Bioenerg. Res. (2011) 4:22–35
DOI 10.1007/s12155-010-9098-y

# Feasibility of Spectroscopic Characterization of Algal Lipids: Chemometric Correlation of NIR and FTIR Spectra with Exogenous Lipids in Algal Biomass

**Lieve M. L. Laurens · Edward J. Wolfrum**

Published online: 27 July 2010


**Abstract** A large number of algal biofuels projects rely on a lipid screening technique for selecting a particular algal strain with which to work. We have developed a multivariate calibration model for predicting the levels of spiked neutral and polar lipids in microalgae, based on infrared (both near-infrared (NIR) and Fourier transform infrared (FTIR)) spectroscopy. The advantage of an infrared spectroscopic technique over traditional chemical methods is the direct, fast, and non-destructive nature of the screening method. This calibration model provides a fast and high-throughput method for determining lipid content, providing an alternative to laborious traditional wet chemical methods. We present data of a study based on nine levels of exogenous lipid spikes (between 1% and 3% (*w*/*w*)) of trilaurin as a triglyceride and phosphatidylcholine as a phospholipid model compound in lyophilized algal biomass. We used a chemometric approach to corrrelate the main spectral changes upon increasing phospholipid and triglyceride content in algal biomass collected from single species. A multivariate partial least squares (PLS) calibration model was built and improved upon with the addition of multiple species to the dataset. Our results show that NIR and FTIR spectra of biomass from four species can be used to accurately predict the levels of exogenously added lipids. It appears that the cross-species verification of the predictions is more accurate with the NIR models ($R^2$=0.969 and 0.951 and RMECV=0.182 and 0.227% for trilaurin and phosphatidylcholine spike respectively), compared with FTIR ($R^2$=0.907 and 0.464 and RMECV=0.302 and 0.767% for trilaurin and phosphatidylcholine spike, respectively). A fast high-throughput spectroscopic lipid fingerprinting method can be applied in a multitude of screening efforts that are ongoing in the microalgal research community.

**Keywords** Microalgae · Biomass · Lipids · Infrared spectroscopy · Multivariate calibration · Chemometrics

L. M. L. Laurens · E. J. Wolfrum (✉)
National Bioenergy Center,
National Renewable Energy Laboratory,
1617 Cole Blvd,
Golden, CO 80401, USA
e-mail: Ed.Wolfrum@nrel.gov

## Introduction

Microalgae have long been recognized as a potential source of biofuels, because of the high biomass productivity and the associated high-lipid yields of large-scale algal cultures [22]. The main conclusions from numerous literature reports are that the biochemical composition of the algal biomass will influence the economics of the algal biofuels economic scenario (for a review of the promises and challenges of algal biofuels, see 9, 12, 21). The lipid content of algal biomass is considered the most important biochemical characteristic for the production of algal biomass-based biodiesel. An accurate method for lipid quantification in algal biomass is necessary for the purpose of selecting optimum species and growth conditions. Throughout the growth of an algal culture, there are a high number of variables that can potentially impact the lipid content and composition of the biomass, for example, the nutrient concentration in the culture medium.

Numerous algal biofuels research projects currently underway require a high-throughput lipid analysis platform. Traditional lipid and fatty acid analyses require relatively large amounts of biomass (>1 g of dry biomass), are time consuming and not particularly effective for the analysis of a large number of algal samples. The aim of this work is to

develop a high-throughput technique that is capable of monitoring a large number of samples with a minimal investment of time. In a bioprospecting project, one needs a way to distinguish and isolate high-lipid containing strains from a large pool of potential strains. Other methods currently available for screening lipid content in algae are based on fluorescent lipophilic dyes such as Nile red, thin layer chromatography (TLC), and mass spectroscopy (MS) [4, 18, 25]. The Nile red screening method has the potential to be a powerful screening tool that is applicable in growing cultures of algae. However, it is known that this fluorometric method is affected by uneven dye uptake and staining and has been known to cause variability between species and between growth conditions due to differences in cell wall composition, limiting the application of this method [3, 6]. Furthermore, in green algae, the presence of high levels of chlorophyll reduces the Nile red fluorescence signal due to high background signal. Despite these limitations, Chen et al. [3] managed to correlate the Nile red fluorescence with the measured neutral lipid content for *Chlorella vulgaris*. The authors reported on an improvement of the Nile red staining method by the addition of solvents as carrier and changing the excitation and emission wavelengths. Even with improvements, no successful demonstration has been reported of the use of Nile red to distinguish between neutral and polar lipid content.

Similarly, HPLC, TLC, and MS are also powerful techniques for lipid analysis [25]. However, these methods rely on a considerable sample preparation step to isolate the lipid fraction prior to analysis and are by default destructive by nature. We focused on infrared (IR) spectroscopy because of the non-destructive nature of the procedure and the application of the technique to whole, homogenized, biomass.

The application of IR spectroscopy to identify and quantify chemical constituents in biomass is based on the chemical bonds of a molecule that absorb energy in the IR region of the electromagnetic spectrum. Near-infrared (NIR) has been developed as a rapid inexpensive method to monitor chemical composition of corn stover [10]. IR spectroscopy has also been important in fundamental lipid research [2] and the oils and fats industry has applied this technique for the determination of trans-isomers in fats.

The use of Fourier transform infrared (FTIR) spectroscopy in algal biomass analysis has been useful in monitoring biochemical changes [5, 11, 19]. Furthermore, the use of chemometrics in combination with FTIR spectroscopy has been shown to be useful for the discrimination of cyanobacterial strains [14].

The NIR absorption of chemical structures results from the overtones and combinations of the same vibrations that play a role in FTIR spectroscopy. In NIR spectra, these bands are much broader and less defined. NIR does not provide the same fine details and structural information as FTIR, however the advantage of NIR spectroscopy is that it allows for the analysis of solid, opaque, and liquid materials with minimal sample preparation requirements [13]. Furthermore, both spectroscopic techniques are non-destructive and fast analytical methods that require only very small amounts of biomass (<100 mg dry biomass for NIR and <10 mg for FTIR analysis).

Infrared spectroscopy combined with multivariate calibration methods are widely used in analytical chemistry disciplines [1, 10, 15]. The biggest advantage of these models is the large number of samples that can be analyzed and screened for specific characteristics, bypassing the need for long and laborious wet chemical analyses. The main advantage of the IR spectroscopic method we are developing is its non-destructive nature, the possibility to develop a real-time monitoring technique for growing and screening algal cultures and chemometrics allowing us to extract the quantitative information from large datasets.

Depending on the application of this screening method, one may want to apply a model that is optimized for a single-species or a multiple-species combined model. Example applications could be to screen for mutant or transgenic lines from one species that have significantly increased their lipid content or to test a range of culture conditions for one species that cause the induction of lipid content of the cells. A combined multiple-species model could be used to screen a large number of algae collected from different places to detect unusually high-lipid producers.

In this manuscript, we are presenting data on the correlation and prediction of lipid content and composition in algal biomass based on spectral information from NIR and FTIR spectra. We prepared a sample set of four species spiked at nine levels with a triglyceride and a phosholipid. The species we used span four major divisions of microalgae, the green algae (Chlorophyceae), Eustigmatophyceae, diatoms (Bacillariophyceae), and blue-green algae (Cyanobacteria). The biomass was collected from open pond large-scale cultures. These four divisions contain good candidates for the potential large-scale algal culture for biofuels production thanks to their inherent high-lipid content and fast growth rates [12]. Specific information on the growth stage, light and temperature profile, and harvest conditions is lacking for this set of biomass samples. The diatom we have used has not been identified, and the exact strain information for the *Nannochloropsis* sp. and *Chlorococcum* sp. biomass is not known. We anticipate that the lack of this information will not reduce the utility of the models we have built, as our aim is to eliminate the species-specific information and develop a model that is applicable to a wide variety of algal biomass.

The spectra taken from these samples were used to build multivariate calibration models to (1) determine the quantitative differences in the oil content of the algae and (2) extract information on the chemical composition of the different lipid species. In this paper we demonstrate multivariate calibration of NIR and FTIR spectra with lipid data from exogenous lipid spikes. We investigate the effect of mathematical pretreatment of the raw spectra on the quality of the prediction models. We also verify the accuracy of prediction of the single-species models as well as the combined multiple-species models.

## Experimental Methods

### Biomass Preparation

Frozen algal biomass was kindly provided to us by Dr. Ami Ben-Amotz (Seambiotic, Israel) and Jim Demattia (Carbon Capture, USA). We chose to work with four species, *Nannochloropsis* sp., *Chlorococcum* sp., *Spirulina* sp. and an unknown Diatom. The frozen biomass was lyophilized and finely ground using a cryo-grinder in the presence of lipid spikes (6770 Freezer/mill Spex Sampleprep, Metuchen, USA).

### Randomized Double Spiking

To generate biomass with sufficient variation in lipid content, solely due to the presence of increasing amounts of triglycerides or phospholipids, algal biomass was spiked at nine randomized levels (0–3% *w*/*w*) with commercially available lipids, trilaurin (Fluka, cat no: 92019) and phosphatidylcholine (Sigma, cat no: P3556) after which the added lipids were mixed thoroughly with the biomass through cryogrinding.

