) for which experimental data had been obtained. Plotting the models' (= "predicted") values versus the experimental (= "true") values enables the models' accuracy which was considered satisfactory given the biological nature of the samples.

[6] Size surfaces after background correction refer to the surface $Y_{\mathrm{size}}(S,c) - \left[\alpha_0(c) + \alpha_1(c)\cdot S + \cdots + \alpha_{Q_S}(c)\cdot S^{Q_S}\right]$. *Normalized* cell counts were introduced in [40] in order to make results which had been obtained from images of replicate samples comparable; this normalization divides the original histogram by the number of cells found in a certain image.

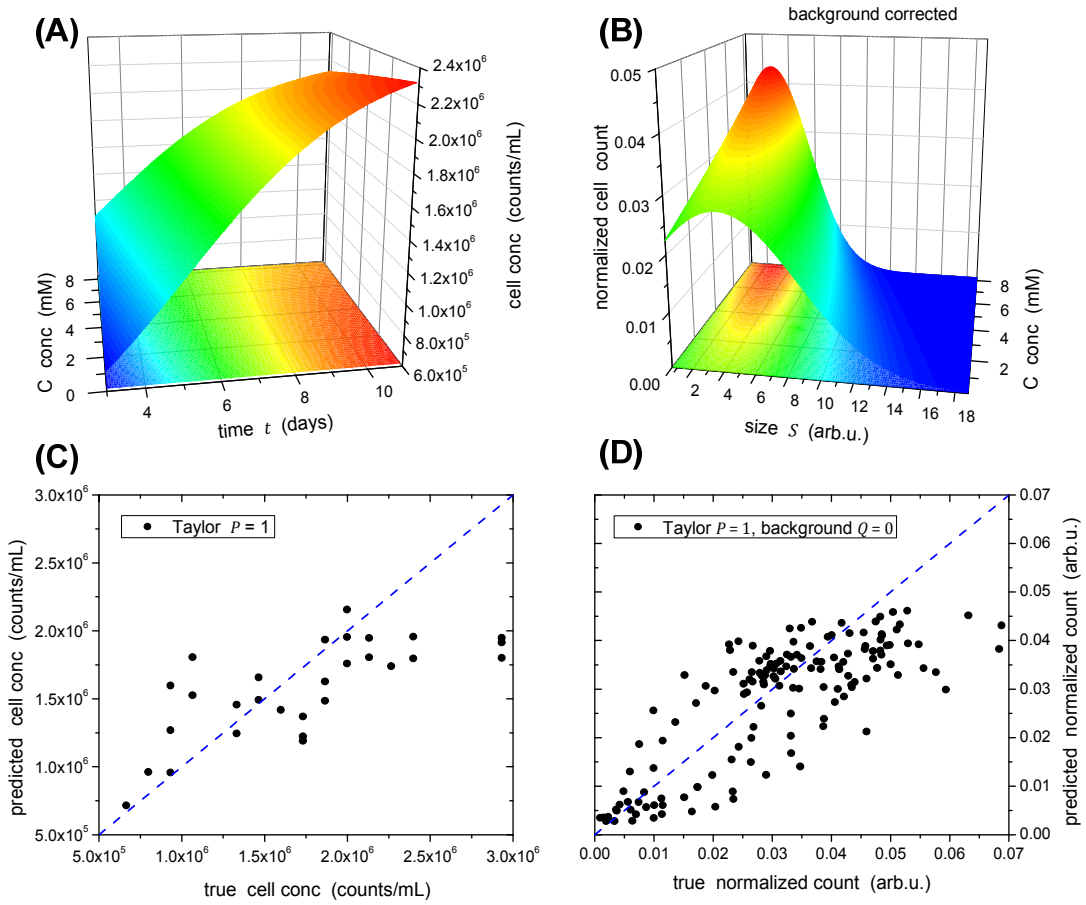


Figure 18. (top row) *Dunaliella salina* grown under varying carbon concentrations - (left column) Growth rate data, (right column) Cell size data [40]; (A) growth rate surface $Y_{\text{count}}(t, c)$ (18), with $P = 1$; (B) size surface $Y_{\text{size}}(S, c)$ (20) with $P = 1$ and $Q = Q_S = 0$ after subtracting the polynomial background surface which is here simply α_0 because $Q = Q_S = 0$; (C) 'predicted', i.e. computed from Y_{count}, versus measured, i.e. 'true', cell concentration - see text; (D) predicted, normalized cell counts versus true counts - see text.

For all investigated models, i.e. algae species, nutrients, and polynomial orders satisfying (22), Figure 19 and Figure 20 depict the mean deviations between model and experimental values. To avoid canceling positive and negative deviations, absolute values were averaged. In both figures, the models with the smallest average error have been marked by a (*) and were considered for subsequent analyses. The fact that all mean errors obtained with different polynomial orders P and (P, Q_S), respectively, are in the same order of magnitude indicates that a change in polynomial order is not completely altering the models; this will be further examined in Figure 23 and Figure 24. It is also noteworthy that the models with the highest polynomial orders were not the most accurate ones and hence it can be concluded that overfitting is not at a critical level.

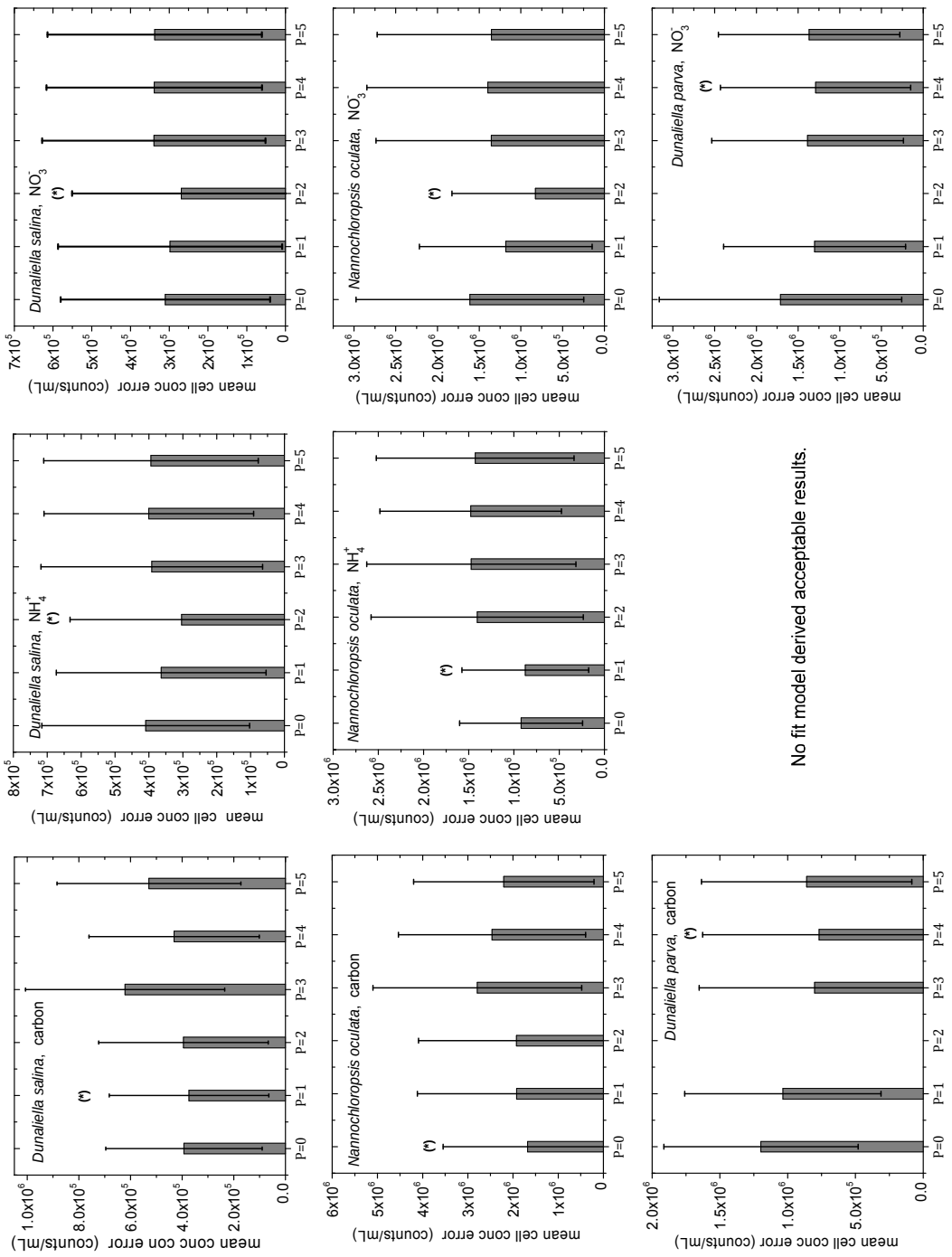


Figure 19. Mean 'prediction errors' for growth curves, i.e. $\overline{\left|y_{t,c} - Y_{\text{count}}(t,c)\right|}$, obtained for the three species cultured under varying concentrations of three nutrients; the most accurate model (polynomial order P) has been indicated by a (*); models not represented here did not result contain acceptable model parameters (22).

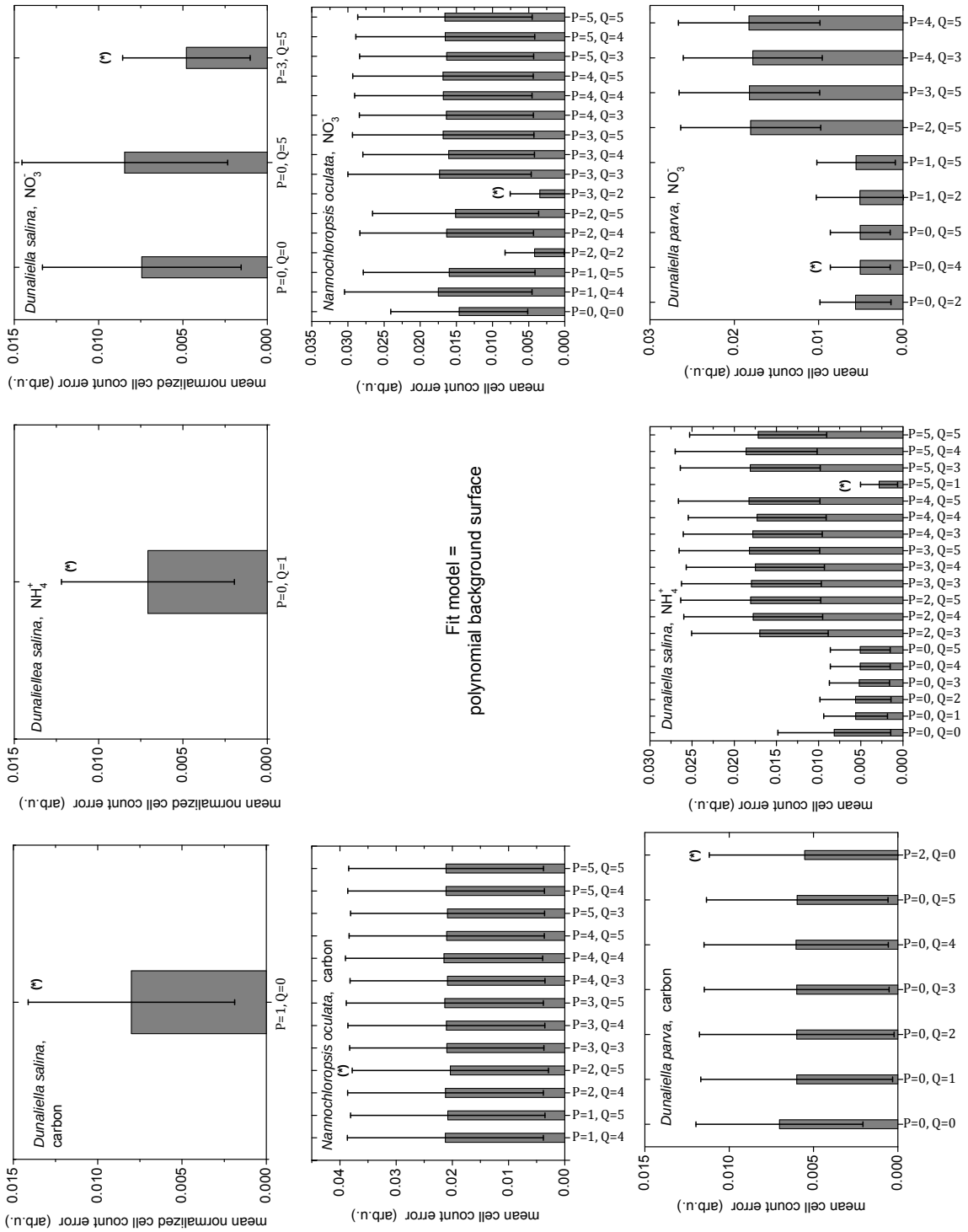


Figure 20. Mean 'prediction errors' for size histograms, i.e. $\overline{\left|y_{S,c} - Y_{\text{size}}(S,c)\right|}$, obtained for the three species cultured under varying concentrations of three nutrients; the most accurate model (polynomial order combination (P, Q_S)) has been indicated by a (*)

The most accurate models $Y_{\mathrm{count}}(t,c)$ and $Y_{\mathrm{size}}(S,c)$ determined based on Figure 19 and Figure 20 are displayed in Figure 21 and Figure 22, respectively. Most concentration dependent growth curves reflect the expectations, i.e. an underlying sigmoidal increase in time which shows none or a linear concentration dependency. The concentration dependency of the growth rates, i.e. the steepness of these surfaces in time direction, is of central importance for this study and will be further investigated below. Two of these surfaces (*Dunaliella salina* with changing NH_4^+ and NO_3^- concentrations, $P = 2$), however, feature a distinct curvature along the concentration axis which may reflect some overfitting of the experimental data. In particular the very similar accuracies (Figure 19) obtained for other polynomial orders $P < 2$ may be an indication that a quadratic term in $s(c)$ may not be significant. Future investigations could augment these modeling steps with Analyses of Variance (ANOVAs) [101], [131], [132] to test the significance of all polynomial terms in (19). For *Dunaliella parva*, $P = 4$ was found to be most accurate – nonetheless, no curvature is observable in the models $Y_{\mathrm{count}}(t,c)$. The reason for this is that $\beta_{p>0,B} \approx \beta_{p>0,s} \approx \beta_{p>0,t_0} \approx 0$ were obtained. For size models $Y_{\mathrm{size}}(S,c)$, similar findings resulted: For most models the overall shape of the surfaces shows either a concentration independent behavior or a linear dependency on the nutrient concentrations. For a few models such as Dunaliella salina (NO_3^-) or *Dunaliella parva* (NH_4^+), a surprisingly curved surface has been computed. Whether these strong curvatures are significant could be tested in future studies by ANOVAs as well.

