To the Graduate Council:

I am submitting herewith a dissertation written by Rebecca B. Horton entitled “A Novel Approach to Assess Environmental Changes in Marine Ecosystems via Spectroscopic Analyses of Microalgae”. I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Frank Vogt, Major Professor

We have read this dissertation and
recommend its acceptance:

Michael Sepaniak

Zi-ling (Ben) Xue

Frank Loeffler

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Novel Approach to Assess Environmental Changes in Marine Ecosystems via Spectroscopic Analyses of Microalgae

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Rebecca Burke Horton

May 2012

Dedication

To my husband, Jonathan, and my family
who have given me the strength and encouragement
to pursue this degree. I am forever grateful for their
sacrifice and support.

Abstract

Chemical analyses for environmental monitoring encounter many challenges which are imposed by a multitude of chemically complex and interrelated processes. For such investigations, innovative analytical methodologies must be developed which characterize chemical shifts of key environmental parameters in order to deduce insights into their ecological relevance. This dissertation is driven by an analytical chemistry perspective to develop chemical sensing techniques with the ultimate goal of gaining a deeper understanding of environmental changes and their chemical origins.

In order to overcome limitations inherent to any chemical sensor designed for a specific task, new paths are pursued which are based on the idea of utilizing microalgae cells as 'biological probes'. It has been observed that microalgae cells sensitively respond to changes in their environment through changes in intracellular chemical composition. Thus, the research hypothesis of this dissertation is to utilize microalgae cells as in-situ 'measurement mediators' to study environmental changes of selected environmental parameters. To develop such analytical methodologies, two complementary research projects were performed:

The first topic focused on quantifying the change in the microalgae cells' chemical composition as a measure of shifts in ambient conditions. To accomplish this, intracellular concentrations of selected lipids, amino acids, proteins, carboxylic acids, mono- and polysaccharides were determined based on FT-IR spectroscopy. This goal required progress in sample preparation methods for solids as well as innovations in multivariate data preprocessing.

The second research topic was geared from an empirical angle and developed prediction models for relating the microalgae's infrared spectroscopic signatures to selected environmental parameters. For this purpose, three algae species were cultured under well-defined and carefully varied conditions such as ambient carbon and nitrogen concentrations. From resulting spectroscopic data sets, nonlinear models of these biological systems were derived by which the ambient growing conditions could be predicted.

Preface

The novel sample preparation methods and data pre-treatment techniques discussed in Chapters 3 and 4 are adapted from a first-author manuscript in Applied Spectroscopy. A subsequent first-author publication discussing the nonlinear modeling (Chapter 6) has been submitted to Analytica Chimica Acta for review and publication. All alterations of these materials are implemented in order to reflect the appropriate style of this dissertation.

Table of Contents

List of Figures

List of Tables

1. Introduction: Utilizing Microalgae as Novel Environmental Sensors

1.1. Objectives

In the last half of the twentieth century, the world's ecosystems have changed more swiftly than any other period in human history, transforming the structure of every ecosystem according to the Millennium Ecosystem Assessment Board [1]. These shifts in ecosystems are largely attributed to the effects of human actions, such as fertilizer run-off, industrial pollution, and increased carbon dioxide levels, among others [2]-[4]. Changes in two major environmental stressors, nitrogen and carbon, have undergone a large increase in concentration. Over the last 250 years, the atmospheric carbon dioxide levels have increased a total of 34%, with approximately 60% of the total increase occurring since 1959 [1], [5]. Similarly, the amount of biologically available nitrogen has experienced a nine fold increase over the last century [1], [5]. These two examples, along with other shifts in atmospheric composition, have impacted marine and terrestrial ecosystems resulting in changes in biodiversity and extinction [6].

Marine ecosystems account for over 70% of the planet's surface and contain 97% of the world's water supply. These ecosystems are the foundation of the food supply of higher trophic levels and play important roles in commerce [7]-[9]. Thus, it is important to investigate the consequences of these changes in environmental conditions in order to preserve their health. However, environmental changes are often intricate and interrelated causing challenges in assessing their origins and their potential impacts on the overall ecosystem. Additionally, marine ecosystems are chemically very complex and standard analytical methods are most often incomplete, evaluating a few parameters rather than an overall status of the aqueous environment. Thus, innovative methods for comprehensive chemical analyses are required.

The objective of this dissertation is to study these environmental changes via new pathways based on the idea of utilizing microalgae cells as 'biological probes'. Microalgae respond quickly to changes in their growing environments and thus may be capable of being indicators of chemical change [10], [11]. [12], [13]. These responses are reflected in their chemical

composition which must be both identified and quantified in order to relate these changes back to their growth environments.

Previous studies have shown that microalgae species not only have a characteristic IR spectrum but that small changes can be seen in the resulting spectra when exposed to different environments [10]. Quantifying these small changes in chemical composition reflecting shifts in ambient conditions requires progress in sample preparation methods as well as innovations in multivariate data preprocessing. Thus, the first portion of this dissertation focuses on developing a comprehensive sensing method utilizing new spectroscopic techniques and novel sample preparation procedures. As a testbed, microalgae cultures were grown under simulated environmental changes; two environmental stimuli, inorganic carbon and nitrogen, were chosen in order to study their impacts on the cell's chemical signature.

The second portion is approached from an empirical perspective, focusing on developing prediction models for relating the microalgae's infrared spectroscopic signatures to selected environmental parameters. It was observed that microalgae's response to changes in its environment were complex, encountering nonlinear and coupled responses between nutrient concentrations and measured absorbances. These intricate responses could not be modeled by linear multivariate least-squares regressions; thus, an expansion of these methods towards quantitative models coined '*Predictor Surfaces*' incorporate these nonlinear responses as well as cross-talk which occurs as a result of interrelated intracellular processes. These predictor surfaces can model the nonlinearities in the cell's IR spectra and predict the concentrations of dissolved nutrients in the microalgae's growth environment.

The next step toward using microalgae as bioprobes involves quantifying the intracellular concentrations of key biochemical analytes to quantify these changes in chemical composition. Therefore, concentration series of 31 selected biochemical analytes were completed and analyzed by FT-IR spectroscopy. This comprehensive spectral database will be used in future projects for building calibration models to be utilized for quantification of nutrient-induced changes in chemical composition of microalgae.

New pathways for quantitative analyses of complex systems and new tools for visualizing these processes are presented in this dissertation. Together, these complementary analytical tools combine the chemical parameters necessary to yield an overall assessment of the impact of environmental conditions on microalgae. As an introduction to this work, a short background on

microalgae's significance as environmental sensors and previous research on their adaptations to environment is presented in this chapter.

1.2. Background and Significance of Microalgae as Novel Sensors of Environmental Health

Microalgae are defined as unicellular organisms of two main divisions: eukaryotic and prokaryotic [8], [14]. Eukaryotic cells contain membrane-bound organelles (such as mitochondria, nuclei, etc.) which control cell function [15]. Prokaryotic cells do not have these organelles and are more like bacteria in biology and physiology, thus, they are often referred to as cyanobacteria [15]. The number of species classified as microalgae estimates from 40,000 to 10 million [16]. Additionally, large variations in size (1 to 300 μm), morphology, metabolism, and life cycle are common within this class [17].

Microalgae typically grow in three main environments, marine, freshwater, and terrestrial, with the most common being marine [18]. The list of requirements for growth for most microalgae species is quite simple. While the concentrations of each nutrient may vary among species for optimum growth, the main components needed to survive include nitrogen, inorganic carbon, potassium, phosphorous, light (for phototrophic algae) and sugars [15]. For example, microalgae which are autotrophic can only convert inorganic carbon dioxide into chemical energy, require a higher salinity, and are also phototrophic [15]. For these species, photosynthesis is key to survival: the chloroplast converts carbon dioxide and solar radiation into oxygen and adenosine triphosphate (ATP) into usable energy for growth and reproduction [19], [20]. Other microalgae species, referred to as heterotrophic, must obtain their organic nutrients from an external source (glucose, for example) as their primary source of energy [21].

Like all plant groups, microalgae have an important role in environmental conservation and sustainability. Over half of the world's photosynthetic activity is attributed to microalgae, recycling carbon dioxide while evolving a large amount of the world's oxygen supply [17], [22]. At least half (by dry weight) of microalgal biomass is composed of carbon which is mostly derived from carbon dioxide [23]. Therefore, microalgae are so efficient that for every 100 tons of algal biomass produced over 183 tons of carbon dioxide can be assimilated [9]. Microalgae constitute over 70% of the world's biomass and are capable of doubling their biomass on a

timeline of every two to five days [9]. For this reason, microalgae form the foundation of the food chain [8].

In addition to their role as natural recyclers, microalgae are used for many of the commercial products manufactured today. Approximately one billion dollars of revenue is generated from over five thousand tons of microalgae biomass annually for commercial products alone [24]. These products range from components in cosmetics, food, and pharmaceuticals [22]. For example, microalgae can be used to supplement important nutrients into the diets of humans and animals due to the high protein, polysaccharide and vitamin content [22], [25]. For pharmaceutical purposes, microalgae can provide high levels of antibiotics, antioxidants and certain toxins which can be used for drug formulations [26]. Perhaps the most prominent use for microalgae is replacing fossil fuels in favor of biofuel production [27]-[29].

While they are the most abundant life form in oceanic ecosystems, in order to utilize microalgae for manufacturing purposes, they must be grown on a commercial level [8]. Two of the most popular ways to grow microalgae for industrial use are open raceway ponds and closed photobioreactors (PBR) [22]. Open ponds are an easy approach to large-scale algae growth since most of their growth parameters are received via the atmosphere, e.g. sunlight, temperature, etc [30]. While open ponds are generally an easy approach, control of the environment and contamination are often challenges. Alternatively, PBRs offer a controlled environment due to the closed system for growth [31]. Different reactor configurations allow for tailoring these systems to a specific species' physiology and nutrient needs, a feature which is unavailable with some open pond designs [31]. PBRs are the more expensive option, however, they can produce an order of magnitude more biomass in the same amount of time as open pond systems [31]. Since most commercial products require some form of extraction to obtain the biochemical analyte of interest, PBRs offer the best advantage to produce large quantities of biomass in a timely manner.

In addition to using PBRs to maximize the amount of biomass grown, there are two techniques which are gaining popularity for increasing the concentrations of analytes of interest within the cells: first, it has been shown that shifts in 'normal' growth conditions can alter the chemical composition of microalgae. Secondly, changes in a microalgae cell's physiology can trigger higher concentrations of these analytes. For this reason, a whole new field of study geared toward genetic engineering of these species has evolved. It was thought that through genetic

engineering, valuable compounds which have potential for commercialization could be produced in higher yields within each cell via alterations in the chemical make-up of these organisms.

Similarly, genetic engineering has become a popular way to modify microalgae species for specific roles in the environment [32]-[33]. While this method is very beneficial, the research involved in developing genetically-modified species is often very time-consuming and costly. Sometimes a microalgae species already contains wanted compounds but in limited quantities [34]. Then, even industrial manufacturing may not make sense economically. In these circumstances, an easier approach than engineering is to alter their growth environments. It is well known that microalgae adapt very quickly to changes in their environmental surroundings [35]–[37]. For example, one of the chosen species for this project, *Dunaliella salina*, is very tolerant of extreme environmental conditions. When exposed to high temperatures, salinity, and even heavy metals, this algal species produces larger amounts of glycerol and β-carotene to maintain an osmotic balance and survive [38].

Utilizing these properties, microalgae have become a topic of interest for many applications; of these applications, perhaps the primary interest involves the search for renewable resources. As human dependence on these natural fuel sources increases, new sources of fuel based on photosynthetic systems have been studied. Via photosynthesis, microalgae take sunlight, carbon dioxide and water and convert it into chemical energy and release oxygen and biomass [15]. This biomass can be converted into biofuel via biological or themo-chemical methods [9]. The advantage to utilizing microalgae as a source for biofuel compared to other biological sources is the overall carbon neutral process: algae biofuels emit carbon dioxide upon burning, however, it can be recaptured and reprocessed by microalgae cultures, thereby reducing the amount of carbon dioxide in the atmosphere [39]. Therefore, a reduction of over 75% in carbon dioxide emissions combined with half the particulate matter byproduct of other sources makes this environmentally friendly source a very attractive replacement. Many other products can be extracted from microalgae including biodiesel, biohydrogen, biomethane, and bioethanol [22], [40]-[42]. One species chosen for this project, *Nannochloropsis oculata,* has been a major interest for these biofuel studies [34].

While microalgae have the components needed for biofuel production, there are many requirements which must be considered in order to make this process cost efficient. Microalgae species must be selected by several criteria; an appropriate choice would have a high lipid content, be very adaptable to more harsh conditions such as photobioreactors, have a low

nutrient requirement, and a fast life cycle [15]. External stimuli can be introduced for this purpose as an option to create a species which meets most of the criteria. By varying several growth parameters such as temperature, light intensity, nitrogen levels, salinity, and changing the harvesting process, higher concentrations of key biofuel products can be achieved [9], [15]. A two-stage growing process was designed to increase lipid production that involved growing the microalgae first under normal conditions then transferring them into a nitrogen-limited environment while in the early growth phase. Through this novel growth process, the lipid concentration was 2.82 times higher than under normal growing conditions.

Since microalgae cells are known to sensitively react to chemical changes in their ambient conditions, they could serve as environmental *in-situ* sensors. Thus, understanding the physiology and ecology of these species can open up new pathways and products [24]. Extracting the characteristic shifts in chemical composition can provide novel insight into the result of microalgae interactions with changing environmental conditions.

1.3. Background of Chemical Sensing on Microalgae

Several biomonitoring techniques have been developed for studying the effects of environmental stressors on microalgae. There are two dependent areas of study to consider when determining the effects of changing environmental conditions: the ambient chemical parameters (temperature, pH, salinity, nutrient status, etc.) and the shifts in a cell's composition as a result. Current techniques examine only a few chemical parameters which do not yield an overall assessment of changes in marine ecosystems and thus lack the overall ability to discern the complete effects on the biochemical make-up of these species. For example, the effects of salinity and pH as a result of changing nitrogen and carbon conditions in selected microalgae species were studied [43]. Notably, this study obtained the optimal growing conditions for these species and yielded insight into the chemical effects of the growing environment.

Other methods for studying the effects of chemical parameters involve the use of bioassays [44]. Most commonly, bioassays, which are simple techniques for propagating the effects of a chemical or virus from an infected species to an uninfected species, are utilized for field studies [14]. For example, a microalgae species was studied as a bioassay for determining the toxicity in contaminated estuaries [45]. The algae species was grown both in the estuary and the

laboratory as a control to determine if salinity, light levels, and temperature in the estuary had a significant impact on algae growth.

Similarly, novel measurement techniques have been developed for studies of very specific shifts in chemical composition. Microalgae are a potential source for biofuel production, however, a high lipid accumulation must be obtained in order to make them a viable option [15]. For investigating the triglyceride levels of test species, a microfluidic device was developed for *in-situ* measurements of lipid concentration [46]. In order to utilize microalgae for biofuel production, certain environmental conditions which are known to stimulate an increase the lipid accumulation within microalgae cells must be investigated. The appeal to growing and studying microalgae in a microfluidic device is the rapid screening potential due to the small size of the device which is advantageous when several different growth situations are being tested. Using this technique, microalgae were successfully grown on-chip and a higher concentration of lipids was achieved.

Biomarkers have been utilized to identify the physiological and sub-organismal stressors associated with exposure to changing environments [47]. These approaches combine the chemical, biological, and physical parameters necessary to yield an overall assessment; however, utilizing these methods as a continuous biomonitoring technique is difficult due to the long preparation time needed for analysis [48]. Additionally, using a single biomarker to make an assessment on overall health is nearly impossible; thus, several biomarkers must be utilized and complex statistical methods are needed to make any conclusions on overall health [48].

For some studies and applications, only a few chemical parameters or compounds are pertinent to the experiment. However, microalgae's response to changes in environment is not constant over all populations, even within the same ecosystem (or even flask) due to slight variations in the microenvironment to which each cell is exposed [49]. Therefore, for the purpose of studying overall environmental health, more comprehensive methods are desired. Optical spectroscopy has become the focus of innovative methods for studying microalgae as indicators of environmental change. The two spectroscopic methods at the forefront of environmental monitoring are Fourier Transform Infrared (FT-IR) Spectroscopy [50], [51] and Raman spectroscopy [37], [50]. These two techniques offer high selectivity and sensitivity via whole-cell analysis and imaging techniques making them ideal for these applications. Moreover, compared to previous methods of analysis, FT-IR and Raman require far less biomaterial for rapid measurements, with single cell analysis being feasible [52], [37]. Studies have also

been conducted on living cells within their microenvironments allowing for detection of ecological changes in real time [49].

For aqueous biological samples, Raman spectroscopy offers many advantages over other spectroscopic methods. Perhaps the most advantageous for this kind of sampling is the low sensitivity to water [53]. Additionally, concentrations as low as 10^{-8} M can be detected to facilitate the low concentrations of biochemical analytes contained in microalgae cells [54]. Raman spectroscopy has often been used as a complementary technique to infrared spectroscopy for determining the chemical composition of biomaterials [54]. For example, the effects of nitrogen starvation on the carotenoid and chlorophyll *a* concentrations of microalgae species *Dunaliella tertiolecta* have been studied using Raman spectroscopy [49]. These pigments can be used as indicators of shifts in environmental conditions and also signal a shift in the triglyceride content of a cell with applications toward biofuels [54]. Additionally, composition mapping of the triglyceride content within a single cell can be accomplished with Raman [54]. With composition mapping, spectra obtained from a large number of sequential locations are studied over a small area [55]. When these spectra are filtered for a particular wavelength of interest, or characteristic peak, the location of the analyte of interest can be determined via the high intensity areas on the spectral 'map'. This technique is particularly beneficial for determining the triglyceride, β-caretenoid, and chlorophyll concentrations in microalgae as a result of environmental stimulation [54].

Some metabolic processes, such as lipid metabolism, have been studied using Coherent anti-Stokes Raman Scattering (CARS) which is based on resonant scattering [54]. This method can further be utilized for lipid quantification and does not require any labeling and minimal sample preparation. These studies have shown promise as an *in-situ* method for studying the environmental effects on microalgae. Several studies have also been conducted on *in vivo* sampling of microalgae cells using Raman. One study in particular combines an acoustic levitation technique with a portable Raman spectrometer for single cell analyses. Utilizing an ultrasonic levitator, the single cell can be analyzed without any interference from a substrate [50].

A central consideration when performing Raman spectroscopy is the level of laser power utilized for measurements. Many of the chemical components in microalgae are subject to degradation by light, causing shifts in the chemical composition from flask to measurement [56]. Furthermore, the heat the laser produces can also cause discrepancies in chemical composition

from sample-to-sample and can also rupture the cells producing inconsistent results [54]. However, if the laser power is too low, a sufficient signal-to-noise ratio cannot be achieved [57]. Finally, the algae's pigmentation can cause interfering fluorescence when studying other macromolecular classes of compounds but can be avoided by choosing the correct laser wavelength [58]. Since chlorophylls *a* and *b* absorb in the regions of 400-800 nm, robust data processing techniques are often required [58].

The other spectroscopic method which has been utilized for studies on microalgae cells is FT-IR. It has been argued that FT-IR has more potential for studying a broader range of macromolecular compounds than Raman [50]. Additionally, FT-IR has been applied to measure the relative concentrations of many analyte classes, such as carbohydrates and proteins, with as much success as non-spectroscopic techniques without requiring large preparation time and extraction of each class [37]. While water is strongly absorbing in the mid-infrared region, FT-IR is still a viable option for these studies through dehydration and minimal data pre-processing [11]. Unlike Raman, variations in pigmentation concentrations (i.e. carotenoids and chlorophyll *a* and *b*) can be studied along with fluctuations in other biochemical classes simultaneously without the interference of fluorescence [50].

FT-IR (Fourier Transform Infrared) spectroscopy acquires information about the molecular structure and the chemical bonds of a given analyte [59]. If the analyte is infrared active, the chemical bonds vibrate at specific frequencies of matching wavelength. When the infrared radiation frequency matches the vibrational modes and the dipole moment of the analyte undergoes a change, the infrared radiation is absorbed. The detector records the transmitted radiation at consecutive wavenumbers which results in a spectrum from which functional groups and analyte concentrations are determined. A Fourier Transform (FT) is used to determine the frequencies and the amplitudes of the measured spectrum. Once the single beam spectrum is given, the information is plotted for each wavenumber against the respective absorbance or transmittance yielding a comprehensive spectroscopic signature for the given analyte.

Initial studies utilizing FT-IR focused on identifying the spectroscopic signatures of multiple microalgae species for discrimination. Since each microalgae species, whether green algae, diatom, or cyanobacteria, have a distinct chemical composition, even within the same genus, FT-IR was able to discriminate one species from another [60]. Furthermore, spectral band in the fingerprint regions can be assigned to major functional groups for studying the effects of shifts in their growth environments [61]. While spectra in the mid-infrared region are very

complex due to the contributions of all the IR-active cellular components, each molecular class does have a distinct absorbance in the fingerprint region (1750-900 cm^{-1}) [62]. Band assignments for these groups of interest are as follows: lipids and fatty acids can be assigned at 1740 cm^{-1}, the amide I band contribution of proteins is assigned at 1650 cm^{-1}, the amide II band manifests at 1540 cm^{-1}, the main protein band occurs at 1455 cm^{-1}, the amide III band at 1320 cm^{-1}, the phosphodiester foundation of nucleic acids absorbs at 1240 cm^{-1}, the stretch from 1200-900 cm^{-1} can be generically assigned to polysaccharides (monosaccharides ≈ 1000 cm^{-1}), and finally microalgae species which have cell walls composed of silica (frustules) absorb at 1080 cm^{-1} [50].

Infrared microspectroscopy has become a very popular option since it provides single-cell analysis with high spatial resolution making this technique advantageous to studying the response of biological samples to external stimuli [11], [63]. This high spatial resolution also allows for spectroscopic determination of subcellular localizations of each major biochemical class within a single cell. The distribution of these chemicals and the way in which the cells store them can change when exposed to different nutrient conditions.

The success of FT-IR can be heightened by using a different radiation source, such as a synchrotron light source [12]. With larger microalgae species especially, a more powerful radiation source is needed to transmit through the sample in order to obtain a high signal-to-noise ratio [50]. Studies which have utilized synchrotron radiation source often analyze living cells via flow-through cells which keep the microalgae in the medium throughout the entire analysis [12]. When high intensity radiation sources are not available, reflection/absorption utilizing attenuated total reflectance (ATR) spectroscopy is a good alternative to traditional transmission analyses [50].

Although different species have distinct infrared features, the addition of FT-IR imaging techniques can further enhance classification. Microspectroscopic images of microalgae cells acquired with a focal plane array multi-element detector have been used to successfully discriminate between species [64]. With high spatial resolution and a high signal-to-noise ratio, species can be discriminated via the morphology and allocation of functional groups within the cells.

Due to the many benefits of FT-IR, it has been selected as the technique-of-choice in this dissertation for studying the impacts of a cells' nutritional condition on their mid-IR spectra. In these studies, analytical methods are presented which establish advances in FT-IR

experimental design combined with innovative chemometric methods for developing microalgae bioprobes as *in-situ* sensors of the status of marine ecosystems.

1.4. Dissertation Overview

This dissertation is motivated by the general goal of developing new analytical sensing techniques for gaining a deeper understanding of environmental changes within marine ecosystems and their chemical origins. For this purpose, new paths to study these changes are pursued which are based on the idea of utilizing microalgae cells as 'biological probes' which undergo measureable and reproducible chemical adaptations as a function of environmental parameters.

Microalgae have an important role in environmental conservation and sustainability. Constituting over half of the world's biomass, microalgae naturally form the foundation of the food chain [8], [9] and evolve a large amount of the world's oxygen supply [17], [22]. It is well known that microalgae adapt very quickly to changes in their environmental surroundings [35]–[37]. However, these changes in growth conditions can have profound consequences for the entire ecosystem. Thus, the significance of utilizing microalgae as bioprobes of their environment is to facilitate new understanding into the physiology and ecology of these species and open up new pathways for studying the impacts of environment [24]. Furthermore, extracting the characteristic shifts in chemical composition as a result of environmental change can provide novel insight into the result of microalgae interactions with changing environmental conditions.

Determining the effects, whether positive or negative, of these changes is often very difficult due to the chemical complexity of these microorganisms. Additionally, microalgae's response to changes in environment is not constant over all populations, even within the same ecosystem (or even flask) due to slight variations in the microenvironment to which each cell is exposed [49]. Therefore, for the purpose of studying overall environmental conditions, more comprehensive methods are desired.

For these reasons, the chosen technique for studying microalgae's adaptations to environment is Fourier Transform Infrared (FT-IR) Spectroscopy [50], [51]. It has been proven that FT-IR

transmission spectroscopy is a promising technique for studying microalgae since most relevant analytes are IR active. Additionally, this technique offers high selectivity and sensitivity via whole-cell analysis and imaging techniques and yields a comprehensive spectrum making FT-IR ideal for these applications.

Thus, in order to study the feasibility of using microalgae as bioprobes, three species, *Dunaliella salina, Dunaliella parva, Nannochloropsis oculata*, were grown under different nutrient conditions to simulate variations in ambient conditions. Two main nutrients, inorganic carbon and nitrogen sources, were provided in the ESAW at different concentrations to simulate several starving, normal, and excess situations. To enable investigations of changes in chemical composition of microalgae cells in response to shifting environmental conditions, microalgae cells subjected to these conditions were extracted from their growth environments. For a deeper understanding of the changes in their chemical composition, these microalgae cells were fixed and gently dried for spectroscopic analysis. A sample preparation method was developed the algae samples were mixed with KBr powder and pressed into pellets for analysis. FT-IR transmission spectra were recorded for a minimum of 5 replicate flasks of each nutrient condition and compiled for later data analysis.

From these spectra, novel analytical methods were developed to relate the changes in IR spectra to the growing conditions in which they were exposed. Thus, novel sample preparation and chemometric methods were required to model these changes and extract the relevant chemical information. For this purpose, two complementary research projects were performed: the first project was to investigate the changes in chemical composition towards quantification of key biochemical components that are commonly found within microalgae cells. To accomplish this, concentration series of 31 biochemically relevant compounds were analyzed and placed into a spectroscopic calibration database for further investigations into the changes microalgae undergo as a result of their environments. This goal required progress in sample preparation methods for solids as well as innovations in multivariate data preprocessing.

The second project aimed to empirically relate the IR spectrum to ambient parameters, such as nutrient concentration. Upon closer investigation, it was determined that microalgae cells react sensitively but nonlinearly to chemical changes in their ambient conditions; it is also known that microalgae need certain nutrient mixes to thrive. However, restrictions of conventional, linear multivariate models originate from the assumption that the measured data are a simple, proportional superposition of several input parameters. For quantitative and qualitative analyses

of coupled, nonlinear multivariate systems, such as microalgae's response to nutrient availability, this requirement is not given and hence '*Predictor Surfaces*' were developed. Predictor Surfaces are based on approximating a nonlinear multivariate model which facilitates quantitative analyses of nonlinear systems as well as creating visual inspections of complex chemical systems. It was shown that by means of Predictor Surfaces, environmental parameters could be quantitatively predicted based on the microalgae cell's IR spectra. Such models not only facilitate improved accuracy and precision but also enable in-depth studies of complex chemical systems. This application of Predictor Surfaces demonstrates a new approach for quantitative analyses of systems too complex for current, linear chemometric methods.

The ultimate goal of this project is to monitor *in-situ* and in real-time the impacts of changing environmental conditions on microalgae species. Since microalgae are at the bottom of the tropic level, this information could yield an overall picture of the effects of environment on the marine ecosystem as a whole. Therefore, the following chapters explain the novel approaches taken to study microalgae's response to environmental conditions through novel sample preparation techniques and chemometric methodologies. These steps serve as proof-of-principle for utilizing microalgae as bioprobes of environmental conditions. The results obtained from the work included in this dissertation demonstrate the feasibility of this project for assessing the impacts of environment on marine ecosystems.

2. Culturing Microalgae Species and Preparation for Spectroscopic Analyses

2.1. Introduction

The ultimate goal of this dissertation is to develop chemical sensing techniques for the purpose of gaining a deeper understanding of environmental changes. To this end, microalgae were investigated as novel, *in-situ* sensors since they reactive so sensitively to changing environments. In order to use them in such a role, microalgae must first be cultured under well-defined laboratory conditions. Using established culturing methods, three microalgae species were selected and cultured for this study.

To examine the effects of environment on microalgae cell's chemical composition, different environmental parameters had to be simulated under well-defined conditions. For this purpose, two known environmental stressors, carbon and nitrogen, were introduced to study the effects of nutrient source and concentration. In order to quantify these nutrient-induced changes in chemical composition, FT-IR spectroscopy was chosen since many biochemical components within microalgae cells feature distinct mid-infrared spectroscopic signatures. An established method for extracting the microalgae cells from their growth environments were used in order to prepare the microalgae for spectroscopic analysis.

2.2. Microalgae Culturing Techniques

Stock solutions of algae cultures *Dunaliella salina, Dunaliella parva, Nannochloropsis oculata* were obtained from the UTEX laboratory at the University of Texas, Austin. The following steps were implemented upon receiving starting cultures to ensure consistency and prevent contamination. First, all glassware, utensils, and media used for culturing were properly cleaned with Contrex AP powdered labware detergent (Fisher Scientific), rinsed thoroughly with deionized water, and allowed to dry completely. The second step of sterilization involves the

use of an autoclave (Getinge Vacuum/Gravity Steam Sterilizer, 733LS). All items for autoclaving were wrapped in aluminum foil and tagged with autoclave tape to ensure they reached the proper sterilization temperature. The autoclave cycle exposed all items to 121°C at 2.0 atm steam pressure for 20 minutes [14].

Inoculation of algae cultures was performed inside an AirScience Purair VLF laminar flow cabinet/PCR (polymerase chain reaction) cabinet as a final step toward preventing any contaminations. The laminar hood suppresses airborne particulate matter and provides an ultraviolet sterilization of all surfaces and items to be in contact with the microalgae. Additionally, an ultra-low penetration air (ULPA) filter and a down-flow air circulator greatly reduced the number of contaminants entering the flow hood. Prior to use, all surfaces were wiped down with a 70% ethanol solution. Gloves and clean sleeves were worn at all times while working in the hood.

Microalgae cultures were provided by The Culture Collection at the University of Texas, Austin (UTEX). Three species were selected for study: *Dunaliella parva* (#LB1983)*, Nannochloropsis oculata* (#LB2164)*, Dunaliella salina* (#LB200) [65]-[66]. Approximately 15 mL aliquots of these cultures were received suspended in Erdschreiber's [67] medium. Upon arrival, the screw-cap tubes were slightly unscrewed for two hours prior to inoculating to allow for air exchange without contaminating the cultures. Although the algae cultures arrived in Erdschreiber's medium, they were transferred into a similar medium, Enriched Artificial Seawater (ESAW) [68]-[69] which will subsequently enable the variation of nitrogen and carbon nutrient sources and concentrations.

The ESAW medium was made according to specifications as given in reference [14]. The major nutrients in the medium were prepared individually as stock solutions. A large portion of the medium was made in two separate flasks: salt solution I and II. Salt solution I was prepared in 600 mL of deionized water with the reagents and masses listed in

Table 1. Similarly, salt solution II was prepared in 300 mL of deionized water as listed in Table 2. After separate preparation, salt solutions I and II were combined in a one liter PYREX bottle. In addition to these solutions, the major nutrients and vitamins were prepared as separate stock solutions (Table 3).

Table 1. Reagents used to make Salt Solution I for ESAW medium [14].

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 [14].

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 [14].

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

Small amounts of the nutrients were then added: one mL nitrate, one mL of phosphate, two mL of silicate, and one mL of each metals stock. In order to ensure the proper pH of the medium, approximately 8.4, and to prevent any precipitation after autoclaving [69], ten mL of tris hydrochloride solution (1M, Fisher Scientific) was added. The medium was brought to a final volume of 999 mL by adding 83 mL of deionized water. At this point, the medium can be autoclaved for sterilization purposes.

The medium must cool for 24 hours after autoclaving prior to adding the vitamin stock. This stock cannot be added prior to autoclaving since the high temperature will cause the reagents to precipitate out of solution [68]. Therefore, the vitamin stock is filtered separately using a mixed cellulose filter (Fisher Scientific) with a pore size of 0.22 micrometers into a clean, sterile flask inside the sterilized laminar flow hood; after filtration, one mL of vitamin stock was added to the sterile medium. Upon completion of the medium, each batch was tested to ensure a pH of approximately 8.4. The filtered stock solution was then stored inside the laminar flow hood and was only utilized once the cabinet was sterilized.

Since the growth medium was different from the original algae stock, the cultures had to be transferred into their new medium. Prior to extraction, the screw-cap was removed inside the flow hood and the top of the vial was flamed with a Bunsen burner to ensure all possible contaminants were eliminated. The cultures were extracted from the vials and poured into a clean, sterile beaker, which was flamed, for further use. The reason for the transfer is to preserve the culture source in case the cultures do not grow in their new environments and need to be inoculated again. The received cultures can be kept for a few weeks after they are received if the screw-cap is slightly loosened and stored in a sufficient growth environment.

The cultures placed in the clean, sterile beaker were used to begin all new cultures in the ESAW medium. Starting cultures were placed in a clean, sterile 250 mL Erlenmeyer flask. Each Erlenmeyer was flamed, filled with approximately 100 mL of prepared ESAW, and flamed again. Using a Finnpipette adjustable volume pipette (Fisher Scientific) and a sterile tip, exactly two mL of algae cultures were placed in the Erlenmeyer flask, flamed again, and plugged with cotton to allow for sterile air exchange (Figure 1, left).

Figure 1. (left) Newly inoculated cultures with cotton plugs. (right) Cultures in exponential growth phase.

The new cultures were placed in an environmental growth chamber (Precision 818, Thermo Scientific) at 20°C with continuous illumination until ready for harvest, Figure 1 (right).