The sample set consisted of nine levels for four species, resulting in 36 independent samples (see Table 1). Each of the 36 samples was analyzed as four replicates, yielding a total of 144 spectra for both NIR and FTIR. Preparation of the biomass was optimized before spectroscopic measurements were taken. We found lyophilized biomass was best finely ground in liquid nitrogen to obtain a homogeneous powder. This also proved to yield the best mixing of the spiked lipids with the biomass, ensuring no visible differences between the spiked and non-spiked biomass samples.

### NIR/FTIR Spectra Collection

NIR spectra were collected using a Foss NIR Systems model 6500 Forage Analyzer with a transport reflectance module. For each sample, prepared in a circular sample cell with a 2.5 cm insert, a total of four spectra were collected and averaged (three scans per spectrum of four replicate prepared samples). The spectra were collected in the range 400–2,500 nm (at a 2 nm resolution). WinISI software (Foss, USA) was used for collection, standardization, and export of the spectra.

FTIR attenuated total reflectance spectra were collected on a Nicolet 6700 (Thermo Scientific) FTIR instrument using a diamond Smart iTR reflectance cell with a DTGS detector. Algal biomass did not require any preparation and was pressed against the diamond cell prior to scanning. A total of 32 scans were taken for each spectrum, for four replicate samples. The spectra were collected in the range of 4,000 to 500 $cm^{-1}$ (at 4 $cm^{-1}$ resolution) and data were exported using Omnic 8.0.342 Software (Thermo Scientific).

### Multivariate Calibration

Principal component analysis (PCA) was carried out in R 2.9.0 (R Development Core Team [24]), using the NIPALS algorithm specified in the pcaMethods R package [23].

Partial least square (PLS) multivariate calibration models were built using the Unscrambler v9.7. Two different types of PLS models were built; PLS1 and PLS2. The optimum number of principal components used for the PLS regression is shown in the text accompanying the figures and was selected by an Unscrambler algorithm based on an apparent minimum in root mean square error of the cross-validation (RMECV) of the validation of the models. For all models, PLS regression was performed using the NIPALS algorithm, using full cross-validation on a centered dataset. We investigated the effect on the statistics of the calibration models of eliminating part of the visible spectrum (for the NIR spectra), the application of different mathematical spectral pretreatments and spectral derivatives. The algorithms we used were multiplicative scatter correction (MSC), and 3-point 1st and 2nd derivatives (S. Golay). All mathematical algorithms were applied in The Unscrambler software. The prediction uncertainty intervals shown in Table 4 were calculated using the U-deviation [27] algorithm present in The Unscrambler software.

To compare the prediction models after spectral trimming and pretreatment, we applied the Fisher's *z*- transformation ($z = a\tanh(r)$) and then calculated the confidence intervals around the *z*-transformed variable [8]. To compare prediction uncertainties, we squared the RMECV values and compared the value of this variance produced by each calibration equation using the standard *F*-test [8].

**Table 1** Overview of the 9 levels of triglyceride and a phospholipid spike concentrations in algal biomass from four species

| | Theoretical concentration (%) | | Actual concentration (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Triglyceride | Phospholipid | *Nannochloropsis* sp. | | *Chlorococcum* sp. | | Diatom | | *Spirulina* sp. | |
| | TG | PL | TG | PL | TG | PL | TG | PL | TG | PL |
| Level 1 | 0.0 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Level 2 | 0.0 | 0.5 | 0.00 | 0.55 | 0.00 | 0.50 | 0.00 | 0.13 | 0.00 | 0.49 |
| Level 3 | 0.1 | 1.5 | 0.11 | 1.40 | 0.13 | 1.42 | 0.08 | 1.52 | 0.09 | 1.45 |
| Level 4 | 0.5 | 1.0 | 0.52 | 0.99 | 0.49 | 1.01 | 0.52 | 1.04 | 0.43 | 0.92 |
| Level 5 | 1.0 | 3.0 | 0.94 | 2.87 | 0.97 | 2.85 | 0.96 | 2.74 | 0.98 | 2.80 |
| Level 6 | 1.5 | 2.5 | 1.39 | 2.32 | 1.43 | 2.24 | 1.39 | 2.35 | 1.49 | 2.48 |
| Level 7 | 2.0 | 0.0 | 2.01 | 0.00 | 1.97 | 0.00 | 1.97 | 0.00 | 1.87 | 0.00 |
| Level 8 | 2.5 | 2.0 | 2.41 | 1.87 | 2.40 | 1.84 | 2.44 | 1.95 | 2.27 | 1.79 |
| Level 9 | 3.0 | 0.1 | 2.87 | 0.09 | 2.95 | 0.13 | 2.90 | 0.12 | 2.90 | 0.11 |

The concentration range of the spikes was between 0.1% and 3% (*w*/*w*) of the total biomass. The random, independent, spiking between the two spikes allows for independent regression of both types of lipids

## Results

### NIR and FTIR Fingerprinting of Algal Biomass

We prepared algal biomass samples of four species were prepared as nine different levels of spikes, between 0% and 3% (*w*/*w*) of the biomass. We chose these relatively low levels of spike concentration to keep within the anticipated intra-species natural lipid level variation if these cells were grown under varying growth conditions.

The pure compound lipid spectra (NIR and FTIR) are shown in Fig. 1 for trilaurin (a triglyceride) and phosphati-

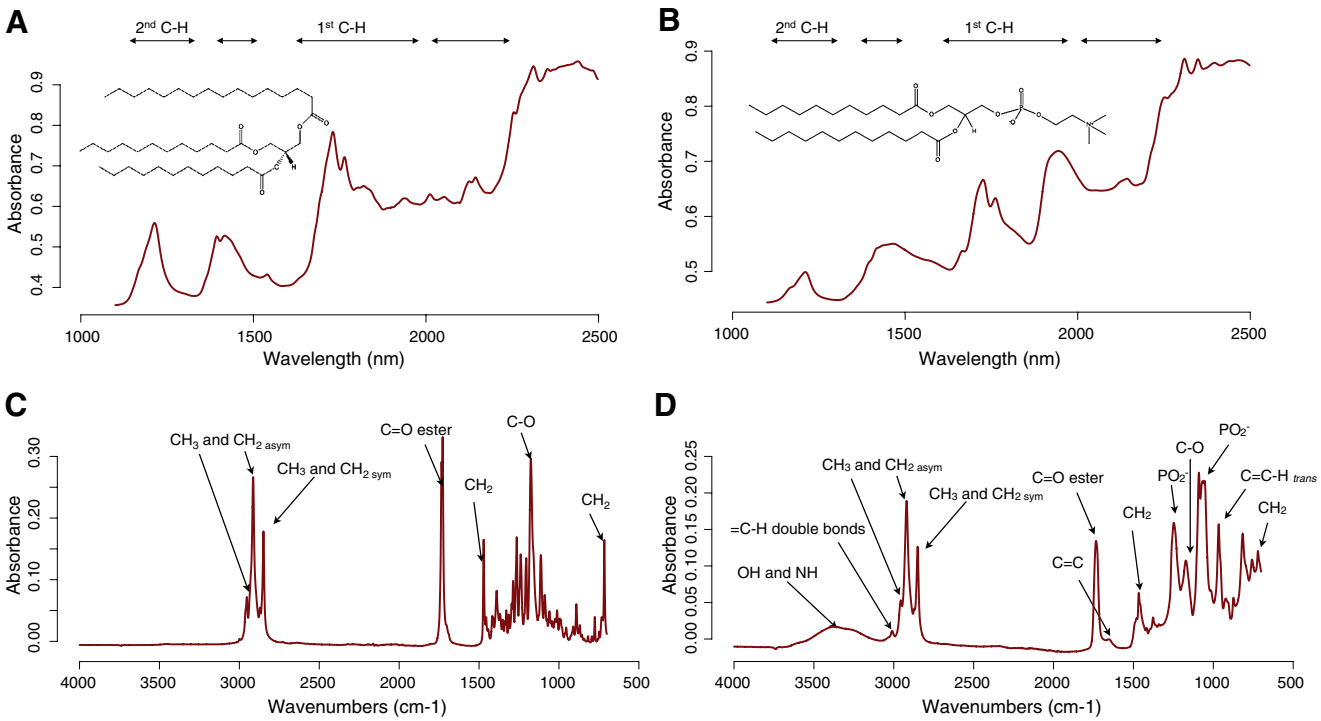


**Fig. 1** NIR (**a**, **b**) and FTIR (**c**, **d**) spectra for pure triglyceride (trilaurin, **a**, **c**) and phospholipid (phosphatidylcholine, **b**, **d**). In the NIR spectra, the highlighted regions indicated are identified to be the characteristic absorption bands (overtones) of lipids, showing some overlap with the more intense areas of the pure triglyceride and phospholipid component spectra. The three main, lipid-specific, spectral overtone regions are indicated with brackets on the NIR spectra. In the FTIR spectra (**c**, **d**), the individual peaks are annotated based on information on specific absorption bands from the literature. A large number of peaks in the 1,500–1,000 $cm^{-1}$ wave number region could not be annotated in detail. The *insets* show the chemical structure for trilaurin (**a**) and phosphatidylcholine (**b**)

dylcholine (a phospholipid). The four brackets indicate the lipid-specific spectral overtone regions in the NIR spectra (Fig. 1a, b) and correspond with the most prominent peaks in the spectra. The characteristic absorption bands of oils in the NIR spectrum are (1) the first overtones of C–H stretching vibrations (1,600–1,900 nm), (2) the region of second overtones of C–H stretching vibrations (1,100–1,250) and (3) two regions (2,000–2,200 nm and 1,350–1,500 nm) which contain bands due to combinations of C–H stretching vibrations and other vibrational modes [13].