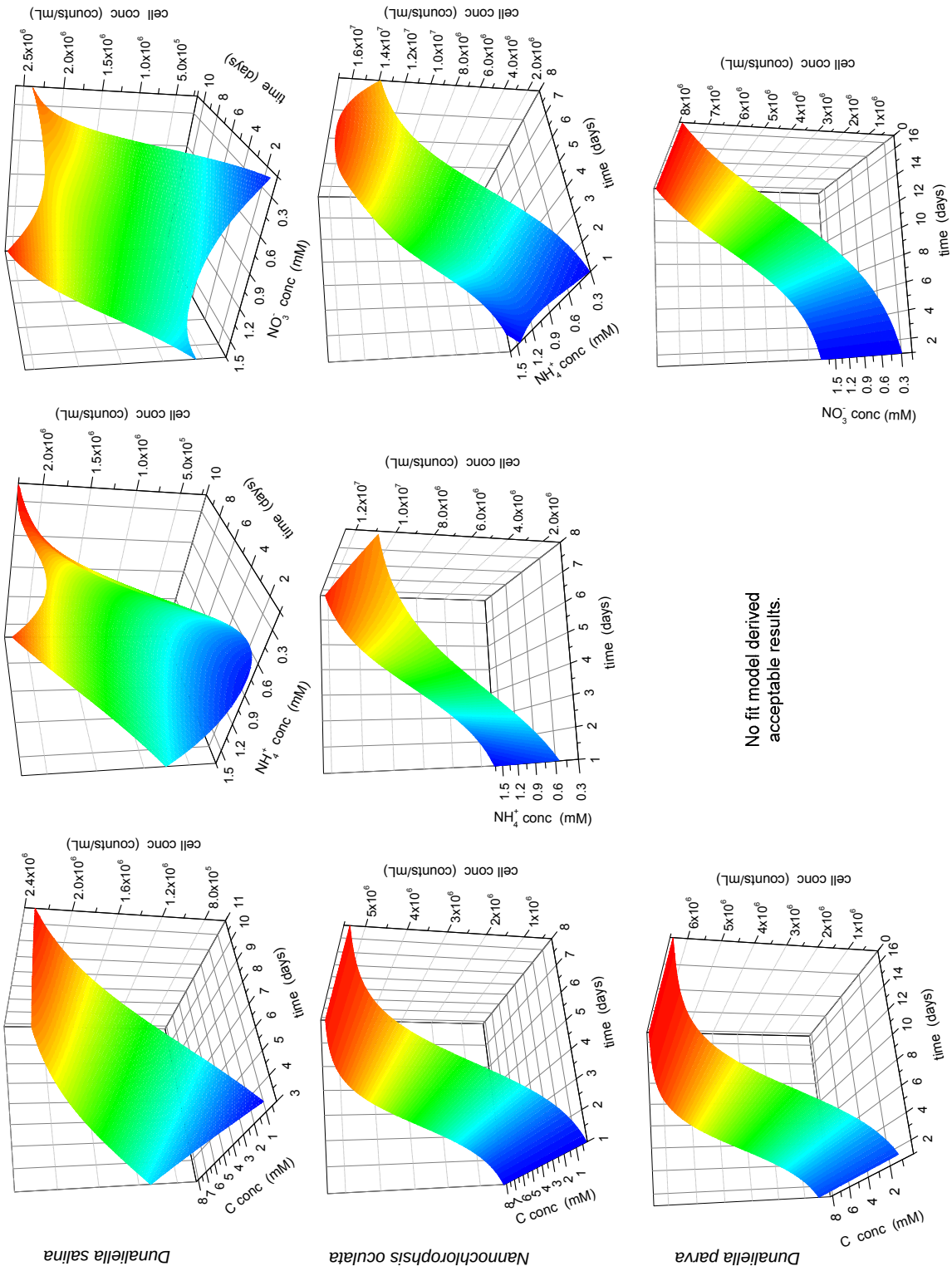


Figure 21. Growth surfaces $Y_{\text{count}}(t, c)$ which describe the experimental data best; see models marked by (*) in Figure 19.

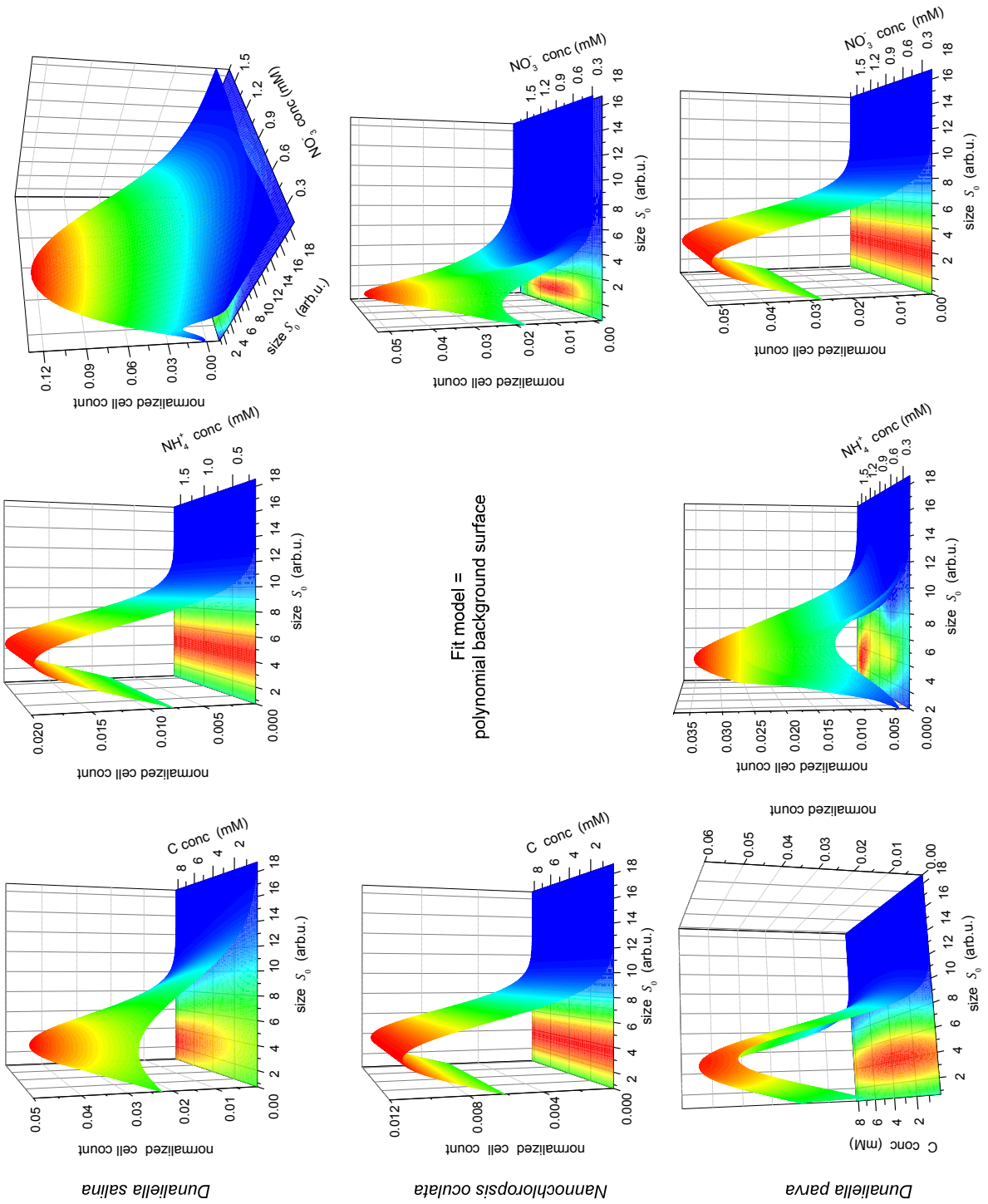


Figure 22. Size histogram surfaces $Y_{\text{size}}(S, c)$ after subtracting the background polynomial surface which describe the experimental data best; see models marked by (*) Figure 20.

Figure 23 and Figure 24 depict the concentration dependent growth rates $s(c)$ and the most common cell size $S_0(c)$, respectively, for all species and all nutrients for which realistic models (22) had been obtained. Among the growth rates, two main types of concentration dependencies were found: For higher polynomial orders P and by definition for $P = 0$, the growth rate $s(c)$ did not show a concentration dependence. For low polynomial orders, the growth rates overall tend to be decreasing with increasing nutrient concentrations. The cell sizes $S_0(c)$ on the other hand were either found to be concentration independent or featured an increased trend with increasing nutrient concentrations. One notable exception is *Dunaliella salina* when cultured with varying carbon concentrations.

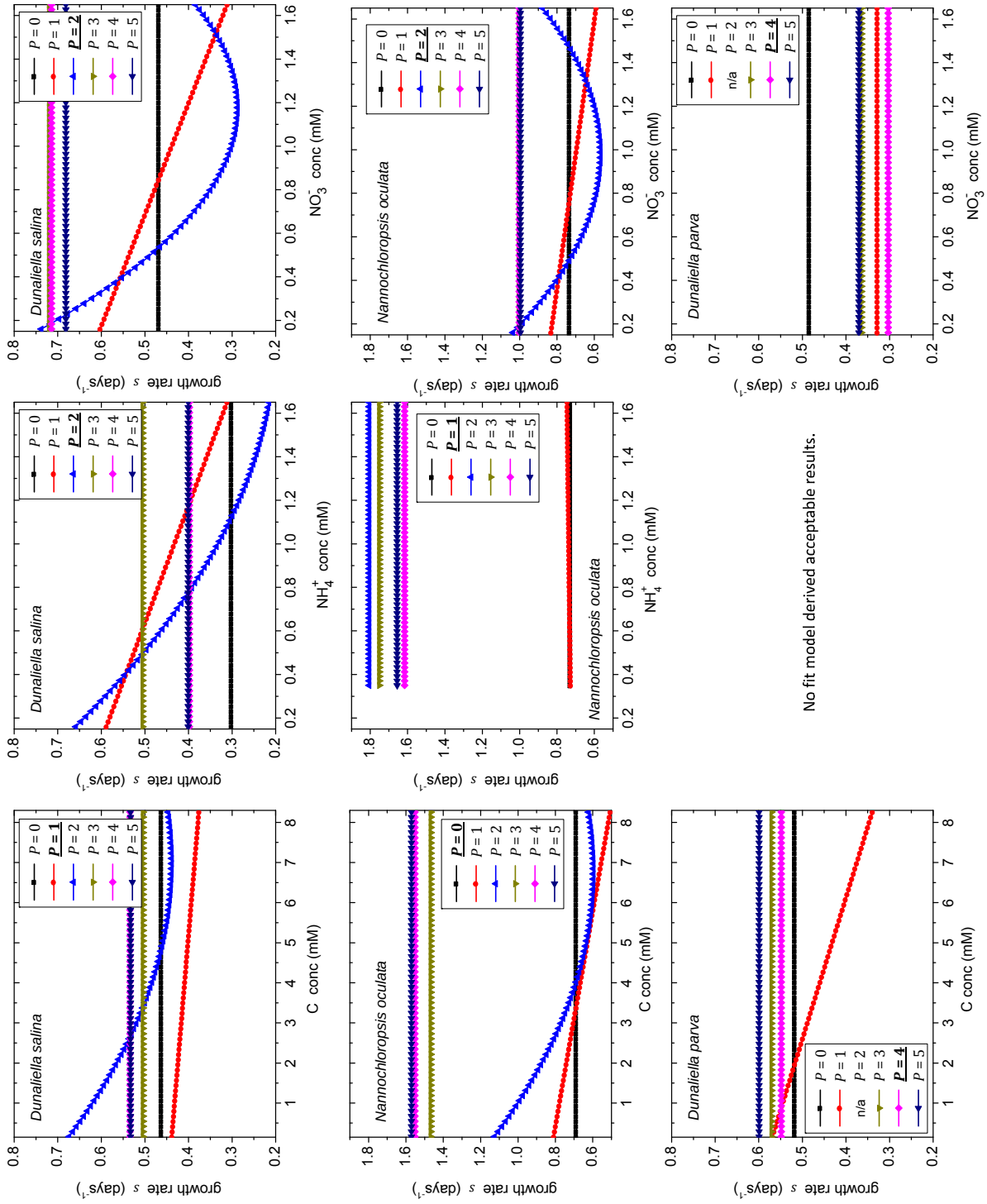


Figure 23. Growth rates $s(c)$ as computed from models $Y_{\mathrm{count}}(t,c)$ (18) such as shown in Figure 18 (A); the legends for the most accurate models (cp. Figure 19) were underlined and printed in bold face.

