Fourier Transform Infrared Spectroscopy Combined with Partial Least Squared Regression and Standard Addition-Net Analyte Signal Method for Quantification of Chitin in Insect-based Fish Feeds

By

Tsegu Lijalem Gezahegn

Thesis Submitted for the Degree of European Master's in Quality in Analytical Laboratories, Department of Chemistry, University of Bergen

Bergen, Norway

September 2018

Fourier Transform Infrared Spectroscopy Combined with Partial Least Squared Regression and Standard Addition-Net Analyte Signal Method for Quantification of Chitin in Insect-based Fish Feeds

By

Tsegu Lijalem Gezahegn

Thesis Submitted for the Degree of European master's in quality in Analytical Laboratories, Department of Chemistry, University of Bergen

Supervisors

Bjorn Grung, PhD

Professor, Department of Chemistry, University of Bergen

Egil Noodland, PhD

Chief Engineer, Department of Chemistry, University of Bergen

Pedro Araujo, PhD

Senior scientist, Institute of Marine Research

Bergen, Norway

September 2018

Abstract

Fourier transform infrared spectroscopy (FTIR) spectroscopy is a rapid and non-destructive technique applied with little or no sample preparation step. This work presents the use of FTIR spectroscopy combined with partial least squares (PLS) regression and standard addition based on the net analytical signal (SANAS) methods of multivariate calibrations to predict the quantity of chitin in fish diets. Besides, the attenuated total reflectance (ATR)-FTIR spectroscopy characterization study was conducted on chitin powders purified from shrimp and black soldier flies (BSF) exhibited close similarity in all infrared (IR) vibrational bands of the spectrum.

The estimation accuracy of PLS model was based on the root mean squared error of cross-validation (RMSECV), root mean squared error of prediction (RMSEP) and correlation coefficient (R^2) values. Precision (coefficient of variation, 5.64%), linearity, limit of detection (LOD = 0.026 % w/w) and limit of quantification (LOQ = 0.084 % w/w) were also estimated to validate the PLS multivariate calibration model. The correlation coefficient (R^2) and RMSECV between the predicted and measured percentages of chitin calibration sets were 0.982 and 0.220. RMSEPs for purified, raw, and demineralized BSF insect prediction samples were 20.7, 11.2, 1.53 % chitin (w/w) respectively. The % of chitin (w/w) predicted for purified, raw and demineralized BSF insect samples were 66.8, 31.9 and 16.9. In addition to this, 8.90, 8.60 and 10.51 % chitin (w/w) were estimated in three random insect-based fish diet test samples. Throughout the study, second-derivative Savitzky-Golay (15-point window size) third order polynomial expression was applied as a spectra data pre-processing method.

The net analyte signal (NAS) of analyte component is obtained by projecting the spectral data to the subspace orthogonal to the other components (interferents) in the mixture. The NAS vector of the analyte in the unknown sample and standard spiked with varying concentrations of the analyte were calculated. Plots of Euclidean norms of NAS vectors against the spiked concentration of standards have been done. The analyte in the unknown sample is estimated by extrapolating at the opposite of abscissa intercept. This linear standard addition plot is similar to the univariate calibration model which made it very easy to interpret the results. Several sources of chitin and types of fish feed were investigated. The correlation coefficients for all calibration models were above 0.9900, with the exception of purified chitin from black soldier fly (0.9641). The predicted weight percentage of chitin obtained in purified and raw black soldier fly powders were 9.32 and 1.66 whereas in one fish feed sample without insect added and other three insect-included fish feed samples were estimated, 1.07,1.65, 3.12 and 3.82 % chitin (w/w) respectively. The percentage of chitin quantified in the purified and raw black soldier flies by the use of net analyte signal method were not in agreement with the acid detergent fiber method.

Keywords: chitin; BSF insect-based fish feed; FTIR spectroscopy; multivariate calibration; PCA; PLS; SANAS.

Contents

Acknowledgments

This master thesis has been carried out at the department of chemistry, university of Bergen, Norway between October 2017 and September 2018. It is part of the Erasmus Mundus program master's projects in Quality in Analytical Laboratories. I would like to thank the following peoples and Universities contributed for the successful completion of the study.

First, I would like to thank Almighty God and his mother, Blessed Virgin Mary who has been my help and strength throughout my entire life.

Second, my appreciation goes to my supervisors, Bjorn Grung and Egil Noodland (Professors at the department of chemistry, University of Bergen) and Pedro Araujo (Senior Scientist at the Norwegian Institute of Marine research, Bergen) for their continuous advise in each step of the study. I am very thankful to Professor Bjorn Grung for his help starting from the admission and visa application to the completion of the thesis work. He allowed me to discuss with him on the progress of the study. Professor Egil Noodland has trained me about FTIR instrument working principles from the scratch. His office was always open when I have questions to ask him and to discuss the results and get the guidance in each step of the study. Professor Pedro Araujo has told me the overview of the project in the beginning and boosted my confidence to conduct the research independently. His friendly approach was very important to discuss the results openly with him and to ask questions. I am also very thankful for his continuous advises, suggestions and fast feedback during the write-up of the thesis paper.

I would also like to extend my special gratitude to Professor Clara Costa, director of 9th edition of EMQAL at the University of Algarve for nicely coordinating the program during the coursework. I would like to acknowledge the department of chemistry executive officers and members at the university of Bergen for giving me the services and assistances throughout my laboratory work.

It is also my pleasure to express my appreciation to classmates of the 9th edition of EMQAL for their good friendships and carrying out activities and group assignments together.

Finally, I would like to thank my loving wife, Tigst Gidey and my mother, Yajeb Getu for providing me moral and unending inspiration in the duration of two years.