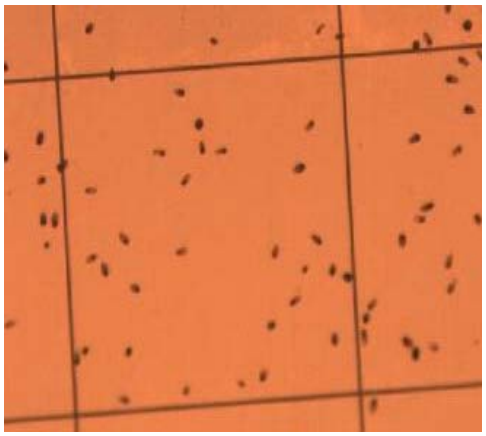
For each culture, it is important to obtain growth curves for each species under a given set of conditions [14]; these growth curves were utilized for determining when the cultures were in exponential growth phase for harvest [70]. In order to obtain growth curves, microalgae species were counted every 24 hours using a hemocytometer (Cole-Palmer) with a 0.1 mm chamber depth. The cultures were placed inside the flow hood and approximately 5 mL of culture was poured into a sterile, flamed beaker for further analysis. Since some of the microalgae cultures have the tendency to settle to the bottom of the flask, each flask was swirled sufficiently prior to obtaining an aliquot of the culture in order to ensure a homogeneous distribution of the algae for counting [71]. Using a Finnpipette and a sterile tip, 200 µL of algae culture was transferred into a small, sterile glass tube. To preserve the microalgae species, 20 µL of Lugol's mixture (Sigma-Aldrich) was added to the cultures; this prevented the microalgae species from moving once inside the hemocytometer in order to obtain an accurate count [72].

Using a 200 µL Finnpipette, one drop of the preserved algae was placed in the hemocytometer cavity and was drawn into the chamber via capillary action. As can be seen in Figure 2, each microalgae species has a different shape and size thus the hemocytometer has different counting areas within the grid. The microalgae cells were counted using a 10x microscope objective. A total of ten counts from ten independent squares were taken and averaged to obtain a mean value of counts for a given species on a given day.

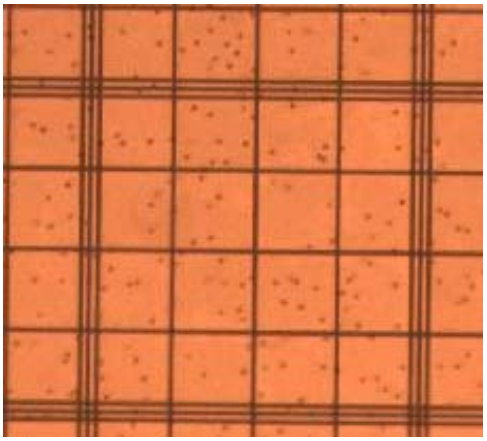
The volume calculations (number of cells/mL of solution) were dependent on the size of the grid (0.25 nL for small squares vs. 6.25 nL for large squares) [72]. In order to extrapolate the average count to the total algae population in the flask, the following equation was used:

$$\left(\frac{\text{Average Number of Cells}}{\text{Square Volume } (nL)}\right) \times \left(\frac{10^9 nL}{1\ L}\right) \times \left(\frac{1\ L}{1000\ mL}\right) \quad (1)$$

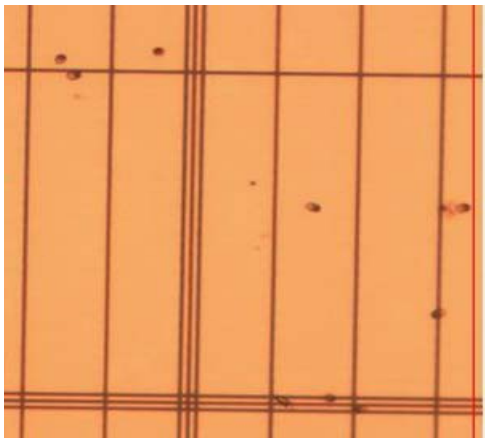
Cell counting continued until the culture was no longer in exponential growth phase.



Dunaliella parva



Nannochloropsis oculata



Dunaliella salina

Figure 2. Images of three microalgae species, *Dunaliella parva, Nannochloropsis oculata, Dunaliella salina*, in the hemocytometer chamber using a 10x objective.

Figure 3 displays the growth curves for each species over two conditions: sodium nitrate (0.549 mM) and ammonium chloride (0.873 mM). Each growth curve was averaged over three flasks of the same species and condition. These growth curves were utilized throughout the entire project in order to successfully harvest the cells in their exponential phase.

2.3. Inducing Changes in Microalgae Environments

Two major contributors to the changes in environment experienced by these ecosystems are nitrogen and carbon [2], [13]-[73]. The uptake of these nutrients can occur from many different sources of nitrogen- and carbon-containing compounds, however, optimal growth conditions for each species seem to be compound-specific, e.g. some species prefer urea as a nitrogen source over sodium nitrate [74]. For example, urea is a poor nitrogen source for the microalgae *Ellipsoidion* sp. which favors nitrate and ammonium containing compounds [75]. Inappropriate levels of these nutrients, often accelerated by human activities, are potentially toxic to some microalgae and may cause changes or losses in the biodiversity of these ecosystems [76]-[77]. Limiting nitrogen-containing nutrients in microalgae cultures can result in an increase of lipid and carbohydrate storage but a reduction in overall protein structures [62].

Similarly, extreme concentrations of carbon-containing nutrients, whether atmospheric or dissolved, have major impacts on the health of microalgae, e.g. changes in morphology, limited mobility, sharp increase in mortality [46], [78]. If the concentration of carbon dioxide in seawater is too low, certain microalgae species will experience major reductions in photosynthesis [79].

Increased concentrations of carbon-containing nutrients in microalgae cultures can invoke increased production of transparent exopolymer particles (TEPs) which increase cell aggregation and leads to clusters of cells sinking to the bottoms of the ecosystem—a consequence which may disrupt the food cycle of higher trophic levels [80].

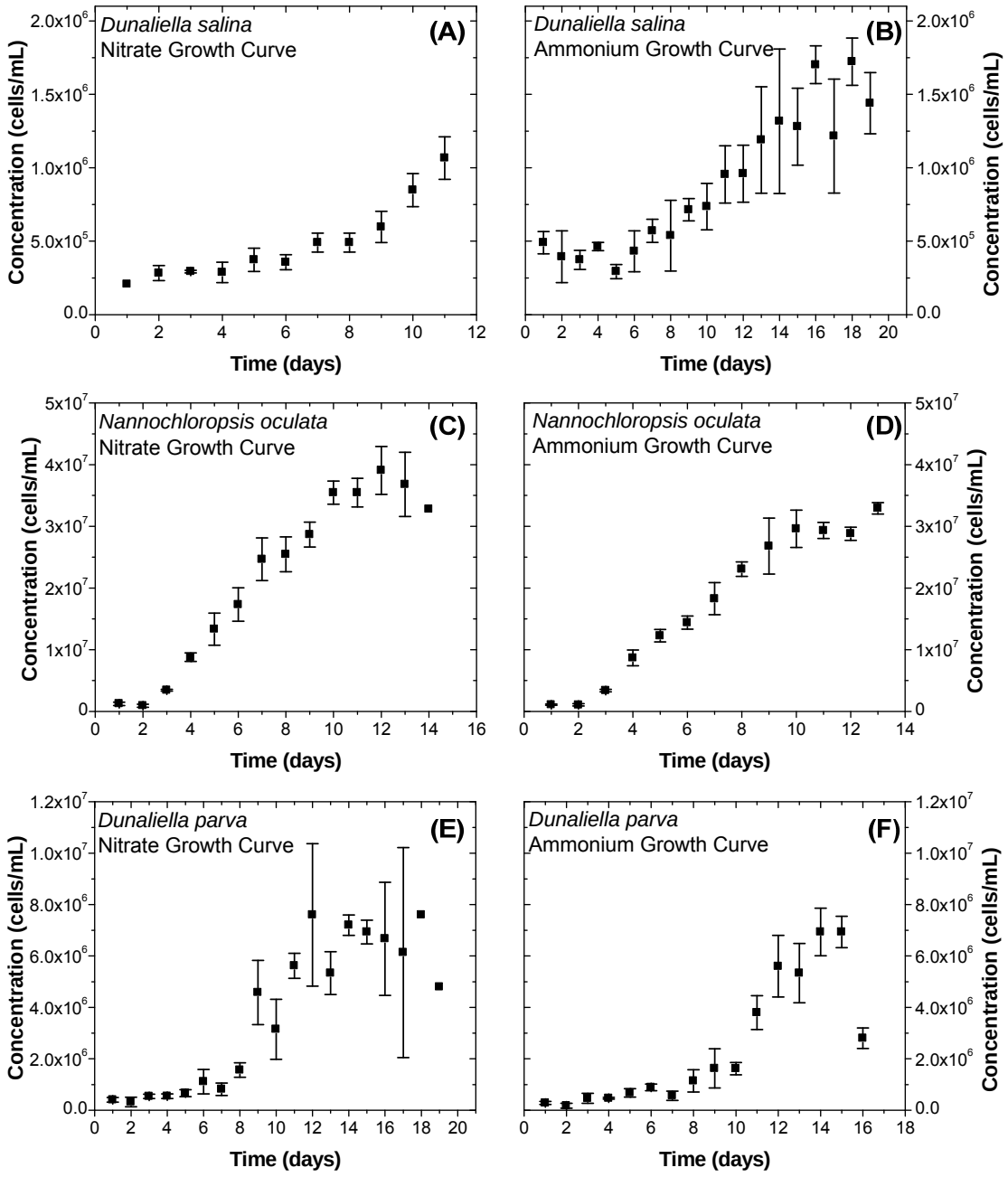


Figure 3. Growth curves for three microalgae species, *Dunaliella salina* (A, B)*, Nannochloropsis oculata* (C, D)*, Dunaliella parva* (E, F), under different nitrogen nutrient sources: (left) Nitrate source (sodium nitrate, 0.549 mM), (right) Ammonium source (ammonium chloride, 0.873 mM).

Several studies have proven that microalgae respond quickly to changes in their growing conditions which are reflected in their chemical composition. Although some effects of changing environmental conditions on marine ecosystems have been studied, there exists a need to understand more holistically the changes that occur to the chemical composition. Therefore, two main nutrient sources, nitrogen and carbon, were altered in order to better understand the effects of these conditions. In order to study the impacts of the nitrogen source, two different sources were utilized: sodium nitrate (nitrogen source in ESAW, 0.549 mM) and ammonium chloride. It was not only important to determine the effect of different sources of nitrogen but also different concentrations. Therefore, an experiment was set-up such that varying concentrations of each nitrogen source were studied. The following concentrations were prepared covering situations from limiting to excess of nutrients: 0.16 mM, 0.35 mM, 0.549 mM, 0.873 mM, 1.28 mM, 1.47 mM, 1.65 mM.

In a similar fashion, the effects of the carbon nutrient source were studied. There are two primary sources of carbon uptake for microalgae cells: sodium bicarbonate ($NaHCO_3$) and carbon dioxide (CO_2) [79]. It is well known that each species has a specific pathway, and thus requires a particular nutrient source, for carbon uptake [81]-[82]. Most microalgae species prefer dissolved inorganic carbon (DIC), however, some species use a CO_2 concentrating mechanism (CCM) [74]. Not all species are capable of assimilating carbon using a CCM; for this reason, only sodium bicarbonate was utilized. Therefore, the following concentrations of sodium bicarbonate were studied: 0.16 mM, 1.11 mM, 3.63 mM, 5.18 mM, 6.72 mM, 8.26 mM, and the normal concentration of 2.07 mM in ESAW.

2.4. Microalgae Extraction and Preparation of Cells for Spectroscopic Analyses

Before any extraction or analysis can begin, the cell concentration of each culture must be determined to ensure it is in exponential growth phase as determined by the previous growth curves. If the cultures were not in the proper growth phase, the cell concentration was checked again every 24 hours until the proper population was obtained and were ready to be harvested for spectroscopic measurements (Figure 4, A). Prior to fixation, each culture must first be re-inoculated into a new culture in order to continue the culture for growth under new conditions.

For this purpose, a small amount of algae culture was transferred into a sterile, flamed flask. Exactly two mL of the cultures in exponential phase were pipetted into a sterile, flamed 250 mL Erlenmeyer flask filled with approximately 100 mL of modified ESAW medium. The new cultures were plugged with cotton to allow for air exchange while preventing contamination and placed inside the growth chamber for further experiments.

The cultures which were used to re-inoculate the new cultures remained in the flow hood to begin the harvesting and fixation process. Approximately 40 mL of algae from each flask was transferred into a sterile, flamed flask. Borosilicate glass centrifuge tubes (Fisher Scientific) with an outer diameter of 16 mL and a height of 100 mm were used to harvest the algae cultures. Exactly ten mL of microalgae was pipetted into the centrifuge tubes with ten µL of Lugol's solution added for fixation. The centrifuge tubes were placed in an Eppendorf 5702 centrifuge and spun for five minutes on 4400 rotations per minute. The supernatant was discarded leaving the algae pelleted in the bottom of the tubes. This procedure was repeated a minimum of three times, pouring the new batch of algae and Lugol's mixture on top of the pelleted algae, in order to have enough material for spectroscopic analysis. When a large enough pellet of algae was obtained, two mL of 0.5 M ammonium formate (Alfa Aesar) was used to wash the algae pellet. The washing step removes any residual salts on the remaining biomass [83]; while most of these salts are infrared transparent, these salt crystals can cause scattering of the IR light causing interference and baseline effects in the spectroscopic signatures of microalgae species [50]. Two separate washings with ammonium formate were performed. After the second washing, the supernatant was again discarded and the tubes containing the pelleted algae were covered with foil. Prior to analysis, the tubes containing the algae were placed in an oven and dried gently at 60°C for 2-3 days (Figure 4, B). The remaining cultures left in the flask after harvest were analyzed to determine the pH of the cultures. After all steps were performed, the remaining algae were properly discarded.

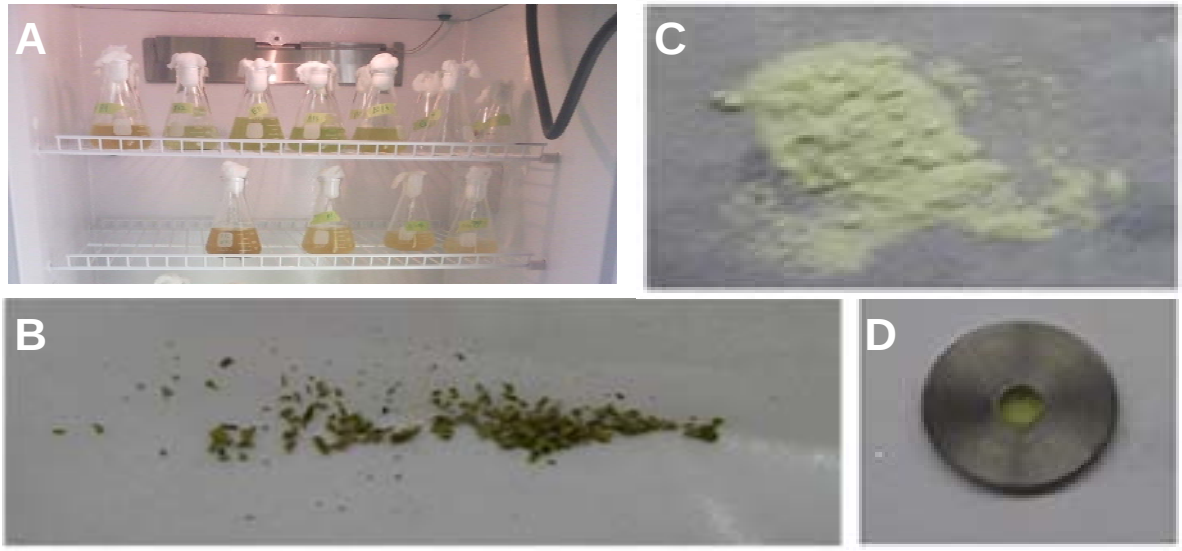


Figure 4. (A) Three microalgae species were cultured using different nutrient sources and concentrations to mimic environmental changes in a controlled growth chamber; (B) algae cultures were fixed using Lugol's solution and dried; (C) samples were prepared in concentrations of 0.6 weight percent in KBr powder for concentration definition; (D) washers with inner-diameters of 3 mm were filled with the mixture and pressed.

2.5. Conclusions

Culturing microalgae cells within the lab for spectroscopic analyses required several steps. Three microalgae species were chosen, *Dunaliella salina*, *Nannochloropsis oculata*, *Dunaliella parva*, since studies have shown they react quickly to changes in their growing conditions [65], [84]. Due to this fact, special care was taken to ensure as much consistency in their growing environments as possible. After establishing a proper procedure for growing the cultures under well-defined conditions, two known environmental stressors, carbon and nitrogen, were studied for their effects on the chemical composition of microalgae cells. Varying concentrations of each stressor was added to the growth medium and a different nitrogen source was also studied. Once harvested, the chemical composition of the algae cells was fixed and dehydration of the cells was performed in order to study the samples spectroscopically.

3. FT-IR Analysis of Algae Samples

3.1. Introduction

In order to determine the effects of changing environments, an analytical method had to be developed based on measuring microalgae's chemical composition as a novel chemical probe for detecting changes in ecosystems. Several different methods and instrumentation have previously been utilized for studying microalgae's adaptation to changing environments [43], [47], [48]. However, these techniques often focus on quantifying one chemical parameter rather than conducting a more comprehensive examination of the overall chemical composition of microalgae cells. Therefore, a method for analyzing these adaptations to their environment using a more holistic approach was desired. Thus, for this project, FT-IR was the chosen method of analysis since many analytes contained in microalgae absorb in the fingerprint region. FT-IR also offers short measurement times, versatility and multi-component capabilities which are required for analyzing complex biological samples like microalgae.

Since the microalgae samples had to be extracted from the medium prior to analysis, a new sample preparation method had to be developed based on quantifying solid analytes using their IR signatures. This novel method was based on well-defined concentrations of microalgae cells in infrared-transparent KBr. Using this preparation technique, spectra of all three species were acquired for various nutrient sources and conditions in order to study the impacts of environmental conditions.

The ultimate goal was to be able to relate changes in the microalgae cell's IR signatures back to their growing environments. However, some spectral artifacts due to the solid nature of these samples were observed which prevented replicate samples from being reproducible. In order to correct for these fluctuations, novel chemometric methods for baseline removal and pathlength correction were developed and applied.

3.2. Sample Preparation

In order to conduct quantitative investigations of solid samples using FT-IR, several innovations in sample preparation were required. Since the microalgae cells were dehydrated prior to analysis (section 2.4), the following preparation steps were performed to analyze these samples spectroscopically. First, dried algae samples were ground individually into fine powders and mixed with potassium bromide (KBr) for concentration definition [85] (Figure 4, C, section 2.4). The hygroscopic KBr was not ground in order to prevent increasing the surface area exposed to ambient moisture whose features would impact the FT-IR spectra. For the same reason, KBr was stored in a desiccator under a low vacuum and the algae samples remained in the oven until prepared for analysis. For all species and conditions, concentrations of 0.6 weight percent (wt %) algae in KBr were prepared (Figure 4, C, section 2.4). These resulting KBr-algae mixtures were then filled into the center hole of a washer and pressed into IR-transparent pellets (~1.5 mm thickness, 3 mm diameter) where they remained for measurements (Figure 4, D, section 2.4). At this concentration, absorbance spectra of the pellets showed a sufficient signal-to-noise ratio (SNR) and the detector was far from being saturated. For acquiring background spectra, pure KBr pellets were utilized.

FT-IR transmission spectra from KBr pellets were acquired with a Bruker Vertex 70 FT-IR spectrometer which was equipped with a deuterated triglycine sulfate (DTGS) detector and controlled by OPUS 6.5 software package. Sixty-four scans covering 7500 – 370 cm^{-1} at 4 cm^{-1} resolution were co-added but only the information-containing range 3500 – 950 cm^{-1} was used for data analyses. From this wavenumber region, the 2700 – 1850 cm^{-1} range was excluded because it did not contain necessary information and also prevented fluctuations in the atmospheric CO_2 concentrations from introducing irreproducible absorption features.

Some algae samples showed a tendency to clump and thus are a potential source of inhomogeneity in the pellets. In order to increase spectroscopic reproducibility, the entire pellet diameter was illuminated rather than a small spot and thus remaining sample inhomogeneities were reduced by spatial averaging over the entire pellet. For this purpose, the spectrometer's aperture had to be selected properly such that the IR beam illuminated the entire pellet and at the same time no light intensity was lost due to reflections of the pellet holder (= washer). For this instrument, the 5mm aperture setting was determined to be the best choice. This aperture setting was determined by stepwise decreasing it and recording the detector signal. As long as

the washer limited the light throughput, decreasing the instrument aperture did not result in decreasing signal levels; for too small instrument apertures, the washer was not the limiting factor anymore and the signal level started to decrease because the IR beam no longer illuminated the entire pellet. This combination of both apertures ensured the aforementioned goal of spatially integrating over the entire pellet and thus averaging out as much sample inhomogeneity as possible.

Despite efforts to be as precise as possible in sample preparation, the solid nature of analytes as well as the limited reproducibility of biomaterials in general required some additional steps in order to ensure a well-defined chemical composition. Additionally, it is also demonstrated here that experimental methods alone are often insufficient to handle cases of low data reproducibility.

3.3. Enhancing the Spectral Quality of Algae Spectra

Each pellet contained approximately 5 mg of powder mixture. For practical purposes, however, mixtures much larger than 5 mg were prepared and only a small portion of this mixture was used for making pellets. To achieve the aforementioned weight percentage (0.6 wt %), dispersing a small amount, ≤ 1 milligram, of sample into ≈100 mg of KBr held the risk of the algae sample not being homogeneously distributed throughout the KBr. Furthermore, placing exactly 5 mg of mixture into the washer was difficult to reproduce manually; 5±1 mg turned out to be realistic. However, varying amounts of powder mixture filled into the washer of a fixed inner diameter directly translated into fluctuating absorption pathlengths, which in return caused washer-to-washer fluctuations in measured absorbance.

Figure 5 demonstrates the effects of different pathlengths on the absorbance values of a mixture of pure glycine and KBr at a concentration of 0.75 weight percent.

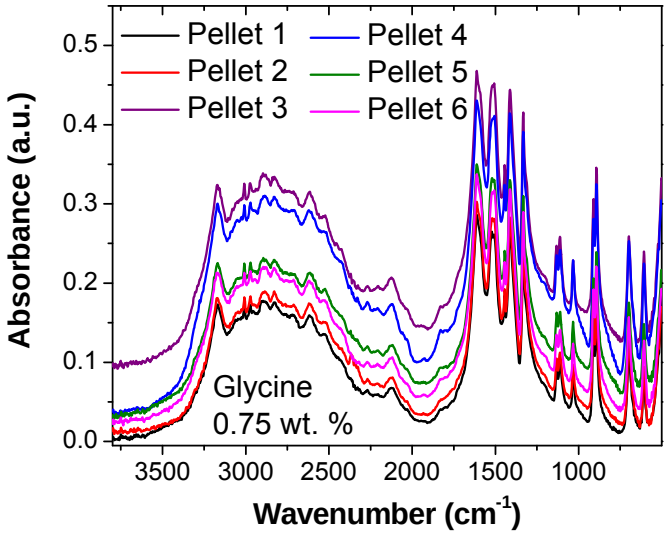


Figure 5. Effects of solid sample preparation on six replicate pellets from the same concentration mixture of Glycine at a concentration of 0.75 wt. %.

Using a pure analyte excludes spectral variation unlike the spectra obtained from microalgae which can have slight fluctuations in algae composition for replicate samples. A further reduction of spectral reproducibility was caused by baseline shifts which were partially assigned to light scattering inside the samples.

Thus, in order to enhance spectral quality, three experimental and chemometric procedures were developed: (i) To ensure a maximum, experimentally-feasible mixture homogeneity, the solid analyte plus KBr powders were shaken for two minutes in a mechanical mixer (WIG-L-BUG, Rinn Crescent). (ii) Pathlength variations (= fluctuations in pellet thickness) were corrected by normalizing the spectra for the pellet weights. (iii) Improvements in baseline shifts were induced by means of a novel baseline correction algorithm which does not require any *a priori* information.

Fluctuations in absorbance pathlength caused by imprecise pellet masses were corrected with the following pre-processing step: The washer used as pellet holder was weighed empty and again after pellet preparation as described above. Since a higher pellet mass is directly correlated with a longer pathlength, and thus a higher absorbance, each spectrum was divided by the corresponding pellet's weight in order to normalize all pellets to the same pathlength. Such a mass normalization was chosen over internal standards because of the complexity of analyte spectra; introducing an internal standard would most likely have overlapping signals with those of the analytes. Hence, the area under an internal standard's absorbance band would be difficult to determine accurately and therefore a weight difference is more simple, robust and less prone to errors.

The second chemometric pre-processing step developed in this study corrects for spectroscopic baseline drifts which are commonly encountered in biological materials (Figure 6, top) and are a source of considerable concentration errors [86]. Baseline shifts in optical spectra are common and often caused by temperature fluctuations, light scattering, alterations of optical components (e.g. aging light source), changing optical properties of the sample (e.g. refractive index) and other causes [87]. Frequent re-calibrations could reduce this problem but would involve time consuming sample preparations and/or would interrupt online monitoring applications.

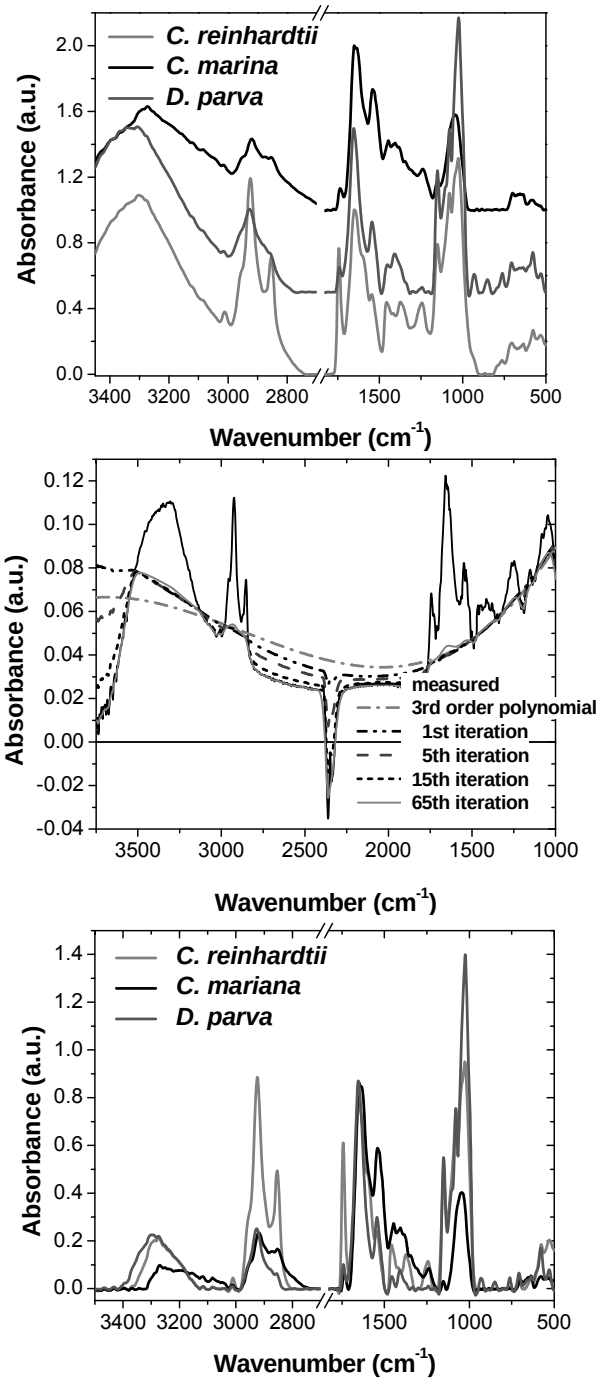


Figure 6: (top) Spectra of three microalgae species impacted by a strong baseline drift; (middle) the estimation of the baseline is iteratively improved; (bottom) after this baseline correction, the drifts interfering with the microalgae spectra shown in the left panel have successfully been removed.

To be generally applicable, a successful baseline correction algorithm has two general requirements: it must be automated in order to handle large amounts of spectra and it should require as few as possible *a priori* assumptions about the shifts' wavelength dependency. The former requirement rules out methods like Bézier curves [88] which require visual inspection and input from a spectroscopist. One automated standard technique is based on computing derivatives [89], [90]-[91]. In reference [92], however, it has been demonstrated that selecting parameters for the Savitzky-Golay algorithm [89] is not without danger as relevant spectral signatures can be suppressed as well. Fourier high- or band-pass filtering has been applied to remove low-frequency components assuming that drifts are much broader in wavelength than real absorption features [93]; a similar idea has been proposed [94]-[95] using wavelet transforms [89], [96]-[97] instead. However, selecting parameters for Fourier or wavelet filtering is subjective since relevant frequency components might also be canceled.

The limitation of such approaches lie in the assumption that baseline shifts are rather broad compared to the real features. In most real-world applications, however, there is no unambiguous differentiation between broad baseline and narrow real features as both have low and high frequency components. Consequently, drift estimation and removal always runs the risk of incomplete removal or over-compensation of drifts and thus disturbance of important features.

Another type of baseline correction builds drifts into chemometric calibration models rather than removing them [86], [92], [98]. This has the advantage of evaluating real features and drifts at the same time while a multivariate regression finds the optimum balance between real features and drifts and thus cannot influence the concentration evaluation. These methods assume that the drift features can be described by means of polynomials and thus are only partially modeling shifts of more general shapes.

The baseline correction algorithm developed in this study is based on a published method [99] as initialization and then iteratively improves the baseline estimate (Figure 6, middle). This original idea used for initialization fits a polynomial to the spectrum to be corrected; it then calculates the residual spectrum by subtracting the fit polynomial from the spectrum and identifies the x% (e.g. 5%) largest residuals (positive and negative values). At wavenumber positions where the x% largest residuals had been found, the spectrum's values are replaced with those of the fit polynomial. In the next step, a polynomial is fit to the modified spectrum. This fit will be closer to the true baseline because the (modified) spectrum's original values have

been replaced with a baseline estimate. This procedure continues back and forth until no significant change in the baseline estimate is achieved and a stable estimation has been obtained. However, it was found that the baseline estimation is often deflected when strong absorbance bands occur at the ends of a spectrum (see Figure 7).

In order to reduce such imperfect corrections, an iterative procedure was utilized to improve this initial baseline estimate. Two antagonizing steps have been implemented: One of which makes the baseline estimate more similar to the spectrum in order to allow the baseline to better follow the underlying baseline. The other tends to make the baseline shorter and thus prevents that the baseline follows true spectral features. When a stable solution is found, the iteration stops. For the tested examples, the algorithm converged after <100 iterations. In conclusion, the novel algorithm for baseline estimation comprises:

- Initialization: Estimate a baseline utilizing the method introduced [99] and define:

$$N = \mathrm{length}(spectrum)$$

- Step 1: Make the current baseline more similar by using:

$$\text{for } k = 1:N$$
$$baseline_{\mathrm{new}}(k) = baseline_{\mathrm{old}}(k) + \beta \cdot (spectrum(k) - baseline_{\mathrm{old}}(k)) \qquad (2)$$

In equation (2) $\beta > 0$ is a parameter that controls how strongly the baseline is made more similar to the spectrum. A rather large value for β puts more weight on Step 1; a small β makes Step 2 preferred.

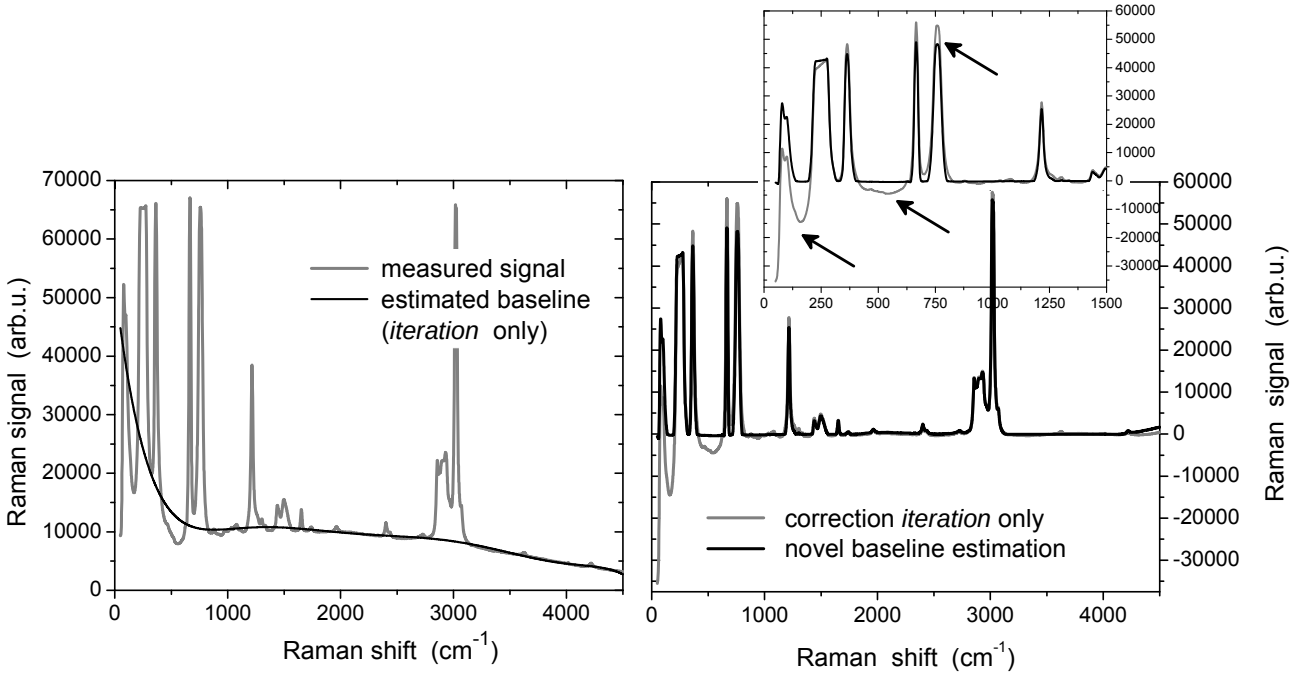


Figure 7: (left) Baseline correction (*initialization* step only) applied to a Raman spectrum of glyceryl tripalmitoleate, a lipid, dissolved in chloroform (785 nm excitation wavelength); (right) the corrected spectrum shown in gray depicts results obtained from the method presented in reference [99] which still shows deviations from a correct baseline marked by arrows in the inset. The black spectrum resulted from the correction method developed in this study, however, and is free of the formerly remaining baseline distortions.