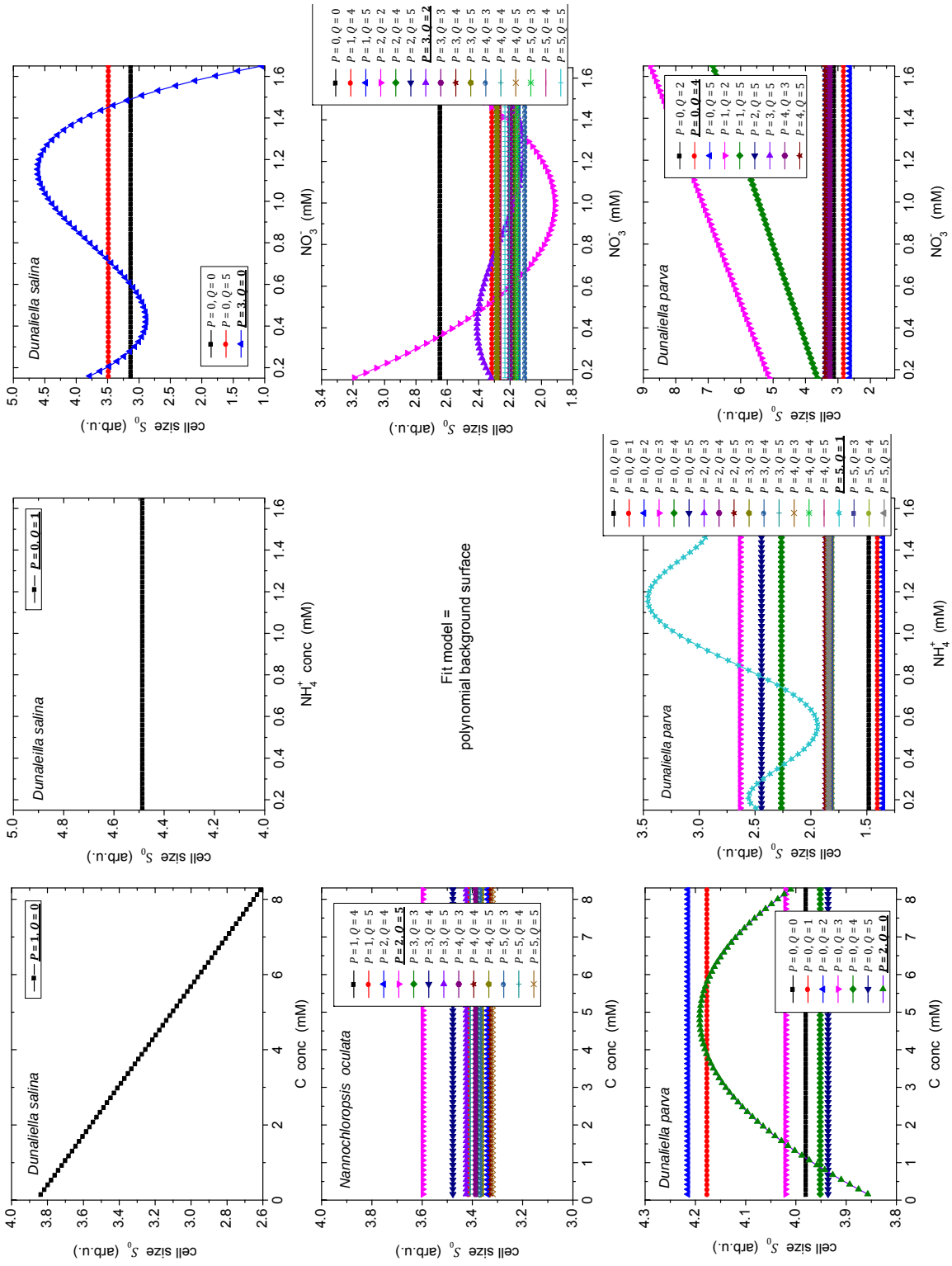


Figure 24. Cell sizes $S_0(c)$ as computed from models $Y_{\text{size}}(S, c)$ (14) such as shown in Figure 18 (B); the legends for the most accurate models (cp. Figure 20) were underlined and printed in bold face.

In order to analyze the relationship between growth rates and cell sizes, $s(c)$ versus $S_0(c)$ has been plotted in Figure 25 separated for species and nutrients. Both functions had been evaluated at one hundred nutrient concentration values c covering the relevant concentration ranges. The blue arrows in the panels indicate the directions in which the growth rate/size relates with increasing nutrient concentration. Three types of relations have been found: (i) a fixed relation (*Nannochloropsis oculata*, carbon; *Dunaliella parva*, NO_3^+); (ii) a concentration dependency in only one parameter with the other being fixed (*Dunaliella salina*, NH_4^+; *Dunaliella parva*, carbon); (iii) a concentration dependent change in both, i.e. growth rate and size (*Dunaliella salina*, carbon and NO_3^-, *Nannochloropsis oculata*, NO_3^-). For the remaining two species/nutrient situations, only one parameter was available which has been shown in Figure 23 and Figure 24. Obviously, the relationship between growth rate and cell size is species and nutrient specific and more species and more nutrient types should be analyzed in order to study whether a general trend across species can be found.

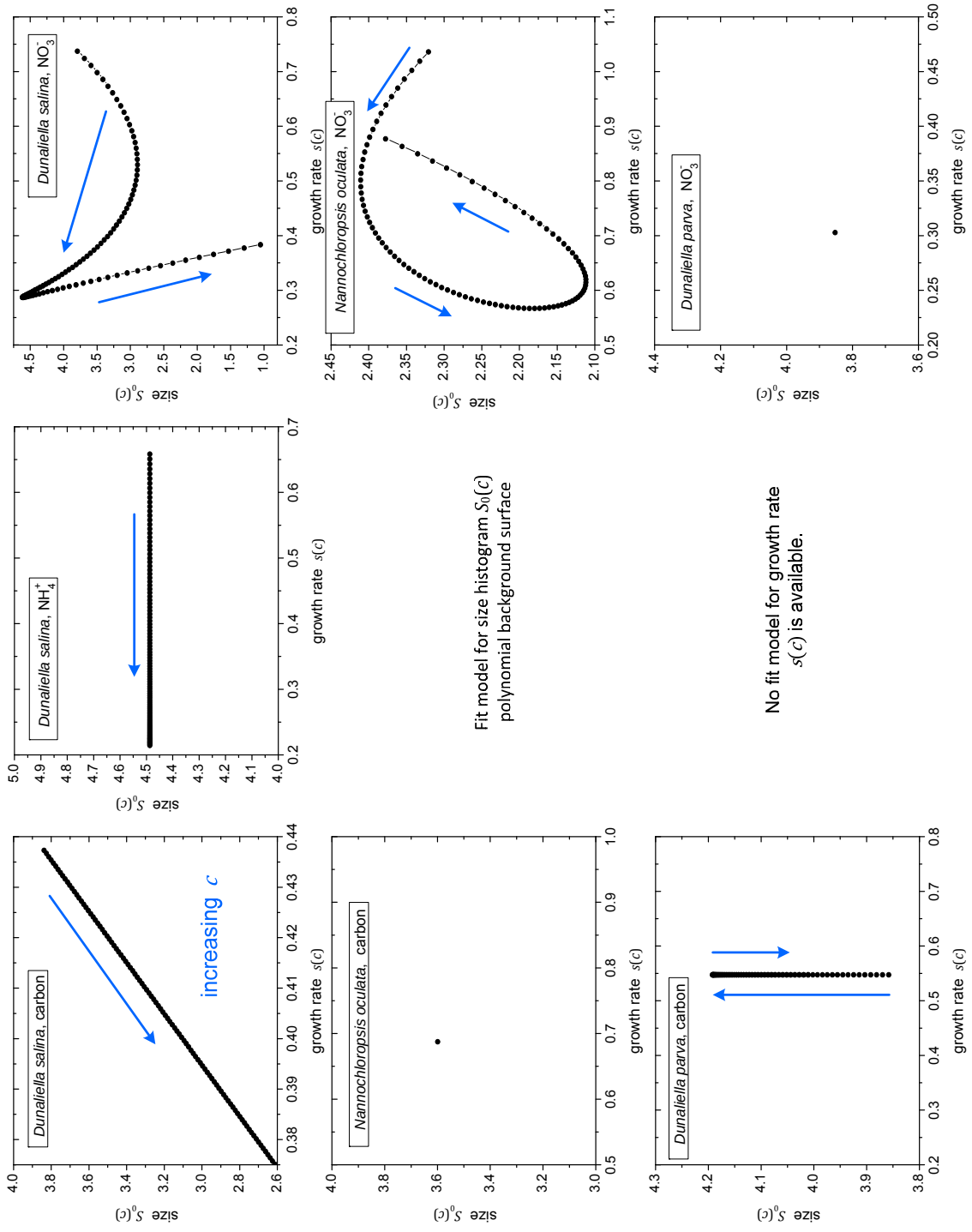


Figure 25. Cell size $S_0(c)$ versus growth rate $s(c)$ for the best models (Figure 20 and Figure 19); the direction of increasing nutrient concentration c is indicated by arrows.

4. Chemical Impacts of Microalgae Mixtures Studied by FT-IR Analysis

The ecosystem is experiencing a drastic change in climate which is due to anthropogenic impacts [133]. Specifically two nutrients nitrogen and carbon, are affecting the ecological community. Carbon in the form of CO_2, has increased and which affects all organisms including microalgae within the ecosystem [1]. The majority of CO_2 that enters the atmosphere dissolves in oceans and increases the amount of the bicarbonate ion as well as dissolved CO_2 [5] concentration levels [6] which effects the distribution of algae. Since each species is affected differently by the raised levels of CO_2 concentration, some species growth rate will increase, while others will not thrive; thus, altering their biochemical composition [1]. Also, due to human involvement, the amount of nitrogen has doubled due to synthetic fertilizer and atmospheric fossil fuel combustion, which are the largest sources of nitrogen contamination [22]. These nitrogen affects reduce the protein content and lipid storage within microalgae cells demonstrating how the chemical composition is altered based on various environmental parameters [23].

Due to the increase in CO_2 rates as well as nitrogen production, there has been a decline in the function of the marine ecosystem [134]. While such changes have been experienced in the past, the rate of change that is happening today is astonishing [133] and these changes cause a loss of biodiversity [22], [135]. In reference [39], microalgae's spectroscopic signatures were depicted to non-linearly react to changes within carbon and nitrogen cultures, however, lacked the real world application of investigating how microalgae species react when grown together.

The availability of resources is one of the main limiting factors of coexisting species and their growth [136]. Experimenters have come to several different conclusions on cellular chemical interactions between algae species. According to Parson and Takashi [128], larger cells are expected to use more nutrients than smaller cells when exposed to nutrient rich conditions. Whereas Zinkel and Irwin [137], used an allometric model and made a prediction that smaller cells should always use more nutrients compared to larger cells. In an additional study, resource imapcts was examined using a vertical mixed water column containing multiple

species under varying concentrations of carbon and was modeled using an allometric equation. The study determined that each species when grown together were able to coexist, but at low nutrient concentrations, the smaller algal strain use more nutrients than the larger algal strain and under higher nutrient concentrations the larger species was able to use more nutrients compared to the smaller species [137].

In another study, three different species (*Ectocarpus, Ulothrix, and Erythrotrichia*) were studied by growing them together in different pairs [138]. The researcher studied how each species would perform when grown together in comparison to all other species combinations when under the same conditions. The goal of the investigation was to measure the growth changes to the three species when using a toxic copper additive. The results determined that all species growth was limited in large copper concentrations, but overall the interactions between the species fluctuated with changing copper concentrations [138].

Tilman [139] utilized two freshwater algal strains, *Asterionellaformosa and Cyclotella meneghiniana*, and empirically modeled their growth to examine if each species is capable of cohabitation when different nutrient sources are limited. For this experiment, several concentrations of phosphate and silicate were studied and analyzed by two prediction models, the Monad equation and an internal storage growth model [140]. The freshwater species' were capable of coexisting in the presence of phosphate and silicate where each species was projected to diminish by a different resource. It was concluded that each species was able to cohabitate when each nutrient source was limited and that predicting the outcome of species chemical interactions is feasible [139].

It was been determined in reference [138] that species grown together are more susceptible to chemical change from the nutrient conditions of their environment when compared to single species. It was also stated that the feasibility of utilizing single species cultures to investigate real-world situations is unlikely. Thus it is necessary to go beyond the study of single species analysis toward mixed cultures to investigate what changes occur when species interact. As a first step for understanding species chemical interactions and further establishing the use of microalgae as an "eco-sensor", analyses will be performed to determine if there is a chemical difference when species are grown together compared to when they are grown in individual cultures. This analysis will assess if a chemical change occurs based on mixed cultures as well as environmental conditions.

This chapter focuses on utilizing FT-IR spectroscopy due to its high sensitivity and selectivity [141] in combination with chemometric tools to determine if there is a change in chemical composition when microalgae species are grown together. Fourier transform infrared spectroscopy is a promising tool for such chemical investigations of microalgae [52], [53], [65], because most macromolecules within microalgae, i.e.: lipids, nucleic acids, carbohydrates and proteins, are infrared active [141]. Although different species have similar chemical compositions, small differences are reflected in their IR-spectroscopic signatures which facilitate their discrimination [65]. Microalgae's response to environmental change is not constant over all ecosystems due to small variations in the ecological community to which the cell is exposed [134]. Thus, for these investigations, microalgae's chemical response to changing conditions has been studied. For detecting chemical shifts in FT-IR spectra, chemometric tools were employed for analysis. This analysis demonstrates the ability to detect and assess chemical changes in marine ecosystems when mixtures of algal strains are grown together under varying nutrient conditions versus unialgal cultures. This in turn will serve as an indicator for determining the changes and affects that occur within the ecosystem.

4.1. Assessment of Chemical Composition of Microalgae Cell Mixtures

The analysis of species chemical interactions will be based on comparing spectroscopic signatures of single cultured species to mixed cultures. The goal is to determine which regions of the spectrum undergo a chemical change within each mixture spectrum.

Data pre-processing is necessary due to pellet-to-pellet fluctuations in the weight percentages as well as in the pellets' thickness. These artifacts, which are difficult to control experimentally, would be interpreted as chemically induced effects. In order to eliminate such misinterpretations, each spectrum has been normalized individually to a maximum peak height of one. For each of a culture's five replicates, five pellets have been prepared for spectroscopic analysis. From ~25 spectra, a mean spectrum and its standard deviation spectrum were computed to represent a culture's chemical information along with the naturally occurring variances (Figure 26).

After averaging all class-specific spectra and calculating the corresponding standard deviations at each wavenumber (i.e. Figure 26), single species spectra are compared to dual species spectra under inorganic carbon, ammonium and nitrate conditions found in section 2.3. Due to the considerable error bars which are the result of spectroscopic replicate-to-replicate fluctuations, statistical testing was deemed necessary. A statistical comparison is used to determine which sections of the spectra are statistically different in order to study if a chemical change occurs when species are grown together. In order to do so, a t-test at each wavenumber was employed for analysis with a significance level of 95% for all analyses.. The sections of the spectra that undergo a change are due to differences in absorbance or chemical composition between single species spectra and mixture spectra. When the wavenumber regions do not have a statistical difference (no change at a particular wavelength), chemical information from both single species are found within the mixture spectrum. The panels (C) and (D) in Figure 26 display such a comparison.

From this point forward the species will now be referred to by Species number (Table 4): Species #1: *Dunaliella salina*, Species #2: *Nannochloropsis oculata*, Species #3: *Dunaliella parva*, Mixture #1: *Dunaliella salina* + *Nannochloropsis oculata*, Mixture #2: *Dunaliella salina* + *Dunaliella parva*, and Mixture #3: *Nannochloropsis oculata* + *Dunaliella parva*.