The FTIR spectra of the pure lipid compounds (Fig. 1c, d) are more complex compared with the respective NIR spectra. Three distinct absorption bands are apparent, of which the $CH_3$ and $CH_2$ (3,025–2,954 $cm^{-1}$) and the C=O ester (1,746–1,654 $cm^{-1}$) are most characteristic for lipids. Furthermore, the hydroxyl and phosphate groups from for example the phospholipids can be distinguished (1,200–500 $cm^{-1}$). From these individual lipid spectra, it is clear that characteristic and distinct fingerprints for triglycerides and phospholipids exist in the FTIR spectrum (also discussed in [13]).

The NIR spectra collected for the four species are shown in Fig. 2. The collected spectra encompass both the visible and the near-IR region of the spectrum (400 to 2,500 nm or 10,000 to 4,000 $cm^{-1}$, respectively). Figure 2 illustrates that the inter-species variation in the visible spectrum (400–1,100 nm) caused by the variation in photosynthetic pigment composition, whereas the variation within each sample set is due solely to the varying content of exogenous lipid spike. When all spectra are subjected to a PCA (Fig. 3), distinct grouping of the spectra occurs along the first two principal components

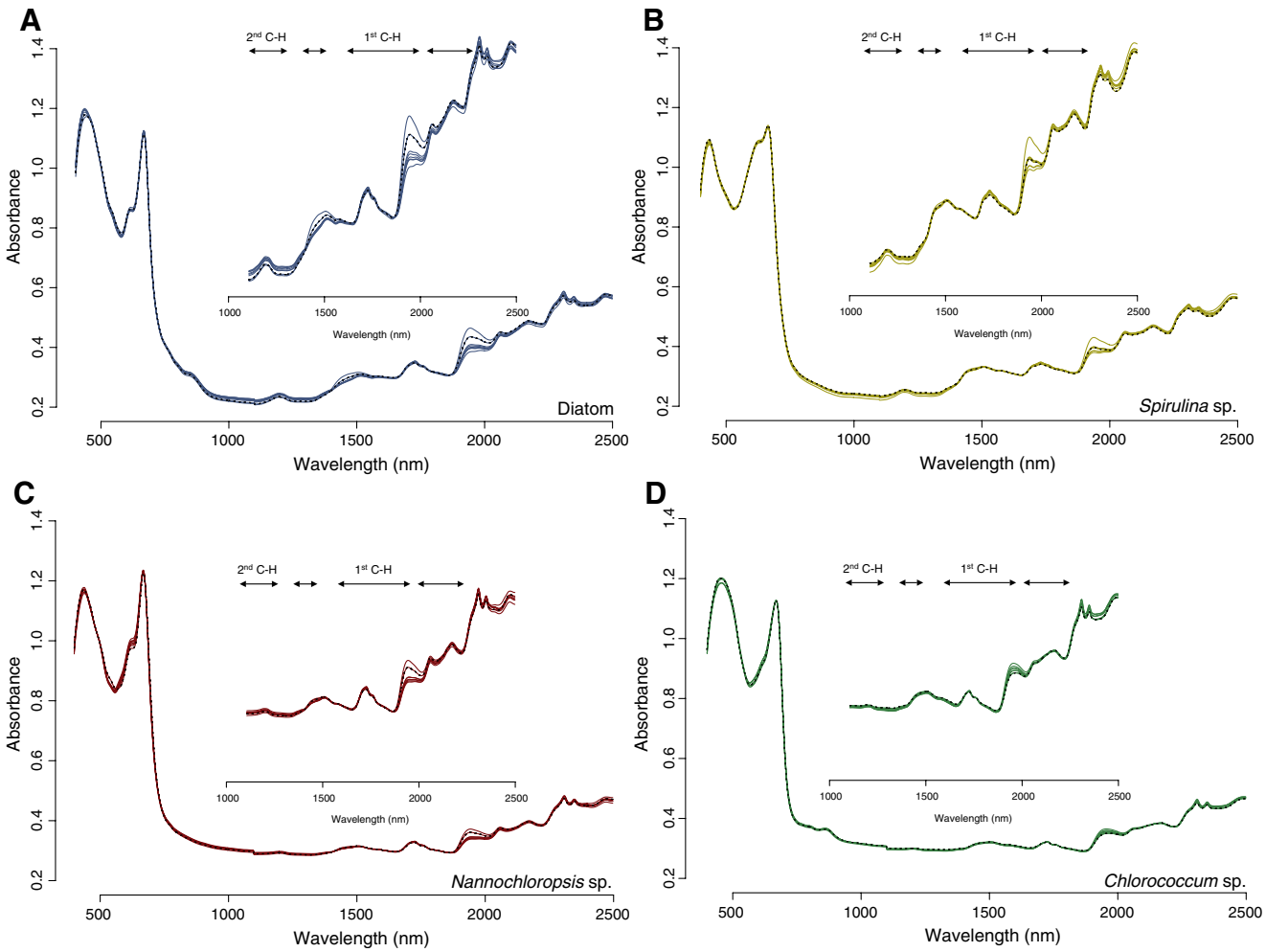


**Fig. 2** NIR spectra for algal biomass for four species showing changes between species and between different levels of spike within a species (**a** Diatom; **b** *Spirulina* sp.; **c** *Nannochloropsis* sp., **d** *Chlorococcum* sp.). Nine spectra (each an average of four replicate spectra per spike level) are shown for each species. The concentration range of the spikes (trilaurin and a phosphatidyl choline) was between 0.1% and 3% of the total biomass. All spiked samples were finely ground with a cryo-grinder and freeze dried prior to spectra collection. The *black dotted line* shows the biomass without any spiked lipids, the remaining eight spectra show the eight spiked biomass samples. *Inset* shows a close-up of the NIR region of the spectrum, with *brackets* indicating the lipid-specific overtones

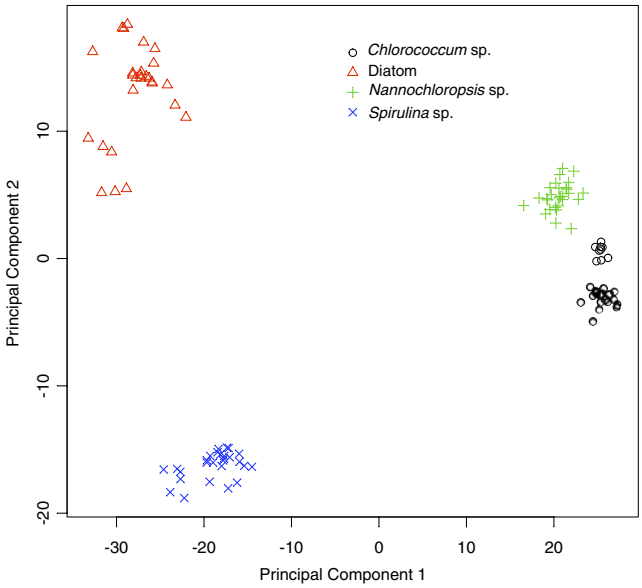


**Fig. 3** Principal component analysis of MSC-corrected NIR spectra (excluding the visible portion of the spectra) from four species with nine levels of spiked lipids. Principal component one (PC1) and PC2 explain 78.8% and 16.7% of the total spectral variation, respectively

(PC1 and PC2). Within each group there is a considerable spread of the variation along PC2 (and PC1) due to the varying spike concentrations in the biomass. PC1 and PC2 explain together 97% of the variation in the spectra (78.8% by PC1 and 16.7% by PC2).

FTIR spectra for each of the four species at the nine spike levels are shown in Fig. 4. FTIR spectra comprise the mid-IR spectrum (between 4,000 and 550 $cm^{-1}$ or 2,500 and 18,182 nm, respectively). The four distinct, lipid-specific absorption band regions are shown in detail for each species, indicating variation due to the spiked lipid concentrations (insets in Fig. 4). Similar to the NIR spectra, there are considerable differences due to inter-species variation. The PCA plot (Fig. 5) illustrates these differences, which cause the spectra of the four species to cluster in distinct groups along PC1 and PC2, explaining 89% of the spectral variation (61% by PC1 and 28% by PC2). There are two sample outliers for *Spirulina* sp. and *Nannochloropsis* sp. that position far removed from the rest of the samples in the PCA scores plot. Upon closer look of their spectra, these samples exhibited spectral anomalies and were therefore excluded from further data analyses.