- Step 2: Without any 'counter measure', Step 1 (equation (2)) would eventually make the baseline identical to the spectrum. Thus, another step is required which makes the baseline shorter again via:

$$\text{for } k = 2 : N-1$$
$$baseline_{\text{new}}(k) = \frac{1}{2} \cdot \left[baseline_{\text{old}}(k+1) - baseline_{\text{old}}(k-1) \right]$$

(3)

In other words, the measurement point k is 'bypassed' by replacing it with the value of a connecting line between points $k-1$ and $k+1$. In equation (3), the two measurement points at each end of $spectrum$, i.e. $k=1$ and $k=N$ have to be treated independently e.g. via extrapolation:

$$baseline_{\text{new}}(1) = 2 \cdot baseline_{\text{old}}(2) - baseline_{\text{old}}(3)$$
$$baseline_{\text{new}}(N) = 2 \cdot baseline_{\text{old}}(N-1) - baseline_{\text{old}}(N-2)$$

(4)

Step 3: Go to step 1 while the current baseline estimate is different from the previous iteration. An example depicted in Figure 6 shows the progress of this iteration equations (2)-(4). Once such a baseline drift has been estimated, it is subtracted from the spectrum. Comparing three corrected algae spectra with and without baseline correction (Figure 6, top versus bottom) demonstrates the capabilities for baseline shift removal of this estimation algorithm resulting in a flat baseline at zero absorbance units. Figure 7 presents an application of this method to Raman spectroscopy and thus outlines its applicability for different optical spectroscopy techniques.

Combining these three techniques corrects for challenges which cannot easily be overcome through sample preparation due to the nature of the samples. Figure 8 shows the final series of corrections for a single biochemical analyte. Regardless of sample type, these challenges originate from the chemical state of the samples (solid) and not from the type of sample. Therefore, these challenges would arise whether the sample was pure or a chemically-complex mixture, like microalgae. These correction methods were successful for pure analytes and are applied to the microalgae samples as well.

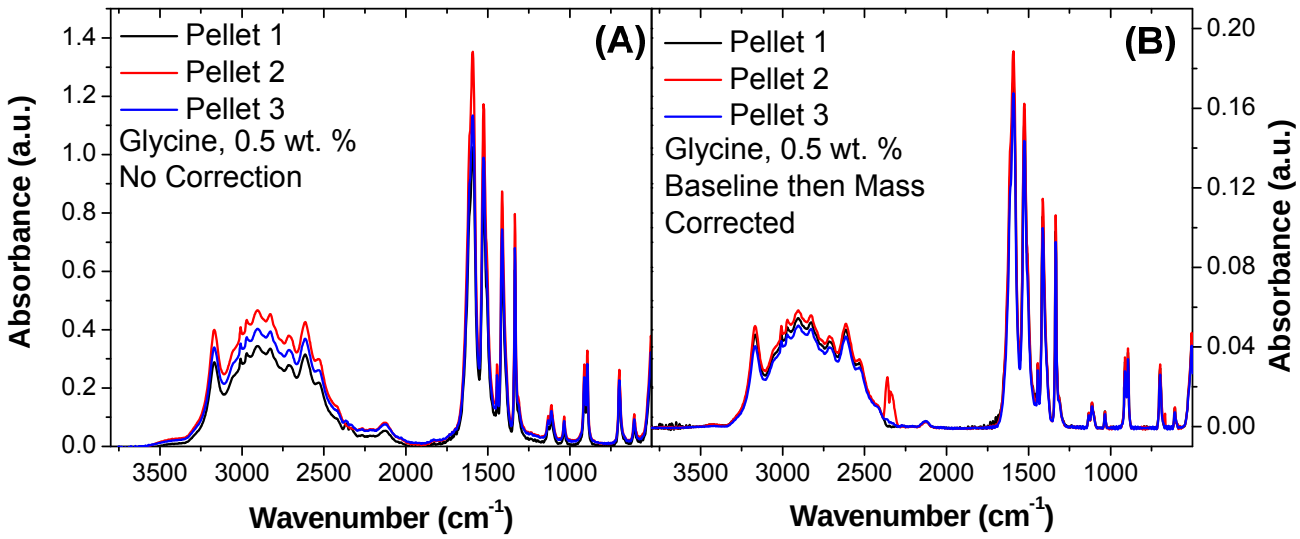


Figure 8. Baseline then pathlength correction for 3 replicate solid samples obtained from the same concentration mixture (Glycine, 0.5 wt. %): (A) raw spectra with no correction, (B) spectra which were first baseline then pathlength corrected.

3.4. Conclusions

Common techniques for studying microalgae are often insufficient in describing the overall effects of changes in their environment since they are typically either a qualitative assessment or focus on the quantification of one particular analyte (group); thus, a more comprehensive method was desired. For this reason, FT-IR was chosen due to the broad range of biochemical components which absorb in the mid-infrared. In order to study microalgae using FT-IR, a novel sample preparation method for analyzing them spectroscopically had to be developed. To this end, well-defined concentrations of microalgae in KBr were prepared and analyzed via FT-IR. After further inspection, it was determined that the spectra acquired from these solid samples were disturbed by light scattering causing baseline shifts, variations in sample thickness, and highly overlapping analyte spectra. These artifacts caused low reproducibility among replicate spectra requiring novel chemometric methods for correction. Thus, a novel baseline removal algorithm and a pathlength correction methodology were developed. Together, these methods reduced the amount of fluctuation between replicate spectra by a factor of two. This style of novel sample preparation and data pre-processing steps were the foundation of all spectroscopic analyses to follow in this dissertation.

4. Improved Principal Component Regression for Quantification of Solid Analytes in Biological Samples

4.1. Introduction

Utilizing microalgae cell's ability to quickly adapt to chemical changes in their environment involves innovative steps in spectroscopic and data analysis [13], [6], [12], [11]. The combination of sample preparation (3.2) and data pretreatment (3.3) methods have been shown to further enhance the comprehensive capabilities of FT-IR and yield new paths for utilizing microalgae as novel environmental sensors. From these cells, which have been introduced to different environmental and nutrient conditions, much insight can be gathered by studying the shifts in major functional groups inside the cells. For this purpose, the quantification of key biochemical analytes is required in order to better understand the nutrient-induced chemical changes occurring inside microalgae cells. Therefore, a reliable quantification method which studies multiple analytes in biomaterials is desired.

Since extracting these key components from algae is tedious and error-prone, an 'artificial' calibration model has been created. In order to test these newly developed baseline and pathlength correction methods (3.3), concentration series of thirty-one mostly solid analytes, chosen from key biochemical groups, were selected and mixed into KBr pellets. Concentration series of these single analytes were prepared and analyzed via FT-IR. To assess the capabilities of these techniques, the selectivity, linearity, and sensitivity of these analytes were tested using PCR calibration models.

4.2. Experimental Section

4.2.1. Preparation of Calibration Samples

Chemometric quantification of target analytes requires a calibration for which a concentration series of well-defined calibration samples must be prepared. Solid compounds were ground individually into fine powders and mixed with potassium bromide (KBr) for concentration definition [85] as described in 'FT-IR Analysis of Algae Samples'. In this study, mixtures were prepared ranging from 0.3 to 1.2 weight percent (wt %) of analyte in KBr. At these concentration ranges (weight percentage or molarity), absorbance spectra of the pellets showed a sufficient signal-to-noise ratio (SNR) and were also far from being saturated.

4.2.2. Preparation of Algae Samples

Stock solutions of alternate algae cultures *Chlamydomonas reinhardtii*, *Chlorella marina*, *Dunaliella parva* (salt water), *Neochloris oleoabundans*, and *Scenedesmus subspicatusi* (fresh water) were obtained from the UTEX laboratory at the University of Texas, Austin (www.utex.org). These species were acquired for testing purposes and were not utilized throughout this dissertation. All species were grown in semi-continuous cultures, at a continuous photon flux density of 100 µmol m^2 s^{-1} photosynthetically active radiation (PAR), at 20°C. ESAW medium at pH 8.2 was used for saltwater species [69]; freshwater species were cultured by UTEX in BG11 medium at pH 7.1. The concentration of each algae culture was determined via direct counts using a hemocytometer. For IR spectroscopic analyses, cells were fixed by the addition of 3 µL of Lugol's solution per 1 mL of algal suspension. Cells were harvested through centrifugation and washed afterwards twice with an isoosmotic solution of ammonium formate to minimize medium carryover. After gently drying for 4-5 days at 60˚C, microalgae samples were mixed with KBr and pressed into pellets as described in 'Preparation of Calibration Samples'. Spectra of microalgae were acquired with the spectrometer's DTGS detector, i.e. the same detector used for analyzing the single-analyte calibration samples.

4.3. Results and Discussion

Inhomogeneities of powdery mixtures especially for low concentrations have been mediated through the use of a mechanical mixer and careful selection of aperture settings enabling a spatial integration of the absorbed light. The challenge of different absorption pathlengths due to different pellet masses has been resolved via pathlength normalization and in a third step a baseline correction algorithm has been implemented. To quantify these improvements, three samples of 0.5 weight % Glycine in KBr were prepared and their relative sample-to-sample fluctuations of absorbance values at two selected wavenumbers (2901 cm^{-1} and 1595 cm^{-1}) were compared before and after corrections. Without corrections, the lowest signal and highest signal levels deviated by ±15% at 2901 cm^{-1} and +19%/-10% at 1595 cm^{-1} from the medium spectrum (Figure 9, A). After corrections, the relative deviations in absorbance values are reduced to ±7% at 2901 cm^{-1} and +8%/-4% at 1595 cm^{-1} (Figure 9, B).

As previously demonstrated, the selected analytes possess similar spectroscopic signatures especially below 1700 cm^{-1}, but nonetheless, enough differences are evident to ensure selectivity in multi-component PCR calibration models [100]-[101], [102]-[104]. In the next step, the accuracy of this approach in multi-component analyses was investigated. For this purpose, concentration series (molarity) for all analytes contained in the database were incorporated into a twenty-one-component calibration model. This calibration contained over 600 calibration spectra from which 48 relevant principal components (PCs) were extracted based on an empirically chosen cut-off ratio (largest singular value divided by smallest accepted to be 1000). For a first assessment, this empirical measure was considered sufficient; for model fine-tuning, more sophisticated methods like cross-validation [100] and F-testing calibration models [105] are available. Since the number of relevant PCs (# = 48) is larger than the number of present analytes (# = 21), apparently imperfections like water vapor or non-linearity in concentrations [106] required modeling by additional PCs.

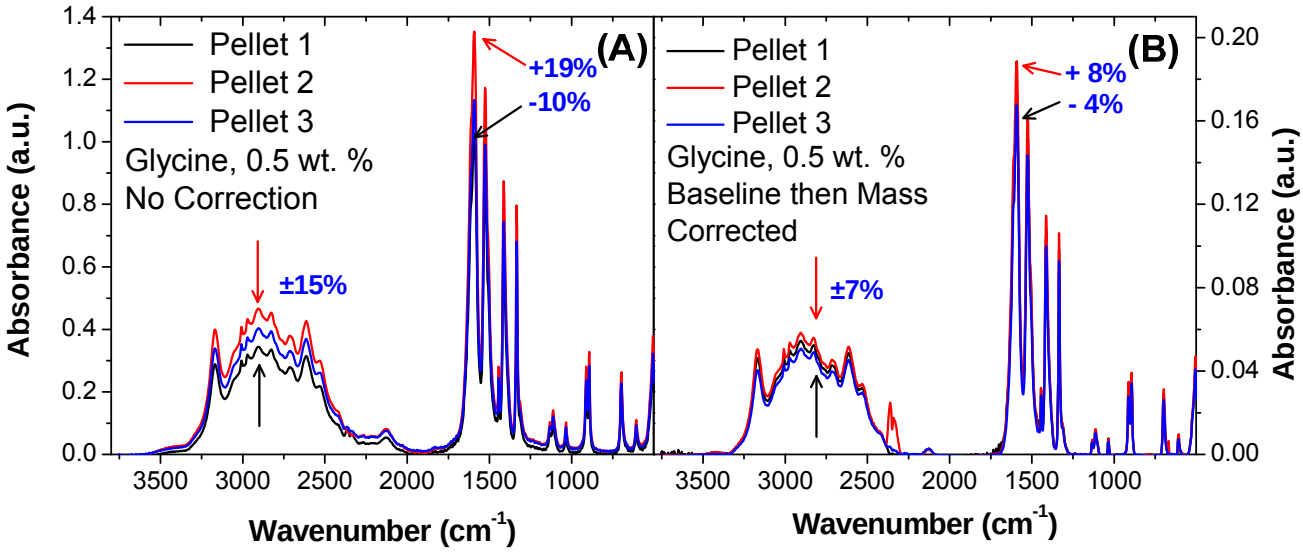


Figure 9. Method of correction for three replicate solid samples obtained from the same concentration mixture: (A) no correction, the lowest signal and highest signal levels deviated by ±15% at 2901 cm^{-1} and +19%/-10% at 1595 cm^{-1} from the medium spectrum. (B) Baseline then mass (pathlength) correction, the relative deviations in absorbance values are reduced to ±7% at 2901 cm^{-1} and +8%/-4% at 1595 cm^{-1}.

For four selected analytes (Figure 10 (top) arginine, cysteine, fructose, glucose), half of the concentrations included in the series were excluded from calibration for independent testing of the resulting PCR calibration model. The predicted concentrations for the independent test samples were plotted as black dots versus the true concentrations known due to sample preparation; the given error bars were obtained from evaluating three replicate measurements. As these samples contained only one analyte but the calibration model was built for 21 compounds, the remaining 20 analyte concentrations were in this case expected to be zero. These results are concluded in the gray dots also depicted in the four panels.

Here, however, the gray dots were obtained from averaging the 20 zero concentration values and the error bars were determined accordingly. As shown in the panels of Figure 10, most of the 21 predicted concentrations per sample are in good agreement with the expected values. A similar calibration model was built containing *thirty-one* analytes which was applied to predict the concentration of two polymeric analytes for which no molar masses are available; thus, in such cases the analysis was restricted to samples containing one analyte only whose concentration was given as weight percentages of analyte in KBr.

In the next step, this calibration was applied to model spectroscopic signatures obtained from two freshwater microalgae species (*Scenedesmus subspicatus*, *Neochloris oleobundans*). This was performed to assess the comprehensiveness of the database for studies of biological materials; if the aforementioned 31 analytes are indeed well-chosen representatives for the key compounds in biomaterials, calibration models derived from them are expected to describe the spectroscopic features of these biosamples well. The chosen approach for testing completeness was based on decomposing the biomaterials' spectra into the PCs which were obtained from single analyte samples; then, the spectra were reconstructed via calculating scores times PCs. Figure 11 shows the originally measured spectra in gray and the reconstructed spectra in black. The spectroscopic information contained in the database can describe most of the spectroscopic signatures of microalgae samples. However, the residual spectra (measured minus reconstructed) not only contained random noise – there are also some real features; therefore, it was concluded that for future studies additional analytes need to be incorporated into the database in order to enhance the accuracy of spectroscopic modeling of biomaterials and thus concentration predictions. Nevertheless, the current database is complete enough for quantitative investigations of several key components in microalgae.

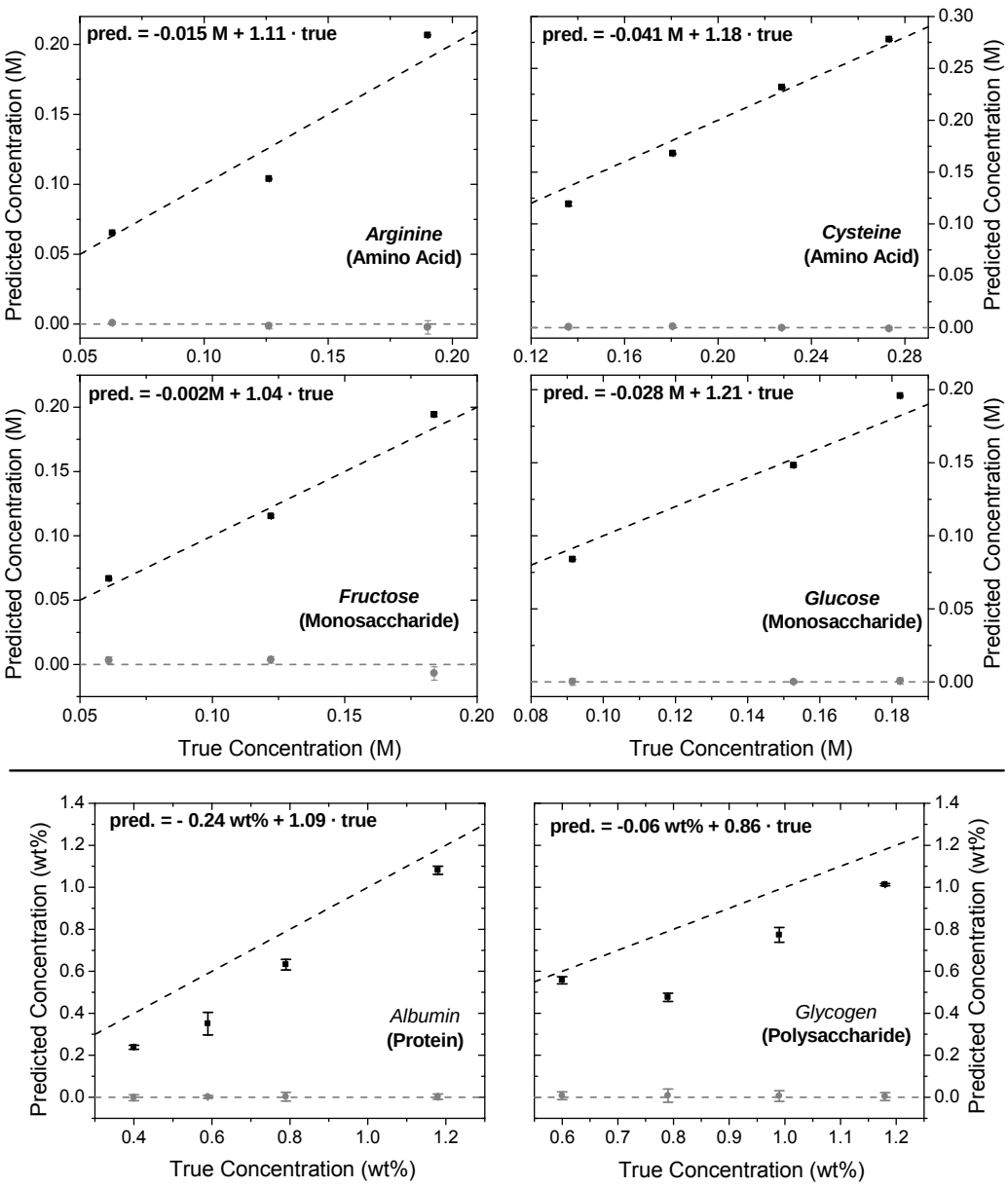


Figure 10: (top) Twenty-one analytes have been calibrated by means of a PCR; the black dots indicate predicted versus true concentrations (molarity) for four selected analytes (error bars were obtained from three replicates); since the remaining 20 analytes were not contained in these samples, their predicted concentration was expected to be zero (gray dots); the black and gray dashed line indicate predicted = true concentrations; (bottom) Thirty analytes (including the 21 from top, excluding glutamic acid) have been calibrated by means of a PCR requiring the use of weight percent as concentration unit.

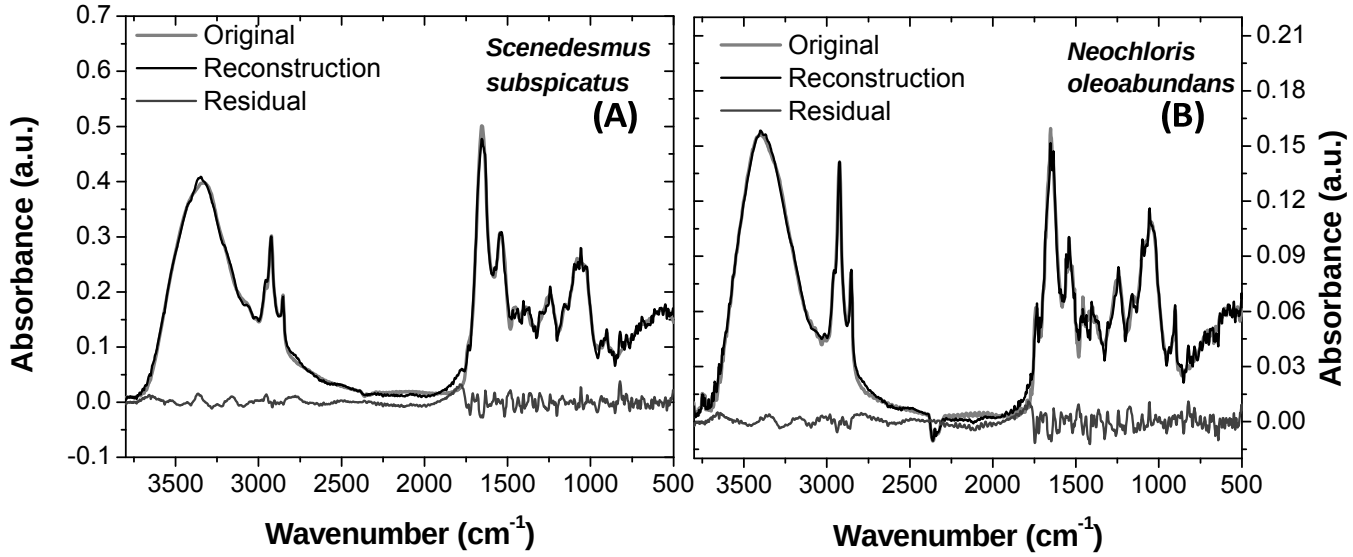


Figure 11: Reconstruction of spectra obtained from two microalgae species: (A) Scenedesmus subspicatus and (B) Neochloris oleoabundans.

4.4. Conclusions

Biological samples are often complex mixtures consisting of a large number of analyte groups including proteins, carbohydrates, lipids, carboxylic acids, and amino acids most of which are solids. Characterizing and studying chemical processes like the response of microalgae to changing environmental conditions requires more comprehensive investigations than are currently feasible based on highly selective single analyte measurements. For such applications, FT-IR transmission spectroscopy supported by chemometrics is a promising technique as most relevant analytes are IR active (Appendices

Appendix 1). For all selected analytes, concentration series were prepared in the form of KBr pellets.

However, solid samples, and especially biomaterials, often have a limited reproducibility on top of challenges inherent in their chemical complexity. Thus, innovative sample preparation steps (3.2), data pre-processing (3.3), and calibration procedures were developed to reduce spectroscopic artifacts and to enhance multi-component quantifications. Both data pre-processing steps combined resulted in a considerable improvement of the spectroscopic quality and the prediction power of multi-component PCR calibration models.

Selectivity, linearity, and sensitivity of 31-compound PCR calibration models have been demonstrated based on KBr pellets. To assess the database's completeness, FT-IR spectra from microalgae samples were reconstructed based on the synthetic, single analyte calibration spectra. Aside from rather small non-noise features, which will be further reduced by adding analytes to the database, residual spectra mostly consisted of noise. Thus, a novel, quantitative analysis tool for solid biological samples has been presented which is based on FT-IR spectroscopy supported by data pre-processing methods for enhancing the spectral quality.

5. Environmental Impacts on the Spectroscopic Signatures of Microalgae Cells

5.1. Introduction

Assessing whether nutrient-induced changes in the chemical composition of microalgae cells can be utilized as novel, *in-situ* sensors of environmental conditions involves many steps. Previous studies have shown that small changes in environment can be seen in the resulting spectra when grown under different nutrient conditions [10]. However, spectra acquired from the same nutrient and concentration have been shown to possess large replicate-to-replicate fluctuations. Often the shifts in chemical composition as a result of changing environmental status are very slight; therefore, ensuring the replicate-to-replicate fluctuations are smaller than the nutrient concentration-to-concentration fluctuations is required in order to continue investigating microalgae as potential bioprobes. A method of correction is then needed in order to extract the nutrient-induced changes in the spectrum

While the three microalgae spectra have a distinct spectrum (Figure 12), variations in replicate spectra due to physical properties of the samples make comparing different environmental parameters more challenging. As shown in Figure 13, FT-IR spectra were acquired from six independent samples of *Dunaliella salina* grown under 0.549 mM sodium nitrate. It was assumed that spectra obtained from the same nutrient and concentration should have the same absorbance values at a given wavenumber. Although previous data preprocessing steps (section 3.3) were applied, the spectra still show remaining in-replicate fluctuations. Thus, an additional normalization was needed to distinguish replicate-to-replicate fluctuations from real chemical change. Building off a standard multivariate least-squares regression, a spectral normalization step was developed in order to gain an overall representation of the changes occurring in the biochemical composition of these cells. With increasing reproducibility of replicate spectra, slight changes in the spectra were then seen due to changes in growing conditions.

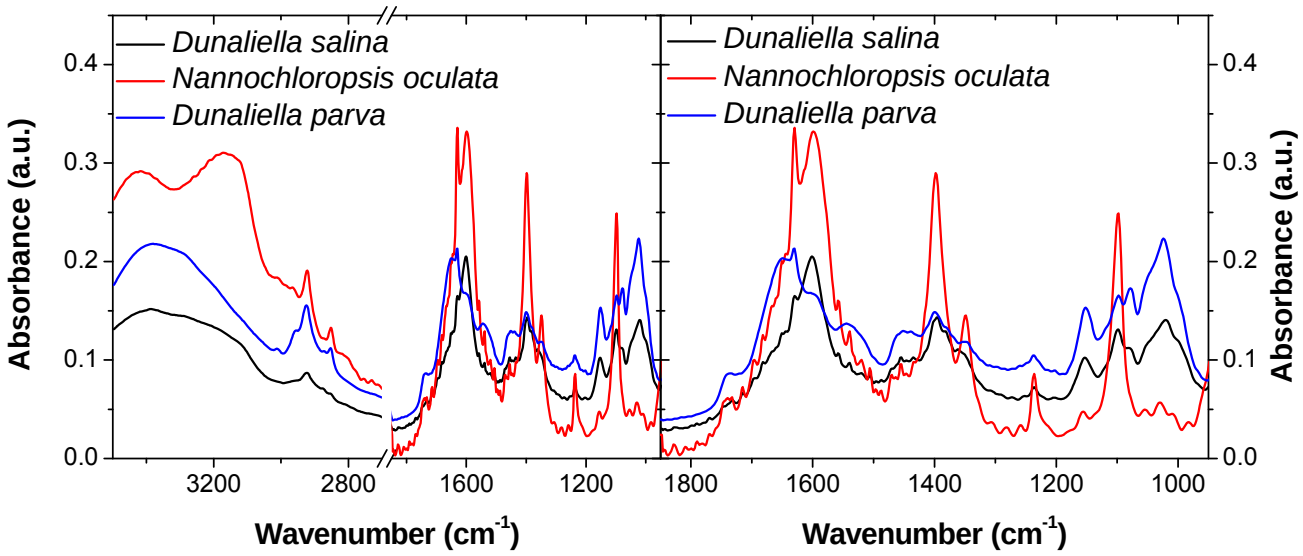


Figure 12. Spectroscopic signature of three microalgae species: (left) full spectral range (3500 – 950 cm^{-1}, excluding 2700 – 1850 cm^{-1}), (right) fingerprint region (1850 - 950 cm^{-1}).

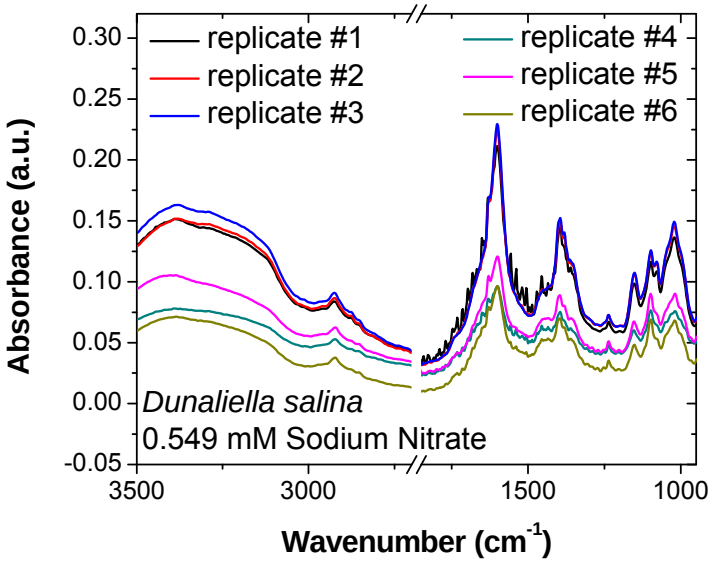


Figure 13. Replicate spectra acquired from six independent samples of *Dunaliella salina* grown under 0.549 mM sodium nitrate. Spectra have been corrected for pathlength variations and baseline shifts (section 3.3) but still show remaining in-replicate fluctuations. Thus, a further correction step is needed.

In the interest of qualitatively assessing the effects of growing environment on microalgae's spectroscopic signatures, investigations to determine which portions of the spectra are statistically different are required. For this purpose, these variations were extracted using a t-test.

5.2. Spectral Normalization

Similar to single biochemical analytes, minute fluctuations in algae concentration within each pellet can cause spectral fluctuations. While other sources of variations which have been previously discussed (3.3) contribute, for microalgae samples, the natural "clumping" of the algae within the KBr mixture introduces additional sample-to-sample variations. In addition, the concentrations of minor (not including nitrogen or carbon) ESAW nutrients could also cause variations in the chemical composition of each sample. Thus, a method based on multivariate least-squares regression was designed in order to distinguish replicate-to-replicate fluctuations from real chemical change. In the remainder of this discussion, the following notation will be utilized:

$\mathbf{x}_{(N\times 1)}$ a lower case bold letter indicates a (column) vector with N elements $\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$; the subscript is not always given

$\mathbf{X}_{(N\times K)}$ a capital bold letter indicates a matrix with N rows and K columns.

$X_{n,k}$ the matrix element of $\mathbf{X}$ at the position row number n and a column number k.

When qualitatively investigating the microalgae spectra acquired from different nutrient conditions, changes in the spectral shape as a result of environment are desired rather than quantifying absolute absorbance values for concentration. However, low reproducibility among spectra prevents these minor changes from being obvious. These fluctuations can be normalized by means of a multiplicative factor (c) and an additive spectral shift, which is approximated by a wavenlength-dependent polynomial $\mathbf{r} = \mathbf{r}(\lambda)$ (Figure 14).

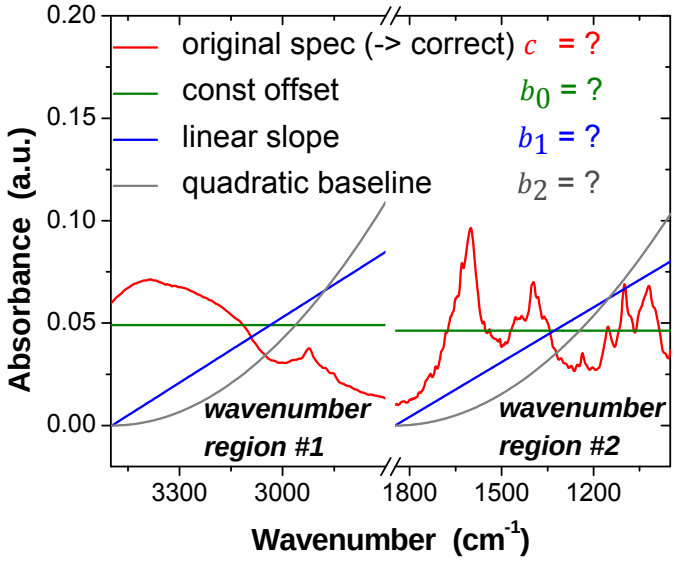


Figure 14. Model demonstrating the spectral normalization for replicate spectra: the original spectrum which needs correction is shown in red. Utilizing the scaling factor, (*c*), and a polynomial (here, quadratic) wavelength-dependent baseline correction (**r**), replicate spectra from the same class were normalized for comparison. This correction is performed in two independent wavenumber regions since the polynomial baseline is slightly different in each region.

A representative spectrum, or 'gold standard', must be chosen which best describes the representative features of a replicate class. For example, Figure 15, (A) shows two spectra of the same species, nutrient, and nutrient concentration whose main spectral features are very similar. However, these two spectra suffer from variations in algae concentration and some polynomial wavelength dependence. Before any investigations of nutrient induced spectroscopic changes can be performed, these spectra must be normalized such that systematic errors are removed and only random measurement errors remain. For this purpose, a chosen representative spectrum, $\mathbf{s}_1$, would be the model to which all remaining replicate spectra, $\mathbf{s}_{k=2,\cdots,K}$ would be normalized. Thus, $\mathbf{s}_k$ can be expressed as a scaled version of $\mathbf{s}_1$ plus some unknown polynomial baseline ($\mathbf{r}$).

Since the polynomial baseline is different in each of the chosen, relevant wavenumber regions, this correction is performed in both wavenumber regions independently (3500 cm^{-1} – 2700 cm^{-1}, 1850 cm^{-1} – 950 cm^{-1}). Polynomials from first to third order were tested to obtain the best fit of the underlying baseline. Figure 15, B-D demonstrates the effect of each polynomial baseline fitting. The first and second polynomial orders are appropriate choices for assimilating the two test spectra. The third order polynomial tends to over-fit in some areas causing large differences between the two spectra at several wavenumber positions.