As an example, mean spectra of Species #1 and Mixture #2 (Figure 26 (A)) grown under 6.72 mM HCO_3^- are compared. t-test results for Species #1 and Mixture #2 are depicted in Figure 26 (C) and show the wavenumber regions that differ. After a significant difference is found, a value of either +0.1 or -0.1 is given. When the output value of a t-test is 0.1, there is a significant difference found between the single species spectra and the mixture spectra and the contribution is from the single species. If the output value is -0.1, there is a significant difference between the single species spectra and the mixture spectra and the contribution favors the mixture species. If no significant difference is found the t-test value is zero and all spectroscopic information from both single species is contained in the Mixture spectrum [142].

Table 4. Species reference numbers

Species #	Functional Groups
Species #1	*Dunaliella salina*
Species #2	*Nannochloropsis oculata*
Species #3	*Dunaliella parva*
Mixture #1	*Dunaliella salina* + *Nannochloropsis oculata*
Mixture #2	*Dunaliella salina* + *Dunaliella parva*
Mixture #3	*Nannochloropsis oculata* + *Dunaliella parva*

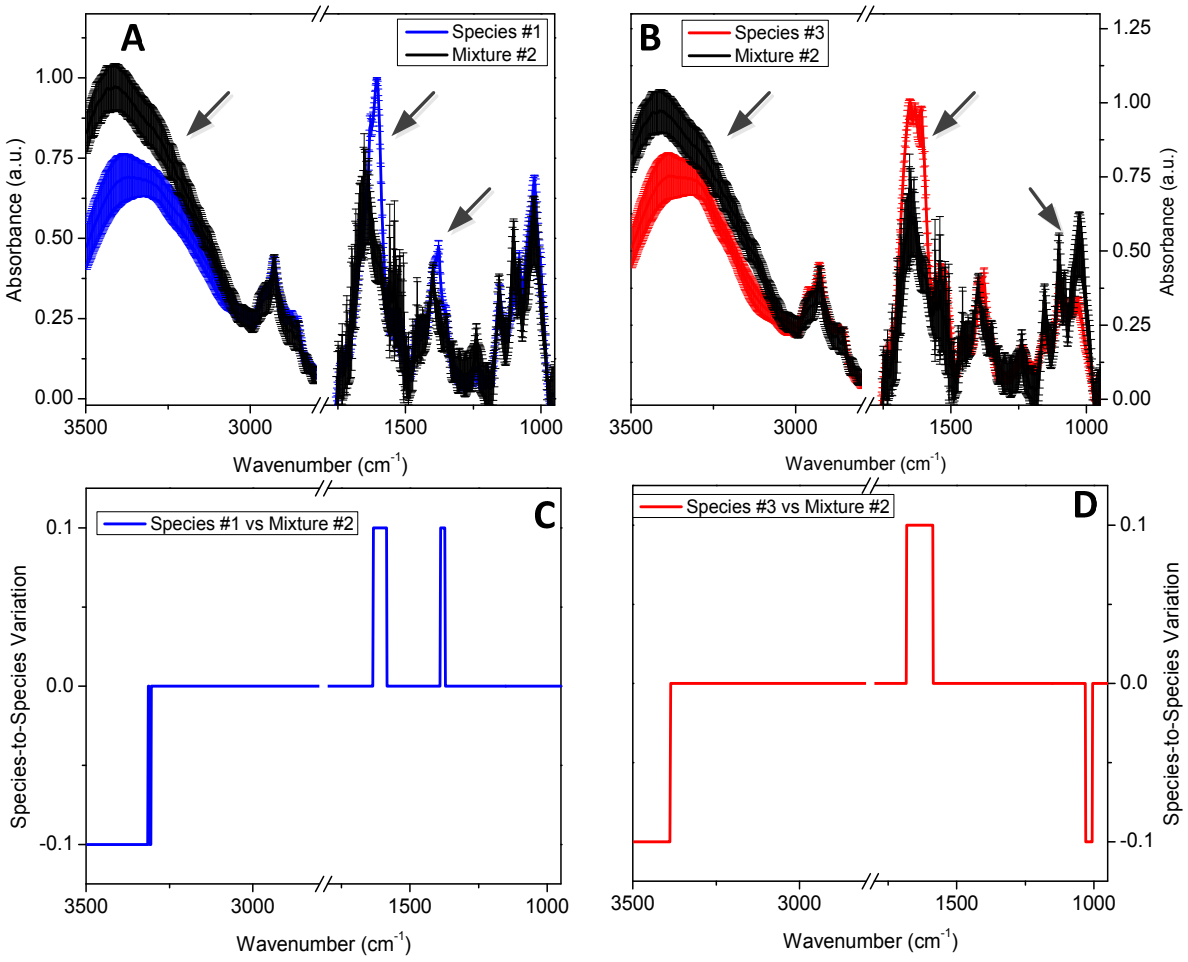


Figure 26. Replicate spectra obtained from 6.72 mM HCO_3^- are averaged to obtain mean spectrum with error bars within the wavenumber region of 3500-2800 cm^{-1} and within the fingerprint region from 1800-950 cm^{-1} [100]. The region of 2800-2200 cm^{-1} has been cut out due to atmospheric CO_2 as well as that region did not contain any pertinent chemical information. (A) and (B) mean spectra and standard deviations. (C) Result of t-test of Species #1 vs. Mixture #2 (D) Result of t-test between Species #3 vs. Mixture #2.

Nonetheless, a direct assessment of any single species spectrum with a mixture's spectrum cannot be achieved because a mixture spectrum will contain additional information that is not present in either species. Thus, a t-test would not be able to discriminate between features obtained from single species or chemically induced chemical changes.

For further analysis, a comparison test was performed between two resultant t-tests. This test is to determine what is occurring at each wavenumber position within the mixture spectrum i.e. are chemical contributions found from both single species, only found in the mixture spectrum or is chemical information from single species reduced within the mixture spectrum. From this comparison method, nine situations are possible and are outlined below in Table 5. This outline will refer to Figure 26 (C) and (D) for illustration purposes and the comparison test result is displayed in Figure 27. Examples of all species and nutrient concentrations are depicted in Figure 30. The analysis method utilized in this project is based on comparing spectroscopic signatures of mixed algal cultures to an entire set of single species spectra.

Table 5. Nine situation outcomes for Mixture spectrum depicted in Figure 27. This analysis is performed at each wavenumber position.

Situation # and color assignments	Absorbance of Species #1	Absorbance of Species #3	Interpretation of Chemical Impacts
Situation #0	No diff	No diff	Species #1 and Species #3 found in Mixture #2
Situation #1	Higher than mixture	Higher than mixture	Chemical info. from Species #1 and Species #3 is reduced in Mixture #2
Situation #2	Higher than mixture	No diff	Chemical info. from Species #1 reduced is in Mixture #2
Situation #3	No diff	Higher than mixture	Chemical info. from Species #3 reduced is in Mixture #2
Situation #4	Higher than mixture	Lower than mixture	Some chemical info. from Species #1 is reduced in Mixture.
Situation #5	Lower than mixture	Higher than mixture	Some chemical info. from Species #3 reduced in Mixture.
Situation #6	Lower than mixture	No diff	Species #3 chemical info. found in Mixture #2
Situation #7	No diff	Lower than mixture	Species #1 chemical info. found in Mixture #2
Situation #8	Lower than mixture	Lower than mixture	New chemical info. found in Mixture #2

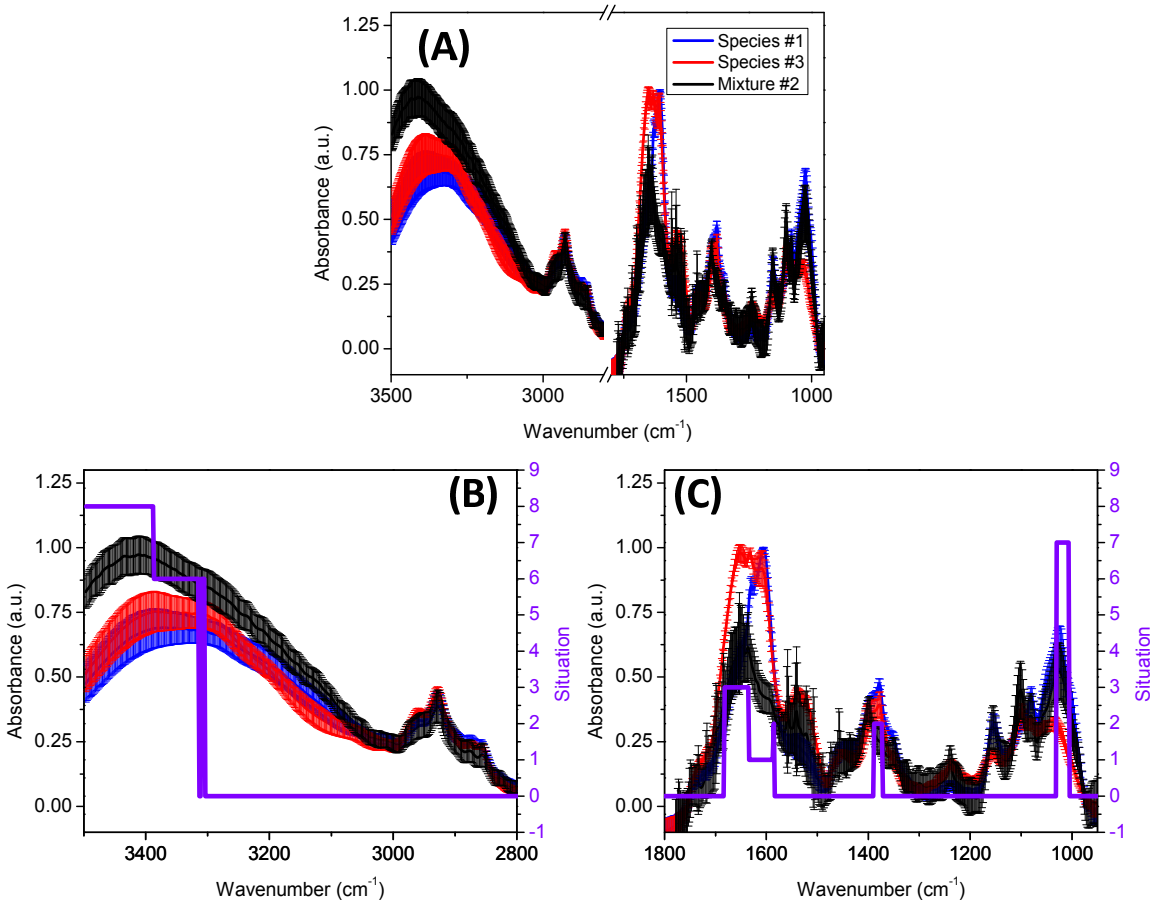


Figure 27. (A) Averaged spectra from Species #1, Species #3 and Mixture #2. Resultant comparison test depicts the chemical contributions contained within Mixture #2 at each wavenumber position: (B) Zoomed in from 3500 – 2800 cm^{-1} (C) Zoomed in from 1800 – 950 cm^{-1}

A similar analysis was performed in Figure 28 (A) and (B), mean spectra of Species #1 and Species #3 are both individually plotted vs. Mixture #2 at 2.07 mM HCO_3^- and there are visual indications of change. While these arrows are pointing to a visual assessment of spectroscopic change, this is not a reliable evaluation. To determine whether the visual assessment was correct, the mean spectra are t-tested to determine if there is a true significant difference at all wavenumber positions. There is a significant statistical difference between Species #1 and Mixture #2 as well as between Species #3 and Mixture #2. After the mean spectra were t-tested the comparison test was performed in Figure 28 (E). From the comparison test, the analysis determined that there is new chemical information, Situation #8, found at: ~1700 cm^{-1} and 1500 cm^{-1}. At ~1700 cm^{-1} it was found references [53], [59], [95], [143], [144] and (Table 6) that these are where lipids, fatty acids and carboxylic acid are present. At ~1550 cm^{-1} is the area of the Amide I peak.

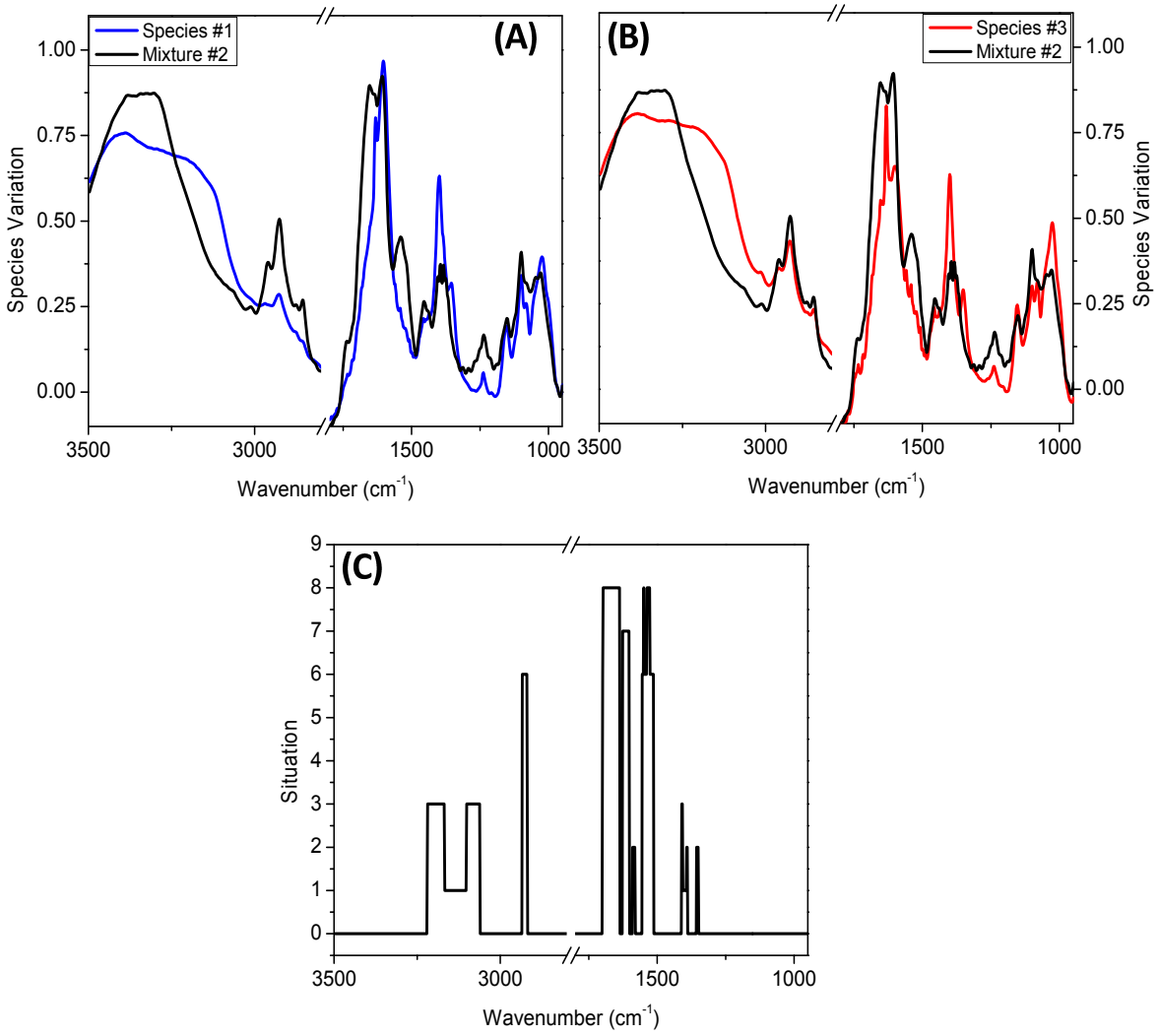


Figure 28. 2.07 mM HCO_3^- analysis of: (A) Species #2 and Mixture #3 mean spectra (B) Species #3 and Mixture #3 mean spectra (C) Results of comparison test