## Multivariate Calibration of Exogenous Lipid Spike

For each species in the spike data set we used PLS regression methods to build multivariate calibration models of the exogenous spike concentrations. PLS regression is a linear regression formula (or model) describing the relationship between a response variable $Y$ (e.g., lipid content) and a set of predictor variables $X$ (e.g., IR spectra). To verify (or cross-validate) the results, data subsets are applied to intermediate models to determine robustness and potential high leverage samples of the datasets. The results from these cross-validation calculations illustrate the quality and 'robustness' of the prediction. In our study, we report on the quality of the calibration using the $R^2$ value and the RMECV of our cross-validated models.

To demonstrate the effect on the model statistics we have created single-species models. The statistics of the single-species NIR models are shown in Table 2. For all models, an $R^2>0.92$ and RMECV<0.296% was obtained for both triglyceride and phospholipid spike concentrations. The cross-validation results indicate that there is not a significant difference in performance between these models. In the context of this article, we are specifically interested in developing a model that is capable of predicting the spike concentration across different species or divisions of algae; therefore, we have focused our further discussions on the combined four-species model.

We compared PLS1 and PLS2 regression analysis for both the triglyceride and the phospholipid spike. PLS1 independently calibrates the dependent variables (spectral information) against the independent variables (chemical constituents, triglyceride, and phospholipid); PLS2 calibrates all independent variables simultaneously [16, 17]. When we compared PLS1 and PLS2 prediction models, we found no statistically significant differences ($p=0.05$) in the quality of the prediction models built by either PLS1 or PLS2 (data not shown). All prediction models shown were calculated from PLS2 models.

It is common to mathematically transform spectral data prior to building calibration models [20, 26]. These pretreatments can help to reduce spectral variation due to instrument or sample variability. We have investigated the effect of four mathematical pretreatments on the performance of the calibration models. Table 3 shows the statistical summary of the results obtained for both NIR and FTIR spectra. The data were collected from fully cross-validated combined multiple-species PLS2 regression models. We statistically compared the RMECV and $R^2$ values obtained for each calibration as described in the methods section ($p=0.05$). For NIR, we found that we can significantly improve the calibration models by mathematical spectral pretreatment, whereas the FTIR models did not significantly improve, but rather reduced the regression quality.

For the NIR calibrations (Table 3), we found a significant improvement of the trilaurin spike regression (but not of the phoshatidylcholine spike) by using the 1st derivative of the full spectra (both visible and NIR, i.e., 400–2,500 nm); $R^2=0.961$ and RMECV=0.204%. Interestingly, when using the 2nd derivative, we noticed a significant reduction in the

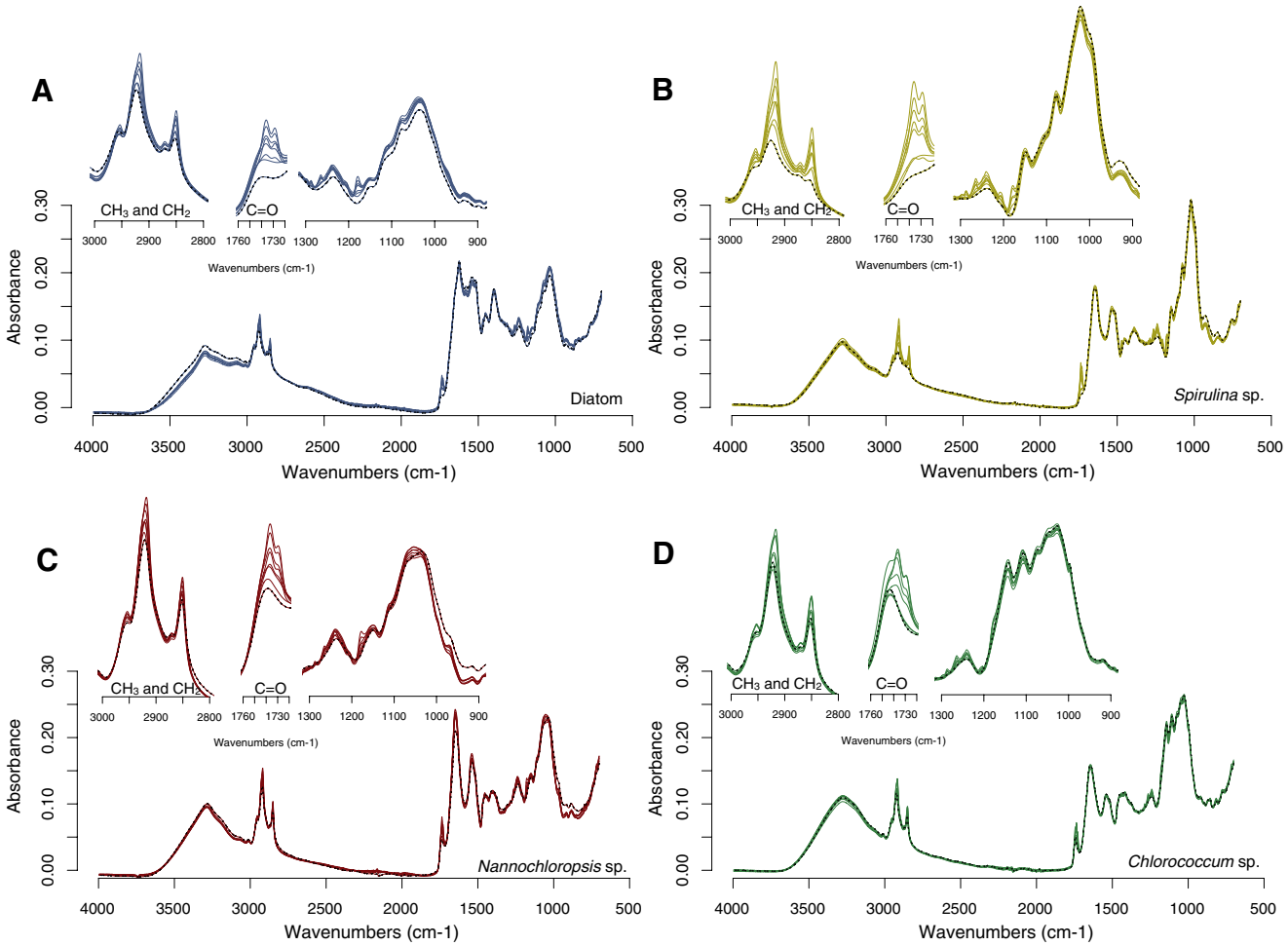


**Fig. 4** ATR-FTIR spectra for algal biomass illustrating changes between species and between different levels of spike (**a** Diatom; **b** *Spirulina* sp.; **c** *Nannochloropsis* sp., **d** *Chlorococcum* sp.). Nine spectra are shown for each species (after multiplicative scatter correction of the spectra), reflecting the nine different levels of spike concentration. *Insets* show a close-up of lipid-specific bands indicating changes between species and between different levels of spike within a species. FTIR spectra were collected between 4,500 and 500 $cm^{-1}$ as an average of 32 scans per sample for four replicate samples per spike level. The spectra shown are the average of the four replicate spectra for each spike level and show the region between 4,000 and 700 $cm^{-1}$. The *black dotted line* shows the spectrum of the biomass for each species without any spiked lipids, whereas the other eight colored spectra shown indicate the eight levels of spike concentration

quality of the model of the phosphatidylcholine spike ($R^2$=0.838 and RMECV=0.412%), but not for trilaurin. A significant improvement of the regression was found for both phosphatidylcholine and trilaurin spike concentrations ($R^2$=0.969 and RMECV=0.182% for trilaurin and $R^2$=0.951 and RMECV=0.226% for phosphatidylcholine) after MSC. Note that this improvement is only present when the visible region (400–1,100 nm) of the spectrum was excluded from the model. The high number of optimum number of principal components (>10 principal component (PC)) could indicate that the PLS calibration models are overfitting the data and will therefore perform worse in the prediction. The optimum number of components is chosen by the software package we use and usually coincide with a minimum in the RMECV.

The models built using the FTIR spectra (Table 3) did not significantly improve with mathematical pretreatment, rather MSC, 1st and 2nd derivation of the spectra reduced the quality of the calibration compared with the statistics of regression of the no-treatment spectra. There is a remarkable difference in the optimum number of PCs between the four models (18 versus 4, 5, and 6, respectively). The 18 PCs selected for the no-treatment regression indicates that the software algorithms could not find a clear minimum in RMECV, but rather a gradual decline as more components are added to model. By investigating the regression coefficients of this model at 18 PCs it appeared that a lot of the spectral noise was modeled along with the spike concentrations (overfitting). This usually leads to better statistics ($R^2$ and RMECV) but the model will

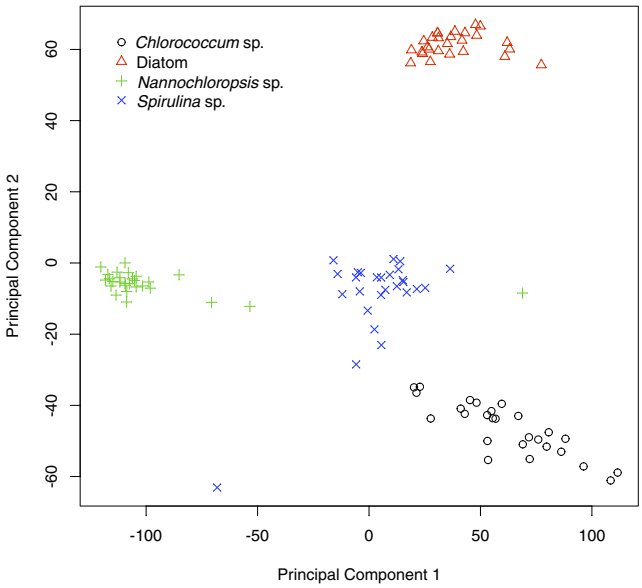


**Fig. 5** Principal component analysis of MSC-corrected FTIR spectra from four species with nine levels of spiked lipids. Principal component one (PC1) and PC2 explain 60.7% and 27.8% of the total spectral variation, respectively

perform worse with an unknown sample dataset (which will vary in the noise portion of the spectra). We decided to select eight PCs for this model to keep the model in line with other regression analyses performed on this dataset. When comparing the statistics at eight PCs ($R^2$=0.907 and RMECV=0.302% for trilaurin and $R^2$=0.464 and RMECV=0.766% for phosphatidylcholine) to the models obtained using mathematically pretreated spectra we could no longer detect statistically significant changes in the performance of the models.