While visually there are few differences between the first and second polynomial fit, the second order polynomial was determined to best correct for baseline deviations between spectra. Using this polynomial order, the unknown baseline ($\mathbf{r}$) can be explained by the following parameters: $\mathbf{r} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$. Given this information, the following equation will be used as a model, where $\mathbf{s}_{\mathrm{k=2,\cdots,K}}$, for estimating the multiplicative scaling factor (c) and polynomial fit parameters ($\mathbf{r}$) for correcting additive shifts,:

$$\mathbf{s}_k = c \cdot \mathbf{s}_1 + b_0 \cdot \mathbf{1} + b_1 \cdot \boldsymbol{\lambda} + \mathrm{b}_2\boldsymbol{\lambda}^{\mathbf{2}} \qquad (5)$$

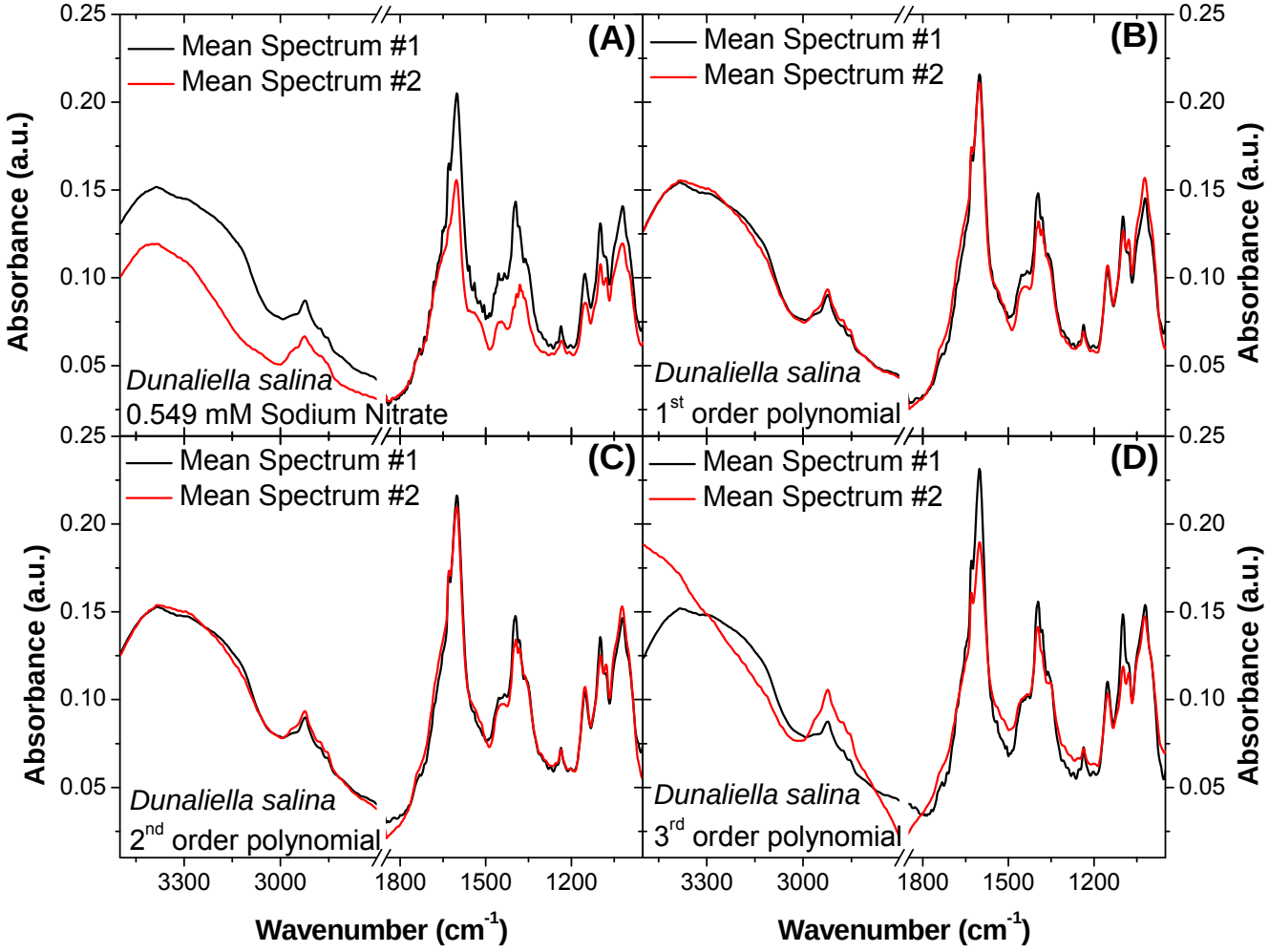


Figure 15. Microalgae spectra utilized for spectral normalization: (A) original spectra acquired from two independent samples of the same species, nutrient, and nutrient concentration, (B)-(D) first through third order polynomials were tested to determine the best fit. Second order polynomial was chosen for normalizations.

However, measurement errors, δ, impact the measurements and thus (5) is not accessible:

$$\mathbf{s}_k = \mathrm{c} \cdot \mathbf{s}_1 + b_0 \cdot \mathbf{1} + b_1 \cdot \boldsymbol{\lambda} + b_2 \boldsymbol{\lambda}^{\mathbf{2}} + \delta$$

(6)

With multiple spectra in each class, (6) expands to:

$$\begin{aligned}
\mathrm{s}_{k,1} &= c \cdot \mathrm{s}_{1,1} + b_0 \cdot 1 + b_1 \cdot \lambda_1 + b_2 \lambda_1^2 + \delta_1 \\
\mathrm{s}_{k,2} &= \mathrm{c} \cdot \mathrm{s}_{1,2} + b_0 \cdot 1 + b_1 \cdot \lambda_2 + b_2 \lambda_2^2 + \delta_2 \\
\mathrm{s}_{k,3} &= \mathrm{c} \cdot \mathrm{s}_{1,3} + b_0 \cdot 1 + b_1 \cdot \lambda_3 + b_2 \lambda_3^2 + \delta_3 \\
&\vdots \\
\mathrm{s}_{k,i} &= \mathrm{c} \cdot \mathrm{s}_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_4 + b_2 \lambda_4^2 + \delta_K
\end{aligned}$$

(7)

These measurement errors, δ_k, impose an unsolvable problem because there are $4 + k$ unknowns but only 4 equations. Therefore, *good overall parameters*, b, c_0, c_1, c_2, must be estimated. Least-squares regression is such a fit procedure which estimates parameters via minimizing the sum of squared errors (SSE).

$$\mathrm{SSE} = \sum_{i=1}^{K} \delta_i^2$$

(8)

$$= \sum_{i=1}^{K} \left[\mathrm{s}_{k,i} - \left(\mathrm{c} \cdot \mathrm{s}_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}} = \min$$

(9)

In order to determine this minimum, partial derivatives are computed with respect to the wanted parameters and then set equal to zero:

$$\text{(I)}\ \frac{\partial \mathrm{SSE}}{\partial \mathrm{c}} = 0 \qquad \text{(II)}\ \frac{\partial \mathrm{SSE}}{\partial b_0} = 0 \quad \text{(III)}\ \frac{\partial \mathrm{SSE}}{\partial b_1} = 0 \quad \text{(IV)}\ \frac{\partial \mathrm{SSE}}{\partial b_2} = 0$$

(10)

From equation (10), an equation system results for whose solution can be defined into a matrix, $\mathbf{\Lambda}$, and vector, $\mathbf{d}$ (derivations for this normalization can be found in (Appendix 2)):

$$\mathbf{\Lambda}_{(Kx4)} = \begin{pmatrix} s_{1,1} & 1 & \lambda_1 & \lambda_1^2 \\ s_{1,2} & 1 & \lambda_2 & \lambda_2^2 \\ s_{1,3} & 1 & \lambda_3 & \lambda_3 \\ \vdots & \vdots & \vdots & \vdots \\ s_{1,K} & 1 & \lambda_K & \lambda_K^2 \end{pmatrix} \qquad \mathbf{d}_{(4x1)} = \begin{pmatrix} c \\ b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

(11)

Using this matrix-vector notation, the following can be expressed:

$$\mathbf{\Lambda^T \cdot \Lambda \cdot d} = \mathbf{\Lambda^T \cdot s}_k$$

(12)

Thus, the unknowns, $\mathbf{d}$, are estimated via:

$$\left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \left(\mathbf{\Lambda^T \cdot \Lambda}\right) \cdot \mathbf{d} = \left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \mathbf{\Lambda^T \cdot s}_{\mathrm{k}}$$

$$\mathbf{d} = \left(\mathbf{\Lambda^T \cdot \Lambda}\right)^{-1} \cdot \mathbf{\Lambda^T \cdot s}_{\mathrm{k}}$$

(13)

Figure 16 portrays the success of this regression technique for normalizing spectra from the same class. With this correction, replicate-to-replicate fluctuations are minimized allowing for true chemical changes which occur as a result of different nutrient conditions to be studied.

5.3. Statistical Analysis of Changes in Spectroscopic Signatures

With this method applied to remove sample-to-sample artifacts from spectra, the changes induced by the ambient conditions (nutrients) can be analyzed (Figure 17). Before any quantitative analyses can be performed, statistical analyses to confirm these changes in microalgae's spectroscopic signatures must be conducted.

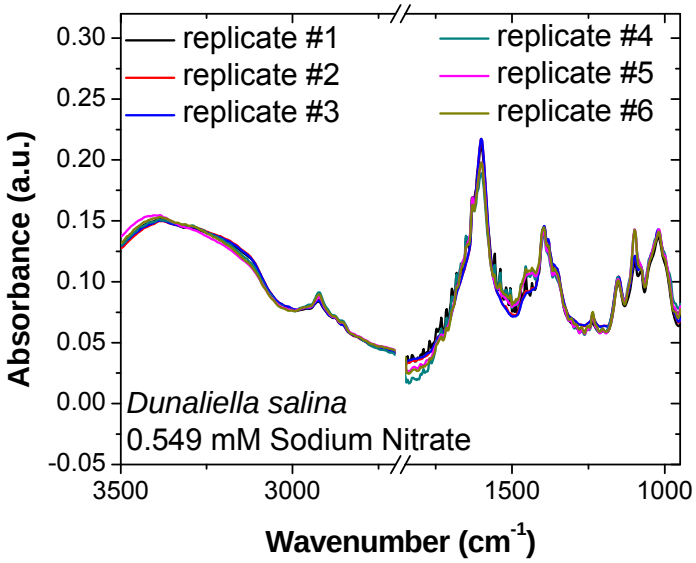


Figure 16. Results of spectral normalization on six replicate spectra obtained from independent *Dunaliella salina* samples.

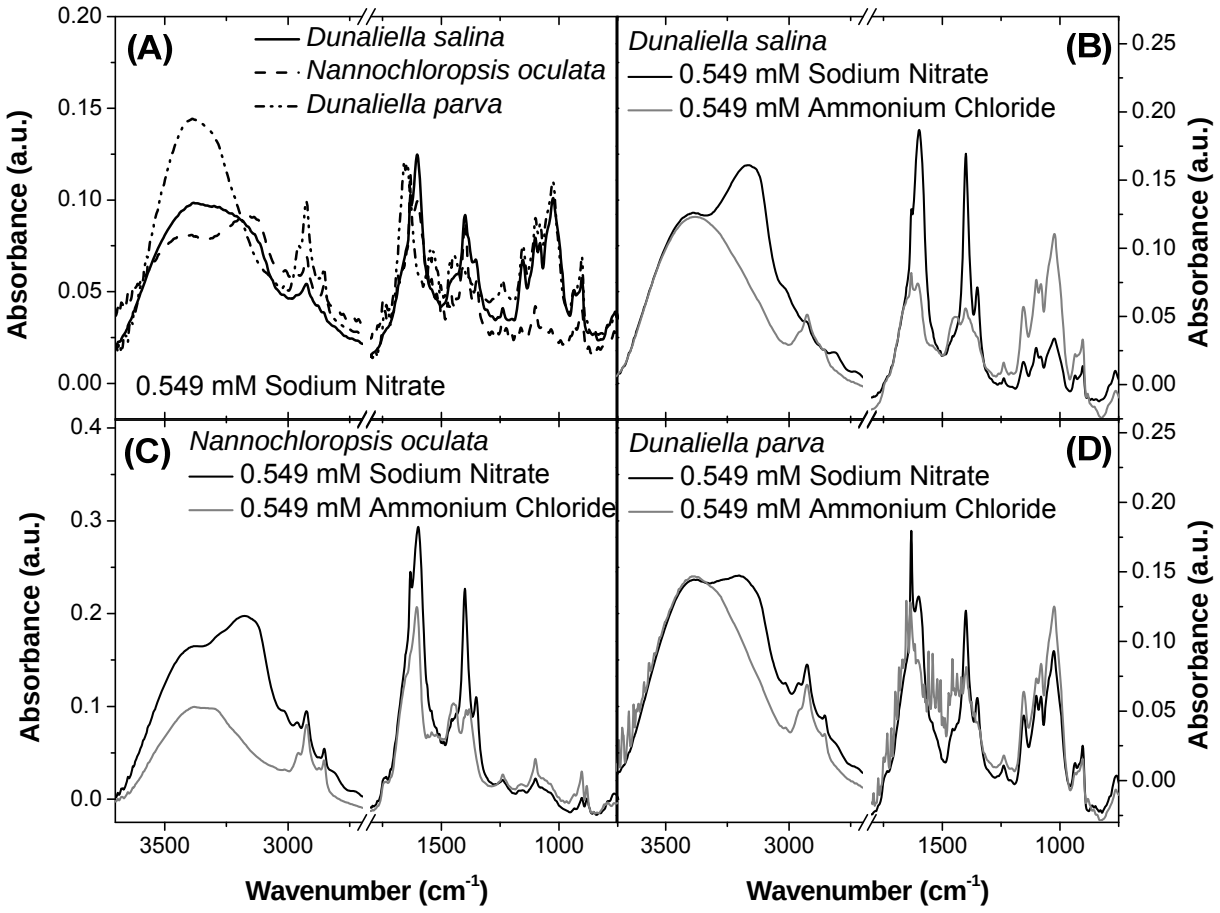


Figure 17: Microalgae spectra after correction demonstrating the effects of varying nutrient sources and concentrations: (A) Signature FT-IR spectra of three different microalgae species under the same nutrient and concentration: *Dunaliella salina, Nannochloropsis oculata, Dunaliella parva*; (B)-(D) Microalgae species grown under two different nitrogen sources. One region was cut-out due to the absence of chemical features and presence of atmospheric carbon dioxide (1800-2700 cm^{-1}).

The statistical method of choice for this application was a t-test; for this purpose, average spectra from two different nutrient conditions were calculated. These two spectra, including their standard deviations, were compared at every wavelength position to determine where the two mean spectra differ in absorbance i.e. where does the second mean spectrum increase or decrease as compared to mean spectrum one. These portions of the spectra which undergo an increase or decrease are indicative of nutrient-induced changes in chemical composition. Those wavenumber regions which are not significantly different, i.e. no change in absorbance is observed, are not impacted by the change in nutrient concentration.

For example, in Figure 18, (A), two mean spectra $\pm$ standard deviations of *Dunaliella salina* grown under different sodium bicarbonate concentrations, (2.07 mM vs. 8.26 mM) are compared. The t-test results, as shown in Figure 18 (B), illustrate the wavenumber regions which underwent a chemical changes as a result of nutrient concentration. If no significant difference is determined between the two classes, the t-test output is zero. When a significant difference is calculated, the t-test output determines whether an increase or decrease in absorbance occurred at that wavenumber position. A t-test result of 1 indicates the first mean spectrum experienced a decrease in absorbance as compared to the second mean spectrum; likewise, t-test result of -1 indicates the first mean spectrum experienced an increase in absorbance as compared to the second spectrum.

t-test results for varying concentrations of sodium bicarbonate were used as an example; trends from all nutrient sources were established using this method. According to the t-test results, changes in the nutrient concentration within the microalgae growth environment directly correlate to shifts in chemical composition.

While significant differences were desired between two concentrations of the same nutrient source, it was also desired for there to be no significant difference between replicate spectra acquired from the same nutrient source and concentration. Thus, a t-test was performed for this purpose as shown in Figure 19. The t-test result indicated there was no significant change between the two sets of replicate spectra.

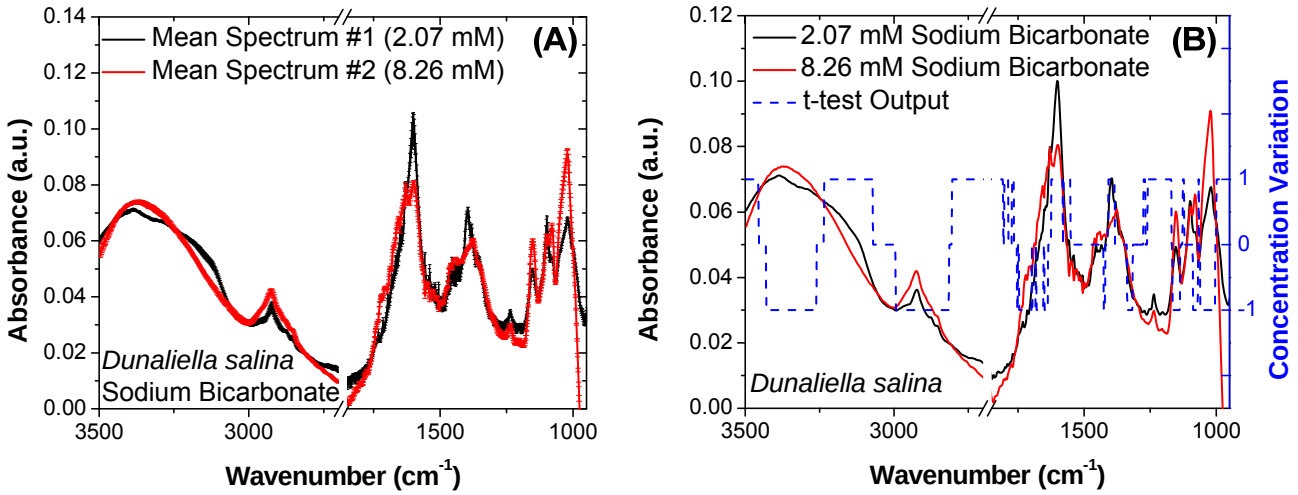


Figure 18. (A) Average spectra from two different carbon nutrient concentrations (2.07 mM and 8.26 mM sodium bicarbonate) with error bars (shown). (B) t-test result obtained by comparing both mean spectra at every wavenumber position to determine a statistical difference. Concentration variation is defined by the given rules: 0 = no difference at given wavenumber position, >0 = red < black, <0 = red > black.

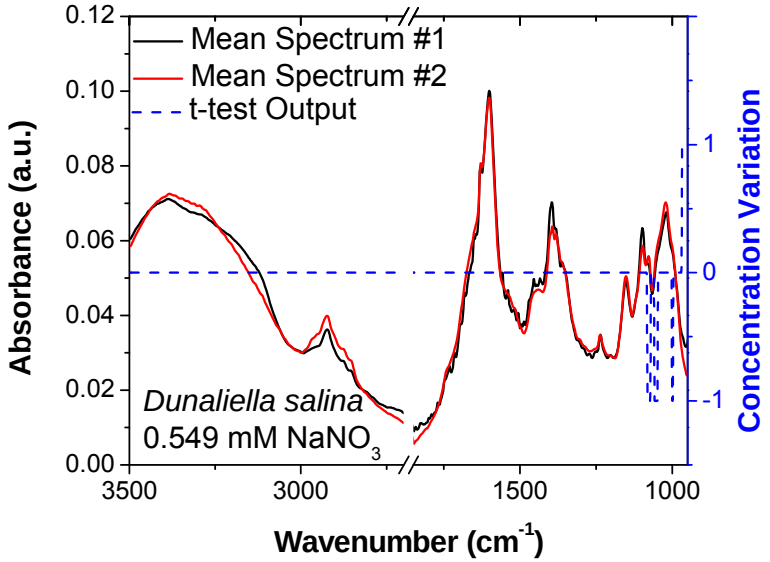


Figure 19. Demonstrating the effects of the spectral normalization. Replicate spectra from the same class was divided into two "classes": mean spectrum 1 and 2 both acquired from the same nutrient source and concentration. t-test output confirms minimal differences between the two average spectra.

These results identify the regions of the spectrum which undergo the most change as a result of growth environments. In order to use this information, the next step was to develop models for relating these changes in spectroscopic signatures to selected environmental parameters. Therefore, five of the seven concentrations, both limiting and excess, of each nutrient source were studied for calibration purposes.

During investigations of these concentration series, it was observed that these microalgae species do not always react to changes in their environment in a linear fashion. In fact, Figure 20 demonstrates the highly nonlinear response of *Dunaliella salina* to three concentrations of sodium nitrate. For example, when comparing corrected spectra averaged over 15 replicates under the normal ESAW concentration of sodium nitrate (0.549 mM) with average spectra from an excessive (0.873 mM) and limiting (0.35 mM) concentration, the result is a nonlinear relationship between nutrient concentration and measured absorbance. Furthermore, the middle concentration (0.549 mM) in Figure 20 appears to be the lowest absorbance of all three concentrations whereas the lowest and highest concentrations yield nonlinear responses at several different wavenumber positions, i.e. at some wavenumber positions, the lowest concentration yields a higher absorbance value and vice versa.

For this reason, modeling changes in the cell's chemical composition with the goal of quantifying ambient parameters requires nonlinear prediction algorithms. As a result, linear methods such as PCA-PCR, MLR, and CLS, become invalid. Therefore, in order to model the response of these biological systems, a nonlinear method was derived by which the ambient growing conditions could be predicted (this topic is pursued in "Nonlinear Modeling of Microalgae Cells Responses to Ecological Parameters" to follow).

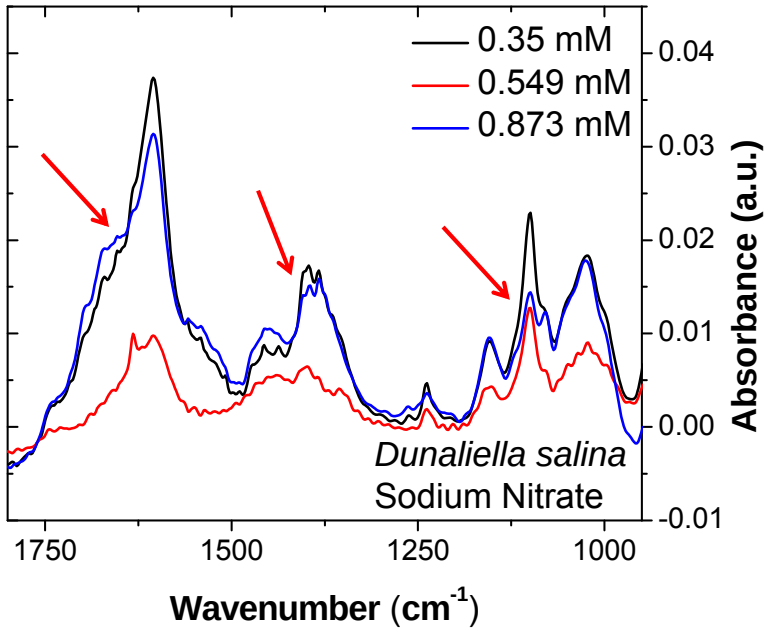


Figure 20. Nonlinear result of varying the concentration of a nitrogen nutrient, sodium nitrate, in ESAW on the chemical composition of *Dunaliella salina*: 0.35 mM (limiting), 0.549 mM (normal condition), 0.873 mM (excess).

5.4. Conclusions

Much insight into the effects of different nutrient situations can be gathered from studying the shifts in major functional groups inside microalgae cells which have been introduced to different environment and nutrient conditions. Thus, qualitative investigations of these changes were performed in order to determine the effects on the spectroscopic signatures of the microalgae cells. However, further challenges in reproducibility beyond baseline shifts and pathlength variations were observed upon inspection of these replicate spectra. In order to discriminate between replicate-to-replicate fluctuations and real chemical change, these replicate spectra needed to be normalized to one another in order to later compare them to other nutrient concentrations.

Two sources of fluctuations were found: the first source, most likely due to variations in algae concentration in each pellet, manifests as a shift in the y-direction or fluctuating absorbances. This is corrected by a multiplication constant which is derived for each spectrum. The other source which caused irreproducible spectra was an additive shift similar to a baseline drift. Thus, a spectral normalization based on least-squares regression was devised to estimate these two components.

While the changes in chemical composition as a function of the growing conditions selected for this dissertation were slight, comparisons of different nutrient concentrations revealed large regions of change in the spectroscopic signatures. In order to further process this information, the changes in chemical composition must be extracted to determine which regions of the spectrum were significantly different. A t-test was implemented for this purpose which compared the spectra from one concentration to that of another at each wavenumber position. The t-tests comparing these two concentrations resulted in significant differences in many different wavenumber regions.

From these results, all nutrient concentrations were studied for an overall impact of environmental variations. Upon further inspection, it was determined that there was a nonlinear dependence between nutrient concentration and absorbance value in several different wavenumber regions. In order to utilize the spectral changes induced by ambient parameters to quantify these changes, nonlinear calibration models are required which are presented in the

next chapter. This nonlinear method was developed in order to study these effects for the purpose of predicting future trends.

6. Nonlinear Modeling of Microalgae Cells Responses to Ecological Parameters

6.1. Introduction

Chemical analyses of complex processes often encounter nonlinear relationships between chemical parameters of interest and measured signals. Such nonlinearities and coupling of several parameters can originate, for example, from interactions of analytes within samples [107], light scattering from nanoparticles [108] or chromophores [109], or –as demonstrated in this study- from biological requirements. For building quantitative prediction models to describe such nonlinear data, standard multivariate least-squares regression (MLR) [110], [111] based techniques like Principal Component Regression (PCR) [100] – [103] and Partial Least-Squares (PLS) [100], [120] are limited as they can empirically approximate nonlinear behavior to a certain extent but do not shed light on the underlying chemical processes. Generally, if the measured signals are not a simple sum of the individual analytes' contributions, or if the signal levels are not directly proportional to the parameters generating them, the calibration model is flawed and considerable prediction errors can occur.

Methods based on neural networks [112], for instance, may be capable of empirically predicting the parameters in a nonlinear context but, due to their design, they cannot provide information about the underlying chemical relations. Kernel Principal Component Analysis (KPCA) [113] - [115] has been introduced for nonlinear modeling and sample classification but the computational expense has been prohibitive for many applications. Response Surfaces [119] have been used to build multivariate polynomial models; however, Response Surfaces are based on Inverse Least Squares (ILS) regression [120] which generally involves a rather large number of predictor variables, W, to derive a small number of response variables, N. Reference [121] for instance applies Response Surfaces for modeling nonlinear and coupled influences of $Q = 6$ instrument parameters to describe a chromatogram's peak resolution $(N = 1)$. While such models are robust due to their large number of adjustable model parameters, W, they also require $K \geq W \gg N$ calibration samples and often become

experimentally unfeasible. This is particularly true for measurement techniques such as optical spectroscopy which produce collinear data sets (N large).

In order to derive a deeper understanding of nonlinear chemical systems while requiring a reasonable level of computer resources and calibration samples, a novel data modeling technique was developed in this study. This method coined 'Predictor Surface' expands the conventional MLR to include coupled and higher-order terms of the predictor variables $\mathbf{x}_{(Q\times 1)} = (x_1 \quad \cdots \quad x_Q)^{\mathbf{T}}$ such as $x_1 \cdot x_2$, or x_5^3, or $x_1 \cdot x_3^2 \cdot x_5^3$. It is shown below that the standard MLR algorithm can be utilized for Predictor Surfaces calibrations as the nonlinearity is built into a modified $\mathbf{x}_{(W\times 1)}$ (with: $W > Q$); for the prediction step, however, nonlinear multivariate equation systems need to be solved for deriving $x_1 \quad \cdots \quad x_Q$.

As a proof-of-principle application, IR spectroscopy of microalgae cells has been chosen. It is hypothesized that microalgae cells can be utilized –here via their FTIR spectrum- as novel *in-situ* 'bioprobes' for chemical shifts in aqueous ecosystems [2], [10], [116]. The advantage of this environmental sensing approach is threefold: (i) Multiple environmental parameters can be quantified and (ii) ecological consequences resulting from these environmental changes can be assessed as the microalgae cells are the beginning of the trophic chain. (iii) Furthermore, based on Predictor Surfaces, the cells responses to chemically coupled ambient parameters can be studied. For these purposes, nonlinear data analyses are required as the microalgae cells' biomass is analyzed rather than the environmental parameters themselves. This biomass production is nonlinear in nature and the cells need certain nutrient concentrations *ratios* to thrive [2], [13].

In this application, concentrations of Q algae nutrients have been chosen as environmental parameters to be monitored; they will play the roles of the predictor variables $\mathbf{x}_{(Q\times 1)} = (x_1 \quad \cdots \quad x_Q)^{\mathbf{T}}$. IR absorbances $A_{n=1,\ldots,N}$ of algae cells measured at N different wavenumbers serve as response variables $\mathbf{y}_{(N\times 1)} = (y_1 \quad \cdots \quad y_N)^{\mathbf{T}}$. However, as Figure 21 (A) depicts, there are considerable nonlinearities between the microalgaes' FTIR spectra and the nutrient concentrations dissolved in the growing medium.

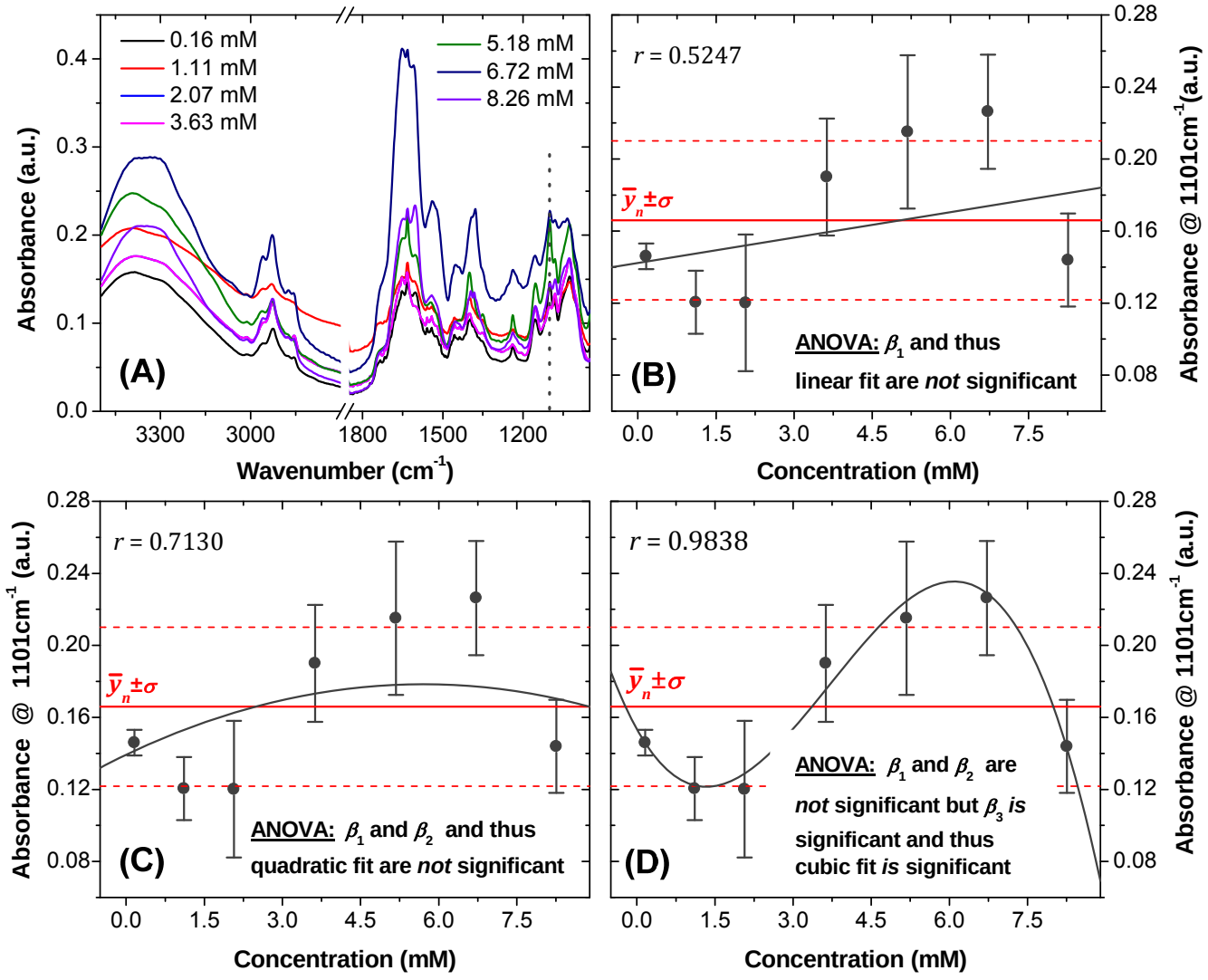


Figure 21: (A) FT-IR spectra acquired from *Dunaliella parva* grown under varying concentrations of sodium bicarbonate as carbon source are changing nonlinearly with the $NaHCO_3$ concentration. For example, the absorbance at $1101\ \mathrm{cm}^{-1}$ was tested whether (B) the inclusion of $\beta_{n,1} \cdot x$ into $y_{n=1101\mathrm{cm}^{-1}} = \beta_{n,0} + \beta_{n,1} \cdot x$ considerably contributes to $\mathrm{var}(y_{n,k})$; (C) the inclusions of $\beta_{n,1} \cdot x$ as well as $\beta_{n,2} \cdot x^2$ into $y_n = \beta_{n,0} + \beta_{n,1} \cdot x + \beta_{n,2} \cdot x^2$ are significant; (D) the significance of $\beta_{n,1} \cdot x$, $\beta_{n,2} \cdot x^2$, and $\beta_{n,3} \cdot x^3$.

Plotting the absorbance $A_{1101\mathrm{cm}^{-1}}(c)$ versus the nutrient concentration c (Figure 21B) revealed that there is no strong correlation $(r = 0.5247)$ and an ANOVA [125], [128], [129] confirmed that for the shown fit function $A(c) = \beta_1 \cdot c + \bar{A}$ the linear term $\beta_1 \cdot c$ does not significantly change the explained data variance compared to the set's mean absorbance $\bar{A}$. Since this mean is concentration-independent, no prediction of c is feasible at this particular wavenumber. When fitting a quadratic polynomial $A(c) = \bar{A} + \beta_1 \cdot c + \beta_2 \cdot c^2$ (Figure 21C), an ANOVA found that neither $\beta_1 \cdot c$ nor $\beta_2 \cdot c^2$ significantly increased the explained data variance compared to the data set's mean $\bar{A}$. For a third order polynomial (Figure 21D), however, it was found that $\beta_3 \cdot c^3$ significantly increased the explained variance and hence -for this wavenumber- a concentration-dependent and thus predictive model function was found. It will be shown below that for a considerable percentage of the wavenumber positions significant higher order terms exist. Furthermore, calibration models for multiple predictor variables $(Q > 1)$ are shown to require coupled predictor terms as well. Analyzing chemically complex systems such as microalgae cells responding to environmental changes requires new chemometrics as conventional, linear algorithms are prone to fail.