Table 6. FT-IR spectroscopy band assignments [53], [59], [95], [143] and [144]

Wavenumber (cm^{-1})	Band Assignment	Functional Groups
~3500-3300	C-H, N-H	Proteins
~3300-2800	saturared CH	Lipids
~1745-1700	C=O esters	Lipids, fatty acids and carboxylic acids
~1655-1638	C=O	Amide I (protein) also C=C stretches
~1545-1540	N-H, C-N	Amide II (protein)
~1455-1450	CH_3 and CH_2	Proteins and lipids
~1460-1370	CH_3, CH_2, C-O and COO-	Carboxylic acids and proteins
~1320	C-H, N-H	Amide III (protein)
~1244-1230	P=O	Phosphodiester backbone of nucleic acid (DNA and RNA)
~1200-900	C-O-C, P=O	Polysaccharides, nucleic acid
~1080	Si-O	Silicate frustules

To evaluate the overall concentration dependency, each Mixture #2 comparison test under carbon concentrations. Figure 29 displays all comparison tests for Mixture #2 under varying carbon concentrations and depicts how the mixture responds differently under each carbon concentration. The impact of the carbon concentration on Mixture #2 is reflected in the different Situations (colored bars) in these comparison tests. All concentrations revealed several yellow peaks depicting a reduction of chemical information from Species #3 that were predominantly found in the lipid region of 3300 – 2800 cm^{-1} (Table 6) and the protein region of the Amide I area between 1655 – 1638 cm^{-1}. Species #1 and Species #3 also reveal reduced spectroscopic features within the protein region of 3600 – 3300 cm^{-1}. Even though the green and yellow hues depict areas of reduced chemical information from single species, overall both single species were able to coexist in all carbon concentrations which are visualized at Situation #0, Situation #6 and Situation #7. Not only were Species #1 and Species #3 able to coexist, but new chemical features were found at various wavenumber positions in all carbon concentrations for Situation #8 except for 1.11 mM HCO_3^-.

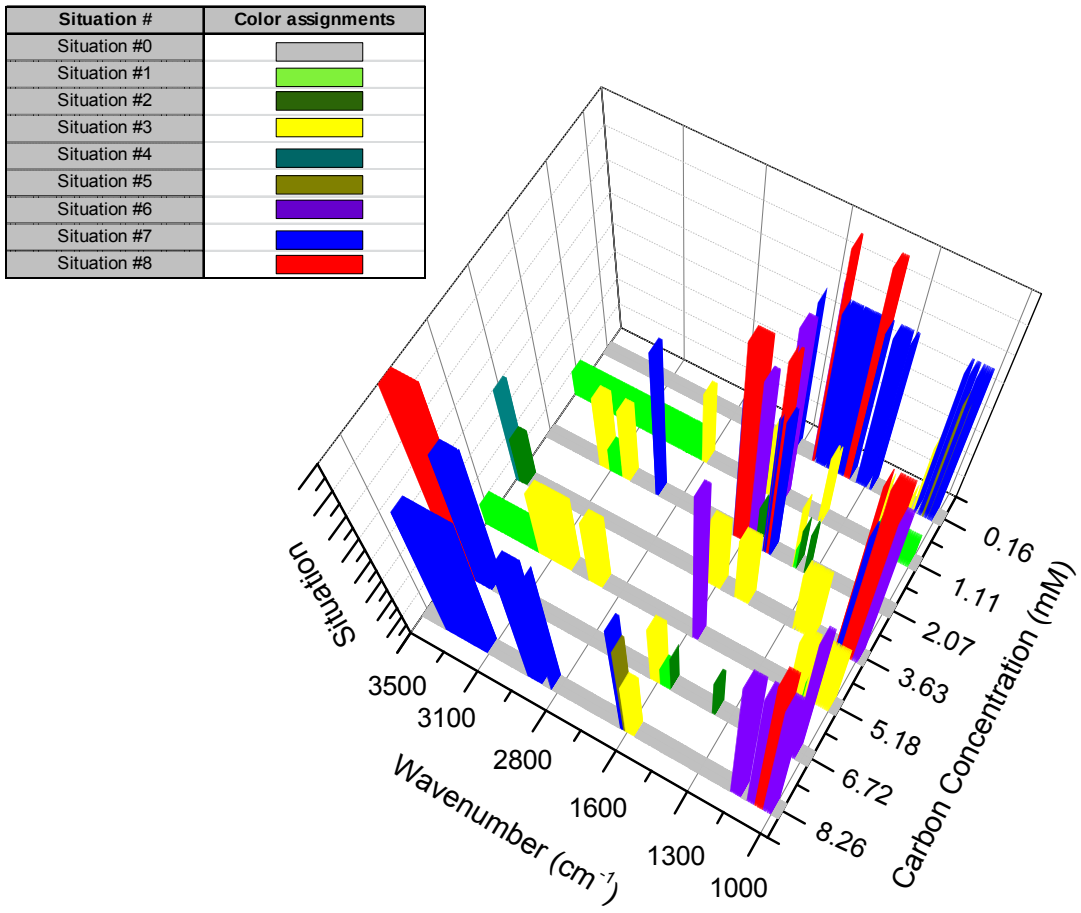


Figure 29. All comparison test results depicted for Mixture #2 displaying all Situations at varying carbon concentrations from 0.16 – 8.26 mM HCO_3^-. The data represents the chemical information present in Mixture#2 between 3500 – 950 cm^{-1}.

4.2. Results

The hypothesis is that the chemical groups (i.e. lipids, proteins and nucleic acids) which interact with their ecological community are the cause for these changes in chemical composition when species are grown together. This analysis through the utilization of FT-IR spectroscopy and statistical analyses were able to provide a visualization of what takes place chemically when species are grown together and was also able to demonstrate the impacts of species interactions.

The results display regions of the spectrum which differ due to growth environments (i.e. dual cultures vs. single cultures as well as nutrients and concentrations). To utilize this information, a comparison test was employed to determine what chemical information is contained within each mixture spectrum. This information will also be utilized for relating spectroscopic information to their ecological surroundings. Figure 30 depicts the information acquired from the comparison test introduced in section 4.1 for all microalgae species under all nutrient sources, both excess and limiting. These analyses methods were developed to gain a better understanding of the cells' chemical adaptation to their ever changing environment.

Within these studies, it was determined that microalgae species do not always react to environmental changes of increasing concentration linearly. Figure 30 visually displays the nonlinear responses for each mixture under all nutrient sources and concentrations. For example, the comparison tests for Mixture #1 under nitrate concentrations 0.16 mM, 0.549 mM and 0.873 mM demonstrate a nonlinear relationship between nutrient concentrations and chemical information present within Mixture #1's spectrum. Additionally, Mixture #1 under nitrate conditions is the only mixture and nutrient source that does not contain any additional chemical information.

When Mixture #1 and Mixture #3 are grown under 0.16 mM carbon and ammonium concentrations, Species #2 consistently contains chemical information within the protein region of 3600 – 3300 cm^{-1}. Under both nitrogen sources at 1.28 mM concentration, when Species #3 is present new chemical features are found in the protein and lipid regions of 3600 – 2800 cm^{-1}. Overall when comparing all dual cultures, if Species #3 is present there are more additional features found within the protein regions of: 3600 – 3300, 1655 – 1638, 1455 – 1450 and 1320 cm^{-1} than for any other functional group area.

From assessing all comparison tests, it is obvious that the spectroscopic variances do not just differ by the nutrient source, but also vary within each nutrient's concentrations. Additional chemical features are found within Mixture #3 under all nitrate concentrations except 0.16 mM. Whereas, Mixture #1 does not contain any additional information when grown within varying nitrate conditions. There are several areas of chemical reduction from Species #1 in Mixture #1 at each nitrate concentration within numerous wavenumber positions. Overall the main trend was found between Mixture #2 and Mixture #3, which are both missing spectroscopic information from Species #3 at various wavenumber positions from at least three different concentrations within all nutrient sources. At each of the following conditions, Species #3 has chemical information that is reduced within Mixture #2 and Mixture #3 i.e. carbon: 1.11 mM, 5.18 mM, 6.72 mM and 8.26 mM, ammonium 0.35 mM, 0.873 mM, and 1.47 mM and finally nitrate 0.35 mM, 1.47 mM and 1.65 mM. The following analysis applies to Mixture #2 and Mixture #3:

At various protein regions Species #3 is reduced in both Mixture #2 and Mixture #3 at the following wavenumber regions ~3600 – 3300 cm^{-1} for the following cultures: 1.11 mM HCO_3^-, 0.35 mM NH_4^+, 0.873 mM NH_4^+, 1.47 mM NH_4^+ and 1.47 mM NO_3^-. Within the Amide I and II protein regions Species #3 is missing chemical information at ~1650 cm^{-1} for 6.72 mM HCO_3^- and 8.26 mM HCO_3^-, ~1575 cm^{-1} for 1.65 mM NO_3^- and ~1500 cm^{-1} for 3.63 mM HCO_3^-. Species #3 was also reduced in the lipid region of 3300 – 2800 cm^{-1} within the following concentrations: 1.11 mM HCO_3^-, 5.18 mM HCO_3^-, 0.35 mM NH_4^+, 0.873 mM NH_4^+, 1.47 mM NH_4^+, 0.35 mM NO_3^- and 1.47 mM NO_3^-. At ~1725 cm^{-1} where lipids, fatty acids and carboxylic acids are found, which are crucial components in the production of biofuels and pharmaceuticals, Species #3 has reduced chemical information in both 3.63 mM HCO_3^- and 1.47 mM NO_3^- cultures. In the range of ~1200 – 900 cm^{-1}, polysaccharides and nucleic acids, spectroscopic information from Species #3 is not as strong from 0.35 mM NH_4^+, 0.35 mM NO_3^- and 3.63 mM HCO_3^- cultures. New chemical information can be found between ~3600 – 3300 cm^{-1} in both Mixture #2 and Mixture #3 when grown in a culture medium of 6.72 mM HCO_3^- as well as in 1.28 mM NH_4^+. At 2.07 mM HCO_3^- and 0.873 mM NH_4^+ in the Amide I region (~1650 cm^{-1}) and at ~1450 cm^{-1} where proteins and lipids are present, there are new spectroscopic features. Situation #8 is found within these two mixture spectrum at ~3300 – 3000 cm^{-1} in the lipid region and also in a polysaccharide and nucleic acid area of ~1075 cm^{-1} in 8.26 mM HCO_3^-. Although chemical information was reduced in dual cultures, each single species demonstrated

their ability to coexist within each mixture culture and is visualized at Situation #0, Situation #6 and Situation #7.

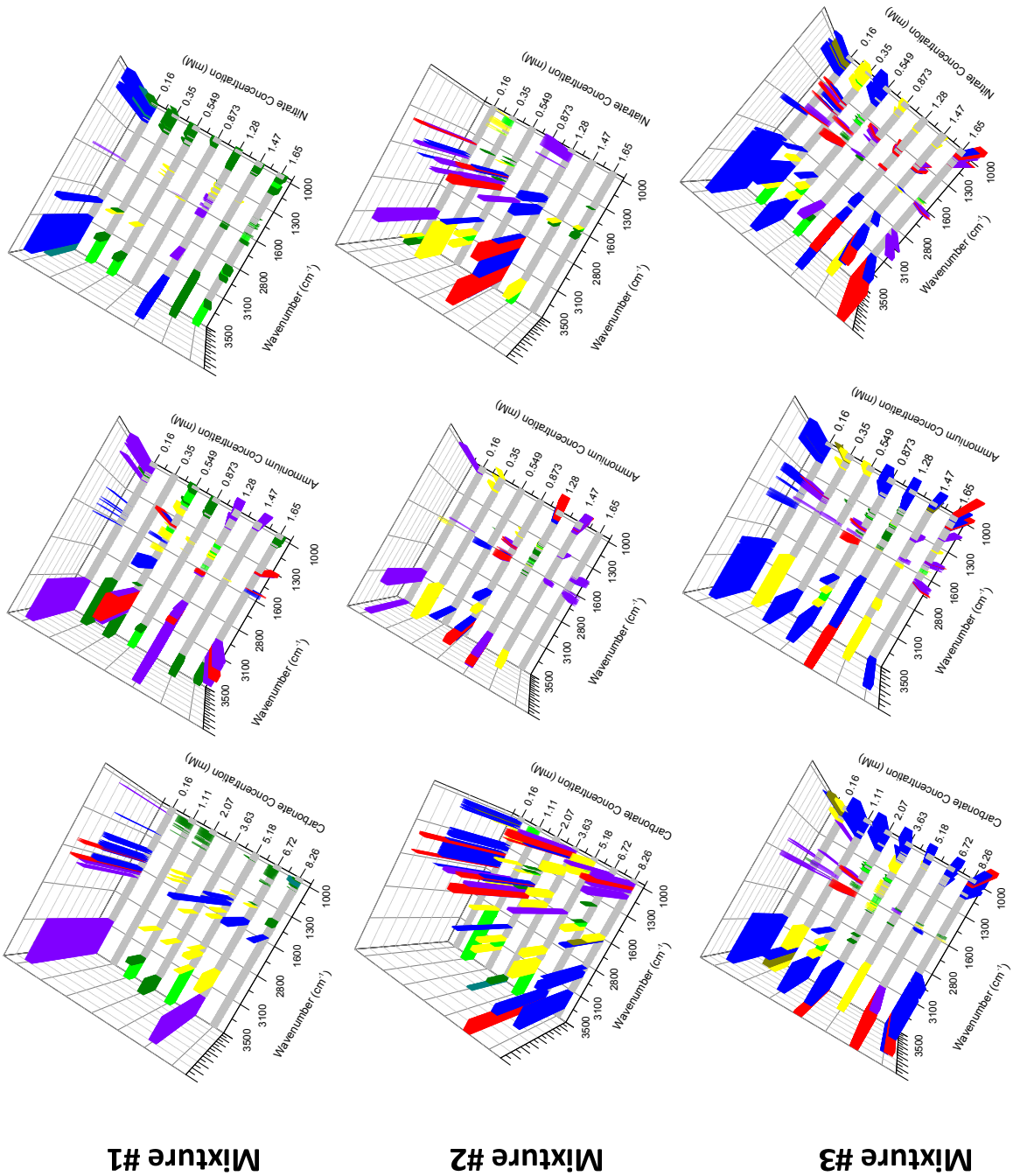


Figure 30. Concentration differences depicted in comparison test results of all three nutrient sources.