We chose to continue working with the best calibration model for both NIR and FTIR. These were obtained by multiplicative scatter correction of the spectra, excluding the visible part of the spectrum (for NIR) and without mathematical pretreatment of the spectra (for FTIR).

The predicted versus measured spike concentrations for the best NIR and FTIR models (selected from Table 3), are shown in Figs. 6 and 7. These figures illustrate the strong correlation ($R^2$>0.95) for both the triglyceride and phospholipid content of the biomass for the NIR prediction models. The FTIR predicted versus measured plots show a relatively good correlation ($R^2$=0.91) for the triglyceride spike but a poor correlation ($R^2$=0.46) for the phospholipid spike predicted and measured concentrations. The regression coefficients, indicating which areas of the NIR spectra are contributing most to the calculations, of the calibration model for both triglyceride and phospholipid for the NIR models (shown in Fig. 6) are shown in Fig. 8a, b.

## Cross-species Prediction of Lipid Content

We tested the accuracy of the prediction across species by using the single-species models as well as the combined species model to predict the concentration of the spikes in biomass from each of the four species. The test spectra were collected from biomass spiked with the same lipids but using a level of spike concentration that was not included in the original calibration models or cross-validation. The collected NIR and FTIR spectra were treated as unknowns and predicted using five models, individual species models (*Nannochloropsis* model (N), *Chlorococcum* model (C), *Spirulina* model (S), Diatom model (D)) as well as with a combined, multiple-species

**Table 2** Statistics of the single-species PLS1 validation models of trilaurin and phosphatidylcholine spikes for both NIR and FTIR spectroscopy

| | | Triglyceride | | | Phospholipid | | |
|---|---|---|---|---|---|---|---|
| Species | Spectrum | RMSECV | $R^2$ | #LV | RMSECV | $R^2$ | #LV |
| NIR | | | | | | | |
| *Nannochloropsis* sp. (N) | Full | 0.140 | 0.983 | 9 | 0.147 | 0.980 | 9 |
| | NIR | 0.122 | 0.987 | 8 | 0179 | 0.970 | 8 |
| *Chlorococcum* sp. (C) | Full | 0.102 | 0.991 | 8 | 0.234 | 0.947 | 8 |
| | NIR | 0.144 | 0.982 | 6 | 0.296 | 0.915 | 6 |
| Diatom (D) | Full | 0.178 | 0.995 | 7 | 0.137 | 0.984 | 7 |
| | NIR | 0.098 | 0.991 | 5 | 0.177 | 0.974 | 5 |
| *Spirulina* sp. (S) | Full | 0.134 | 0.983 | 7 | 0.149 | 0.980 | 7 |
| | NIR | 0.139 | 0.982 | 5 | 0.148 | 0.980 | 5 |
| FTIR | | | | | | | |
| *Nannochloropsis* sp. (N) | | 0.252 | 0.944 | 9 | 0.588 | 0.682 | 9 |
| *Chlorococcum* sp. (C) | | 0.338 | 0.902 | 6 | 0.457 | 0.801 | 6 |
| Diatom (D) | | 0.305 | 0.920 | 4 | 0.545 | 0.740 | 4 |
| *Spirulina* sp. (S) | | 0.272 | 0.933 | 7 | 0.576 | 0.699 | 7 |

Data for the individual species models are the validation results, generated from fully cross-validated (i.e., leave-one-out) models. For the NIR models, models were compared for the full spectrum and just the NIR region (i.e., exclusion of the visible portion, 400–1,100 nm)

*RMECV* root mean square error of cross-validation, *#LV* number latent variables or factors

**Table 3** Summary of the effect of mathematical spectral pretreatment on the quality of NIR and FTIR cross-validated models

| Treatment | Triglyceride | | | Phospholipid | | |
|---|---|---|---|---|---|---|
| | RMSECV | $R^2$ | #LV | RMSECV | $R^2$ | #LV |
| NIR | | | | | | |
| None | 0.239 | 0.947 | 11 | 0.239 | 0.945 | 11 |
| MSC | 0.248 | 0.943 | 14 | 0.234 | 0.947 | 14 |
| 1st Der | 0.204* | 0.961* | 10 | 0.247 | 0.942 | 10 |
| 2nd Der | 0.218 | 0.956 | 9 | 0.412* | 0.839* | 9 |
| No Vis | 0.234 | 0.946 | 12 | 0.323 | 0.914 | 11 |
| MSC—no Vis | 0.182* | 0.969* | 15 | 0.227* | 0.951* | 15 |
| FTIR | | | | | | |
| None | 0.260 | 0.931 | 18[a] | 0.532 | 0.742 | 18[a] |
| | 0.302 | 0.907 | 8 | 0.767 | 0.464 | 8 |
| MSC | 0.482* | 0.763* | 6 | 0.856* | 0.333* | 6 |
| 1st Der | 0.311 | 0.902 | 5 | 0.711* | 0.537* | 5 |
| 2nd Der | 0.303 | 0.906 | 4 | 0.835* | 0.365* | 4 |

Statistics of the combined, four-species prediction models for both NIR and FTIR spectroscopy were generated after mathematical pretreatment of the spectra prior to building the calibration models. The data were generated from fully cross-validated models

*MSC* multiplicative scatter correction, *1st and 2nd Der* 3-point first and second derivative, *No Vis* excluding visible region of spectrum (400–1,100 nm) from the NIR calculation, *#LV* indicate the optimum number of latent variables in the calibration model, i.e., minimum in residual variance. *RMECV* root mean square error of cross-validation

*$p<0.05$, statistically significant different value for $R^2$ and root mean square error of cross-validation (RMECV)

[a] Indicates high number of LV used for calibration compared with other models on same data set, data for eight LV is shown below

model (Table 4). The predicted concentration was compared with the actual concentration of the triglyceride and phospholipid spike in the biomass. The deviations or prediction uncertainties were calculated from the validation variances, residual variances, and the leverage of the variable data in the prediction objects [7, 27]. We are aware that there is spectral variation between replicate samples. Therefore, we have used three and four replicate (i.e., from different parts of the same sample) NIR and FTIR spectra respectively in this cross-species prediction

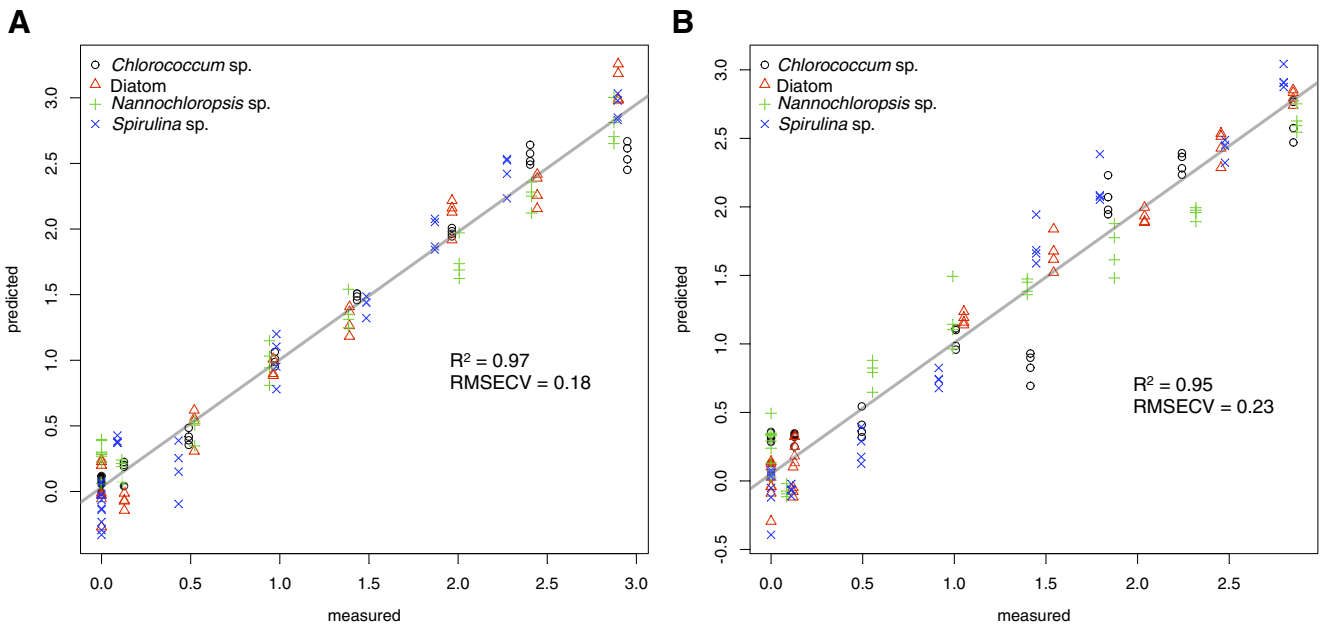


**Fig. 6** Predicted versus measured plots of the validation of the prediction of triglyceride (**a**) and phospholipid (**b**) spike concentrations based on a PLS2 regression model using NIR spectra of four species. $R^2$ and RMECV values illustrate the quality of the correlation