6.2. Theory of Nonlinear Modeling

6.2.1. Expanding Linear Multivariate Least-Squares to Nonlinear Modeling

A sample's physical or chemical properties of interest are often not directly accessible and hence one has to measure a system's response to these properties. For such an indirect measurement, a calibration model $\mathbf{Y}$ is needed which mathematically describes the relation between a sample's properties of interest, the Q predictor variables $\mathbf{x}_{(Q\times 1)} = (x_1 \quad \cdots \quad x_Q)^{\mathbf{T}}$, and N measured response variables $\mathbf{y}_{(N\times 1)} = (y_1 \quad \cdots \quad y_N)^{\mathbf{T}} = \mathbf{Y}_{\boldsymbol{\beta}}(\mathbf{x}) + \boldsymbol{\delta}$. (14), [125] Often, however, the underlying chemical processes are not understood well enough to derive, based on theoretical considerations, a multivariate model function $\mathbf{Y}_{\boldsymbol{\beta}}(\mathbf{x})$. In such cases, the true but inaccessible model function can be approximated by a multivariate Taylor expansion up to polynomial order P:

$$Y_{\boldsymbol{\beta}_n}(\mathbf{x}) = \left(\beta_{n,0} \cdot 1\right)_{p=0} + \left(\sum_{q_1=1}^{Q} \beta_{n,q_1} \cdot x_{q_1}\right)_{p=1} + \left(\sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} \beta_{n,q_1\,q_2} \cdot x_{q_1} \cdot x_{q_2}\right)_{p=2}$$
$$+ \left(\sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} \sum_{q_3=q_2}^{Q} \beta_{n,q_1\,q_2\,q_3} \cdot x_{q_1} \cdot x_{q_2} \cdot x_{q_3}\right)_{p=3} + (\cdots)_{p=3} + \cdots = \boldsymbol{\beta}_{\boldsymbol{n}(1\times W)} \cdot \mathbf{x}_{(W\times 1)}$$

(14)

With $\mathbf{Y}_{\boldsymbol{\beta}}(\mathbf{x}) = (Y_{\boldsymbol{\beta}_{n=1}}(\mathbf{x}) \quad \cdots \quad Y_{\boldsymbol{\beta}_{n=N}}(\mathbf{x}))^{\mathbf{T}}$ and

$$\mathbf{x}_{(W\times 1)} = \left(\underbrace{x_0 = 1}_{p=0} \quad \underbrace{x_1 \quad \cdots \quad x_N}_{p=1} \quad \underbrace{x_1x_1 \quad x_1x_2 \quad x_1x_3 \quad \cdots \quad x_Nx_N}_{p=2} \quad \underbrace{x_1x_1x_1 \quad x_1x_1x_2 \quad \cdots \quad x_Nx_Nx_N}_{p=3} \quad \cdots\right)^{\mathbf{T}}$$

(15)

A multivariate Taylor expansion (14), (15) naturally introduces mixed predictors for modeling cross-talk among predictors (e.g. $x_{q_1} \cdot x_{q_2} \cdot x_{q_3}$ with $q_1 \neq q_2 \neq q_3$) and higher-order predictor terms (with $q_1 = q_2 = q_3$) describing nonlinear responses. For each response variable y_n, a set of W model parameters $\boldsymbol{\beta}_{\mathbf{n}(1\times W)}$ is required and hence the complete model $\mathbf{Y}_{\boldsymbol{\beta}}(\mathbf{x}) = \boldsymbol{\beta} \cdot \mathbf{x}$ contains $N \times W$ model parameters $\boldsymbol{\beta}_{(N\times W)} = [\beta_{n,w}]$ which need to be determined experimentally. In the remainder, mixed and higher-order predictors are treated as independent variables just like x_q and hence there is a total of $W \geq Q + 1$ predictors ($W = Q + 1$ for linear models). The index $q = 1, \ldots, Q$ will refer to a chemical sample property whereas $w = 1, \ldots, W$ will denote a predictor variable. Obviously, restricting equations (14) and (15) to levels $p \leq 1$ derives the standard linear model which is hence a special case of multivariate polynomial models.

6.2.2. Calibration of Predictor Surfaces

Since the nonlinearity between response and predictor variables has been built into $\mathbf{x}$ (15), equation (14) is linear in the model parameters $\beta_{n,w}$ and consequently the calibration of

Predictor Surfaces can be done via a multivariate least-squares regression; a ˆ on top of a variable will indicate a least-squares estimate of the un-hatted item:

$$\widehat{\boldsymbol{\beta}}_{(N\times W)} = \left[y_{n,k}^{\mathrm{cal}}\right] \cdot \left[x_{w,k}^{\mathrm{cal}}\right]^{\mathbf{T}} \cdot \left(\left[x_{w,k}^{\mathrm{cal}}\right] \cdot \left[x_{w,k}^{\mathrm{cal}}\right]^{\mathbf{T}}\right)^{-\mathbf{1}}$$

(16)

The only difference between a conventional multivariate least-squares fit and the calibration of Predictor Surfaces (16) is that for the latter the predictor variables $x_{1,k}^{\mathrm{cal}}, \ldots, x_{Q,k}^{\mathrm{cal}}$ of the k^{th} calibration sample need to be compiled into a vector $\mathbf{x}_{\boldsymbol{k}}^{\mathrm{cal}}{}_{(W\times 1)}$ (15) prior to using it as the k^{th} column in $\left[x_{w,k}^{\mathrm{cal}}\right]$ (16).

Figure 22 depicts for $P = 1, 2, 3$ calibration models $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x}) = \widehat{\boldsymbol{\beta}} \cdot \mathbf{x}$ describing how the carbon concentration c_{C} as a sole predictor variable $(Q = 1)$ determines the IR absorbances of *Dunaliella parva* microalgae cells at $N = 882$ response variables.

In this $Q = 1$ case, $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ is a three-way function [122]-[124] comprising of an axis for the nutrient concentration $x = c$, one axis along which the wavenumber position n increases, and lastly one to display the response variables y_n, i.e. absorbances measured at a given wavenumber position n and nutrient concentration x. In a linear system/model $(P = 1)$, the overall shape of the spectra would not change with the concentration c and the absorbances at all wavenumbers would increase linearly (Figure 22 A) - but based on Figure 21, that was not expected in this application or only as a very imprecise approximation. Quadratic $(P = 2)$ and cubic $(P = 3)$ models held a higher promise to model the data more accurately; the corresponding models $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ are depicted in Figure 22 B and Figure 22 C.

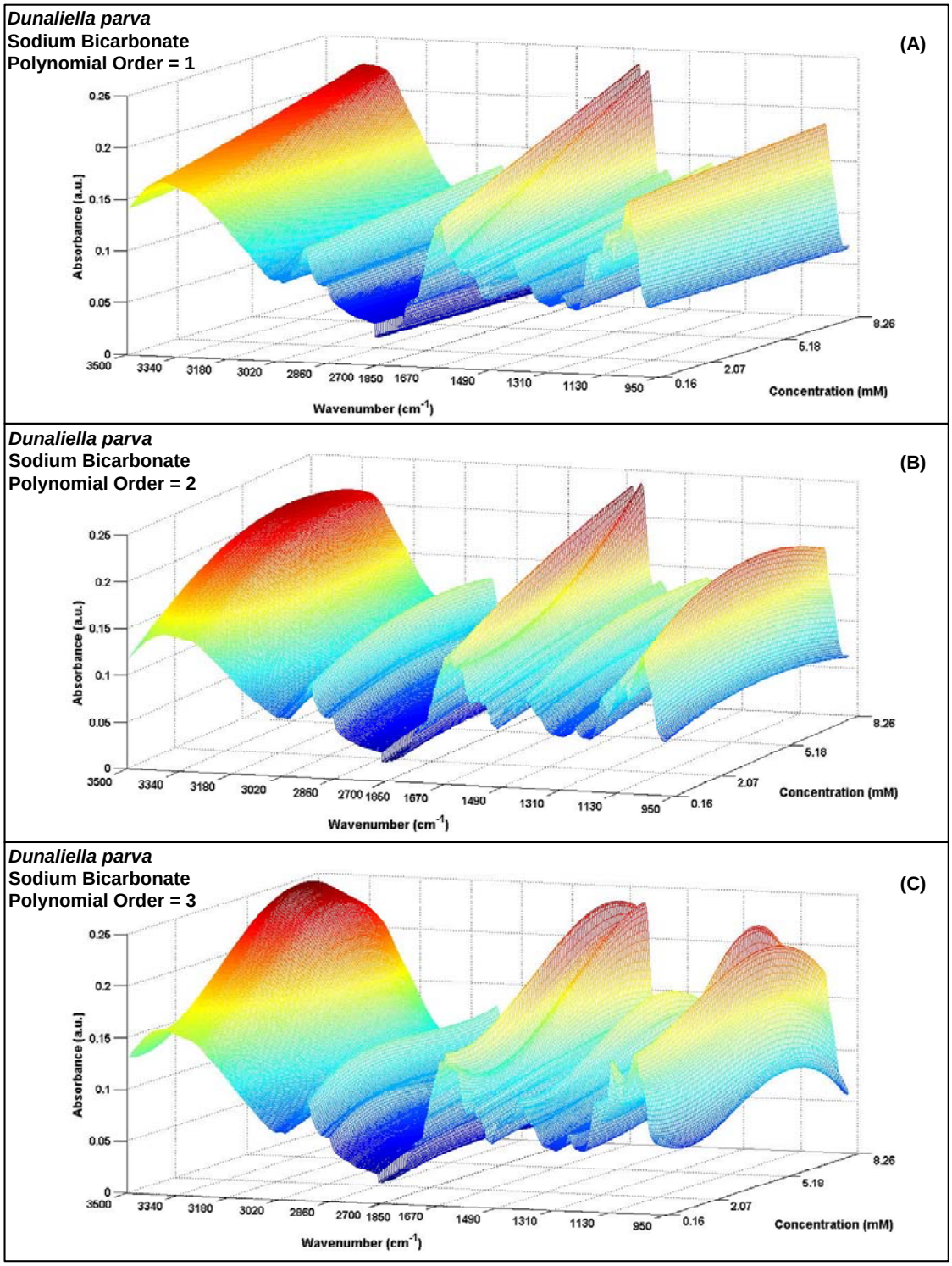


Figure 22. Nonlinear Prediction Surfaces $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ (14) showing the impacts of different polynomial orders on *Dunaliella parva* with carbon as nutrient source $(Q = 1)$: (A) linear $(P = 1)$, (B) quadratic $(P = 2)$, and (C) cubic Predictor Surfaces $(P = 3)$.

Figure 23 displays an example for a multiple predictor model $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ $(Q = 2, P = 2)$ with $x_1 = c_{\mathrm{NO_3^-}}$ and $x_2 = c_{\mathrm{C}}$; nitrate serves here as a nitrogen source. Since this is a four-way model, a graphical representation has been chosen in which one (x_1, x_2) surface per response variable $y_{n=1,\dots,N}$ is shown. Obviously, the system describes a nonlinear relation between the nutrient concentrations and the cells' chemical composition. If the samples would obey a linear model, each of the (x_1, x_2) −surfaces would be a plane linearly increasing with both x_1 and x_2 directions. Instead, the measured absorbance depends nonlinearly on the combination of available $c_{\mathrm{NO_3^-}}$ and c_{C}. For instance, at $\tilde{\nu} = 1020\mathrm{cm}^{-1}$, the absorbance considerably decreases, in a slightly nonlinear fashion with $c_{\mathrm{NO_3^-}}$ and at the same time increases with c_{C}; apparently, the production of $1020\ \mathrm{cm}^{-1}$ −absorbing compounds decreases. The production of analytes absorbing at $2980\ \mathrm{cm}^{-1}$ peaks at medium carbon and nitrate availability. Apparently, the cells thrive best when a certain nutrient *ratio* is available to them – such biological interpretations are an additional value of the Predictor Surfaces over linear algorithms and/or less intuitive data analysis methods such as neural nets. Some wavenumbers such as $1101\ \mathrm{cm}^{-1}$ feature a highly nonlinear behavior whereas wavenumbers such as $1624\ \mathrm{cm}^{-1}$ show a more intuitive nutrient dependency, i.e. increase with both nutrient concentrations albeit nonlinearly. Thus, Predictor Surfaces not only enable prediction of, in this case, nutrient concentrations (see section 6.3) in highly nonlinear chemical systems, they also facilitate fundamental investigations of such systems; these investigations can be based on visualizing models $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ (or parts thereof) or be mathematically based on (14).

6.2.3. The Significance of Nonlinear Terms

When fitting univariate $(Q = 1)$ or multivariate $(Q > 1)$ polynomial models $\mathbf{Y}_{\widehat{\beta}_n}(\mathbf{x})$ to a data set, an appropriate polynomial order P has to be selected. This is not a trivial question especially in presence of considerable fluctuations in the response variables as commonly encountered in chemical analyses of biological material. Of particular interest are situations when one needs to determine whether the response and predictor variables are correlated at all.

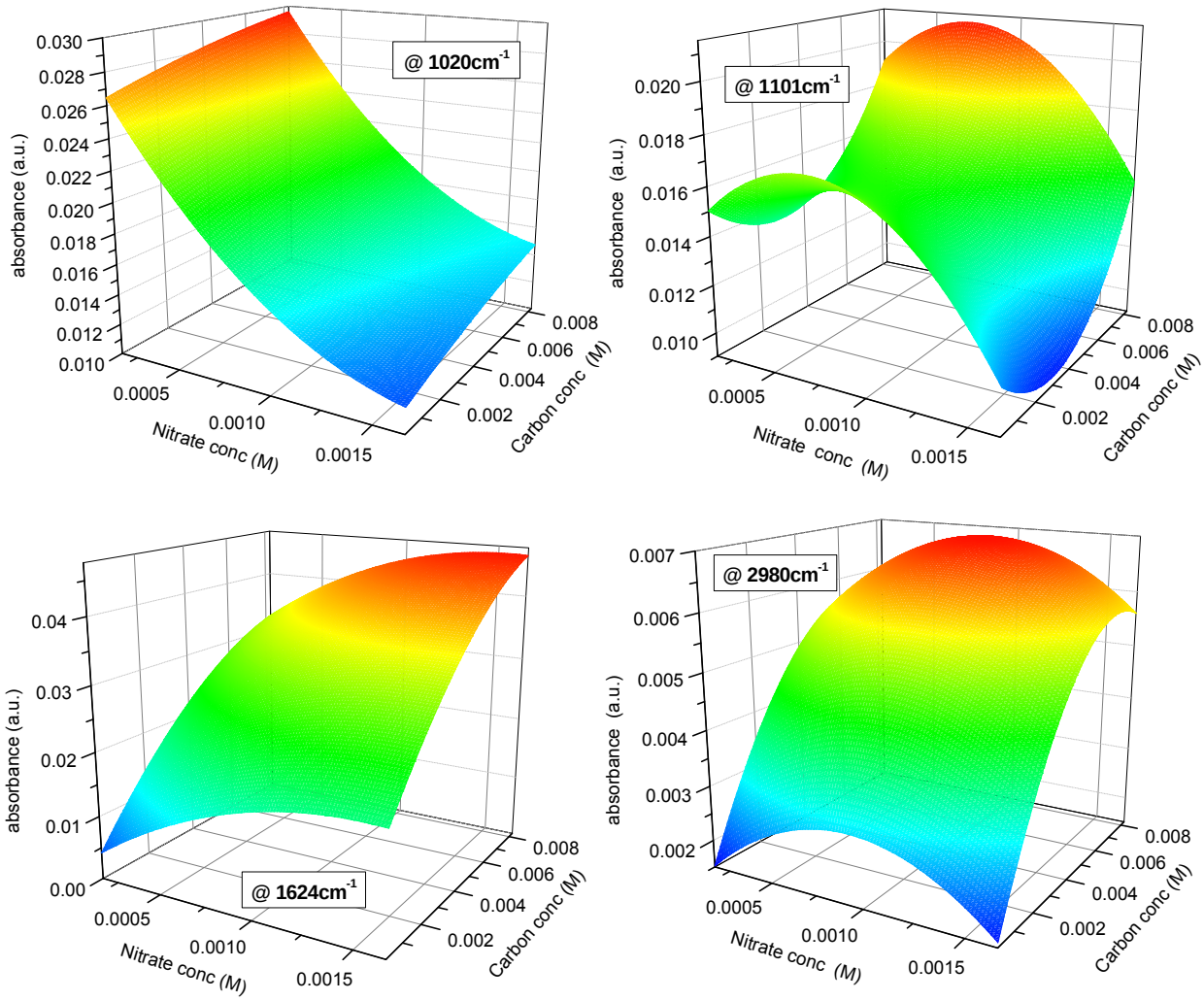


Figure 23: The calibration model $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ (14) ($P = 2$, $Q = 2$) for *Dunaliella parva* is a 4-way data hypercube, i.e. a data set spanned by the four axes: wavenumber $\tilde{\nu}$, nitrate concentration $c_{\mathrm{NO_3^-}}$, carbon concentration c_{C}, and absorbance $A\left(\tilde{\nu}, c_{\mathrm{NO_3^-}}, c_{\mathrm{C}}\right)$. Shown are four 'slices' of the 4-way model, i.e. $A\left(1020\ \mathrm{cm^{-1}}, c_{\mathrm{NO_3^-}}, c_{\mathrm{C}}\right)$, $A\left(1101\ \mathrm{cm^{-1}}, c_{\mathrm{NO_3^-}}, c_{\mathrm{C}}\right)$, $A\left(1624\ \mathrm{cm^{-1}}, c_{\mathrm{NO_3^-}}, c_{\mathrm{C}}\right)$, and $A\left(2980\ \mathrm{cm^{-1}}, c_{\mathrm{NO_3^-}}, c_{\mathrm{C}}\right)$ each of which describes how the *combination* of the carbon and nitrate concentrations in the culturing medium influences the cells' spectroscopic signature at the given wavenumber.

Such a situation was encountered for instance in Figure 21 (B) when the linear fit showed a slope ≠1 and a correlation coefficient ≠0 was obtained. However, due to the sizeable errorbars it was not certain whether this slope (gray linear fit) is actually relevant or whether the concentration-independent mean value plus minus an errorbar (red solid ± dashed lines) would equally well represent the measured data. The difference between both cases is crucial as a real slope (gray line) would establish a predictive model whereas the absence of a real slope (red line) means that no concentration dependent measurement effect was observed. For determining which of the two cases is given for a data set, Analyses of Variance (ANOVA) [125], 127] – [129] have been applied. An ANOVA calculates how much of the response variables' variation is explained by a selected portion of the model $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ such as the linear increase $\beta_{n,1} \cdot x$ in Figure 21 (B). ANOVA then *F*-tests this portion's variance against the full model's residual mean squares, i.e. the sum of squared residuals divided by the corresponding number of degrees of freedom $\|\boldsymbol{\delta}\|_2/(N-W)$. If no significant difference between the two variances is found, the test model portion is not significantly different from the residual variance and thus does not contribute in a relevant way to explaining the measured data. If, however, the two variances are found to be different, the tested predictor explains a significant, additional amount of data variance and is thus relevant for the model. In other words, the hypothesis H_0: $\beta_{n,w} = 0$ is tested, i.e. whether the model parameter w (or parameters) associated with the portion of the variance in question is equal to zero [125]. If the data support the hypothesis H_0, $\beta_{n,w}$ is not significant to the model. Since the N response variables y_n are independent of each other, this testing can be done for all N rows of $[\beta_{n,w}]_{(N\times W)}$ independently. For all *F*-tests onto which these ANOVAs are based, a significance level of 95% was chosen.

For testing the relevance of a parameter $\beta_{n,w}$, ANOVA has been combined for both, $Q = 1$ and $Q > 1$, with the 'extra-sum-of-squares principle' [125]. For the $Q = 1$ case, this extra-sum-of-squares principle starts with the minimum reduced model comprising only of $\beta_{n,0}$ and $\beta_{n,1}$ and tests whether the inclusion of $\beta_{n,1} \cdot x$ significantly increases the amount of explained data variance. If it does, $\beta_{n,1}$ is significant. This procedure is continued with each term $\beta_{n,w} \cdot x^w$ up to $w = P$, i.e. the last possible inclusion. In Figure 21 (B), for example, it was found that $\beta_{n,1} \cdot x$ for a $P = 1$ model does not contribute significantly to the observed data variance $\mathrm{var}(y_{n,k})$. This is not surprising when comparing the gray fit line to the mean plus-minus one standard deviation $\bar{y}_n \pm \sigma$ which covers about the same range on the y-axis. Another test was performed Figure 21 (C) for a $P = 2$ model and it was found that neither $\beta_{n,1} \cdot x$ nor $\beta_{n,2} \cdot x^2$ were significant

in this example. For $P = 3$ models, as depicted in Figure 21 (D), three model parameters were tested and while the first two inclusions were determined to be not significant, $\beta_{n,3} \cdot x^3$ significantly increased the amount of explained variance. As a last test, a $P = 4$ model was fitted (not shown) and an equivalent ANOVA determined that the linear, quadratic, and quartic terms were insignificant but the cubic term was found to be significant.

The statement for the $P = 3$ model that the linear and quadratic terms were not significant does not imply that they are not required. In contrary, it is obvious from the shape of the cubic polynomial which features a rather high correlation coefficient is clearly $\neq 0 \cdot x + 0 \cdot x^2 + \beta_{n,3} \cdot x^2$ or the 'down-and-up swing' could not be included. This statement has to be interpreted that in this case linear and quadratic fits are not sufficient to explain the observed data variance; however, utilizing third order polynomials is the big step towards properly modeling the measurement results. One has to keep in mind that $\beta_{n,2}$ as obtained for a quadratic polynomial has a different value than $\beta_{n,2}$ as calculated in the cubic fit. For further elucidation, the numbers for the $P = 4$ ANOVA are given here: All four contributions to the model's variance are *F*-tested against $\|\boldsymbol{\delta}\|_2/(N - W) = 1.549 \cdot 10^{-4}$; the mean squares contributed by the individual predictor terms $\beta_{n,w} \cdot x^w$ are: $w = 1$: $1.300 \cdot 10^{-4}$, $w = 1$: $2.707 \cdot 10^{-4}$, $w = 3$: $106.120 \cdot 10^{-4}$, and $w = 4$: $1.224 \cdot 10^{-4}$. Obviously, the $w = 3$ contribution is considerably larger than that of the other terms and more importantly much larger than the mean square residual. The contributions from $w = 1,2,4$ are in the same ballpark as the mean squared residual and thus not relevant to explain the data. However, the corresponding powers x^w are required to compute the proper cubic fit polynomial.

Table 4 concludes such ANOVA analyses for Predictor Surfaces with $P = 1,2,3$ computed for *Dunaliella parva* exposed to different carbon, nitrate, and ammonium levels ($Q = 1$ each). Comparing the significance results for $P = 1$ models reveals that only for nitrate linear models were found to be relevant at the majority of the wavenumber positions; for carbon at only 38% of the wavenumber positions and for ammonium at 2% of the wavenumber positions, a linear term was found to contribute in a relevant way to explaining the measured data. Quadratic models $(P = 2)$ considerably improve the amount of explained variance for all nutrients. Especially for ammonium, quadratic terms are most important.

Nutrient	Polynomial Order P	% of wavenumbers with significant *linear* term (β_1 in (14))	% of wavenumbers with significant *quadratic* term (β_2 in (14))	% of wavenumbers with significant *cubic* term (β_3 in (14))
Carbon	1	38%	---	---
	2	40%	41%	---
	3	75%	24%	48%
Nitrate	1	83%	---	---
	2	70%	35%	---
	3	90%	26%	95%
Ammonium	1	2%	---	---
	2	14%	100%	---
	3	22%	100%	61%

Table 4. Results of ANOVA analyses of Predictor Surfaces $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ with $P = 1,2,3$ for three single nutrients $(Q = 1)$ provided to *Dunaliella parva*. For each nutrient and for each polynomial order P, the percentage of wavenumbers $(\%N)$ at which the linear, quadratic, or cubic terms were found to be significant are given (there are no quadratic terms for $P = 1$ or cubic terms for $P = 1,2$).

At a considerable number of wavenumber positions, terms of cubic models ($P = 3$) were found to explain significant amounts of the variability in the response variables. How this translates into accuracy and precision of concentration predictions will be discussed in the remainder. In Table 5, ANOVA results for two $Q = 2$ applications have been compiled. Linear terms for both nutrient concentrations ($x_1 = c_{\mathrm{C}}$ as well as $x_2 = c_{\mathrm{NH_4^+}}$ and $x_2 = c_{\mathrm{NO_3^-}}$, respectively) were found to be significant at almost all wavenumber positions n. Equally interesting is the fact that for all n the coupling between both nutrients is relevant to properly model the data. Obviously, spectroscopic data obtained from the microalgae cells are strongly influenced by coupling effects between carbon and nitrogen source. In both applications, quadratic terms in the predictors are less important but relevant for a considerable number of response variables.

Overall, such a statistical model assessment is of great value to determine the significance of nonlinear and coupled terms in particular for applications which encounter limited data reproducibility. These chemometric tools thus open new perspectives in bioanalytical chemistry of biological materials.

6.3. Nonlinear Prediction

In order to predict $\mathbf{x}^{\mathrm{unk}}$, an unknown response vector $\mathbf{y}^{\mathrm{unk}}$ is fitted to a calibration model $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$. For a $Q = 1$ situation, this fitting procedure can be interpreted as an unknown's spectrum being moved along the concentration axis, i.e. in parallel to the wavenumber axis, until the deviation between the model vector $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x}^{\mathrm{unk}})$ and measured spectrum (response vector) $\mathbf{y}^{\mathrm{unk}} = \widehat{\boldsymbol{\beta}} \cdot \mathbf{x}^{\mathrm{unk}} + \boldsymbol{\delta}$ is minimal. Motivated by this graphical interpretation, the term '*Predictor Surface*' has been devised for calibration models $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$.

Technically, a solution for $\mathbf{x}^{\mathrm{unk}}$ could be estimated via another linear multivariate least-squares regression (MLR), i.e.: $\hat{\mathbf{x}}^{\mathrm{unk}} = \left(\widehat{\boldsymbol{\beta}}^{\mathbf{T}} \cdot \widehat{\boldsymbol{\beta}}\right)^{-\mathbf{1}} \cdot \widehat{\boldsymbol{\beta}}^{\mathbf{T}} \cdot \mathbf{y}^{\mathrm{unk}}$ followed by extracting $\hat{x}_1^{\mathrm{unk}}, \cdots, \hat{x}_Q^{\mathrm{unk}}$ from the resulting $\hat{\mathbf{x}}^{\mathrm{unk}}$ (15). However, this should *not* be utilized because it would consider each of the elements in $\hat{\mathbf{x}}^{\mathrm{unk}}_{(W\times 1)}$ (15) as an independent predictor variable.

nutrient mix	% of N with significant term $\boldsymbol{\beta_1 \cdot x_1}$	% of N with significant term $\boldsymbol{\beta_2 \cdot x_2}$	% of N with significant term $\boldsymbol{\beta_{1\,1} \cdot x_1^2}$	% of N with significant term $\boldsymbol{\beta_{1\,2} \cdot x_1 \cdot x_2}$	% of N with significant term $\boldsymbol{\beta_{2\,2} \cdot x_2^2}$
C & NH_4	96%	57%	17%	100%	3%
C & NO_3	97%	43%	26%	100%	20%

Table 5: Significance of model parameters $\hat{\beta}_{n,w}$ with $P = 2$ for prediction of the two stated nutrients in mixture $(Q = 2)$; the percentages refer to $\%N$.

This is clearly not the case since $\mathbf{x}^{\mathrm{unk}}$ contains $W > Q$ elements which are various products of the predictor variables $x_{q=1,\dots,Q}$ one needs to determine. For the data sets analyzed in this study, it was found that $\hat{\mathbf{x}}^{\mathrm{unk}}$ obtained by means of this MLR step reconstructs $\mathbf{y}^{\mathrm{unk}}$ well via $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\hat{\mathbf{x}}^{\mathrm{unk}})$. However, the values for $x_{q=1,\dots,Q}$ were found to be several orders of magnitude too large; furthermore, higher-order/mixed predictor terms were not the expected products such as $\left(x_1^{\mathrm{unk}\,2}\right)_{p=2} \neq \left(x_1^{\mathrm{unk}}\right)_{p=1} \cdot \left(x_1^{\mathrm{unk}}\right)_{p=1}$. Overall, this failure of the MLR approach is not surprising as the solution is searched for in a W −dimensional space whereas the true solution has to be found in a Q −dimensional space with $Q < W$ or even $Q \ll W$. For such a higher-dimensional solution space, many more solutions are feasible which, however, often have no chemical meaning.

Thus, $\hat{\mathbf{x}}^{\mathrm{unk}}$ is derived via nonlinear least-squares regression which is based on minimizing the sum of squared errors

$$\mathrm{SSE}\left(\hat{x}_1^{\mathrm{unk}}, \dots, \hat{x}_Q^{\mathrm{unk}}\right) = \|\boldsymbol{\delta}\|_2^2 = \left\|\mathbf{y}^{\mathrm{unk}} - \mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\hat{\mathbf{x}}^{\mathrm{unk}})\right\|_2^2 = \sum_{n=1}^{N}\left[y_n^{\mathrm{unk}} - Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})\right]^2$$

(17)

between model $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\hat{\mathbf{x}}^{\mathrm{unk}})$ and measured data $\mathbf{y}^{\mathrm{unk}}$. A minimum of the function is characterized by the necessary condition:

$$\boldsymbol{\nabla}\mathrm{SSE}\left(\hat{x}_1^{\mathrm{unk}}, \dots, \hat{x}_Q^{\mathrm{unk}}\right) = \mathbf{0} = -2 \cdot \sum_{n=1}^{N}\left[y_n^{\mathrm{unk}} - Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})\right] \cdot \boldsymbol{\nabla} Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})$$

(18)

The second equal sign in (18) holds because the y_n^{unk} are measured numbers, i.e. constants, and not functions of $\hat{x}_q^{\mathrm{unk}}$; thus $\boldsymbol{\nabla}\left(y_n^{\mathrm{unk}} - Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})\right) = \boldsymbol{\nabla} y_n^{\mathrm{unk}} - \boldsymbol{\nabla} Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})$ simplifies to $-\boldsymbol{\nabla} Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}(\hat{\mathbf{x}}^{\mathrm{unk}})$.

Below, the prediction steps for the single-predictor case $(Q = 1)$ and for multiple predictors $(Q > 1)$ are derived separately as they are quite different from a software perspective: The former requires finding the roots of a polynomial for which a solution is guaranteed and easy to compute. For $Q > 1$, equation (18) is a system of Q nonlinear equations for which iterative

solvers are available [89], [126]. However, available algorithms cannot guarantee finding a global minimum of $\mathrm{SSE}\left(\hat{x}_1^{\mathrm{unk}},\dots,\hat{x}_Q^{\mathrm{unk}}\right)$ (17) as opposed to finding a local minimum and furthermore the iteration requires the user to supply an initial guess for $\hat{\mathbf{x}}^{\mathrm{unk}}$ which may not always be obvious. For this study, normal distributed random numbers with mean zero and standard deviation of one have been chosen as initialization for $\hat{\mathbf{x}}^{\mathrm{unk}}$.

6.3.1. Single Predictor Variable

For polynomial models (14) with $Q = 1$, equation (17) simplifies to a univariate polynomial of order $2 \cdot P$, i.e. $\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right) = \sum_{n=1}^{N}\left[y_n^{\mathrm{unk}} - \left(\hat{\beta}_{n,0} + \hat{\beta}_{n,1}\cdot\left(\hat{x}^{\mathrm{unk}}\right) + \hat{\beta}_{n,2}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^2 + \cdots + \hat{\beta}_{n,P}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^P\right)\right]^2$. The following redefinition of the polynomial's coefficients $\hat{\beta}_{n,p}$ will make subsequent notations more concise: $b_{n,0} = y_n^{\mathrm{unk}} - \hat{\beta}_{n,0}$ and $b_{n,p} = -\hat{\beta}_{n,p}$ for $p \geq 1$. Furthermore, the gradient operator in (18) is reduced to the univariate differentiation $\frac{d}{d\hat{x}^{\mathrm{unk}}}\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)$. After applying the chain rule of differentiation, the summation over n has been made the inner-most loop in the second step of equation (19) resulting in a polynomial in $\hat{x}^{\mathrm{unk}}$ of order $2P-1$:

$$\frac{d\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)}{d\hat{x}^{\mathrm{unk}}} = 0 = -2\cdot\sum_{n=1}^{N}\left[\sum_{p_1=0}^{P} b_{n,p_1}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^{p_1}\right]\cdot\left[\sum_{p_2=1}^{P} p_2\cdot b_{n,p_2}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^{p_2-1}\right]$$
$$= \sum_{p_1=0}^{P}\sum_{p_2=1}^{P}\underbrace{\sum_{n=1}^{N} b_{n,p_1}\cdot p_2\cdot b_{n,p_2}}_{=B_{p_1,p_2}}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^{p_1+p_2-1} = \sum_{p_1=0}^{P}\sum_{p_2=1}^{P} B_{p_1,p_2}\cdot\left(\hat{x}^{\mathrm{unk}}\right)^{p_1+p_2-1} \qquad (19)$$

The $2P-1$ roots of the polynomial (19) minimize $\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)$ and are thus candidates for the wanted $\hat{x}^{\mathrm{unk}}$. Some considerations are required to select the most appropriate of the $2P-1$ roots of (19) though. 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 disregarded as well. Furthermore, those $\hat{x}_{\#p}^{\mathrm{unk}}$ have been removed from the list of possible solutions of (17) for which $\left.\frac{d^2\mathrm{SSE}\left(\hat{x}^{\mathrm{unk}}\right)}{d\hat{x}^{\mathrm{unk}^2}}\right|_{\hat{x}_{\#p}^{\mathrm{unk}}} < 0$ because they are (local) maxima of the

SSE. To pick the best among the remaining roots, the unknown's response vector $\mathbf{y}^{\mathrm{unk}}$ is 'reconstructed' from root $\#p$ resulting in a vector $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}\left(\hat{x}_{\#p}^{\mathrm{unk}}\right)$. The correlation coefficient $\mathrm{corr}\left(\mathbf{Y}_{\hat{\boldsymbol{\beta}}}\left(\hat{x}_{\#p}^{\mathrm{unk}}\right), \mathbf{y}^{\mathrm{unk}}\right)$ between reconstructed and originally measured response vector is used as a metric to find the reconstruction that most closely resembles the measured data. The corresponding $\hat{x}_{\#p}^{\mathrm{unk}}$ is then used as analysis outcome; additional constraints can be applied to assess the reasonability of the found solution such as requiring a minimum correlation coefficient (applied here: $r > 0.90$).