4.3. Conclusions

The goal of this investigation was to study the chemical interaction between microalgae species when grown within a mixed species environment. To study these interactions chemometric tools were developed to identify microalgae's chemical response to varying chemical parameters via FT-IR spectroscopy.

In order to mimic chemical changes that occur in marine ecosystems, three microalgae dual cultures were grown under a series of well-defined nutrient conditions. This analysis assessed the ability to detect and analyze chemical changes in marine ecosystems when mixtures of algal strains are grown together under varying nutrient conditions versus unialgal cultures. This in turn will serve as an indicator for determining the changes that occur within the ecosystem.

The analysis was based off of the cells' mean spectroscopic signatures via t-test and comparison analysis. Studies revealed at which wavenumber position each spectrum were statistically different. Within each mixture spectrum, a comparison test was utilized to determine the chemical information that is present. It was discovered, that some mixture spectrum contained additional chemical information that was not present in either of the single species spectrum. Finally it was found that when species are grown together, some spectroscopic features are reduced.

For a more in depth study of the biomolecules, a references measurement should be applied to confirm and quantify these findings. These results are a first step into investigating changes in chemical composition due to species chemical impacts.

5. Summary and Conclusions

For the past few decades, the Earth has been greatly impacted by anthropogenic CO_2 and nitrogen by way of industrialization and fossil fuel combustion. This shift has caused drastic, harmful changes to occur to the world's ecosystems, especially the aquatic ecosystems. Because these systems cover 71% of the Earth's surface, there is an obvious need to study these impacts in order to detect the changes that occur. For this reason, microalgae were chosen due to their ability to rapidly adapt to their environment and were utilized as proof-of-principle. Therefore, analytical methodologies were needed for rapid analysis to study microalgae for obtaining a comprehensive assessment of marine environments.

Thus, this dissertation developed innovative analytical and mathematical methodologies based on microscopic imaging, FT-IR chemical signatures as well as novel chemometric techniques to study how microalgae respond under various environmental conditions for utilization as an *in-situ* sensor to detect change.

Three microalgae species, *Dunaliella salina*, *Nannochloropsis oculata* and *Dunaliella parva*, were chosen due to their ability to rapidly adapt to their environment. Each species was cultured by *in*-vitro analysis under varying nutrient (inorganic carbon and nitrogen) conditions and concentrations to simulate real-world environmental parameters.

Within this dissertation, an images analysis was employed due to its rapid ability to analyze a large number of cells. It was shown that the method developed was capable of discriminating different cell species based on their cell shapes and sizes. It is also demonstrated, that through the utilization of nonlinear multivariate methods, the nutrient availability has a quantifiable impact on cell size.

Since microalgae cells' size distribution determines the amount of available biomass and cell size and growth rate influence the aquatic habitant's response to environmental changes as well as impacts on the food web structures further analyses were performed. A study was implemented to determine how cell size and growth rate correlate. While no general trend was found between growth rate and cell size, it was proven that the novel modeling method was able to model both growth rate and cell size accurately. Therefore this tool can be utilized for future

analysis of concentration predictions of nutrient induced physical impacts which occur within the ecological community.

For chemical investigations, an analytical methodology was developed to detect chemically induced shifts in the microalgae's chemical composition. The hypothesis was: (1) the chemical compounds inside microalgae cells (i.e. lipids, proteins and nucleic acids) interact with their external environment are the causes of these changes in chemical composition when species are grown together. Within a mixture spectrum, wavenumber regions in which significant changes in the spectroscopic signatures occurred helped to identify which chemical compounds are involved in chemically induced interactions. Mixture spectrum displayed new chemical information that was not present within the single species spectra. Species specific chemical impacts also revealed the reduction of spectroscopic features at various wavenumber regions. The results have demonstrated that chemical change is present in the biomass and are nutrient source and concentration dependent. These methodologies have demonstrated the ability to detect chemical shifts within changing environments via the utilization of a biological sensor.

Given the findings of this dissertation, the foundation has been laid for assessing: (1) the prediction of the chemical state of the ecosystem, (2) the cells' nutritional value for the advancement of biofuels and pharmaceutical research and (3) the assimilation of atmospheric CO_2 as well as nitrogen. The studies performed in this dissertation demonstrate novel analytical approaches for assessing the effects of chemical change on marine ecosystems.

List of References

[1] J. Beardall and S. Stojkovic, ScienceAsia **32**, (2006), 1-10.

[2] G. Peters, G. Marland, C. Le Quéré, T. Boden, J. Canadell, M. Raupach. Nature Climate Change. **2**, (2012), 2-4.

[3] M. Eby, K. Zickfeld, A. Montenero, D. Archer, K. Meissner, A.Weaver. Journal of Climate **22**, (2009), 2501-2511.

[4] J. Blunden, D. Arndt, M. Baringer (Eds.). Bulletin of the American Meteorological Society **92**, (2011), S1–S266.

[5] L. Brennan and P. Owende. Renewable and Sustainable Energy Reviews. 14, (2009), 557–577.

[6] G. C. Hayes, A. J. Richardson, C. Robinson. Trends in Ecology & Evolution. **20**, (2005), 337–344.

[7] E. W. Nester, D. G. Anderson, C.E. Roberts, M.T. Nester. *Microbiology A Human* Perspective. (McGraw-Hill, New York, NY, 2007).

[8] M. Y. Menetrez . Environ. Sci. Technol. 46, (2012), 7073−7085.

[9] R. A. Andersen. *Algal Culturing Techniques* (Elsevier, Burlington, 2005).

[10] J. Raven, M . Giordano, J. Beardall, S. Maberly. Photosynthesis Research. **109**, (2011), 281–296.

[11] M. Behrenfeld, R. O'Malley, D. Siegel, C. McClain, J. Sarmiento, G. Feldman, A. Milligan, P. Falkowski, R. Letelier, E. Boss. Nature. **444**, (2006), 752 – 755.

[12] C. Field, M. Behrenfeld, J. Randerson, P. Falkowski. Science. **281**, (1998), 237-240.

[13] G. Hays, A. Richardson, C. Robinson. Trends in Ecology and Evolution, **20**, (2005), 337 -344.

[14] J. S. Erickson, N. Hashemi, J. M. Sullivan, A. D. Weidemann, F. S. Ligler. Anal. Chem. **84**, (2012), 839–850.

[15] E. Huertas, G. Navarro, S. Rodríguez-Gálvez, L. Prieto. Canadian Journal of Botany. **83**, (2005) 929–940.

[16] R. Bardgett, C. Freeman, N. Ostle. The ISME Journal. **2**, (2008), 805–814.

[17] J. Zhou, K. Xue, J. Xie, Y. Deng, L. Wu, X. Cheng, S. Fei, S. Deng, Z. He, J. Van Nostrand, Y. Luo. Nature Climate Change. **2**, (2012), 106-110.

[18] C. Quéré, C. Rödenbeck *et al.* Science. **316**, (2007), 1735–1738.

[19] P. T. J. Johnson, A. R. Townsend, C. C. Cleveland, P. M. Gilbert, R. W. Howarth, V. J. McKenzie, E. Rejmankova, M. H. Ward. Ecological Applications **20**, (2010), 16–29.

[20] P. M. Gilbert, D. M. Anderson, P. Gentien, E. Graneli, K. G. Sellner. Oceanography. **2**, (2005), 130-141.

[21] K. G. Sellner, G. J. Doucette, G. J. Kirkpatrick. J. Ind. Microbiol Biotechnol. **30**, (2003), 383-406.

[22] R. W. Howarth, Harmful Algae. **8**, (2008), 14–20.

[23] W. Rodhe. Algal Cult. Limnol., **21, (**1976), 7-20.

[24] J. Berges, P. Falkowski. Limnology & Oceanography. **43**, (1998), 129-135.

[25] M. Giordano, G. Bowes. Plant Physiology. **11**, (1997), 1049-1056.

[26] S. Burkhardt, I. Zondervan, U. Riebesell. Limnology and Oceanography. **44**, (1999), 683-690.

[27] S. Sanudo-Wilhelmy, A.. Kustka, C. Gobler, D. Hutchins, M. Yang, K. Lwiza, J. Burns, D. Capone, J. Raven, E. Carpenter. Nature. **411**, (2001), 66 – 69.

[28] J. Raven. Photosynthesis Research. **77**, (2003) 155–171.

[29] J. Beardall, S. Roberts, J. Raven. Canadian Journal of Botany. **83**, (2005), 859–864.

[30] J. Raven. Photosynthesis Research. **77**, (2003), 155–171.

[31] M. Giordano, J. Beardall, J. A. Raven. Annu. Rev. Plant Biol. **56**, (2005), 99-131.

[32] M. Behrenfeld, K. Halsey, A. Milligan, *Phil. Trans. Royal Soc. B.* **363**, (2008), 2687-2703

[33] D. Bilanovic, A. Andargatchew, T. Kroeger, G. Shelef. Energy Conversion and Management. **50**, (2009), 262–267.

[34] K. Halsey, A. Milligan, M. Behrenfeld. Journal of Phycology. **47**, (2011), 66-76.

[35] J. Raven, M. Giordano, J. Beardall, S. Maberly. Phil. Trans. R. Soc. B. **367**, (2012), 493–507.

[36] C. Harley, R. Hughes, K. Hultgren, B. Miner, C. Sorte, C. Thornber, L. Rodriguez, L. Tomanek, S. Williams. Ecology Letters. **9**, (2006), 228 - 241.

[37] F. Schläpfer, B. Schmid. Ecological Applications. **9**, (1999), 893–912.

[38] A. Longhurst. Limnology and Oceanography. **36**, (1991), 1507-1526.

[39] R. B. Horton, M. McConico, C. Landry, T. Tran, F. Vogt. Analytica Chimica Acta, **746**, (2012), 1-14.

[40] M. McConico, R. Horton, K. Witt, F. Vogt. J Chemometrics. **26**, (2012), 585-597.

[41] M. McConico and F. Vogt. J. Chemometrics. (2013), Accepted.

[42] J. Beardall, S. Stojkvic, S. Larsen. Plant Ecology and Diversity. **2**, (2009), 191-205.

[43] J. Beardall, C. Sobrino, S. Stoijkovic. Photochemical & Photobiological Sciences. **8**, (2009), 1257-1265.

[44] D. R. Vardon, M. M. Clark, D. A. Ladner. Desalination. 277, (2011), 193-200.

[45] V.I. Slaveykova, C. Guignard, T. Eybe, H.N. Migeon, L. Hoffmann. Analytical Bioanalytical Chemistry. **393**, (2009), 583-589.

[46] Y. Iwata. Journal of Radioanalytical and Nuclear Chemistry. **2**, (2001), 343-348.

[47] R. Cesareo, A. Castellano, A.M. X-Ray Spectrometry. **27**, (1998), 257-264.

[48] Y.Y. Huang, C.M. Beal, W.W. Cai, R.S. Ruoff, E.M. Terentjev. Biotechnology and Bioengineering. **105**, (2010), 889-898.

[49] D.M. Haaland, E.V. Thomas. Analytical Chemistry. **60**, (1988), 1193-1202.

[50] F. Vogt, personal communication (11/2012)

[51] P. Heraud, J. Beardall, D. McNaughton, B.R. Wood. FEMS Microbiological Letter. **275**, (2007) 24-30.

[52] C. Hirschmugl, B. Zuheir-El Bayarri, M. Bunta, J. Holt, M. Giordano. Infrared Physics & Technology, **49**, (2006), 57-63.

[53] M. Giordano, M. Kansiz, P. Heraud, J. Beardall, B. Wood, D. McNaughton. Journal of Phycology. **37**, (2001), 271-2 79.

[54] P. Heraud, B. Wood, M. Tobin, J., Beardall, D. McNaughton. FEMS Microbiology Letters, **249**, (2005), 219-225.

[55] M.Giordano, A. Norici, D. Gilmour, J. Raven. Functional Plant Biology, **34**, (2007), 925-934.

[56] A. Domenighini and M. Giordano. Journal of Phycology. **45**, (2009), 522-531.

[57] A. Beardall, T. Berman, P. Heraud, M. O. Kadiri, B.R. Light, G. Patterson, S. Roberts, B. Sulzberger, E. Sahan, R. Uehlinger, B. Wood. Aquatic Sciences. **63**, (2001), 107-121.

[58] P. Heraud, S. Stojkovic, J. Beardall, D. McNaughton, B.R. Wood. Journal of Phycology. **44**, (2008), 1335-1339.

[59] J. N. Murdock, W. K. Dodds, J. A. Reffner, D. L. Wetzel. Spectroscopy. **25**, (2010), 32-41.

[60] R. E. Lee. *Phycology*. (Cambridge University Press, Cambridge, UK, 1999).

[61] L. Barsanti, P. Gualtieri. *Algae*. (CRC Press, Boca Raton, 2006).

[62] J. P. Grover. J. Phycol. **25**, (1989), 402-405.

[63] H. Hillebrand, C. Dürselen, D. Kirschtel, U. Pollingher, T. Zohary, **35**, (1999), 403–424.

[64] L. Arin, X. A. G. Moran, M. Estrada. Journal of Plankton Research. **24**, (2002), 1019-1033.

[65] M. Giordano, S. Ratti, A. Domighini, F. Vogt. Plant Ecology and Diversity. **2**, (2009), 155-164.

[66] M. Frac, S. Jezierska-Tys, J. Tys. African Journal of Biotechnology. **9**, (2010) 9227 -9236.

[67] D. Nelson, M. Cox, *Lehninger Principles of Biochemistry*. (W.H. Freeman and Company, New York, NY, 2004).