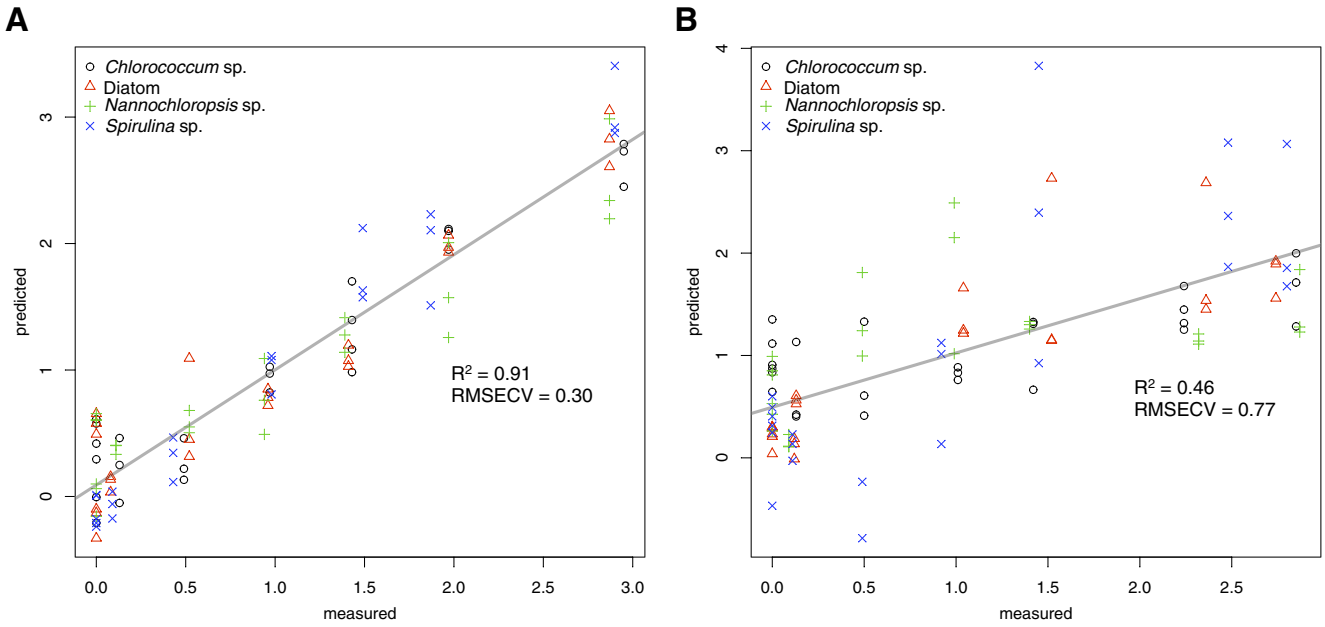


**Fig. 7** Predicted versus measured plots of the validation of the prediction of triglyceride (**a**) and phospholipid (**b**) spike concentrations based on a PLS2 regression model using FTIR spectra of four species. $R^2$ and RMECV values illustrate the quality of the correlation

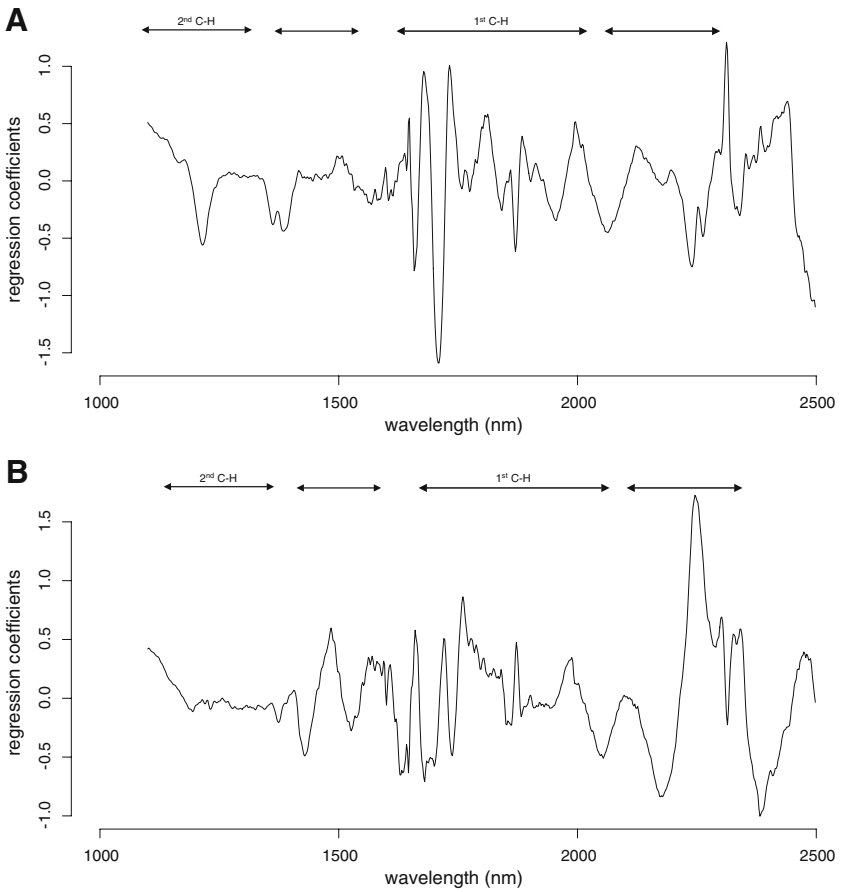


**Fig. 8** Weighted regression coefficients for the NIR calibration models, for the triglyceride (**a**) and phospholipid (**b**) spike concentration. The relative intensity of the peaks indicates the importance of these spectral regions in the contribution of those specific wavelengths to the overall calibration model. The spectral overtones that are overlapping with prominent peaks in the pure trilaurin and phosphatidylcholine (see Fig. 1) are indicated with *brackets*

**Table 4** Cross-species prediction of the lipid content in four different species using either individual species models or a combined, four-species model (COMB)

| Species | Model | Triglyceride | | | Phospholipid | | |
|---|---|---|---|---|---|---|---|
| | | Spike conc | Predicted±deviation | | Spike conc | Predicted±deviation | |
| | | | NIR | FTIR | | NIR | FTIR |
| *Nannochloropsis* sp. (N) | N | 2.41 | 2.71±0.17 | 2.77±0.17 | 1.87 | 1.51±0.24 | 0.82±0.48 |
| | C | 2.41 | 1.45±3.14 | 6.18±2.74 | 1.87 | 10.31±6.83 | 4.02±3.07 |
| | S | 2.41 | 9.48±4.73 | 5.39±1.13 | 1.87 | 14.07±5.90 | 2.12±2.05 |
| | D | 2.41 | 4.86±3.37 | 4.98±3.55 | 1.87 | 8.28±2.92 | 14.13±6.34 |
| | COMB | 2.41 | 2.49±0.11 | 2.49±0.15 | 1.87 | 1.73±0.15 | 0.79±0.35 |
| *Chlorococcum* sp. (C) | N | 2.40 | 2.63±3.10 | −6.58±4.76 | 1.84 | −6.70±4.31 | −2.93±13.37 |
| | C | 2.40 | 2.58±0.11 | 2.88±0.48 | 1.84 | 1.65±0.24 | 0.62±0.54 |
| | S | 2.40 | 9.95±4.15 | 4.86±1.71 | 1.84 | 5.36±5.18 | 2.89±3.10 |
| | D | 2.40 | 5.80±4.04 | 4.02±5.60 | 1.84 | 0.78±3.50 | 9.52±10.01 |
| | COMB | 2.40 | 2.26±0.15 | 1.95±0.34 | 1.84 | 2.25±0.20 | 0.74±0.78 |
| *Spirulina* sp. (S) | N | 2.27 | −17.6±11.83 | −12.08±6.55 | 1.79 | −6.90±16.47 | −8.71±18.41 |
| | C | 2.27 | −23.1±7.61 | 3.52±2.74 | 1.79 | 0.29±16.57 | 2.90±3.07 |
| | S | 2.27 | 2.80±0.48 | 2.33±0.24 | 1.79 | 1.16±0.60 | 1.90±0.44 |
| | D | 2.27 | −1.55±5.98 | −0.01±4.03 | 1.79 | 7.43±5.18 | 9.15±7.21 |
| | COMB | 2.27 | 3.41±0.50 | 2.90±0.28 | 1.79 | 2.00±0.67 | 1.74±0.64 |
| Diatom (D) | N | 2.44 | −13.2±6.61 | −3.31±6.95 | 1.95 | 11.52±9.20 | 0.44±19.54 |
| | C | 2.44 | −14.7±6.67 | 7.46±5.36 | 1.95 | 41.08±14.52 | −6.17±6.01 |
| | S | 2.44 | −2.90±7.33 | 6.62±1.98 | 1.95 | 18.95±9.15 | −7.58±3.59 |
| | D | 2.44 | 2.62±0.15 | 2.09±0.27 | 1.95 | 1.47±0.13 | 1.56±0.48 |
| | COMB | 2.44 | 2.34±0.31 | 2.07±0.23 | 1.95 | 1.18±0.42 | 1.43±0.53 |