6.3.2. Multiple Predictor Variables

When incorporating multiple analytes $(Q > 1)$, a system of Q nonlinear equations is derived from (18) whose solution are the wanted $\hat{x}_{q=1,\ldots,Q}^{\mathrm{unk}}$. Deriving an equation for the required gradient $\boldsymbol{\nabla} Y_{\hat{\boldsymbol{\beta}}_{\mathbf{n}}}\left(\hat{\mathbf{x}}^{\mathrm{unk}}\right)$ for multivariate polynomial models (14) is somewhat tedious and has thus been moved into Appendix 3; see equation (25). Algorithms for iteratively solving nonlinear equation systems of type $\mathbf{f}\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right) = \begin{pmatrix} f_1\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right) \\ \vdots \\ f_1\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right) \end{pmatrix} = \mathbf{0}$ were presented, for instance, in references [89], [126]. Here, these algorithms are utilized to solve $\mathbf{f}\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right) = \boldsymbol{\nabla}\mathrm{SSE}\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right) = \mathbf{0}$ (18), i.e. to find $\hat{\mathbf{x}}^{\mathrm{unk}}$ that minimizes $\mathrm{SSE}\left(\hat{\mathbf{x}}^{\mathrm{unk}}\right)$ (17). For initializing this iteration, an initial guess of $\hat{\mathbf{x}}^{\mathrm{unk}}$ is required; here, normal distributed values with zero mean and a standard deviation of one have been utilized.

6.4. Experimental

As an experimental test bed for the proposed nonlinear Predictor Surfaces, spectroscopic analyses of microalgae have been chosen. The goal was to establish data models which enable the prediction of selected environmental parameters from the cells' spectroscopic signatures. Starting cultures for three sea water species, i.e. *Dunaliella salina, Dunaliella parva*, and *Nannochloropsis oculata*, were obtained from The Culture Collection of Algae at the University of Texas, Austin. Cultures were inoculated in 'enriched seawater, artificial water' medium (ESAW, pH 8.2) [14], [68], [69] and exposed to continuous illumination while maintained at 20°C. In order to simulate changing environmental conditions, ≥ 5 independent replicate cultures per condition were grown under different concentrations of two important algae nutrients, i.e. inorganic carbon and nitrogen. Inorganic carbon concentrations were adjusted via dissolving different amounts of sodium bicarbonate ($NaHCO_3$) in the growing medium; as nitrogen sources, sodium nitrate ($NaNO_3$) and alternatively ammonium chloride (NH_4Cl) have been dissolved in ESAW at different concentrations. In order to generate starving, normal, and excess situations, the following carbon concentrations were realized for these studies: 0.16 mM, 1.11 mM, 2.07 mM (normal condition), 3.63 mM, 5.18 mM, 6.72 mM, 8.26 mM; as nitrogen concentrations, 0.16 mM, 0.35 mM, 0.549 mM (normal condition), 0.873 mM, 1.28 mM, 1.47 mM, and 1.65 mM were chosen. In order to ensure high reproducibility among replicate cell cultures, care was taken that all parameters other than intentionally varied nutrient concentrations, i.e. light level, temperature, etc., remained as constant as experimentally possible. ESAW contains a number of additional, minor nutrients, which were all composed from the same stocks in order to minimize fluctuations in the chemical composition of the growing medium. With changing concentrations of the major nutrients, however, the concentration of Na^+, HCO_3^-, NH_4^+ and/or Cl^- ions in the culture medium changed which might have had an indirect impact on the pH. Therefore, the pH was recorded at harvest to record any variation due to the nutrient concentrations. For *Dunaliella parva* cultures grown under different nitrate concentrations for example, the pH averaged over all samples remained in the range 8.04 ± 0.14; the average pH over all concentrations of sodium bicarbonate and ammonium chloride yielded very similar results, i.e. 8.13 ± 0.20 and 7.93 ± 0.10, respectively. Comparable pH fluctuations were observed for the other two microalgae species and considered negligible in this study.

Cell concentrations of each algae culture were determined via direct counts using a hemocytometer. Guided by the resulting sigmoidal growth curves, cultures were harvested while still in the exponential growth phase such that a sufficient yield of 'acclimated' cells was achieved. For harvest, 1 µL of Lugol's solution (Sigma-Aldrich) was added per 1 mL of algal suspension in order to fix the cells in their current state. Cells were then extracted from their solutions through centrifugation (4400 rpm) and afterwards washed twice with an isoosmotic solution (0.1 M) of ammonium formate (Alfa Aesar) to minimize medium carryover. Finally, after gently drying (4-5 days at 60˚C), the algae material was mixed with IR-transparent KBr powder (0.6 weight % of algae) and pressed into a pellet. Given the biological nature of the samples, a considerable spectroscopic fluctuation in replicate samples was expected and found. In order to incorporate these unavoidable replicate-to-replicate fluctuations into calibration models, at least five independent replicate cultures were grown; from each of these cultures, three algae-KBr pellets were prepared and spectroscopically analyzed. From such pellets, FT-IR transmission spectra were recorded [51] covering the wavenumber range 3500 – 950 cm^{-1}; for subsequent chemometric data analyses, the region 2700 – 1850 cm^{-1} was excluded for $Q = 1$ situations as it did not contain relevant information but rather fluctuations in atmospheric CO_2 concentrations. For $Q = 2$ experiments, the wavenumber range was further restricted to 3500 – 2900 cm^{-1} combined with 1620 – 980 cm^{-1} as the excluded wavenumber regions in particular 1750 – 1620 cm^{-1} may be important for biochemical studies but introduced strong replicate-to-replicate fluctuations which interfered with the concentration predictions.

During preliminary investigations it was found that these calibration models for multiple analytes ($Q > 1$, cp. Figure 23) are more sensitive to data artifacts than models with $Q = 1$. While baseline preprocessing was not required for single analyte data sets, a Fourier highpass filter was found to remove baseline shifts [93], 130] and thus to improve the models accuracy for the prediction results shown below for two $Q = 2$ cases. This highpass filter's parameters were chosen such that only a constant offset (constant over wavenumber) was removed from the spectra as well as spectral features very broad in wavenumber; this filter is not to be confused with removing IR frequencies. Further prediction improvements could be achieved by dividing each spectrum by the mass of IR absorbing algae material contained in that particular KBr-algae pellet. This step normalized for the mass contained in a sample and thus reduced spectroscopic fluctuation from replicate to replicate which are induced by imperfections in manual sample preparation and the amount in the samples [51].

6.5. Results and Discussion

The relevance of polynomial Predictor Surfaces with $P > 1$ has been demonstrated by means of Figure 22, Figure 23, and section 6.2.3. For microalgae cells to be feasible as environmental bioprobes, different environmental parameters must have discriminable impacts on each species' chemical signature and different species must be discriminable when grown under the same conditions. Otherwise, it would be uncertain whether two spectra are different because they were acquired from different species which were exposed to the same ambient conditions or whether the data were obtained from the same species but the environmental parameters changed between measurements. In order to assess the cells' selectivity for different stimuli such as nutrient concentrations, Predictor Surfaces $(P = 2, Q = 1)$ were computed for *Dunaliella parva* for all three nutrients incorporated into this study, i.e. carbon, nitrate, and ammonium. From comparing Figure 22 (B), Figure 24 (A), and Figure 24 (B) it is obvious that the Predictor Surfaces not only change considerably with each nutrient's concentration but are also defined by the nutrient type. The spectroscopic changes induced by carbon are much more distinct than among the two nitrogen sources but nonetheless, sufficient differences are imposed on the Predictor Surfaces by nitrate versus ammonium to be clearly visible (cp. Figure 24 A versus B). Since more than one microalgae species are expected in environmental samples, it was investigated whether different species can be discriminated when exposed to the same ambient conditions.

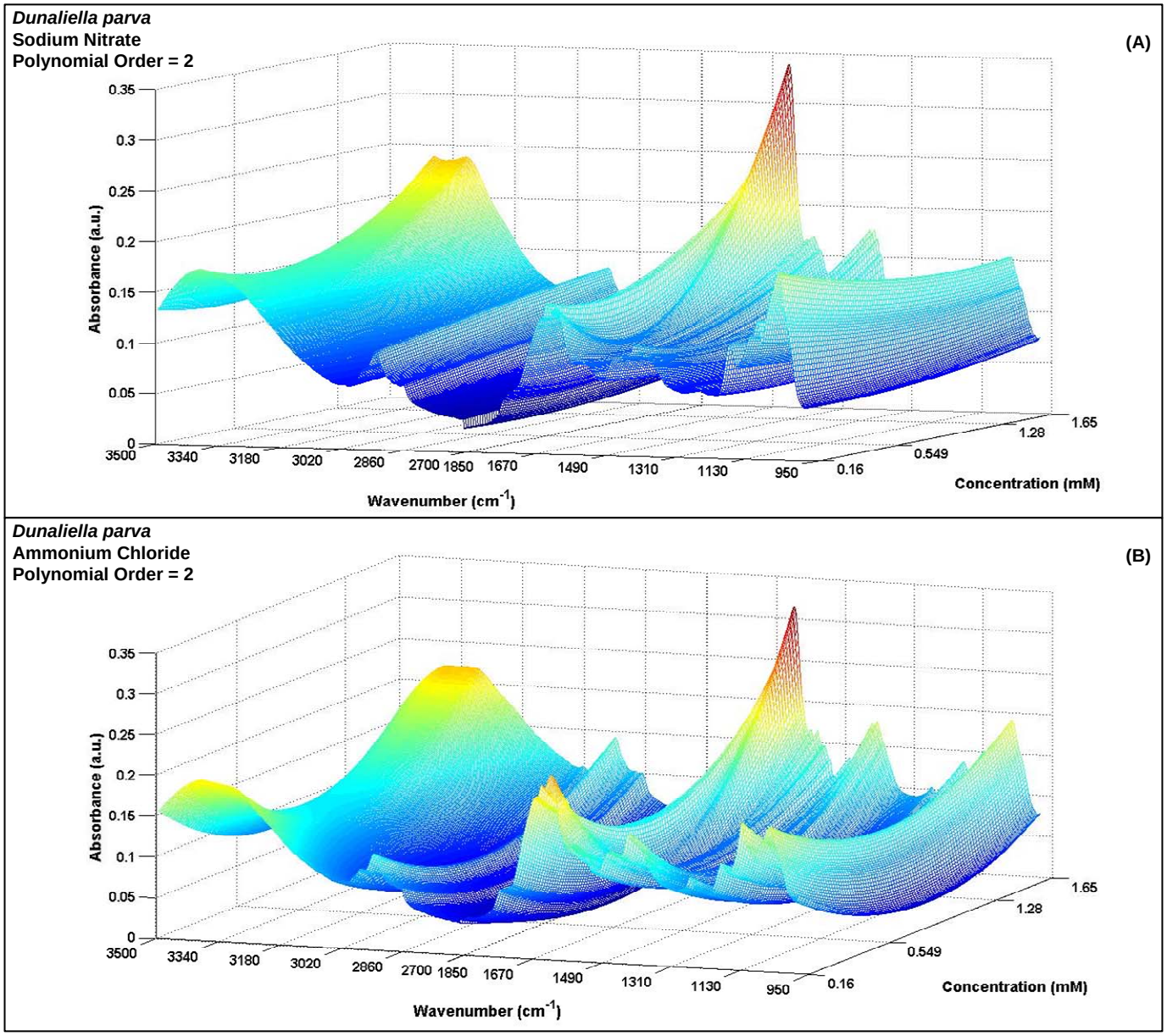


Figure 24. Nonlinear Prediction Surfaces $\mathbf{Y}_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ (1) $(P = 2, Q = 1)$ for different nutrient sources supplied to *Dunaliella parva* visualize spectrochemical impacts of the (A) nitrate and (B) ammonium concentrations, respectively.

For this purpose, Predictor Surfaces for *Dunaliella parva*, *Dunaliella salina*, and *Nannochloropsis oculata* were determined for different carbon concentrations which show considerably different features (Figure 22 (B), Figure 25 (A) and (B)) and thus it can be concluded that species discrimination is possible.

As a proof-of-principle for the quantitation of environmental parameters via the microalgaes' chemical signatures, the concentration of inorganic carbon as well as two nitrogen sources (nitrate and ammonium ions) had been chosen which are important cell nutrients. For the subsequent discussion, one has to keep in mind that these concentrations do *not* refer to analytes contained in the algae themselves; the nutrients were only available in the culturing medium during the cells' growth and impacted during that time the chemical composition of the cells' biomass. The nonlinearity in data and the coupling among environmental parameters (predictor variables) originate from this experimental setup. This arrangement had been devised to enable *in-situ* analyses of multiple environmental parameters and to measure ecological impacts of chemically changing environments via changing trophic values of microalgal biomass. Lastly, this novel analytical methodology will also enable bioanalytical investigations of the microalgae's interactions with their environments.

For comparing the accuracy and precision of the novel, nonlinear Predictor Surfaces to a standard linear regression method, Principal Component Regression (PCR) [100] – [103] was chosen. While PCR per se is a linear algorithm, it has been shown in a different context [106] that nonlinear effects such as band broadening in UV spectroscopy and slightly nonlinear absorbances can be modeled; in another application [92], PCR was capable to empirically approximate baseline fluctuations in spectra by a set of principal components whose number exceeded the number of analytes present. In this respect, PCR is often superior to MLR and in many applications PCR or PLS have a sufficient accuracy and precision; however, their lack of model interpretability limits the insight into chemically complex systems they can gain. Choosing the appropriate number of principal components has here been based on limiting the condition number of $\left[y_{n,k}^{\mathrm{cal}}\right]$, i.e. the ratio of the largest singular values [89], [126] divided by the smallest one to be accepted. It was found here that a threshold of 1000 retained too much noise and resulted in a rather low accuracy of the predicted nutrient concentrations. Reducing this ratio to 100 improved the predicted concentration considerably and this value has thus been applied in the remainder.

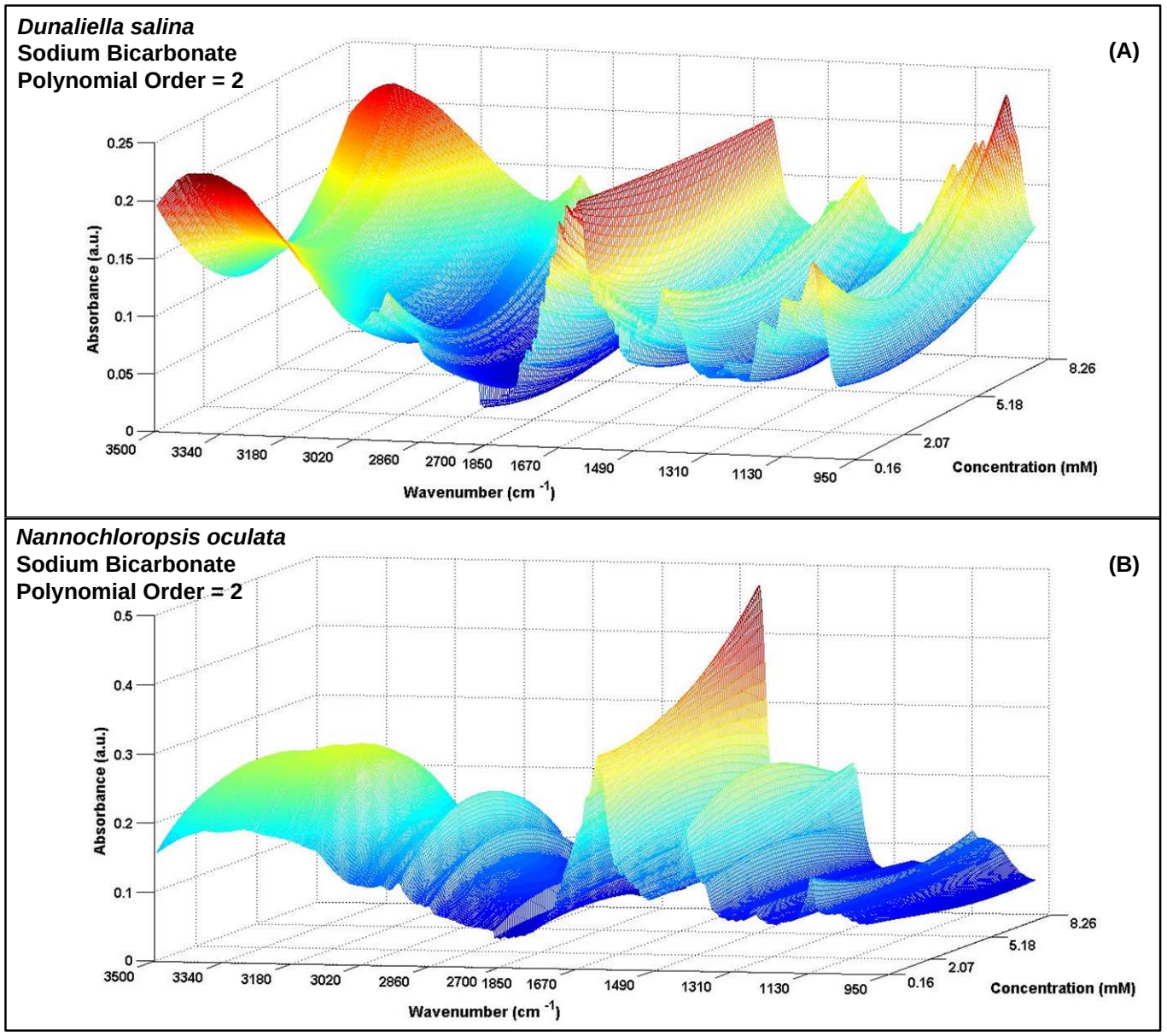


Figure 25. The considerable differences in the concentration dependency of the Predictor Surfaces $\mathbf{Y}_{\widehat{\boldsymbol{\beta}}}(\mathbf{x})$ (14) $(P = 2, Q = 1)$ (A = *Dunaliella salina*) and (B = *Nannochloropsis oculata*) demonstrate that cells of different species undergo different chemical processes depending on the carbon concentration in the culturing medium.

Depending on the microalgae species and nutrient, this threshold translated into six to ten relevant principal components for experiments with one nutrient $(Q = 1)$ and into thirteen principal components employed for $Q = 2$ tasks; hence, a considerable number of non-noise principal components were available for modeling nonlinear effects in the data.

For all quantitative experiments, two set of samples have been prepared – one for calibration and one for assessing this method's prediction power. Both sets comprised of completely independent samples, i.e. different concentrations were used in the calibration set compared to the test sets rather than using some replicate samples for calibration and the remaining for testing the prediction. Therefore, the data modeling is assessed rather than the reproducibility of the sample preparation. From the above list of nutrient concentrations, the following were used for calibration of single nutrient Predictor Surfaces $(Q = 1)$ in which only the concentration of one nutrient was varied whereas all other nutrients were kept at the standard ESAW concentrations [14]. For calibration purposes, bicarbonate at concentrations of 0.16 mM, 2.07 mM, 5.18 mM, and 8.26 mM were used; bicarbonate concentrations of 1.11 mM, 3.63 mM, and 6.72 mM were used in the test set. For building Predictor Surfaces for either nitrogen source, the following concentrations were used: 0.16 mM, 0.549 mM, 1.28 mM, and 1.65 mM; the concentrations 0.35 mM, 0.873 mM, and 1.47 mM served for testing purposes. With four different values for the predictor variable incorporated into the calibration sets, $P = 3$ is the highest polynomial order (14) that should be realized even with $K > P$. Three Predictor Surfaces were built for $P = 1, 2,$ and 3 all of which were based on the same spectra; the same calibration data were also used for the comparison method PCR. Figure 26 (A-C) depicts predicted versus true bicarbonate concentrations. The blue lines in these graphs indicate where predicted equals the true concentrations; straight fit curves shown as red dashed lines measure how close overall the predicted concentrations are to the predicted ones. Obviously, the linear, $P = 1$ Predictor Surface, i.e. a standard MLR, is incapable of reliably predicting the carbon level in the cells' environment; this failure is assigned to the inappropriateness of linear models for these nonlinear data (cp. Figure 21 (A) and Figure 22 (A)). PCR performed somewhat better (Figure 26 D) – apparently, the number of incorporated principal components being $> Q = 1$ helped to empirically compensate for nonlinear relations between predictor and response variables.

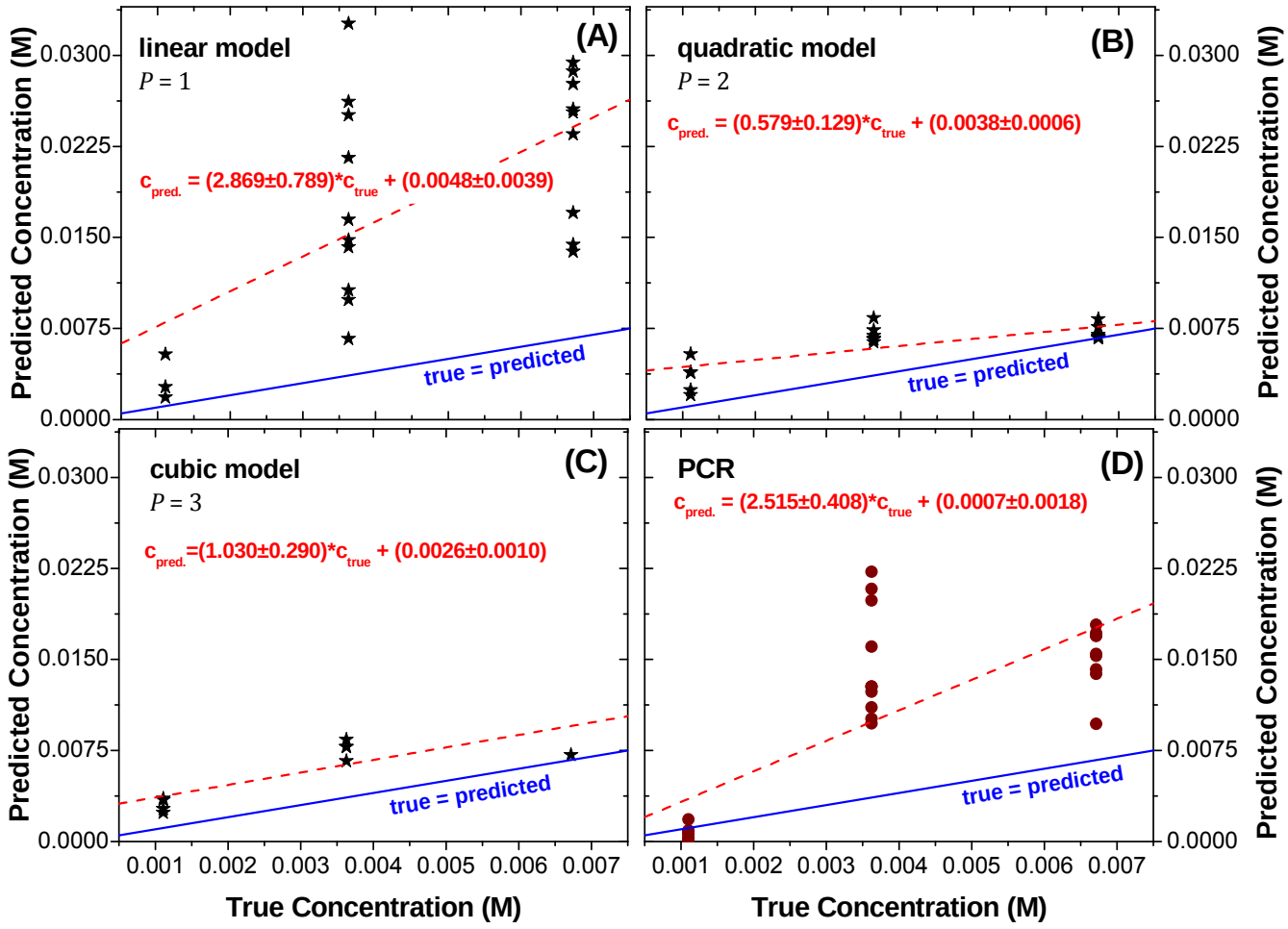


Figure 26. Demonstrating the impacts of the polynomial order P on the concentrations predicted by nonlinear predictor surfaces: (A) $P = 1$, (B) $P = 2$, (C) $P = 3$ (cp. Table 4); (D) predictions by Principal Component Regression (PCR) for comparison; compare Table 6 for results on nitrate and ammonium.

When expanding the Predictor Surface models to quadratic ($P = 2$) and cubic models ($P = 3$), the prediction quality considerably improves, with $P = 3$ being most precise and reproducible. In Figure 27, the prediction results for nitrate (top row) and ammonium (bottom row) are concluded and compared to PCR's performance. For both nitrogen sources, $P = 2$ resulted in the best predictions. However, only the nitrate predictions are meaningful as neither Predictor Surfaces nor PCR could derive a fit curve for ammonium whose slope was found to be significant by an ANOVA. For nitrate predictions, PCR achieved a higher reproducibility than the $P = 2$ Predictor Surface but PCR's accuracy as measured by the fits' slopes of $\beta_1^{\mathrm{PCR}} = 0.411$ is more than a factor of two worse than the accuracy obtained by a $P = 2$ Predictor Surface of $\beta_1^{\mathrm{Pred.Surf.}} = 0.901$ which compares well to $\beta_1^{\mathrm{ideal}} = 1.0$ given the nature of the samples.

Table 6 concludes these findings for all three analytes and polynomial orders: Carbon and nitrate can be quantified via nonlinear Predictor Surfaces whereas standard PCR fails because it either highly over- or under-estimated the true concentrations. For ammonium though, Predictor Surfaces and PCR fail to derive a meaningful prediction model as indicated by an insignificant linear fit of predicted versus true concentrations.

In order to demonstrate the capabilities of Predictor Surface for predicting multiple environmental parameters, microalgae species were exposed to mixtures of two different nutrients at varying concentrations ($Q = 2$). Two situations were analyzed, i.e. mixtures of carbon and nitrate as well as carbon and ammonium. As calibration sets, carbon concentrations of 0.16 mM, 1.11 mM, 2.071 mM, 3.63 mM, and 6.72 mM were utilized in various combinations of nitrogen concentrations, i.e. 0.16 mM, 0.549 mM, 0.873 mM,1.28 mM, 1.47 mM, and 1.65 mM; nitrogen was provided at these concentrations via ammonium and via nitrate but in separate cultures. As prediction samples served combinations of 0.16 mM, 5.18 mM, and 8.26 mM carbon with 0.16 mM, 0.549 mM, and 1.28 mM of either via nitrate or ammonium. Figure 28 concludes the prediction results for carbon and nitrate mixtures; Figure 29 depicts the results for carbon and ammonium mixtures. In preliminary tests, it was found, that in this particular application the precision and accuracy of $Q = 2$ Predictor Surfaces were very sensitive towards including certain, limited wavenumber regions.

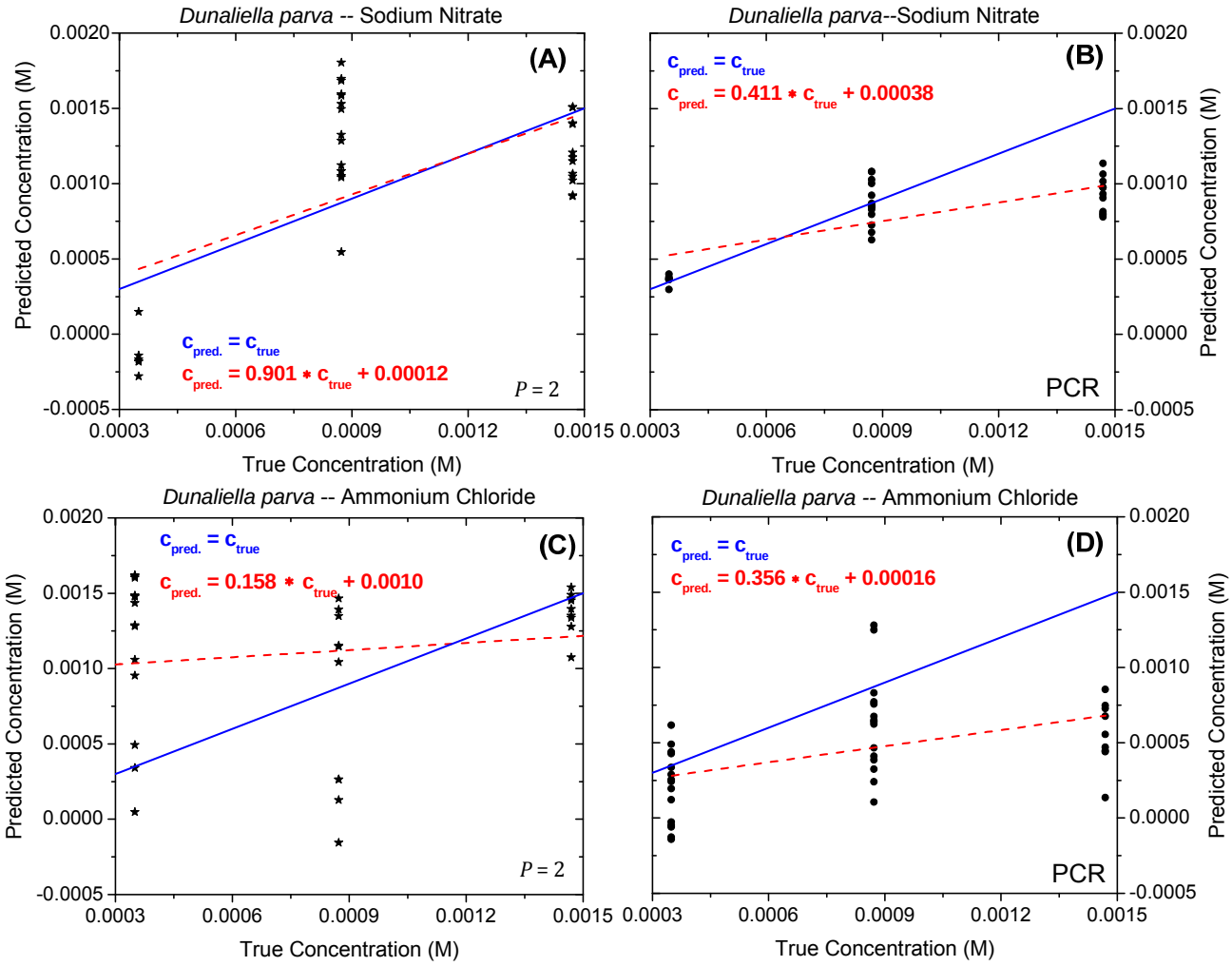


Figure 27. Concentration predictions for nonlinear modeling ($P = 2$, cp. Table 4) and PCR for two different nitrogen sources (two separate $P = 2, Q = 1$ cases): (A, B) sodium nitrate; (C, D) ammonium chloride. As a reference, the blue line indicates where predicted and true concentrations are equal; the dashed red line is a least-squares fit which in the ideal case would have a slope of one and a zero intercept.

Nutrient	Polynomial Order P	Slope (expected = 1)	Intercept (expected = 0)
Carbon	1	2.86899	0.0048
	2	0.579386	0.0037
	3	1.03026	0.0026
	PCR	2.51542	0.00074
Nitrate	1	-0.114895	0.00121
	2	0.900594	0.00012
	3	0.395177	0.00072
	PCR	0.411005	0.00038
Ammonium	1	0.169253	0.00098
	2	0.158027	0.00098
	3	-0.258632	0.00099
	PCR	0.356409	0.00016

Table 6. Results from fitting a straight line to plots of predicted versus true concentrations of a single nutrient ($Q = 1$) provided to *Dunaliella parva* (cp. e.g. Figure 26 and Figure 27).

Despite being of general interest to biochemical investigations of the algae cells themselves, the regions 1600-1735 cm^{-1} and 2750-3100 cm^{-1} were excluded as the spectra showed a considerable randomness within them. The chemical origins for this will require more careful future studies. Hence, for concentration predictions only the 980-1600 cm^{-1} and 2900-3500 cm^{-1} ranges have been incorporated. Furthermore, a few samples were visually identified as obvious outliers and were removed from the data sets.

The prediction results for $Q = 2, P = 2$ Predictor Surfaces are concluded in the top rows of Figure 28 and Figure 29 and compared to their PCR equivalents shown in the bottom rows of these figures. Overall, the accuracy and precision of $Q = 2$ Predictor Surfaces are lower than for the $Q = 1$ examples possibly because the rather strong sample-to-sample fluctuations are amplified by varying a second ambient parameter. Nonetheless, linear fits of predicted versus true concentrations were found by ANOVAs to be significant in three out of four cases, i.e. carbon in both mixtures and nitrate. For carbon, both slopes of the predicted versus true fits indicate that the recovery of the true carbon concentrations falls short by 12% and 22%, respectively. Despite a significant linear fit of predicted versus true nitrate concentrations, Predictor Surfaces missed to recover a considerable portion of the nitrate concentrations. Compared to these $Q = 2, P = 2$ Predictor Surfaces (top rows of Figure 28 and Figure 29), PCR (bottom rows of Figure 28 and Figure 29) derived considerable inferior calibration models regarding precision and accuracy.

6.6. Conclusions

Many popular chemometric calibration methods such as Principal Component Regression (PCR) are based on the assumption that measurement data are governed by linear models. However, linear models are not incorporating higher-order terms of the predictor variables $x_{q=1,\dots,Q}$ and neither are they able to model coupling among chemical parameters of interest.