[68] T. L. Brown, H. E. Lemay, B.E. Bursten, C.J. Murphy, *Chemistry: The Central Science* (Prentice Hall, Upper Saddle River, NJ, 2009).

[69] J. McMurry. *Organic Chemistry*. (Brookes-Cole, Belmont, CA, 2004).

[70] M. A. Borowitzka, Microbiol Sci. **3**, (1986), 372-5.

[71] G.W. Prescott. **22**, (1956), 167-240.

[72] M. A. Borowitzka and C. J. Siva. J Appl Phycol. **19**, (2007), 567-590.

[73] A. A. Ramos, J. Polle, D. Tran, J. C. Cushman, E. Jin, J. C. Varela. Algae **26,** (2011), 3-20.

[74] M.A. Hajibagheri, D.J. Gilmour, J.C. Collins, and T.J. Flowers. Journal of Experimental Botany. **37**, (1986), 1725-1732.

[75] S. Rabinowitch, N.B. Grover, B.Z. Ginzburg. J. Membrane Biol. **22**, (1975), 211-230.

[76] K. P. Fawley and M. W. Fawley. Protist. **158**, (2007), 325-336.

[77] E. N. Sharifah, M. Eguchi. PLoS ONE. 6, (2011), 1-8.

[78] J. G. Day, E. E. Benson, R. A. Fleck. In Vitro Cell. Dev. Biol.—Plant. **35**, (1999), 127-136.

[79] O. Pulz. Appl. Microbiol. Biotechnol. 57, (2001), 287-293.

[80] S. D. Varfolomeeva and L. A. Wasserman. Applied Biochemistry and Microbiology, **47**, (2011), 789–807.

[81] Y. Chisti. Biotechnology Advances. **25**, (2007), 294–306.

[82] M. A. Borowitzka. **7**, (1995), 3-15.

[83] J.L. Zevenbergen and M. Rudrum. Fat Sci. Technol. **12**, (1993), 456-460.

[84] S. Lowry. Spectroscopy. 26, (2011), 30-36.

[85] V. Patil, K. Tran, H. R. Giselrød. Int. J. Mol. Sci. **9**, (2008), 1188-119.

[86] Y. Y. Braga. Marine Ecology. **3**, (1980), 157-161.

[87] P. Harrison, R. Waters, F. Taylor F. J. Phycol. **16**, (1980), 28-35.

[88] J. Berges, D. Franklin, J. Harrison. J. Phycol **37**, (2001), 1138-1145.

[89] M. Giordano, personal communication (8/2009)

[90] M. D. Rolfe, C. J. Rice, S. Lucchini, C. Pin, A. Thompson, A. D. S. Cameron, M. Alston, M. F. Stringer, R. P. Betts, J. Baranyi, M. W. Peck, J. C. D. Hinton, Journal of Bacteriology. **194**, (2012), 686-701.

[91] M. K. Danquah, B. Gladman, N. Moheimani, G. M. Ford. Chemical Engineering Journal. **151**, (2009), 73-78.

[92] B. M. Auinger, K. Pfandlh, J. Boenigk. Appl. Environ. Microbiol. **74**, (2008), 2505-2510.

[93] Anita F. da Silva, Sergio O. Lourenc., Ricardo M. Chaloub. Aquatic Botany. **91**, (2009), 291-297.

[94] C. J. Zhu,Y. K. Lee, Journal of Applied Phycology. **9**,(1997), 189-194.

[95] K. Kneipp, R. Aroca, H. Kneipp, E. Wentrup-Byrne. *New Approaches in Biomedical Spectroscopy*. (Oxford University Press, Washington, 2007).

[96] J. Rodriguez, G. Reina, J. Rodriguez. Biomed. Chromatogr. **11**, (1997), 335-336.

[97] A. Derrien, L. Coiffard, C. Coiffard, Y. De Roeck-Holtzhauer. J. Appl. Psychol. **10**, (1998), 131–134.

[98] M. Rigobello-Masini, J. Penteado, C. Liria, M. Miranda, J. Masini. Anal. Chim. Acta. **628**, (2008), 123–132.

[99] J. Penteado, M. Rigobello-Masini, C. Liria, M. Miranda, J. Masini. J. Sep. Sci. **32**, (2009), 2827–2834.

[100] K. Stehfest, J. Soepel, C. Wilhelm. Plant Physiology and Biochemistry. **43**, (2005), 717-726.

[101] N. Draper, H. Smith. *Applied Regression Analysis. (*John Wiley & Sons, New York, 1998).

[102] A. Sen, M. Srivastava. *Regression Analysis – Theory, Methods, Applications*. (Springer, New York, 1990).

[103] S. Weisberg. *Applied Linear Regression.* (John Wiley & Sons, Hoboken, 2005).

[104] G. Box, N. Draper. *Empirical Model-Building and Response Surfaces*. (John Wiley & Sons: New York, 1987).

[105] D. Naumann, D. Helm, H. Labischinski. Nature.**351**, (1991), 81-82.

[106] J. N. Murdock and D. L. Wetzel. Appl. Spec. Rev. **44**, (2009), 335 – 361.

[107] A. Gelman, J. Carlin, H. Stern, D. Rubin. *Bayesian Data Analysis*. (Chapman & Hall/CRC, Boca Raton, 2004).

[108] W. Press, S. Teukolsky, W. Vetterling, B. Flannery. *Numerical Recipes in C++*, (Cambridge University Press: New York, 2002).

[109] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi. *GNU Scientific Library Reference Manual*, (Network Theory Ltd, 2009).

[110] F. Vogt. J. Chem. Educ. **88**, (2011), 1672–1683.

[111] D. Haaland, E. Thomas. Anal. Chem. **60**, 1988, 1193-1202.

[112] H. Martens. T. Naes. *Multivariate Calibration*. (John Wiley & Sons, New York, 1991).

[113] http://www.gnu.org/software/gcc/ accessed on 06/12/2012

[114] http://www.gnu.org/software/gsl/ accessed on 06/12/2012

[115] http://www.netlib.org/lapack/index.html accessed on 06/12/2012

[116] http://math-atlas.sourceforge.net/ accessed on 06/12/2012

[117] M. McConico, L. White, K. Witt, F. Vogt. in preparation.

[118] F. Vogt, F. Gritti, G. Guiochon. J. Chemom. **25**, (2011), 575-585.

[119] E. P. Y. Tang. Journal of Plankton Research. **17**, (1995), 1325-1335.

[120] Y. Sin, R. L. Wetzel, I. C. Anderson. Journal of Plankton Research. **22**, (2000), 1945-1960.

[121] A. E. Wilson, R. B. Kaul, O. Sarnelle. PLoS ONE. **5**, (2010), 1-7.

[122] C. L. Nunn, R. A. Barton. The American Naturalist. **156**, **(**2000), 519-533.

[123] M. Kagami, J. Urabe. Limnology. **2**, (2001),111-117.

[124] K. Banse. J. Phycol. **12**, (1976), 135-140.

[125] D. A. Schlesinger, L. A. Molot, B. J. Shuter. Can. J. Fish. Aquat. Sci. **38**, (1981), 1052- 1058.

[126] S. L. Nielsen and K. Sand-Jensen. Limnology and Oceanography. **35**, (1990), 177-181.

[127] K. J. Niklas and B. J. Enquist. PNAS. **98**, (2001), 2922-2927.

[128] Z. V. Finkel and A. J. Irwin. J. theor. Biol. **204**, (2000), 361-369.

[129] S. L. Nielsen. Journal of Phytoplankton. **28**, (2006), 489-498.

[130] S. Rubinow. *Introduction to Mathematical Biology*. (Dover Publications, Mineola, NY, 2003).

[131] J. Turner and J. Thayer. *Introduction to Analysis of Variance*. (Sage Publications: Thousand Oaks, CA, 2001).

[132] E. Crow, F. Davis, M. Maxfield. *Statistics Manual*. (Dover Publishing, Mineola, NY 2011).

[133] J. Beardall and J. A. Raven. Journal of Phycologia. **43**, (2004), 26-40.

[134] P. Heraud, B. R. Wood, J. Beardall, D. McNaughton. Journal of Chemometrics. **20**, (2006), 193-197.

[135] G. Klein P. Perera, European Commission. *Eutrophication and health*. *(*Luxembourg: Office for Official Publications of the European Communities, 2002)

[136] R.H. MacArthur. Pro. NaH. Acad. Sci. **64**, (1969), 1369-1375.

[137] O. Kerimoglu, D. Straile, F. Peeters. Journal of Theoretical Biology. **300** (2012), 330-343.

[138] A.H. Fielding and G. Russell. Journal Of Applied Ecology, **13**, (1976), 871-876.

[139] D. Tilman. Ecology. **58**, (1977), 338-348.

[140] M. R. Droop. J. Mar. Biol. Assoc. U.K. **54**, **(**1974), 825-855.

[141] R. B. Horton, E. Duranty, M. McConico, F. Vogt. Applied Spectroscopy. **65**, (2011), 442-453.

[142] M. McConico and F. Vogt. Analytica Chmica Acta. xx, (2013), xxxx.

[143] L. D'Souza, P. Devi, S. Shridhar, C. G. Naik. Analytical Chemistry. **3**, (2008), 135-143.

[144] G. V. Venkataramana, J. K. Kumar, A. G. D. Prasad, P. Karimi. Romanian J. Biophys. **20**, (2010), 315-322.

[145] M. K. Gilbert, R. D. Luttrell, D. Stout, F. Vogt. Journal of Chemical Education. **85**, (2008), 22-39.

[146] F. Vogt, and K. Booksh. *Chemometrics. Kirk-Othmer Encyclopedia Chemical* Technology. (Wiley- Interscience, Hoboken, NJ 2004).

[147] F. Vogt. Submitted 2012.

Appendices

Appendix 1

Chemometrics is utilized for mathematical analyses of measurement data. Within analytical chemistry, chemometrics has been used for calibrations in order to predict concentrations of unknown samples [145]. For this to be accomplished, it is important to have instrumentation that is able to follow such figures of merit: selectivity, precision and sensitivity. While these are important, they are sometimes very challenging to obtain. To account for these difficulties, chemometric analyses are employed in order to extract the desired information [146]. Several of these chemometric algorithms are based on least-squares regression techniques [145].

These methods have a calibration step in order to utilize the data and its concentration information for concentration prediction of an unknown sample [49].

Beer's Law, a univariate model, utilizes least-square regression. Beer's Law states that absorption at a specific wavenlength is given by the concentration of the sample, the molar absorptivity and the pathlength.

$$A(\lambda) = \varepsilon(\lambda) \cdot L \cdot c$$

(23)

Based on Beer's Law, the molar absorptivity times the absorption pathlength times the concentration of an analyte will give the absorption at a certain wavelength. Because the molar absorptivity and the length do not change over time, they can be determined in a calibration and stored for future analyses of unknowns. In real-world situations, this often cannot be successfully applied due to random measurement errors and overlapping absorbers contained in the samples causing cross-sensitivities and concentration errors. The solution is to incorporate absorbance measurements at several wavelength positions; this will help to reduce random errors and more important, an equation system can be built for evaluating multiple compounds and thus avoid cross-sensitivities. Thus, the originally univariate Beer's Law is being replaced by the multivariate Classical Least-Squares regression (CLS). However, CLS requires that all analytes contained in the samples must be known or concentration errors will still result [49]. If any unknown analyte is contained in the calibration, the calibration model will be flawed; if any appears in the unknown samples, the spectrum will be inconsistent with the calibration model. A technique that accounts for unknowns in the calibration samples (a very common situation) is Principal Component Regression (PCR). PCR is a two-step process first

using principle component analysis and followed by a regression step [49]. PCR does not make the assumption that everything is known within the sample. All spectroscopic information will be contained within the calibration model in order to correctly analyze the calibration spectra. With the use of PCR, analytes that are not expected will not lead to concentration errors because they will be implicitly contained within the calibration model [145]. Partial Least Squares is another technique used in quantitative spectral decomposition and is also very similar to PCR [49]. The difference is that PLS utilizes concentration information within the decomposition process [50].

Appendix 2

In order to introduce the difference between linear and nonlinear methods, a short background on regression analysis is supplied. The basis for regression analysis is found in the model function, Y_a, which is obtained from the data that was measured. This function depicts the relationship between the measured data and the information that is of interest. This model function comes with a number of model parameters. The prediction step then will use the calibration model, Y_a, to predict the properties of the unknown sample. Each experimental data point, $y_{i=1,\dots,N}^{cal}$, is impacted with a measurement error and the measured data and the model are described by $y_{n,i} = Y_a(x_i) + \delta_i$.

The basis of least-squares (^) is to look for an approximation of the true model parameters. These estimations are made by minimizing the sum-of-squared-errors that are obtained from the calculated model values and the measured values which can be seen in sections 3.1.1 - 3.1.4.

$$\mathrm{SSE}(\hat{\mathbf{a}}) = \sum_{i=1}^{N} \delta_i^2 = \sum_{i=1}^{N} \left[y_{n,i}^{\mathrm{cal}} - Y_{\hat{\mathbf{a}}}\left(\mathbf{x}_i^{\mathrm{cal}}\right) \right]^2$$

(24)

The parameters for which the surface has the lowest point is the one best adapted to the measured data and hence selected for the model parameters. The requirement for a minimum is, $\boldsymbol{\nabla}\mathrm{SSE}(\hat{\mathbf{a}}) = \mathbf{0}$ [147].

$$\frac{\partial \mathrm{SSE}(\hat{\mathbf{a}})}{\partial \hat{a}_p} = -2 \sum_{i=1}^{N} \left[y_{n,i}^{\mathrm{cal}} - Y_{\hat{\mathbf{a}}}\left(\mathbf{x}_i^{\mathrm{cal}}\right) \right] \cdot \frac{\partial Y_{\hat{\mathbf{a}}}\left(\mathbf{x}_i^{\mathrm{cal}}\right)}{\partial \hat{a}_p} = 0$$

(25)

When the gradient, $\boldsymbol{\nabla}$, is equal to zero, this means that $\mathrm{SSE}(\hat{\boldsymbol{a}})$ is no longer changing and that the extremum has been reached.