The spike concentration (spike conc) is shown together with the predicted concentration (±deviation) of the spike level in algal biomass using either NIR or FTIR models for triglyceride and phospholipid content of the biomass. The deviations shown are calculated from the validation variances, residual variances, and the leverage of the variable data in the prediction objects, based on an empirical formula (U-deviation) built in the software package

*N* nannochloropsis model, *C Chlorococcum* model, *S Spirulina* model, *D* diatom model

analysis. We found that the predicted values using the NIR replicate spectra varied on average 4% and 9% of the predicted value of triglyceride and phospholipids respectively. For FTIR predictions, this variation between replicate spectra increased to 11% and 31% of the triglyceride and phospholipid predicted values. This larger variation in FTIR predictions indicates that FTIR may be more sensitive to sample homogeneity. Also, the poor phospholipid calibration model (Fig. 6) could contribute to the large variation seen in predicted values for the phospholipid concentration, indicating the variation is mostly due to inconsistencies in the performance of the model rather than due to problems with the homogeneity of the biomass samples.

## Discussion

The work presented here illustrates the feasibility and potential for using NIR and FTIR spectroscopic fingerprinting of algal biomass for predicting the lipid content and composition. As far as we know, this study is a first in the application of IR spectroscopy coupled with chemometrics for lipid analysis in algal biomass. We have used biomass from four phylogenetic divisions of algal species to test the applicability of a prediction model across species and divisions of algae.

Two lipids, a triglyceride (trilaurin) and a phospholipid (phosphatidylcholine) were used for this study. Both lipids have distinct fingerprints in both the NIR and FTIR spectra, indicated by characteristic lipid-specific absorption bands.

We demonstrate the possibility of a chemometrics approach to calibrate lipid spikes in algal biomass. We used PLS regression algorithms for calculating our calibration models. When comparing PLS1 and PLS2 prediction models we found no significant differences in quality of the prediction. The ability to distinguish both lipids simultaneously from one set of data (as is the case in PLS2

regression calculations) indicates that both lipids have sufficiently different fingerprints in both the NIR and FTIR spectra.

The variation present in the NIR spectra was analyzed by PCA (Fig. 3), which indicated that spectra from *Nannochloropsis* sp. and *Chlorococcum* sp. are more closely related along PC1 and PC2 compared with the diatom and *Spirulina* sp. biomass. This observation is consistent with *Nannochloropsis* sp. and *Chlorococcum* sp. having a more closely related phylogenetic relationship. This will become important when we discuss the possibilities for cross-species prediction of the lipid content (Table 3). This close relationship was not observed in the FTIR spectra PCA plot.

The single-species models indicate that it is possible to correlate the exogenous lipid content in biomass of all four species. Furthermore, eliminating the visible region of the NIR spectra does not significantly change the model statistics ($F$ test) for all species apart from the Diatom TG correlation. It is not clear why this TG correlation is affected by the elimination of the visible region of the spectrum.

When we investigated the effect on the prediction quality of different pretreatments and derivatives we found significant changes in the accuracy of the prediction of both NIR and FTIR spectra. We found that for NIR calibration models, MSC in combination with the exclusion of the visible region of the spectrum yielded the best performing calibration models ($R^2$= 0.969 and RMECV=0.182% for trilaurin and $R^2$=0.951 and RMECV=0.226% for phosphatidylcholine). Whereas for FTIR models, no mathematical pretreatment improved from the models built using the raw spectra.

The spectral derivatives (1st and 2nd 3-point) did not affect the quality of the model in the same way for all NIR and FTIR models. The NIR triglyceride model significantly improved and the FTIR model did not significantly change when the 1st derivative of the spectra was used (Table 3). The phospholipid models significantly deteriorated when using derivatives of the spectra for both NIR and FTIR. An explanation for the reduced accuracy is the loss of spectral information with every derivative from the original spectrum. The influence of single wavelengths on the calibration computation could be lost upon taking a derivative of the spectra. We did notice that the complexity of the models changed with the derivative spectra, i.e. the optimum number of components reduced significantly when the models were built using 1st and 2nd derivative spectra.

We have reduced the species-specific influences on the prediction model by combining all spectra and calculated a multiple-species calibration model. Although the individual species models were shown to be highly accurate and show a robust prediction, we were most interested in developing a model that can predict lipids across algal species from different phylogenetic divisions. Based on the inter-species variation in the NIR and FTIR spectra shown in Figs. 2 and 4, it is likely that this could make it difficult to predict lipids across species. Therefore, we tried to reduce the species-specific influences as well as increasing the influence of the lipid fingerprints on the model by combining the datasets from all four species prior to building the calibration model. To further reduce the species-specific influences we have excluded the visible region of the NIR spectrum from regression calculations and found a slight improvement of the correlation. Algal photosynthetic pigments cause the majority of the NIR reflectance in the visible spectrum. The lack of statistical improvement of the prediction with the exclusion of the visible spectrum indicates that this part of the spectrum does not contribute significantly to the prediction model.

We obtained $R^2$ values of 0.97 and 0.95 for NIR models for triglyceride and phospholipid respectively. For FTIR models, these values are 0.91 and 0.46 for trilaurin and phosphatidylcholine respectively. It is not clear why the phospholipid correlation is significantly lower compared to the triglyceride spike in the FTIR models. One possible explanation is spectral interference from triglycerides in the same region of the spectrum, which has been reported to mask the detection and quantification of phospholipids [15]. The strongest absorption bands of triglycerides and phospholipids are found in the carbonyl region of the spectrum, 1,742 and 1,737 $cm^{-1}$, respectively (Fig. 1). The proximity of these wavelengths causes each of the lipids to be identified and quantified individually, but multi-component mixtures are practically impossible to separate [15]. In our case, we used computationally powerful methods, the PLS2 multivariate calibration; however, these were not able to accurately quantify the phospholipid concentration in our double spiking experiments.

We identified important spectral regions by plotting the regression coefficients, illustrating the areas of the spectrum that contribute most to the regression calculations. The regression coefficients for the triglyceride and phospholipid spike are distinct, indicating characteristic regions of the spectra are responsible for the calibration of either type of lipid. This is consistent with the previously demonstrated distinct NIR and FTIR fingerprints of trilaurin and phosphatidylcholine (Fig. 1). Interestingly, the highest peaks in the regression coefficients plot correspond to peaks in the pure component spectra for both trilaurin and phosphatidylcholine. These regions in the spectra allow us to discriminate quickly between triglycerides and phospholipids in algal biomass. Very different lipid types will likely generate more distinct respective lipid fingerprints and more robust calibration models. Furthermore, when comparing the regression coefficients of the FTIR models it is clear that distinct regions of the NIR spectrum are

contributing to the quantitative prediction model of triglyceride and phospholipid content. These distinct coefficients overlap with characteristic lipid absorption band of the pure compound spectra (Fig. 1). The regression coefficients for both spiked lipids (Fig. 9) show significant similarities for the FTIR models, which could partly explain the difficulties in distinguishing the two spikes.

To verify the quality of the prediction models generated, we predicted test spectra obtained from spiked biomass that was not included in the calibration models. The results show that individual species models are accurate in predicting the level of spike for the respective species spectra. However, if alternative species models are used, the cross-species prediction has considerable errors in the prediction (large confidence interval). In some cases the deviations are an order of magnitude greater than the predicted value. This limitation is true when using both NIR and FTIR calibration models. However, we found that using N and C NIR models, it is possible to cross-predict both *Nannochloropsis* sp. and *Chlorococcum* sp. spectra (Table 4). This is not the case for FTIR Nannochloropsis and Chlorococcum models, where only the respective species models provided adequate predictions (Table 4). The redundancy in prediction between *Nannochloropsis* sp. and *Chlorococcum* sp. could be explained by NIR spectral similarities illustrated in the PCA plot of Fig. 3 where the scores plot of first two principal components (PC1 vs PC2) is shown for all four species. The overlap could be based on morphological and pigment similarities between the two species. Surprisingly, this same overlap is not present when the FTIR spectra are compared (Fig. 5).