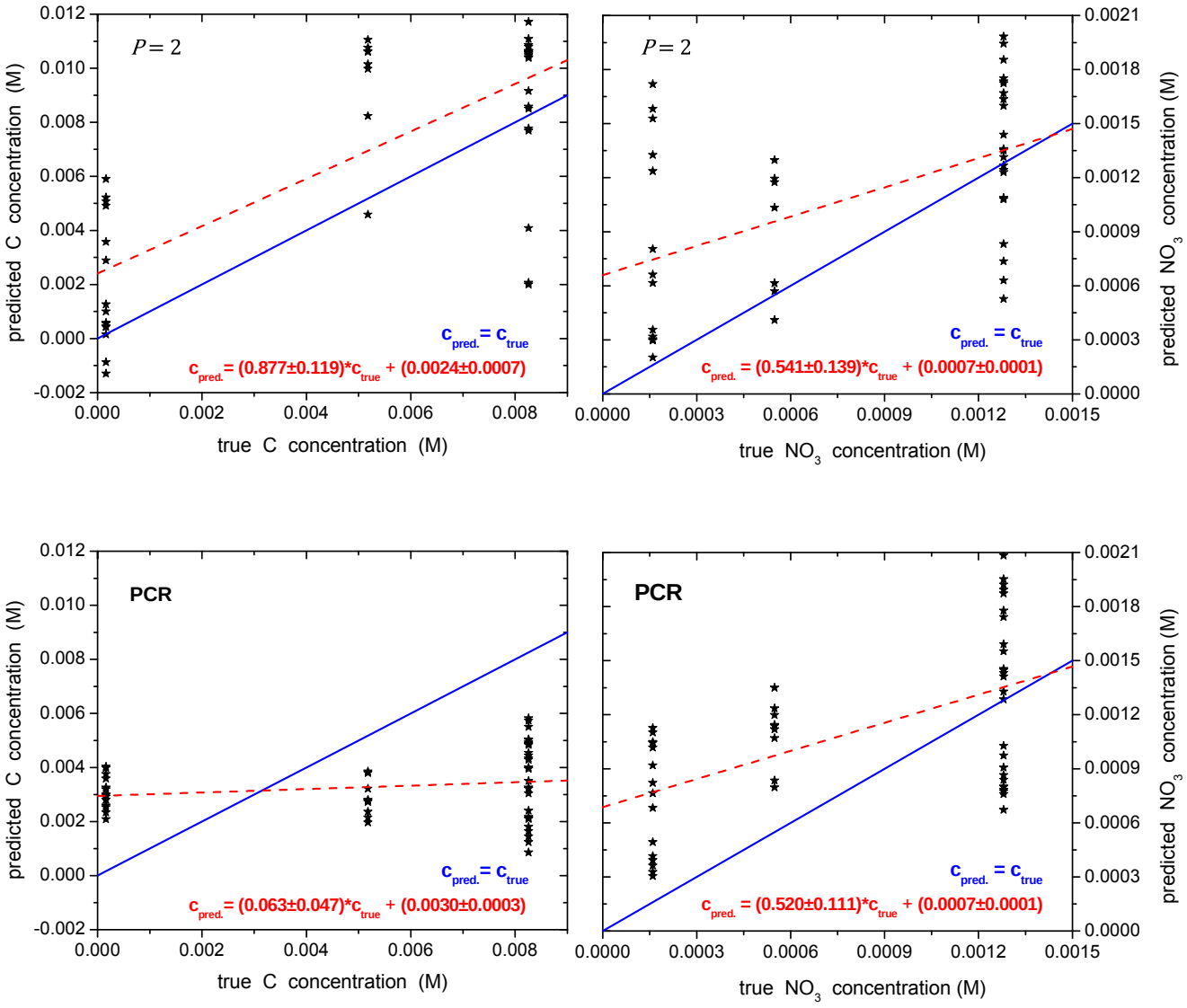


Figure 28: (top tow). Predictions of two nutrient concentrations (carbon and nitrate) for quadratic Predictor Surfaces in mixtures ($P = 2, Q = 2$). The blue lines indicate where $c_{\text{predict}} = c_{\text{true}}$; the red dashed lines are fits of predicted versus true concentrations both of which are found to be significant by an ANOVA; (bottom row) PCR results for comparison – for the carbon (bottom left), the linear fit is not significant and thus no prediction of the carbon concentration is feasible in this particular case; the linear fit (bottom right) is significant for nitrate, though.

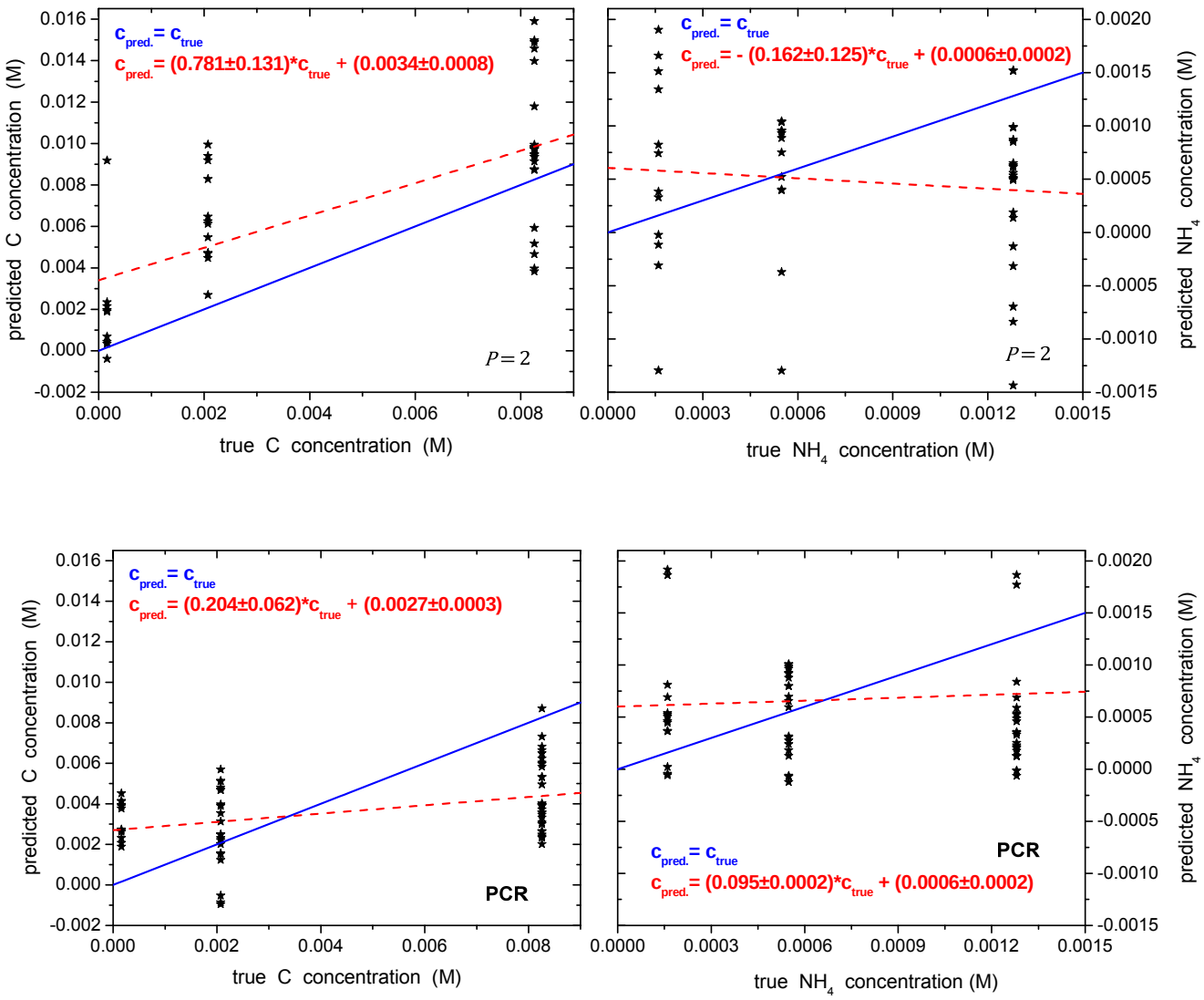


Figure 29: Same as Figure 28 but for carbon and ammonium mixtures; (top row, left) carbon concentrations predicted by a $P = 2$ Predictor Surfaces; (top row, right) the fit of predicted versus true ammonium concentrations was found to be not significant; (bottom row) the PCR results for the same data sets; here, the fit of the carbon concentrations was found to be significant but considerably lacks recovering the true concentrations; the fit of the ammonium concentrations was determined to be insignificant.

The focus of this study is expanding linear multivariate least-squares regression towards quantitative models coined '*Predictor Surfaces*' which incorporate nonlinear responses of the studied chemical system as well as cross-talk among the predictor variables. These Predictor Surfaces are based on multivariate Taylor approximations of an unknown, multivariate relation between predictors and responses and hence the needed nonlinear terms are naturally incorporated up to a user-selected polynomial order. The only factor which limits the approximation of these nonlinear multivariate model functions is the amount of available calibration data.

The chosen proof-of-principle application investigated the feasibility of microalgae cells as '*in-situ* bioprobes' for monitoring multiple environmental parameters. The hypothesis for this measurement task is that microalgae cells chemically interact with their environment and adapt to changing ambient conditions (predictor variables). These adaptations are then reflected in the cells' chemical composition and therefore their IR spectra (response variables). Preliminary investigations found that the relation between the cells' IR signatures and environmental parameters such as nutrient concentrations and sources were highly nonlinear and thus this application mandated the development and use of the proposed innovations in chemometrics. In order to simulate chemical changes in ecosystems, three algae species were cultured under series of well-defined nutrient situations. Quantitative experiments measuring the concentration of inorganic carbon, nitrate, and ammonium confirmed that the concentrations of the algae nutrients dissolved in the growing medium could be predicted from the cells' IR spectra by means of the innovative nonlinear Predictor Surfaces. Linear multivariate least-squares regression and Principal Component Regression (PCR), a representative for linear algorithms, failed to produce a sufficient accuracy and precision of predicting values for the aforementioned environmental parameters. In a subsequent step, the Predictor Surfaces were built to incorporate multiple predictor variables (nutrient concentrations) simultaneously. It was found that the nonlinear model modeled the nonlinearities in the response variables (IR spectra) and could predict the concentrations of dissolved nutrients fairly well. Thus, this innovative chemometric tool has superseded the capabilities of linear methodologies for successfully modeling nonlinear responses of many complex chemical systems.

In addition to considerably improving prediction results of chemically complex samples such as biomaterials, the nonlinear Predictor Surface models themselves are an innovative tool to visualize and study chemical processes. For example, it was found that not only a certain

nutrient level is required for the microalgae cells to produce biomass of certain chemical composition –nutrient *ratios* are of key importance.

7. Summary and Conclusions

Over the last century, the status of marine ecosystems has been on the decline as a result of the detrimental effects of human activities. These ecosystems are essential as they are the basis of the food chain, play important roles in global commerce, and are imperative for sustainability. For these reasons, a deeper understanding of environmental changes and their chemical origins is necessary. However, chemical analyses for bioanalytical applications encounter many challenges which are imposed by a multitude of chemically complex and interrelated processes. Current techniques often focus on a single chemical or physical parameter independently which does not yield an overall understanding of the alterations in chemical composition occurring as a result of these changes. A method which yields an overall picture of the effects of environment on the marine ecosystems is desired. Therefore, this thesis describes a novel, comprehensive path for studying these effects based on the combination of novel FT-IR measurement techniques with innovative chemometric algorithms in order to utilize microalgae cells, which chemically adapt to their environment, as 'biological probes'.

In order to utilize microalgae in this capacity, microalgae cultures were grown *in-vitro* under controlled conditions to mimic changes in environmental conditions. Three microalgae species were chosen, *Dunaliella salina*, *Nannochloropsis oculata*, *Dunaliella parva*, since studies have shown they react quickly (days) to changes in their growing conditions. After establishing a proper procedure for growing cultures under well-defined conditions, variations of two well-known environmental triggers, carbon and nitrogen, were imitated via changes in concentration and, in the case of nitrogen, the source (i.e. nitrate and ammonium). It has been demonstrated that FT-IR is a promising technique for studying microalgae since most relevant analytes are IR active. Thus, these microalgae samples were analyzed via FT-IR in order to study the nutrient-induced impacts on their spectroscopic signatures.

To extract the relevant chemical information needed to correlate the cell's IR spectrum to the nutrient condition, two main approaches were taken: the first approach focused on quantifying the change in the microalgae cells' chemical composition as a measure of shifts in ambient conditions. Characterizing and studying chemical processes like these responses require more comprehensive investigations than are currently feasible based on highly selective but single

analyte measurements. Thus, a novel, quantitative tool for analyzing solid samples based on FT-IR spectroscopy and supported by data pre-processing methods for enhancing the spectral quality was shown. Concentrations series of selected lipids, amino acids, proteins, carboxylic acids, mono- and polysaccharides were determined based on FT-IR spectroscopy. Utilizing these novel methods, biochemical spectra from 31 relevant compounds were placed in a calibration database for further quantitative investigations.

The second approach focused on gathering a more empirical understanding of the effects of nutrient concentration on the chemical composition of a microalgae cell. Qualitative investigations previously determined which portions of the spectra were statistically different using a t-test. Upon further investigation of these spectral regions, it was determined that these microalgae species do not always react to changes in the environment in a linear fashion. It is believed that the biochemical analytes which interact with one another within the microalgae sample as well as cross-talk of several environmental parameters are the source of these nonlinear shifts in chemical composition.

Given these large deviations from linearity, qualitative studies using standard MLR based techniques cannot be successfully applied to model these responses. Therefore, for such investigations, innovative chemometric methodologies were developed which characterize and model these nonlinear responses in order to derive deeper understanding of the chemical interactions. For this purpose, quantitative models coined '*Predictor Surfaces*' which incorporate nonlinear response variables, i.e. measurement data $y_{n=1,\ldots,N}(x_{q=1,\ldots,Q})$ as well as cross-talk among the predictor variables were developed. The chosen proof-of-principle application investigated the feasibility of microalgae cells as '*in-situ* bioprobes' for monitoring multiple environmental parameters. In order to simulate chemical changes in ecosystems, three algae species were cultured under series of well-defined nutrient situations.

Using these innovative Predictor Surfaces, quantitative experiments measuring the concentration of inorganic carbon, nitrate, and ammonium confirmed that the concentrations of the algae nutrients dissolved in the growing medium could be predicted from the cells' IR spectra. Representatives for linear algorithms, i.e. linear multivariate least-squares regression and Principal Component Regression (PCR), failed to produce a sufficient accuracy and precision of predicting values for the aforementioned environmental parameters. Since nutrient ratios are required for these species to survive, the Predictor Surfaces were built to incorporate multiple predictor variables (nutrient concentrations) simultaneously. It was found that the

nonlinear model modeled the nonlinearities in the response variables (IR spectra) and could predict the concentrations of dissolved nutrients fairly well. In addition to considerably improved prediction results, the nonlinear Predictor Surface models themselves are a valuable, innovative tool to visualize and study chemical processes which are applicable to different nonlinear data sets, allowing for a wide range of use.

The goal of this dissertation was to develop new analytical methodologies which would facilitate insights into the ecological relevance of the effects of changes in environmental conditions on marine ecosystems. These studies present a FT-IR spectroscopic technique which combines innovations in sample preparation and chemometrics. Together, these novel techniques have enabled the extraction of quantitative information from FT-IR spectra. Additionally, chemometric pre-processing steps have been developed for handling sample-to-sample fluctuations of absorption pathlengths and baselines, a common byproduct of spectra acquired from biological samples. Furthermore, novel nonlinear models which facilitate improved accuracy and precision and enable in-depth studies of complex chemical systems have been developed. These new methods in spectroscopy along with novelties in data analysis allow for predicting the nutrient status of unknown microalgae samples while quantitatively determining how these algae species are changing chemically with their environment. Results from this work outline a new analytical approach toward utilizing microalgae as novel *in-situ* probes for determining the status of the ecological community. This combination of novel spectroscopic methods along with innovative chemometric algorithms presented in this dissertation describe a new methodology for studying the impacts of environmental conditions on marine ecosystems. Using this work as a foundation, further strides in method development toward an *in-situ* measurement technique in real-time can be accomplished for utilizing microalgae as bioprobes of the environmental state.

8. List of References

List of References

1 W. V. Reid, et. al. *Millennium Ecosystem Assessment. Ecosystems and Human Well-being: Synthesis.* (Island Press, Washington, DC., 2005).

2 J. Beardall, S. Beer, J. A. Raven. Botanica Marina. **41**, (1998), 113-123.

3 R. W. Howarth, Water Science and Technology. **49**, (2004), 7-13.

4 R. W. Howarth, Harmful Algae. **8**, (2008), 14-20.

5 S. H. Julius, J. M. West, J. S. Baron, et. al. *Preliminary Review of Adaptation Options for Climate-Sensitive Ecosystems and Resources*. U.S. Environmental Protection Agency, Washington, DC, 2008,

6 F. B. Metting, Jr. Journal of Industrial Microbiology. **17**, (1996), 477-489.

7 T.P. Boyer, J. I. Antonov, O. K. Baranova, H. E. Garcia, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, D. Seidov, I. V. Smolyar, M. M. Zweng, Ed. S. Levitus. *NOAA Atlas NESDIS 66, World Ocean Database*. (U.S. Gov. Printing Office, Washington, D.C., 2009).

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

9 J. A. V. Costa, M. Greque de Morais. Bioresource Technology. **102**, (2011), 2-9.

10 M. Giordano, S. Ratti, A. Domenighini, F. Vogt, Plant Ecology & Diversity. **2**, (2009), 155-164.

11 J. Murdock and D. Wetzel, Appl. Spec. Reviews **44**, (2009), 335.

12 C. Hirschmugl, Z. Bayarri, M. Bunta, J.B. Holt, and M. Giordano, Infrared Physics & Technology **49**, (2006), 57.

13 M. Giordano, J. Plant Physiol. **158**, (2001), 577-581.

14 R.A. Andersen. *Algal Culturing Techniques* (Elsevier, Burlington, 2005).

15 L. Brennan, P. Owende. Renewable and Sustainable Energy Reviews. **14**, (2010), 557-577.

16 D. I. Hawksworth, L. A. Mound *The biodiversity of microorganisms and insects.* (CAB International, Wallingford, 1991).

17 T. Matsunaga, H. Takeyama, H. Miyashita, H. Yokouchi. Adv. Biochem Engin / Biotechnol. **96**, (2005), 165-188.

18 N. Hicks, M. T. Bulling, M. Solan, D. Raffaelli, P. C. L. White, D. M. Paterson. BMC Ecology. **11**, (2011), 1-8.

19 P. G. Falkowski, J. A. Raven. *Aquatic photosynthesis*. (BlackwaterScience, London, 1997).

20 B. G. Zilinskas. *Algal physiology and biochemistry*. (Blackwell Scientific Publications, Oxford, 1974).

21 R. E. Lee. *Phycology*. (Cambridge University Press, New York, 1980).

22 R. Harun, M. Singh, G. M. Forde, M. K. Danquah. Renewable and Sustainable Energy Reviews. **14**, (2010), 1037-1047.

23 Y. Chisti. Trends in Biotechnology. **26** ,(2008), 126–131.

24 O. Pulz, W. Gross. Appl. Microbiol. Biotechnol. **65**, (2004), 635-648.

25 E.C. Carballo-Cardenas, P. M. Tuan, M. Janssen, R. H. Wijffels. Biomolecular Engineering. **20**, (2003), 139–147.

26 M. N. Garcia-Casal, J. Ramirez, I. Leets, A. C. Pereira, M. F. Quiroga. British Journal of Nutrition. **101**, (2009), 79–85.

27 G. Andrich, U. Nesti, F. Venturi, A. Zinnai, R. Fiorentini R. European Journal of Lipid Science and Technology. **107**, (2005), 381–386.

28 A. Demirbas. Energy Sources Part A, Recovery Utilization and Environmental Effects. **31**, (2009), 163–168.

29 F. Pernet, R. Tremblay. Lipids. **38**, (2003), 1191–1195.

30 J. Masojidek, G. Torzillo. Encyclopedia of Ecology. **3**, (2008), 2226–2235.

31 Y. Chisti. Biotechnology Advances. **25**, (2007), 294–306.

32 H. Kojima, Y. Kawata, S. Yano. Recent Research Developments in Biotechnology & Bioengineering. **2**, (1999), 119-139.

33 T. G. Dunahay, E. E. Jarvis, S. S. Dais, P. G. Roessler. Applied Biochemistry and Biotechnology. **57/58**, (1996), 223-231.

34 C. Su, L. Chien, J. Gomes, Y. Lin, Y. Yu, J. Liou, R. Syu. J. Appl. Phycol. **23**, (2011), 903-908.

35 M. A. Torres, M. P. Barros, S. C. G. Campos, E. Pinto, S. Rajamani, R. T. Sayre, P. Colepicolo. Ecotoxicology and Environmental Safety. **71**, (2008), 1-15.

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

37 M. Giordano, M. Kansiz, P. Heraud, J. Beardall, B. Wood, D. McNaughton. Journal of Phycology. **37**, (2001), 271-279.

38 Y. Zhu, J. Jiang. Journal of Toxicology and Environmental Health, Part A. **71**, (2008), 474-477.

39 K. Skjanes, P. Lindblad, J. Muller. Biomol. Eng. **24**, (2007), 405-413.

40 A. Vergara-Fernandez, G. Vargas, N. Alarcon, A. Velasco. Biomass and Bioenergy. **32**, (2008), 338–344.

41 K. Tsukahara, S. Sawayama. Journal of the Japan Petroleum Institute. **48**, (2005), 251–259.

42 B. Hankamer, F. Lehr, J. Rupprecht, J. H. Mussgnug, C. Posten, O. Kruse. Physiologia Plantarum. **131**, (2007), 10–21.

43 J. C. Goldman, M. R. Dennett, C. B. Riley, J. Biotechnology and Bioengineering. **24**, (1982), 619-631.

44 D. J. Chappie, G. A. Burton Jr. Eviron. Toxicol. Chem. **16**, (1997), 559-564.

45 M. Moreira dos Santos, I. Moreno-Garrido, F. Goncalves, A. M. V. M. Soares, R. Ribeiro. Environ. Toxicol. Chem. **21**, (2002), 567-574.

46 R. E. Holcomb, L. J. Mason, K. F. Reardon, D. M. Cropek, C. S. Henry. Anal. Bioanal. Chem. **400**, (2011), 245-253.

47 A. Knap, et. al., Environmental Health Perspectives. **110**, (2002), 839-845.

48 T. S. Galloway, R. J. Brown, M. A. Browne, A. Dissanayake, D. Lowe, M. H. Depledge, M. B. Jones, Marine Pollution Bulletin. **53**, (2006), 186-194.

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

50 K. Kneipp, R. Aroca, H. Kneipp. *New Approaches in Biomedical Spectroscopy.* (Oxford University Press, Washington, 2005).

51 R. B. Horton, E. Duranty, M. McConico, F. Vogt. *Appl. Spec.* **65***,* (2011), 442-453.

52 P. Heraud, B. R. Wood, M. J. Tobin, J. Beardall, D. McNaughton. FEMS Microbiol. Lett. **249**, (2005), 219-225.

53 S. Frank, E. R. Parker. *Applications of infrared, Raman, and resonance Raman spectroscopy in biochemistry.* (Springer, New York, 1983).

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

55 A. Demirbas. Energy Policy. **35**, (2007), 4661–4670.

56 D. Bumann, D. Oesterhelt. Proc. Natl. Acad. Sci. **92**, (1995), 12195-12199.

57 B. R. Wood, P. Heraud, S. Stojkovic, D. Morrison, J. Beardall, D. McNaughton. Anal. Chem. **77**, (2005), 4955-4961.

58 J. R. DeEll, P. M. A. Toivonen. *Practical Applications of Chlorophyll Fluorescence in Plant Biology*. (Kluwer Academic Publishers, Norwell, 2003).

59 B. C. Smith. *Fundamentals of Fourier-transform Infrared Spectroscopy,* 2nd Ed*.* (CRC Press, Boca Raton, 2011).

60 M. Kansiz, P. Heraud, B. Wood, F. Burden, J. Beardall, and D. MacNaughton,

Phytochemistry **52**, (1999), 407.

61 J. Beardall, T. Berman, P. R. Heraud, M. Kadiri, B. Light, G. Patterson, S. Roberts, B. Sulzberger, U. Uehlinger, B. R. Wood. Aquatic Science. **63**, (2001), 44-67.

62 K. Stehfest, J. Toepel, C. Wilhelm, Plant Physiol. and Biochemistry. **43**, (2005), 717-726.

63 E. Levenson, P. Lerch, M. C. Martin. Infrared Phys. Tech. **49**, (2006), 45-52.

64 M. J. Nasse, E. C. Mattson, R. Reininger, T. Kubala, S. Janowski, Z. El-Bayyari, C. Hirschmugl. Nuclear Instruments and Methods in Physics Research A. **649**, (2011), 172-176.

65 A. Oren, Saline Systems, **1**, (2005), 1.

66 D.J. Hibberd, Botanical Journal of the Linnean Society, **82**, (1981), 93.

67 E. Schreiber, Wiss. Meeresuntersuch, N.R., **10**, (1927), 1.

68 P.J. Harrison, R. E. Waters, F.J.R. Taylor, J. Phycol. **16**, (1980), 28.

69 J.A. Berges, D.J. Franklin, J. Phycol., **37**, (2001), 1138.

70 R.P. Trenkenshu, L.V. Vopilova. Seriya Biologicheskikh Nauk, **2**, (1984), 3.

71 M. K. Danquah, B. Gladman, N. Moheimani, G. M. Forde, Chemical Engineering Journal. **151**, (2009), 73.

72 B. M. Auinger, K. Pfandl, J. Boenigk, Applied and Environmental Microbiology. **74**, (2008), 8.

73 A. F. da Sliva, S. O. Lourenco, R. M. Chaloub. Aquatic Botany. **91**, (2009), 291-297.

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

75 N. Xu, X. Zhang, X. Fan, L. Han, C Zeng, J. Appl. Phycol. **13**, (2001), 463-469.

76 P. G. Wells, Marine Pollution Bulletin. **39**, (1999), 39-47.

77 P. M. Vitousek, et. al., Ecological Applications. **7**, (1997), 737-750.

78 H. H. Ryu, K. K. Oh, Y. S. Kim, Journal of Industrial and Engineering Chemistry. **15**, (2009), 471-475.

79 S. Ratti, M. Giordano, J. Phycol. **43**, (2007), 693-701.

80 J. Beardall, S. Stojkovic, ScienceAsia. **32** Supplement 1, (2006), 1-10.

81 P. D. Tortell, G. R. DiTullio, D. M. Sigman, F. M. M. Morel. Mar Ecol Prog Ser. **236**, (2002), 37-43.

82 S. Yamasaki, H. Hirata. Aquacultural Engin. **14**, (1995), 357-365.

83 C. J. Zhu, Y. K. Lee, Journal of Applied Phycology. **9**, 189, (1997).

84 S.M. Renaudla, D.L. Parryla, Luong-Van, Thinh, C. Kuo, A. Padovanl, N. Sammy, Journal of Applied Phycology. **3**, (1991), 43-53.

85 G. Peyronel, A. Pignedoli, and W. Malavasi, Spectrochimica Acta **40A**, (1984), 63.

86 F. Vogt, K. Rebstock, and M. Tacke, Chemometrics & Intelligent Laboratory Systems **50**, (2000), 175.

87 I. Salomaa and J. Kauppinen, Appl. Spec. **52,** (1998), 579.

88 A. Koch and J. Weber, Appl. Spec. **52**, (1998), 970.

89 W. Press, S. Teukolsky, W. Vetterling, B. Flannery. *Numerical Recipes in C++*; Cambridge University Press: New York, 2002.

90 F. S. Rojas and C. B. Ojeda, Analytica Chimica Acta. **635**, (2009), 22.

91 C. Ojeda and F. Rojas, Analytica Chimica Acta **518**, (2004), 1.

92 F. Vogt, H. Steiner, K. Booksh, and B. Mizaikoff, Appl. Spectrosc. **58**, (2004), 683.

93 D. Bu, S. Huffman, J. Seelenbinder, and C. Brown, Appl. Spectrosc. **59**, (2005), 575.

94 C. Galloway, E. Ru, and P. Etchegoin, Appl. Spectrosc. **63**, (2009), 1370.

95 L. Shao and P. R. Griffiths, Environmental Science & Technology **41**, (2007), 7054.

96 B. Walczak and D. Massart, Trends in Analytical Chemistry **16**, (1997), 451.

97 G. Strang and T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, MA, 1997).

98 D. Haaland and D. Melgaard, Appl. Spectrosc. **55**, (2001), 1.

99 C. A. Lieber, A. Mahadevan-Jansen. *Appl. Spec.* 57, (2003), 1363-1367.

100 H. Martens, T. Naes. *Multivariate Calibration*, 2nd ed.; John Wiley & Sons: New York, 1991.

101 I. Jolliffe. *Principal Component Analysis,* 2nd ed.; Springer: New York, 2002.

102 E. Malinowski. *Factor Analysis in Chemistry*, 3rd ed.; John Wiley and Sons: New York, 2002.

103 M. Gilbert, R. Luttrell, D. Stout, F. Vogt. *J. Chem. Educat*. **85**, (2008), 135-137.

104 D. Massart, B. Vandegonste, L. Buydens, S. DeJong, P. Lewi, and J. Smeyers, *Handbook of Chemometrics and Qualimetrics* (Elsevier: Amsterdam, 1997).

105 F. Vogt and B. Mizaikoff, J. Chemom. **17**, (2003), 346.

106 F. Vogt, U. Klocke, K. Rebstock, G. Schmidtke, V. Wander, and M. Tacke, Appl. Spec. **53**, (1999), 1350.

107 J. Fabregas, C. Herrero, J. Abalde, R. Liano, B. Cabezas. *Aquaculture*. **53**, (1986), 187-199.

108 R. Sreeja, P. M. Aneesh, K. Hasna, M. K. Jayaraj. *Journal of the Electrochemical Society*. **158**(10), (2011), 187-191.

109 C. J. John, M. Amalanathan, D. Sajan, K. U. Lakshmi, I. H. Joe. *Spectrochimica Acta.* **78**, (2011), 264–272.

110 A. Sen, M. Srivastava. *Regression Analysis – Theory, Methods, Applications*; Springer: New York, 1990.

111 S. Weisberg. *Applied Linear Regression,* 3rd ed.; John Wiley & Sons: Hoboken, 2005.

112 S. Curteanu, H. Cartwright. *J. Chemometrics*. **25**, (2011), 527-549.

113 R. D. Luttrell, F. Vogt. *J. Chemometrics*. **22**, (2008), 510-521.

114 B. Schölkopf, A. Smola, K. R. Müller. *Neural Computation.* **10**, (1998), 1299-1319.

115 B. Schölkopf, A. J. Smola. *Learning with Kernels*. The MIT Press: Cambridge, Massachusetts and London, England, 2002.

116 A. Domenighini, M. Giordano. *Journal of Phycology*. **45**, (2009), 522-531.

117 S. W. Choi, I. B. Lee. *Chemical Engineering Science.* **59**, (2004), 5897-5908.

118 J. M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, I. B. Lee. *Chemical Engineering Science*. **59**, (2004), 223-234.

119 G. Box, N. Draper. *Empirical Model-Building and Response Surfaces*; John Wiley & Sons: New York, 1987.

120 D. M. Haaland, E. V. Thomas. *Anal. Chem.* **60**, (1988), 1193-1202.

121 F. Vogt, F. Gritti, G. Giuochon. *J. Chemometrics*. **25**, (2011), 575-585.

122 R. Bro. *Chemom. Intell. Lab. Syst.* **38**, (1997), 149-171.

123 C. Andersson, R. Bro. *Chemom. Intell. Lab. Syst*. **52**, (2000), 1-4.

124 N. Sidiropoulos, R. Bro. *J. Chemom.* **14**, (2000), 229–239.

125 N. Draper, H. Smith. *Applied Regression Analysis,* 3rd ed.; John Wiley & Sons, New York, 1998.

126 M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi. *GNU Scientific Library Reference Manual*, 3rd ed., version 1.12; Network Theory Ltd, 2009.

127 L. Stahle, S. Wold, Chemometr. Intell. Lab. Syst. **6**, (1989), 259–272.

128 J. Turner, J. Thayer, *Introduction to Analysis of Variance;* Sage Publications, Thousand Oaks, 2001.

129 E. Crow, F. Davis, M. Maxfield, *Statistics Manual*; Dover Publishing: Mineola, 2011.

130 F. Vogt, J. Chem. Educ. **88**, (2011), 1672–1683.

Appendices

Appendix 1

FT-IR Spectroscopic Signatures of Key Biochemical Components

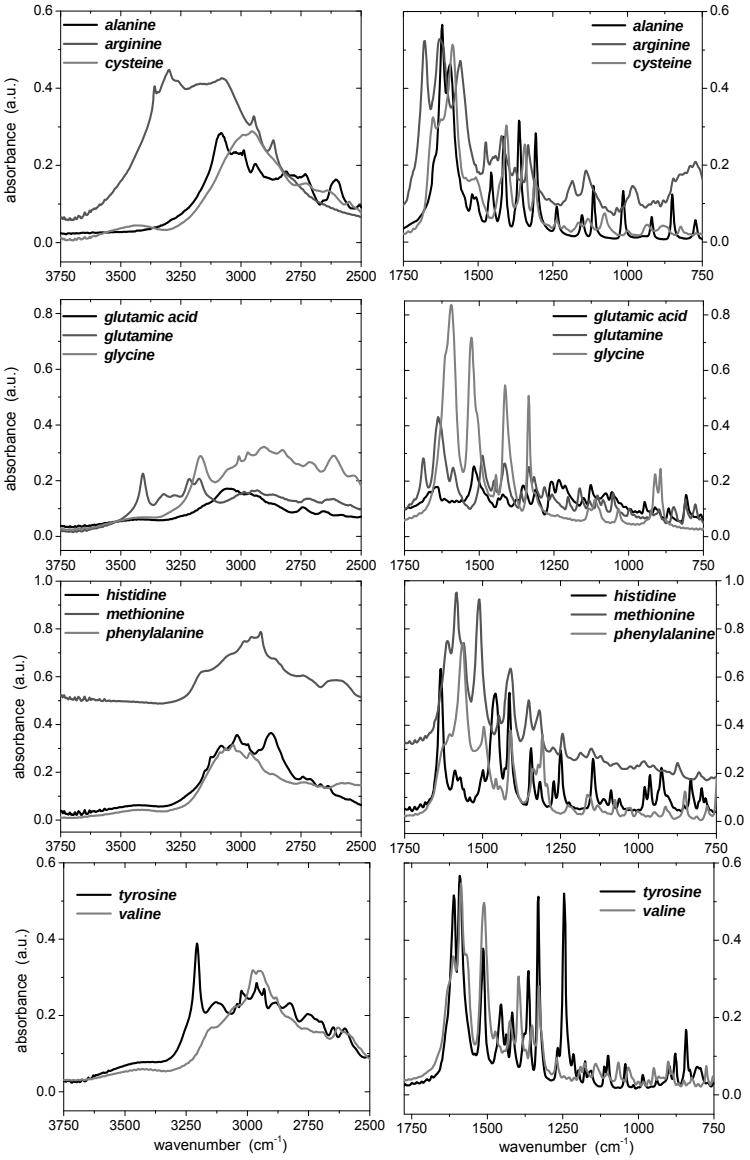


Figure 30: FT-IR spectra of several amino acids incorporated into this study; the clearly different spectroscopic signatures of these and the following analytes (Figure 31-Figure 34) ensure selectivity in multi-component mixtures.