Least-squares regression models utilize equations (24), (25), but there is a difference between linear and non-linear methods. Linear methods are linear when using the sum-of-squared errors and their model parameters. For example for the model $Y = a_0 + a_1 x + a_2 x^2$ and the response (predictor) y. The equation for the sum-of-squared-error is:

$$\mathrm{SSE}\left(\widehat{\boldsymbol{a_0}}, \widehat{\boldsymbol{a_1}}, \widehat{\boldsymbol{a_2}}\right) = \sum_{i=1}^{N} \left(y_{n,i} - Y_{\hat{\mathbf{a}})}\right)^2$$

(26)

For determining this minimum, partial derivatives are calculated with respect to the desired parameters and are then set equal to zero:

$$\text{(I)}\ \frac{\partial \mathrm{SSE}}{\partial a_0} = 0 \qquad \text{(II)}\ \frac{\partial \mathrm{SSE}}{\partial a_1} = 0 \qquad \text{(III)}\ \frac{\partial \mathrm{SSE}}{\partial a_2} = 0$$

(27)

$$\begin{pmatrix} \frac{\partial \mathrm{SSE}}{\partial \hat{a}_0} \\ \frac{\partial \mathrm{SSE}}{\partial \hat{a}_1} \\ \frac{\partial \mathrm{SSE}}{\partial \hat{a}_2} \end{pmatrix} = \begin{pmatrix} -2 \sum_{i=1}^{N} \left(y_{n,i} - Y_{\hat{\mathbf{a}}}\right) \cdot 1 \\ -2 \sum_{i=1}^{N} \left(y_{n,i} - Y_{\hat{\mathbf{a}}}\right) \cdot x \\ -2 \sum_{i=1}^{N} \left(y_{n,i} - Y_{\hat{\mathbf{a}}}\right) \cdot x^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

(28)

(I)

$$0 = 2 \sum_{i=1}^{N} \left[-1 \cdot y_{n,i} + a_0 \cdot 1 \cdot 1 + a_1 \cdot 1 \cdot x_i + a_2 \cdot 1 \cdot x_i^2 \right]$$

$$0 = 2 \left[\sum_{i=1}^{N} -1 \cdot y_{n,i} + a_0 \sum_{i=1}^{N} 1 \cdot 1 + a_1 \sum_{i=1}^{N} 1 \cdot x_i + a_2 \sum_{i=1}^{N} 1 \cdot x_i^2 \right]$$

$$a_0 \sum_{i=1}^{N} 1 \cdot 1 + a_1 \sum_{i=1}^{N} 1 \cdot x_i + a_2 \sum_{i=1}^{N} 1 \cdot x_i^2 = \sum_{i=1}^{N} 1 \cdot y_{n,i}$$

(II)

$$0 = 2\sum_{i=1}^{N}\left[-x_i \cdot y_{n,i} + a_0 \cdot x_i \cdot 1 + a_1 \cdot x_i \cdot x_i + a_2 \cdot x_i \cdot x_i^2\right]$$

$$0 = 2\left[\sum_{i=1}^{N} x_i \cdot y_{n,i} + a_0 \sum_{i=1}^{N} x_i \cdot 1 + a_1 \sum_{i=1}^{N} x_i \cdot x_i + a_2 \sum_{i=1}^{N} x_i \cdot x_i^2\right]$$

$$a_0 \sum_{i=1}^{N} x_i \cdot 1 + a_1 \sum_{i=1}^{N} x_i \cdot x_i + a_2 \sum_{i=1}^{N} x_i \cdot x_i^2 = \sum_{i=1}^{N} x_i \cdot y_{n,i}$$

(III)

$$0 = 2\left[\sum_{i=1}^{N} -x_i^2 \cdot y_{n,i} + a_0 \cdot x_i^2 \cdot 1 + a_1 \cdot x_i^2 \cdot x_i + a_2 \cdot x_i^2 \cdot x_i^2\right]$$

$$0 = 2\left[\sum_{i=1}^{N} -x_i^2 \cdot y_{n,i} + a_0 \sum_{i=1}^{N} x_i^2 \cdot 1 + a_1 \sum_{i=1}^{N} x_i^2 \cdot x_i + a_2 \sum_{i=1}^{N} x_i^2 \cdot x_i^2\right]$$

$$a_0 \sum_{i=1}^{N} x_i^2 \cdot 1 + a_1 \sum_{i=1}^{N} x_i^2 \cdot x_i + a_2 \sum_{i=1}^{N} x_i^2 \cdot x_i^2 = \sum_{i=1}^{K} x_i^2 \cdot y_{n,i}$$

(29)

The unknowns (a_0, a_1, a) are now separate from the knowns for estimating the unknown parameters:

(I) $$a_0 \sum_{i=1}^{N} 1 \cdot 1 + a_1 \sum_{i=1}^{N} 1 \cdot x_i + a_2 \sum_{i=1}^{N} 1 \cdot x_i^2 = \sum_{i=1}^{N} 1 \cdot y_{n,i}$$

(II) $$a_0 \sum_{i=1}^{N} x_i \cdot 1 + a_1 \sum_{i=1}^{N} x_i \cdot x_i + a_2 \sum_{i=1}^{N} x_i \cdot x_i^2 = \sum_{i=1}^{N} x_i \cdot y_{n,i}$$

$$\text{(III)} \quad a_0 \sum_{i=1}^{N} x_i^2 \cdot 1 + a_1 \sum_{i=1}^{N} x_i^2 \cdot x_i + a_2 \sum_{i=1}^{N} x_i^2 \cdot x_i^2 = \sum_{i=1}^{N} x_i^2 \cdot y_{n,i}$$

(30)

The unknowns are a_0, a_1 and a_2 are placed into a vector and the sums are now written as dot products and put into a matrix and multiplied by the vector.

$$\begin{pmatrix} \mathbf{1^T \cdot y_1} & \mathbf{1^T \cdot 1} & \mathbf{1^T \cdot x} & \mathbf{1^T \cdot x^2} \\ \mathbf{x^T \cdot y_1} & \mathbf{x^T \cdot 1} & \mathbf{x^T \cdot x} & \mathbf{x^T \cdot x^2} \\ \mathbf{x^{2T} \cdot y_1} & \mathbf{x^{2T} \cdot 1} & \mathbf{x^{2T} \cdot x} & \mathbf{x^{2T} \cdot x^2} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

(31)

$$\begin{pmatrix} \mathbf{1^T \cdot y_n} \\ \mathbf{x^T \cdot y_n} \\ \mathbf{x^{2T} \cdot y_n} \end{pmatrix}$$

(32)

$$\mathbf{\Lambda^T \cdot \Lambda \cdot a} = \mathbf{\Lambda^T \cdot y}_n$$

(33)

The unknowns, **a**, are estimated via:

$$\left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \left(\mathbf{\Lambda^T \cdot \Lambda}\right) \cdot \mathbf{a} = \left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \mathbf{\Lambda^T \cdot y}_{\mathrm{n}}$$

$$\mathbf{a} = \left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \mathbf{\Lambda^T \cdot y}_{\mathrm{n}}$$

(34)

The previous analysis was a linear equation method that is simple and straight forward which determines the minimum from minimizing the sum-of-squared errors.

Next, starting from (25) a nonlinear equation is depicted in (35).

$$Y_{\hat{\mathbf{a}}}(x) = a_0 \cdot e^{-\frac{(x-a_1)^2}{2 \cdot a_2^2}}$$

(35)

This equation is nonlinear due to the model parameters a_1 and a_2 that are present in $\frac{\partial Y_{\hat{\mathbf{a}}}\left(x_k^{\text{cal}}\right)}{\partial a_p}$ is not a straightforward equation system to solve. For the linear model, all of the parameters are to the first power and do not contain any intertwined variables within the function also none of the parameters are contained within exp. Unlike in non-linear equation systems (i.e. (14)), the non-linearity lies within the model parameters and the solution for the sum-of-squared errors involves an iterative process until $\nabla \text{SSE} = 0$. When this occurs, the nonlinear equation must be solved iteratively [147].

Appendix 3

In the beginning of section 3.1.4, the following equation fits each individual bin and determines how the fit in each specific bin changes.

$$(S, c_1) = A(c_1) \cdot \exp\left\{-\frac{\left(S - S_0(c)\right)^2}{2 \cdot \sigma(c)^2}\right\}$$

(36)

If there is a shift, the shift that is seen is only bin by bin instead of the entire distribution for the specific concentration. This section develops the tools and demonstrates the proof of principle.

For the remainder of section 3.1.4, each parameter ($A, s_0\ and\ \sigma$) are founded by Taylor series expansions which are utilized to expand a function for identifying small parameters. Each γ are found during the fit. This equation is a multivariate Taylor expansion. It is a multivariate equation because of the utilization of two different concentrations. The following equations fit the entire surface. This fit gives information about the entire distribution and how the surface shifts for the particular species over an entire concentration range. This part of section 3.1.4 leads to more real world examples. The following is the expansion of solving for the Predictor Surfaces for concentration dependent surfaces:

$$A(c_1, c_2) = \gamma_{A,0} + \gamma_{A,1} \cdot c_1 + \gamma_{A,2} \cdot c_2 + \gamma_{A,3} \cdot c_1^2 + \gamma_{A,4} \cdot c_1 \cdot c_2 + \gamma_{A,5} \cdot c_2^2$$

(37)

A is the height of the Gaussian.

$$S_0(c_1, c_2) = \gamma_{S_{0,0}} + \gamma_{S_{0,1}} \cdot c_1 + \gamma_{S_{0,2}} \cdot c_2 + \gamma_{S_{0,3}} \cdot c_1^2 + \gamma_{S_{0,4}} \cdot c_1 \cdot c_2 + \gamma_{S_{0,5}} \cdot c_2^2$$

(38)

S_0 is the center position of the size distribution.

$$\sigma(c_1, c_2) = \gamma_{\sigma,0} + \gamma_{\sigma,1} \cdot c_1 + \gamma_{\sigma,2} \cdot c_2 + \gamma_{\sigma,3} \cdot c_1^2 + \gamma_{\sigma,4} \cdot c_1 \cdot c_2 + \gamma_{\sigma,5} \cdot c_2^2$$

(39)

σ is the width of the Gaussian.

s is the bin size.

$$S(c_1, c_2) = \gamma_{A,0} + \gamma_{A,1} \cdot c_{1+}\gamma_{A,2} \cdot c_2 + \gamma_{A,3} \cdot c_1^2 + \gamma_{A,4} \cdot c_1 \cdot c_2 + \gamma_{A,5} \cdot c_2^2$$

$$\cdot \exp(-\frac{(s - (\gamma_{\sigma,0} + \gamma_{\sigma,1} \cdot c_{1+}\gamma_{\sigma,2} \cdot c_2 + \gamma_{\sigma,3} \cdot c_1^2 + \gamma_{\sigma,4} \cdot c_1 \cdot c_2 + \gamma_{\sigma,5} \cdot c_2^2))^2}{2(\gamma_{\sigma,0} + \gamma_{\sigma,1} \cdot c_{1+}\gamma_{\sigma,2} \cdot c_2 + \gamma_{\sigma,3} \cdot c_1^2 + \gamma_{\sigma,4} \cdot c_1 \cdot c_2 + \gamma_{\sigma,5} \cdot c_2^2)^2})$$

(40)

Appendix 4:

Summary of Graduate School Honors, Publications, and Presentations

Honors

1. DuPont Diversity Fellowship, Department of Chemistry, University of Tennessee, Knoxville, TN, January **2011-2013**

2. Graduate Student Senate Travel Award, University of Tennessee, Knoxville, TN, September **2012**.

3. Graduate Student Senate Travel Award, University of Tennessee, Knoxville, TN, February **2012**.

4. Carol Storm Underrepresented Minority Fellowship (CSURM), GRC **2010**.

5. Outstanding Service Award, Department of Chemistry, University of Tennessee, Knoxville, TN, April **2010**.

Publications

1. M. McConico and F. Vogt. *Analytica Chimica Acta*, **2013**, xx, xxx. *Submitted*.

2. M. McConico and F. Vogt. *Journal of Chemometrics*, **2012**, xx, xxx. *Accepted*.

3. M. McConico, R. B. Horton, K. Witt, F. Vogt. *Journal of Chemometrics,* **2012**, 26(11-12), 585-597.

4. R. B. Horton, M. McConico, C. Landry, T. Tran, F. Vogt. *Analytica Chimica Acta,* **2012**, 746, 1-14.

5. R. B. Horton, E. Duranty, M. McConico, F. Vogt. *Applied Spectroscopy*, **2011**, *65*, 442-453.

Presentations

1. M. McConico, K. Witt, F. Vogt. *Monitoring Ecological Changes of Microalgae Cells' Size and Shape based on Image Analysis.* The National Organization for the Professional Advancement of Black Chemists and Chemical Engineers, (NOBCChE), **2012**.

2. M. McConico, R. Horton, F. Vogt. *Monitoring Ecological Changes of Microalgae Cells' Size and Shape based on Image Analysis.* Board of Visitors. (BOV), **2012**.

3. M. McConico, R. Horton, F. Vogt. *Image Analysis of Ecological Changes based on the Size and Shape of Microalgae Cells.* Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Inc. (PITTCON), **2012**.

4. M. McConico, R. Horton, F. Vogt. *Assessment of Ecological Changes based on Image Analyses of Microalgae Cells.* Board of Visitors. (BOV), **2011**.

5. M. McConico, E. Duranty, R. Horton, F. Vogt. *Bayesian Classification of Microalgae FTIR Spectra as Innovative Method to Detect Chemical Changes in Ecosystems.* Federation of Analytical Chemistry and Spectroscopy Societies, (FACSS), **2010.**

6. M. McConico, E. Duranty, R. Horton, M. Robbins F. Vogt. *Spectrochemical Analyses of Microalgae as a Novel Method to Detect Chemical Changes in Marine Ecosystems.* Gordon Conference on Bioanalytical Sensors, (GRC), **2010.**

7. M. McConico, A. LaFramboise, J. Evans. *Analysis of Energy Packed Goo: An Undergraduate Physical Chemistry Experiment.* American Chemical Society Meeting, (ACS), **2008**.

Vita

Morgan Breonne McConico was born in Kansas City, KS on May 25, 1985. She was raised in Bonner Springs, KS attending elementary, middle and high school in Wyandotte County, KS. She graduated from Bonner Springs High School in 2003 and began her undergraduate career at the University of Kansas in the summer. She then transferred to Pratt Community College for a basketball scholarship. She graduated from Pratt with an Associate's Degree in Science in 2005. Morgan transferred to Lee University to continue to play basketball on scholarship. She graduated from Lee with her B.S. in Chemistry and a minor in Religion in spring of 2008. Upon graduation, she began her graduate career in Analytical Chemistry at the University of Tennessee under the direction of Dr. Frank Vogt. Morgan will graduate in May 2013 with her Ph.D. in Analytical Chemistry where she hopes to pursue a career in chemical industry.