As mentioned previously, the aim was to reduce the species-specific influences on the models' accuracy. We achieved this through the creation of a combined multiple-species model (COMB). To test the true species-independent nature of the COMB NIR and FTIR models we applied the COMB model on the independent test spectra we generated. We found that the combined model provided accurate predictions for both spike concentrations in all four species (Table 4), whereas single-species models were limited to their respective biomass samples.

A comparison of the average deviation of the predictions between NIR and FTIR indicates that NIR is more accurate and robust compared with FTIR. It is not clear why this difference exist, one likely explanation is that FTIR spectra are more complex and there are more species-specific differences still present in the spectra making the model less robust across species. This less robust nature of the FTIR models suggests that a NIR model may be preferable to develop further. The good correlations of the individual species models indicate that if only one species is studied, it may be preferable to develop a dedicated prediction model for this application.

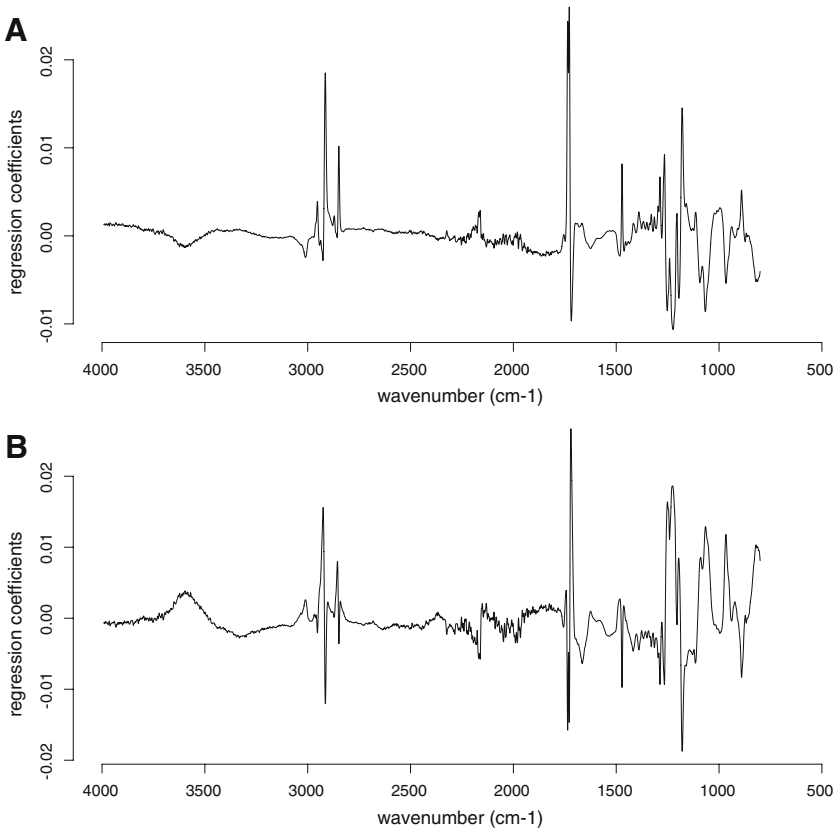


**Fig. 9** Weighted regression coefficients for the FTIR calibration models, for the triglyceride (**a**) and phospholipid (**b**) spike concentration. The relative intensity of the peaks indicates importance in the contribution of those specific wave numbers to the overall calibration

In summary, the work presented here shows that triglycerides and phospholipids have sufficiently different NIR spectral fingerprints to contribute independently to a PLS2 calibration model. We demonstrated that single-species calibration models are accurate in the prediction of spiked levels of both types of lipids. We improved upon the single-species calibration by developing a combined calibration model including datasets from all four species spanning four divisions of microalgae. We used this combined calibration model to predict the levels of spikes in biomass from the four species of algae and found accurate predicted values. All the experiments so far were based on exogenously added lipids to algal biomass. Although the results are promising, it is necessary to report on the possibility of building a prediction model on measured concentrations of lipids in algal biomass.

Future work will include testing and developing similar models on growing algal cultures. It is anticipated that the presence of water will affect the quality of the spectra collected by NIR and FTIR and therefore influence the subsequent multivariate calibration. However, before these models can be used on a routine basis to screen strain collections or monitor growth, more model development is necessary. For each new application, a robust calibration model needs to be developed before it can be applied to unknown samples. This work will form the focus of the next part of this work.

**Acknowledgements** This work was supported by the US Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory, through the NREL Laboratory Directed Research and Development (LDRD) program. We would like to acknowledge the excellent technical assistance from Corinne Feehan with sample preparation and NIR and FTIR spectroscopy, Amie Sluiter and Stuart Black for advice on the NIR and FTIR spectroscopy, respectively, Al Darzins for help in obtaining the algal biomass needed for this study. We are grateful to Eric Jarvis, Al Darzins and Amie Sluiter for their critical review of the manuscript.



## References

1. Burns DA, Ciurczak EW (2001) Handbook of near-infrared analysis. Marcel Dekker, New York
2. Chapman D (1965) The structure of lipids by spectroscopic and X-ray techniques. Wiley, New York
3. Chen W, Zhang CW, Song LR, Sommerfeld M, Hu Q (2009) A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae. J Microbiol Meth 77(1):41–47
4. Cooksey KE, Guckert JB, Williams SA, Callis PR (1987) Fluorometric determination of the neutral lipid content of microalgal cells using Nile red. J Microbiol Meth 6:333–345
5. Dean AP, Sigee DC, Estrada B, Pittman JK (2010) Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. Bioresour Technol 101(12):4499–4507
6. Elsey D, Jameson D, Raleigh B, Cooney MJ (2007) Fluorescent measurement of microalgal neutral lipids. J Microbiol Meth 68 (3):639–642
7. Esbensen KH (2002) Multivariate data analysis—in practice: an introduction to mutlivariate data analysis and experimental design. CAMO Process AS, Oslo
8. Goulden CH (1952) Methods of statistical analysis. Wiley, New York
9. Greenwell HC, Laurens LML, Shields RJ, Lovitt RW, Flynn KJ (2010) Placing microalgae on the biofuels priority list: a review of the technological challenges. J R Soc Interface 7(46):703–726
10. Hames B, Thomas S, Sluiter A, Roth C, Templeton D (2003) Rapid biomass analysis. Appl Biochem Biotechnol 105(1):5–16
11. Hirschmugl CJ, Bayarri ZE, Bunta M, Holt JB, Giordano M (2006) Analysis of the nutritional status of algae by Fourier transform infrared chemical imaging. Infrared Phys Technol 49 (1-2):57–63
12. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M et al (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant J 54:621–639
13. Ismail AA, Nicodemo A, Sedman J, van de Voort FR, Holzbaur IE (1999) Infrared spectroscopy of lipids: principles and applications. In: Hamilton RJ, Cast J (eds) Spectral properties of lipids. CRC Press LLC, Boca Raton, FL
14. Kansiz M, Heraud P, Wood B, Burden F, Beardall J, McNaughton D (1999) Fourier Transform Infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. Phytochemistry 52(3):407–417
15. Kisner HJ, Brown CW, Kavarnos GJ (1982) Simultaneous determination of triglycerides, phospholipids, and cholesteryl esters by infrared spectrometry. Anal Chem 54(9):1479–1485
16. Martens H, Martens M (2001) Multivariate analysis of quality: an introduction. Wiley, New York
17. Martens H, Naes T (1989) Multivariate calibration. Wiley, New York
18. McGinnis K, Dempster TA, Sommerfeld MR (1997) Characterization of the growth and lipid content of the diatom *Chaetoceros muelleri*. J Appl Phycol 9(1):19–24
19. Murdock JN, Wetzel DL (2009) FT-IR microspectroscopy enhances biological and ecological analysis of algae. Appl Spectrosc Rev 44:335–361
20. Naes T, Isaksson T, Fearn T, Davies T (2002) Selection of samples for calibration. A user-friendly guide to multivariate calibration and classifications. NIR, Chichester
21. Pienkos PT, Darzins A (2009) The promise and challenges of microalgal-derived biofuels. Biofuels, Bioprod Biorefin 3:431–440
22. Sheehan J, Dunahay T, Benemann JR, Roessler P (1998) A look back at the US Department of Energy's Aquatic Species Program—Biodiesel from Algae. National Renewable Energy Laboratory, Golden
23. Stacklies W, Redestig H, Wright K (2009) pcaMethods: a collection of PCA methods. R package version 1.22.0
24. R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: (http://www.R-project.org), accessed March 2009
25. Vieler A, Wilhelm C, Goss R, Sub R, Schiller J (2007) The lipid composition of the unicellular green alga *Chlamydomonas reinhardtii* and the diatom *Cyclotella meneghiniana* investigated by MALDI-TOF MS and TLC. Chem Phys Lipids 150(2):143–155
26. Wolfrum EJ, Sluiter AD (2009) Improved multivariate calibration models for corn stover feedstocks and dilute-acid pretreated corn stover. Cellulose 16(4):567–576
27. Zhang L, Garcia-Munoz S (2009) A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): a practitioner's perspective. Chemom Intell Lab Syst 97(2):152–158