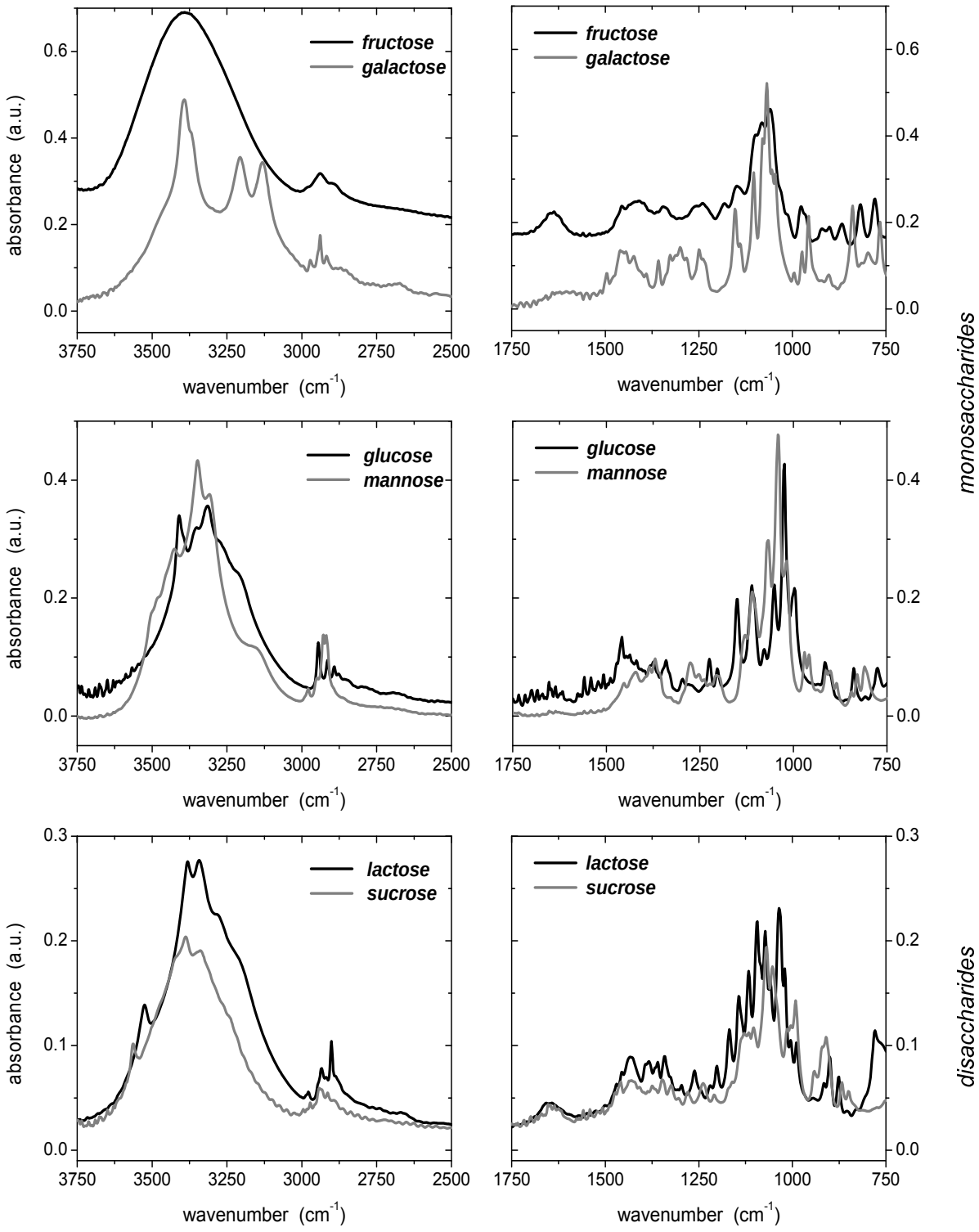


Figure 31: FT-IR spectra of several mono- and disaccharides incorporated into this study.

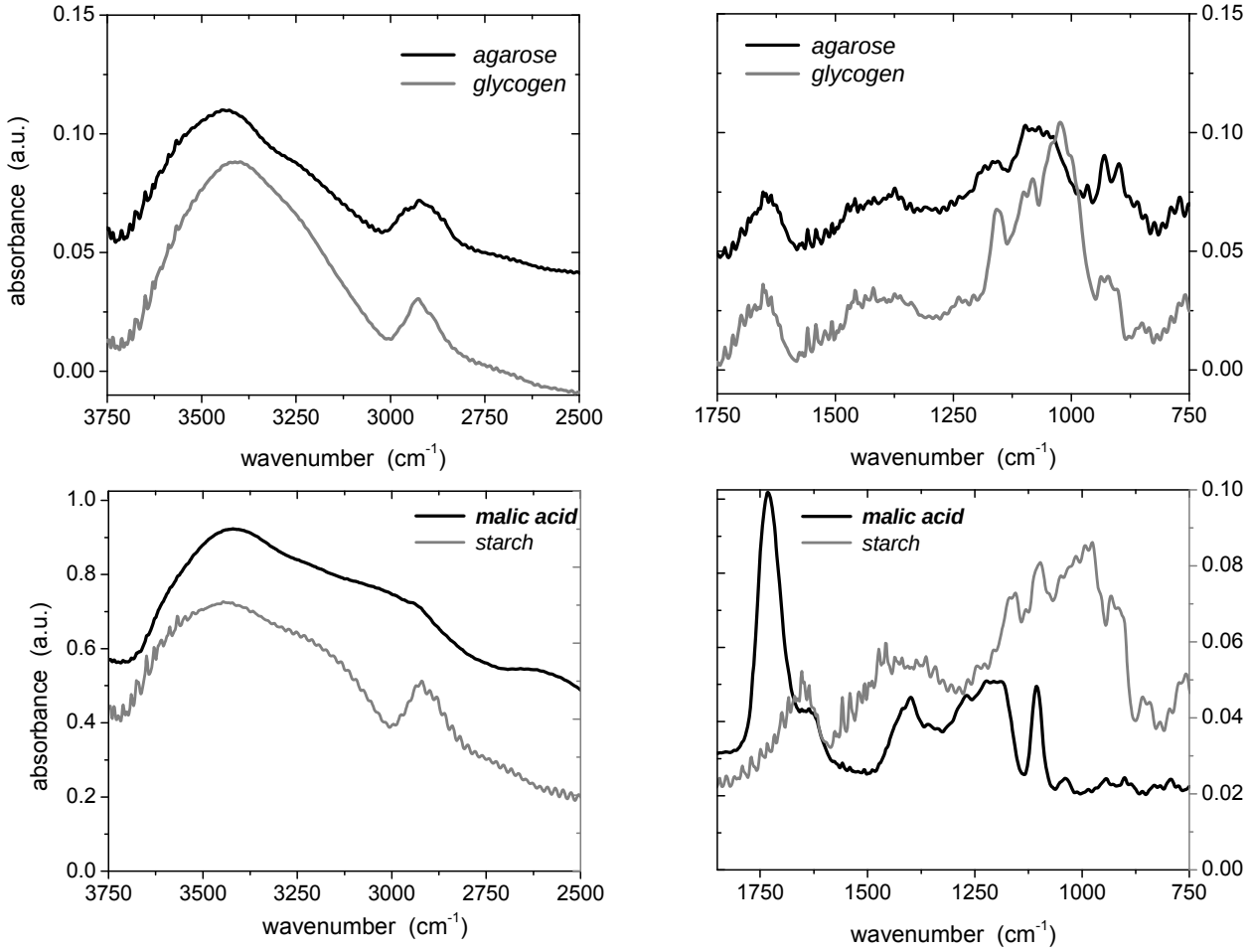


Figure 32: FT-IR spectra of polysaccharides incorporated into this study.

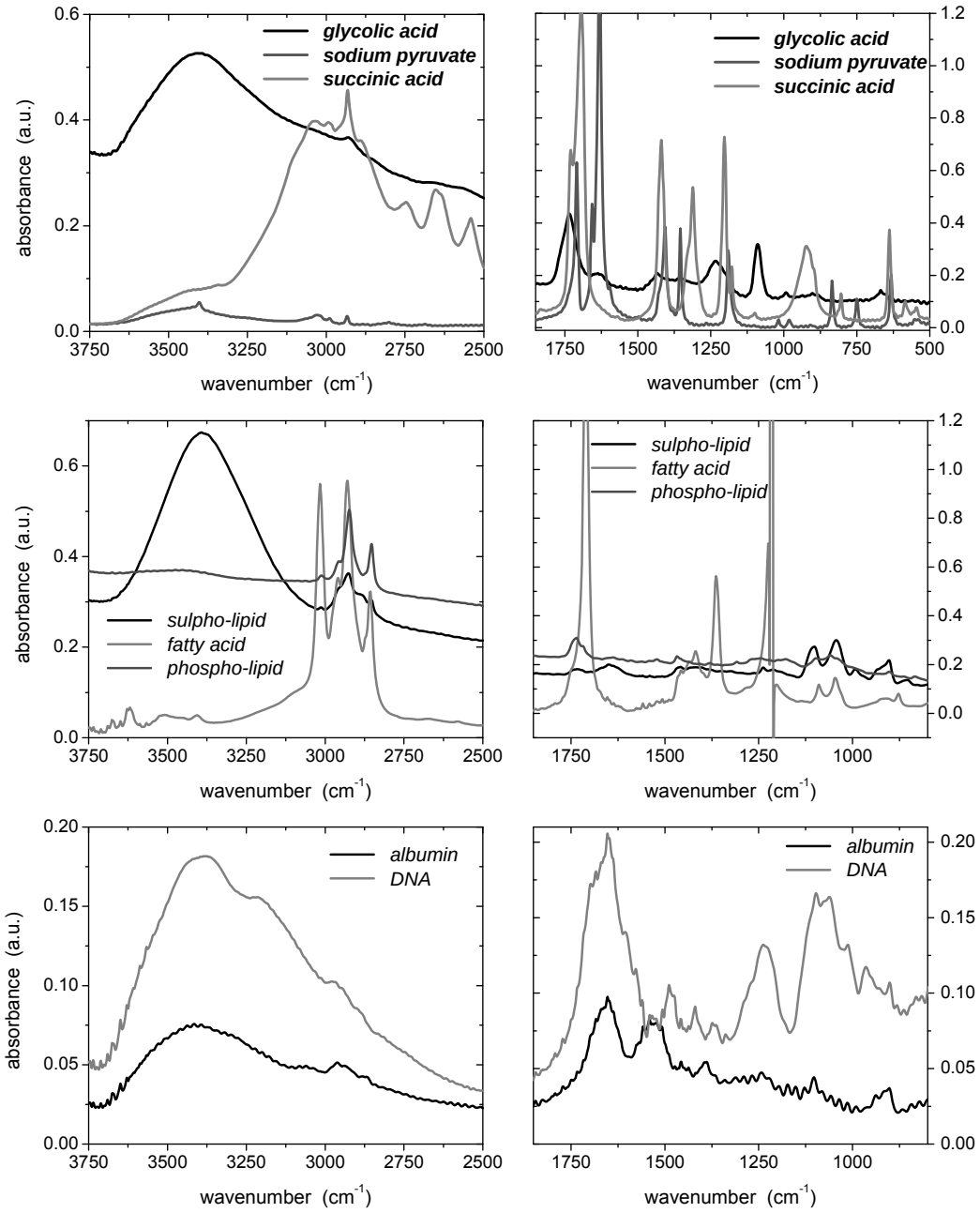


Figure 33: FT-IR spectra of incorporated into this study: (top row) carboxylic acids; (middle row) sulphonoquinovosyl diglyceride (sulpho-lipid), 1,2-diacyl-sn-glycero-3-phospho-L-serine (phospho-lipid), and linoleic acid - this fatty acid is a liquid which was not included in quantitative calibration models; (bottom row) albumin, a representative of proteins, and DNA extracted from salmon sperm.

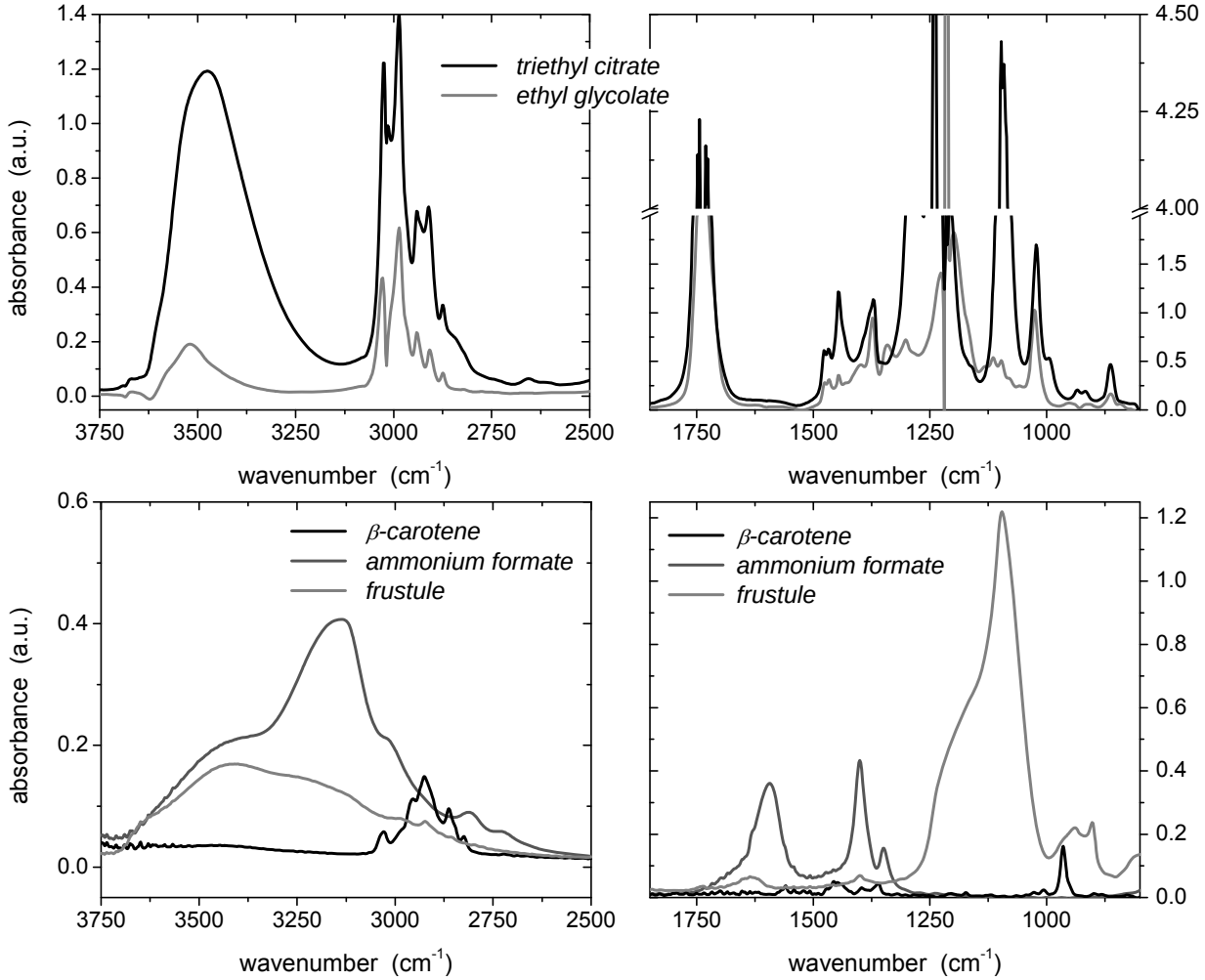


Figure 34: FT-IR spectra of compounds commonly found in microalgae (triethyl citrate and ethyl glycolate are liquids not currently contained in calibration) or required during sample preparation (ammonium formate).

Appendix 2

The following supplemental equations are included as a guide for solving complex least-squares regressions. These derivations were extracted from the spectral normalization, section 54, above.

$$\text{(I)} \quad \frac{\partial \text{SSE}}{\partial \text{c}} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}}$$

$$\text{(II)} \quad \frac{\partial \text{SSE}}{\partial b_0} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}}$$

$$\text{(III)} \quad \frac{\partial \text{SSE}}{\partial b_1} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}}$$

$$\text{(IV)} \quad \frac{\partial \text{SSE}}{\partial b_{\mathbf{2}}} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}}$$

(20)

Each partial derivative is solved step-by-step:

(I)

$$\frac{\partial \text{SSE}}{\partial c} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right]^{\mathbf{2}}$$

$$0 = \sum_{i=1}^{K} 2 \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \right) \right] \left(s_{1,i} \right)$$

$$0 = 2 \left[\sum_{i=1}^{K} s_{1,i} \cdot s_{k,i} - c \cdot \text{s}_{1,i} \cdot \text{s}_{1,i} - b_0 \cdot s_{1,i} \cdot 1 - b_1 \cdot s_{1,i} \cdot \lambda_i - b_2 \cdot s_{1,i} \cdot \lambda_i^2 \right]$$

$$0 = 2\left[\sum_{i=1}^{K} s_{1,i} \cdot s_{k,i} - c\sum_{i=1}^{K} s_{1,i} \cdot s_{1,i} - b_0\sum_{i=1}^{K} s_{1,i} \cdot 1 - b_1\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i - b_2\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i^2\right]$$

$$c\sum_{i=1}^{K} s_{1,i} \cdot s_{1,i} + b_0\sum_{i=1}^{K} s_{1,i} \cdot 1 + b_1\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i + b_2\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i^2 = \sum_{i=1}^{K} s_{1,i} \cdot s_{k,i}$$

(II)

$$\frac{\partial \mathrm{SSE}}{\partial b_0} = 0 = \sum_{i=1}^{K}\left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + c_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right]^{\mathbf{2}}$$

$$0 = \sum_{\boldsymbol{i=1}}^{K} 2\left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right](-1)$$

$$0 = 2\sum_{i=1}^{K}\left[-1 \cdot s_{k,i} + c \cdot 1 \cdot s_{1,i} + b_0 \cdot 1 \cdot 1 + b_1 \cdot 1 \cdot \lambda_i + b_2 \cdot 1 \cdot \lambda_i^2\right]$$

$$0 = 2\left[\sum_{i=1}^{K} -1 \cdot s_{k,i} + c\sum_{i=1}^{K} 1 \cdot s_{1,i} + b_0\sum_{i=1}^{K} 1 \cdot 1 + b_1\sum_{i=1}^{K} 1 \cdot \lambda_i + b_2\sum_{i=1}^{K} 1 \cdot \lambda_i^2\right]$$

$$c\sum_{i=1}^{K} 1 \cdot s_{1,i} + b_0\sum_{i=1}^{K} 1 \cdot 1 + b_1\sum_{i=1}^{K} 1 \cdot \lambda_i + b_2\sum_{i=1}^{K} 1 \cdot \lambda_i^2 = \sum_{i=1}^{K} 1 \cdot s_{k,i}$$

(III)

$$\frac{\partial \mathrm{SSE}}{\partial b_1} = 0 = \sum_{i=1}^{K}\left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right]^{\mathbf{2}}$$

$$0 = \sum_{i=1}^{K} 2\left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right](-\lambda_i)$$

$$0 = 2\sum_{i=1}^{K}\left[-\lambda_i \cdot s_{k,i} + c \cdot \lambda_i \cdot s_{1,i} + b_0 \cdot \lambda_i \cdot 1 + b_1 \cdot \lambda_i \cdot \lambda_i + b_2 \cdot \lambda_i \cdot \lambda_i^2\right]$$

$$0 = 2\left[\sum_{i=1}^{K} -\lambda_i \cdot s_{k,i} + c\sum_{i=1}^{K} \lambda_i \cdot s_{1,i} + b_0\sum_{i=1}^{K} \lambda_i \cdot 1 + b_1\sum_{i=1}^{K} \lambda_i \cdot \lambda_i + b_2\sum_{i=1}^{K} \lambda_i \cdot \lambda_i^2\right]$$

$$c\sum_{i=1}^{K} \lambda_i \cdot s_{1,i} + b_0\sum_{i=1}^{K} \lambda_i \cdot 1 + b_1\sum_{i=1}^{K} \lambda_i \cdot \lambda_i + b_2\sum_{i=1}^{K} \lambda_i \cdot \lambda_i^2 = \sum_{i=1}^{K} \lambda_i \cdot s_{k,i}$$

(IV)

$$\frac{\partial \mathrm{SSE}}{\partial b_2} = 0 = \sum_{i=1}^{K} \left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right]^2$$

$$0 = \sum_{i=1}^{K} 2\left[s_{k,i} - \left(c \cdot s_{1,i} + b_0 \cdot 1 + b_1 \cdot \lambda_i + b_2 \cdot \lambda_i^2\right)\right]\left(-\lambda_i^2\right)$$

$$0 = 2\left[\sum_{i=1}^{K} -\lambda_i^2 \cdot s_{k,i} + c \cdot \lambda_i^2 \cdot s_{1,i} + b_0 \cdot \lambda_i^2 \cdot 1 + b_1 \cdot \lambda_i^2 \cdot \lambda_i + b_2 \cdot \lambda_i^2 \cdot \lambda_i^2\right]$$

$$0 = 2\left[\sum_{i=1}^{K} -\lambda_i^2 \cdot s_{k,i} + c\sum_{i=1}^{K} \lambda_i^2 \cdot s_{1,i} + b_0\sum_{i=1}^{K} \lambda_i^2 \cdot 1 + b_1\sum_{i=1}^{K} \lambda_i^2 \cdot \lambda_i + b_2\sum_{i=1}^{K} \lambda_i^2 \cdot \lambda_i^2\right]$$

$$c\sum_{i=1}^{K} \lambda_i^2 \cdot s_{1,i} + b_0\sum_{i=1}^{K} \lambda_i^2 \cdot 1 + b_1\sum_{i=1}^{K} \lambda_i^2 \cdot \lambda_i + b_2\sum_{i=1}^{K} \lambda_i^2 \cdot \lambda_i^2 = \sum_{i=1}^{K} \lambda_i^2 \cdot s_{k,i}$$

The unknowns $(\mathrm{c}, b_0, b_1, b_2)$ can be separated from the knowns in order to estimate the unknown parameters:

$$\text{(I)} \quad c\sum_{i=1}^{K} s_{1,i} \cdot s_{1,i} + b_0\sum_{i=1}^{K} s_{1,i} \cdot 1 + b_1\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i + b_2\sum_{i=1}^{K} s_{1,i} \cdot \lambda_i^2 = \sum_{i=1}^{K} s_{1,i} \cdot s_{k,i}$$

$$\text{(II)} \quad c\sum_{i=1}^{K} 1 \cdot s_{1,i} + b_0\sum_{i=1}^{K} 1 \cdot 1 + b_1\sum_{i=1}^{K} 1 \cdot \lambda_i + b_2\sum_{i=1}^{K} 1 \cdot \lambda_i^2 = \sum_{i=1}^{K} 1 \cdot s_{k,i}$$

$$\text{(III)}\quad c\sum_{i=1}^{K}\lambda_i\cdot s_{1,i}+b_0\sum_{i=1}^{K}\lambda_i\cdot 1+b_1\sum_{i=1}^{K}\lambda_i\cdot\lambda_i+b_2\sum_{i=1}^{K}\lambda_i\cdot\lambda_i^2=\sum_{i=1}^{K}\lambda_i\cdot s_{k,i}$$

$$\text{(IV)}\quad c\sum_{i=1}^{K}\lambda_i^2\cdot s_{1,i}+b_0\sum_{i=1}^{K}\lambda_i^2\cdot 1+b_1\sum_{i=1}^{K}\lambda_i^2\cdot\lambda_i+b_2\sum_{i=1}^{K}\lambda_i^2\cdot\lambda_i^2=\sum_{i=1}^{K}\lambda_i^2\cdot s_{k,i}$$

(21)

Consider the left side of (21)—the unknowns are c, b_0, b_1 and b_2. These unknowns can be placed in a vector. The sums can be rewritten as dot products and placed in the following matrix which is then multiplied by the unknown vector:

$$\begin{pmatrix}\mathbf{s_1^T\cdot s_1} & \mathbf{s_1^T\cdot 1} & \mathbf{s_1^T\cdot\lambda} & \mathbf{s_1^T\cdot\lambda^2}\\ \mathbf{1^T\cdot s_1} & \mathbf{1^T\cdot 1} & \mathbf{1^T\cdot\lambda} & \mathbf{1^T\cdot\lambda^2}\\ \mathbf{\lambda^T\cdot s_1} & \mathbf{\lambda^T\cdot 1} & \mathbf{\lambda^T\cdot\lambda} & \mathbf{\lambda^T\cdot\lambda^2}\\ \mathbf{\lambda^{2^T}\cdot s_1} & \mathbf{\lambda^{2^T}\cdot 1} & \mathbf{\lambda^{2^T}\cdot\lambda} & \mathbf{\lambda^{2^T}\cdot\lambda^2}\end{pmatrix}\cdot\begin{pmatrix}c\\ b_0\\ b_1\\ b_2\end{pmatrix}$$

(22)

The right side sums can be rewritten in a similar method and replaced by the following vector:

$$\begin{pmatrix}\mathbf{s_1^T\cdot s_k}\\ \mathbf{1^T\cdot s_k}\\ \mathbf{\lambda^T\cdot s_k}\\ \mathbf{\lambda^{2^T}\cdot s_k}\end{pmatrix}$$

(23)

Appendix 3

For the $Q > 1$ case, the prediction step (18) requires calculating the gradient $\boldsymbol{\nabla} Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}\left(\hat{x}_1^{\mathrm{unk}}, \ldots, \hat{x}_Q^{\mathrm{unk}}\right)$ of nonlinear, multivariate polynomial models $Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}} = \widehat{\boldsymbol{\beta}}_{\mathbf{n}} \cdot \hat{\mathbf{x}}^{\mathrm{unk}}$ (14). This gradient's elements $\frac{\partial Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}}{\partial \hat{x}_{q=1,\ldots,Q}^{\mathrm{unk}}}$ can be expressed concisely by means of the Kronecker delta $\delta_{q,p} = \begin{cases} 1 \text{ for } q = p \\ 0 \text{ for } q \neq p \end{cases}$ utilizing $\frac{\partial}{\partial x_q}\left(\hat{x}_{q_1}^{\mathrm{unk}}\right) = \delta_{q,q_1}$, $\frac{\partial}{\partial x_q}\left(\hat{x}_{q_1}^{\mathrm{unk}} \cdot \hat{x}_{q_2}^{\mathrm{unk}}\right) = \delta_{q,q_1} \cdot \hat{x}_{q_2}^{\mathrm{unk}} + \hat{x}_{q_1}^{\mathrm{unk}} \cdot \delta_{q,q_2}$, $\frac{\partial}{\partial x_q}\left(\hat{x}_{q_1}^{\mathrm{unk}} \cdot \hat{x}_{q_2}^{\mathrm{unk}} \cdot \hat{x}_{q_3}^{\mathrm{unk}}\right) = \delta_{q,q_1} \cdot \hat{x}_{q_2}^{\mathrm{unk}} \cdot \hat{x}_{q_3}^{\mathrm{unk}} + \hat{x}_{q_1}^{\mathrm{unk}} \cdot \delta_{q,q_2} \cdot \hat{x}_{q_3}^{\mathrm{unk}} + \hat{x}_{q_1}^{\mathrm{unk}} \cdot \hat{x}_{q_2}^{\mathrm{unk}} \cdot \delta_{q,q_3}$, etc. A few more bookkeeping steps were discussed previously [121] and yield the Q elements of $\boldsymbol{\nabla} Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}$:

$$
\begin{aligned}
\frac{\partial Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}}{\partial \hat{x}_{q=1,\ldots,Q}^{\mathrm{unk}}} = \left(b_{n,q}\right)_{p=1} &+ \left(b_{n,qq} \cdot \hat{x}_q + \sum_{q_1=1}^{Q} b_{n,q_1 q} \cdot \hat{x}_{q_1} \right)_{p=2} \\
&+ \left(b_{n,qqq} \cdot \hat{x}_q \cdot \hat{x}_q + \sum_{q_1=1}^{Q} b_{n,q_1 q\, q} \cdot \hat{x}_{q_1} \cdot \hat{x}_q + \sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} b_{n,q_1 q_2 q} \cdot \hat{x}_{q_1} \cdot \hat{x}_{q_2} \right)_{p=3} \\
&+ \left(b_{n,qqqq} \cdot \hat{x}_q \cdot \hat{x}_q \cdot \hat{x}_q + \sum_{q_1=1}^{Q} b_{n,q_1 qqq} \cdot \hat{x}_{q_1} \cdot \hat{x}_q \cdot \hat{x}_q \right. \\
&\left. + \sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} b_{n,q_1 q_2 qq} \cdot \hat{x}_{q_1} \cdot \hat{x}_{q_2} \cdot \hat{x}_q + \sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} \sum_{q_3=q_2}^{Q} b_{n,q_1 q_2 q_3 q} \cdot \hat{x}_{q_1} \cdot \hat{x}_{q_2} \cdot \hat{x}_{q_3} \right)_{p=4} \\
&+ \cdots
\end{aligned}
$$

(24)

These expressions (24) are then used in (25) to derive the following final Q nonlinear equations $\frac{\partial \mathrm{SSE}\left(\hat{x}_q^{\mathrm{unk}}\right)}{\partial \hat{x}_{q=1,\ldots,Q}^{\mathrm{unk}}} = 0$ (see (18)):

$$0 = \sum_{n=1}^{N} \left[\left(b_{n,0}\right)_{p=0} + \left(\sum_{q_1=1}^{Q} b_{n,q_1} \cdot \hat{x}_{q_1} \right)_{p=1} + \left(\sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} b_{n,q_1\,q_2} \cdot \hat{x}_{q_1} \cdot \hat{x}_{q_2} \right)_{p=2} + \left(\sum_{q_1=1}^{Q} \sum_{q_2=q_1}^{Q} \sum_{q_3=q_2}^{Q} b_{n,q_1\,q_2\,q_3} \cdot \hat{x}_{q_1} \cdot \hat{x}_{q_2} \cdot \hat{x}_{q_3} \right)_{p=3} + (\cdots)_{p=4} + \cdots \right] \cdot \frac{\partial Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}}{\partial \hat{x}_q^{\mathrm{unk}}}$$

(25)

The solution of the equation system (25), i.e. the Q numbers $\hat{x}_{q=1,\ldots,Q}^{\mathrm{unk}}$ (here: nutrient concentrations), minimizes the sum of squares between the assumed model $Y_{\widehat{\boldsymbol{\beta}}_{\mathbf{n}}}$ (14) and measured response variables $\mathbf{y}^{\mathrm{unk}}$ (here: algae spectrum) obtained from an unknown sample.

Appendix 4

Summary of Graduate School Honors, Publications, and Presentations

Honors

1. Eastman Chemical Company Travel Grant, Eastman Chemical Company, Kingsport, TN, May **2011**.

2. Judson Hall Robertson Graduate Award in Analytical Chemistry, Department of Chemistry, University of Tennessee, Knoxville, TN, April **2011**.

3. Board of Visitors Graduate Student Competition Award, Department of Chemistry Board of Visitors, University of Tennessee, Knoxville, TN, October **2010**.

4. C. W. Keenan Outstanding Graduate Teaching Award, Department of Chemistry, University of Tennessee, Knoxville, TN, April **2010**.

5. Outstanding Service Award, Department of Chemistry, University of Tennessee, Knoxville, TN, April **2009**.

Publications

1. R. B. Horton, M. McConico, C. Landry, T. Tran, F. Vogt. *Analytica Chimica Acta,* **2012**, xx, xxx. *Submitted.*

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

3. S. Ratti, C. Olivieri, M. Nasse, E. Duranty, R. Burke, M. Giordano, F. Vogt, C. J. Hirschmugl. *Microscopy and Microanalysis*, **2010**, *16,* 740-741

Presentations

1. R. Horton, M. McConico, C. Landry, T. Tran, F. Vogt. *Utilization of Microalgae Cells as Indicators of Environmental Change via FT-IR and Chemometrics.* American Chemical Society (ACS), **Fall 2011** National Meeting.

2. R. Horton, M. McConico, C. Landry, T. Tran, F. Vogt. *Utilization of Microalgae as Novel Bioprobe for Comprehensive Monitoring of Chemical Changes in Marine Ecosystems.* The Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Inc. (PITTCON), **2011**.

3. R. Horton, E. Duranty, M. McConico, F. Vogt. *Spectroscopic Analyses for Quantification of Key Biological Components in Microalgae as Indicators of Environmental Change.* The Federation of Analytical Chemistry and Spectroscopy Societies, (FACSS), **2010*.***

4. R. Horton, E. Duranty, M. McConico, M. Robbins, F. Vogt. *Quantification of Biochemically Relevant Analytes in Microalgae using FTIR and Chemometric Methods.* The Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Inc. (PITTCON), **2010**.

5. R. Burke, M. K. Gilbert, D. Khandal, M. Giordano, F. Vogt. *Using Microalgae Biodiversity along with FTIR and Chemometric Methods to Detect Environmental Changes in Marine Ecosystems.* The Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Inc. (PITTCON), **2009**.

Vita

Rebecca Burke Horton was born in Kingsport, TN on August 10, 1985. She was raised in Gate City, VA, attending elementary, middle and high school in Scott County, VA. She graduated from Gate City High School in 2003 and began her undergraduate career at Carson-Newman College in the fall. Rebecca graduated with her B.S. in chemistry and a minor in mathematics from Carson-Newman in the spring of 2007. Upon graduation, she began her graduate career in Analytical Chemistry at the University of Tennessee under the direction of Dr. Frank Vogt. Rebecca will graduate in May 2012 with her Ph.D. in Analytical Chemistry where she hopes to pursue a career in chemical industry.