List of abbreviations and acronyms

ADF	Acid detergent fiber
ATR	Attenuated total reflectance
BSF	Black soldier fly
CV	Coefficient of variation
DM	Dry matter basis
DRIFTS	Diffuse reflectance infrared Fourier transform spectroscopy
DTGS	Deuterated triglycine sulfate
EMSC	Extended multiplicative scatter correction
FTIR	Fourier transform infrared
HPLC	High performance liquid chromatography
IR	Infrared
IRE	Internal reflection element
IRM	Insect raw material
IUPAC	International Union of Pure and Applied Chemistry
KM	Kubelka-Munk
LC-MS	Liquid chromatography-mass spectrometry
LOD	Limit of detection
LOQ	Limit of quantification
NAS	Net analytical signal
PCA	Principal component analysis
PLS	Partial least squares
RMSEC	Root mean squared error of calibration
RMSECV	Root mean squared error of cross-validation
RMSEP	Root mean squared error of prediction
SAM	Standard addition method
SANAS	Standard addition-net analytical signal
SNV	Standard normal variate
VIP	Variable importance for prediction

List of tables

List of Figures

1. Introduction

Aquaculture is the fastest growing animal food production sector worldwide and is becoming the main source of aquatic animal food for human consumption (Ottinger, et al., 2016). To satisfy the continuing demand of fish and seafood for the rising global population, the aquaculture industry has relied on the production of fishmeal formulated to meet the nutritional requirement of targeted fishes. Fishmeal is generally added to animal diets to increase feed efficiency and growth through better feed palatability; it enhances nutrient uptake, digestion, and absorption (Ayoola, 2015). High availability of essential amino acids, phospholipids and fatty acids in fishmeal promotes optimum development, growth, and reproduction. However, the steady decline in catches of wild fish and increased demand of aquaculture feeds has resulted in a rapid decrease in availability of fishmeal and fish oil and their concurrent price increase (Henry, et al., 2015). It is estimated that approximately 30% of the total fish catch is converted to fish meal and fish oil for use in animal and fish feeds. Due to the tremendous expansion of aquaculture and limited resource available from the sea, it has been attracting an interest in fishmeal alternatives.

Nowadays, soybeans and other terrestrial plants rich in proteins and lipids have been introduced to replace fishmeal and fish oil (Mandrile, et al., 2018). However, the presence of antinutritional factors, problems of inflammation digestive tract, decreased palatability and price volatility of soybean meal are of a great concern (Henry, et al., 2015). To overcome this problem, insects are becoming the potential candidates as an animal protein source (Barroso, et al., 2014).

Insects are part of the natural diet of freshwater and marine fish, and they are generally rich in amino acids, fatty acids (monounsaturated and polyunsaturated), several micronutrients (Cu, Fe, Mn, Mg, P, Se and Zn) and vitamins (such as riboflavin, pantothenic acid, biotin, and folic acid). Since the nutrient composition of all insects is dependent on the feed and their stage of development, nutrient enrichment via the addition of certain food supplements, such as omega-3 (eicosapentaenoic acid 20:5ω-3; docosahexaenoic acid 22:6ω-3) is very important (Rumpold & Schlüter, 2013). Compared to conventional livestock, insects require less water input and can be reared in organic waste. This leads to high feed conversion efficiency, low greenhouse gas and ammonia emission, lower environmental footprint, few animal welfare issues, and reduced zoonosis transmission risks (Henry, et al., 2015; Magalhães, et al., 2017; Mandrile, et al., 2018).

At the beginning of the century, the European Commission Regulation 999/2001 has banned the use of the processed animal protein (PAP) in animal nutrition to protect from transmissible spongiform encephalopathies (TSE). However, the commission (i.e. Regulation (EU) N0 2017/893) has recently approved to uplift partially the feed ban rules regarding the use of insect processed animal proteins (PAPs) for aquaculture animals. The authorization is only limited to seven insect species: black soldier fly (BSF), house fly, yellow mealworm, buffalo worm, house cricket, banded cricket, and field cricket. From the approved list of seven insects, BSF larvae (*Hermetia illucens*) is the most promising insect species attracting commercial exploitation (Caligiania, et al., 2018). It processes organic waste very quickly, the adult flies do not eat, thus restraining bacterial growth and thereby reducing production of bad odour to a minimum. The larvae of BSF have been reported as a chief source of crude protein (42.1%) and found also about 56.9% in the defatted meal of BSF larvae which is comparable to soybean meal and slightly less than that of fishmeal. In addition, BSF larvae have better amino acid profiles and their fatty acid composition is dependent on the fatty acid composition of the diet (Makkar, et al., 2014).

A study carried out on feeding trail of juvenile turbot fish showed a decreased feed intake and growth performance with increasing incorporation of BSF larvae meal (>33%, DM) due to low palatability. The reduced palatability and growth performance resulted from decreased diet acceptability and feed intake that can be attributed to the presence of chitin, which influences lipid and protein digestibility (Kroeckel, et al., 2012). Li, et al., (2017) have reported that manifestation of dietary stress and intestinal histopathological damage of juvenile Jian carp fed with >75% defatted BSF larvae meal. In addition to this, a feeding trial of Atlantic salmon with 85% of BSF larvae to replace fishmeal protein and soy protein resulted in a significant reduction of apparent digestibility of protein, lipid and all amino acids. In a recent study, it was observed that insectmeal (lower than 600 g/kg) in combination with insect oil in the diets of fresh-water Atlantic salmon did not adversely affected the growth performances, feed utilization, apparent digestibility and whole-body composition (Belghit, et al., 2018). Moreover, the effect of dietary chitin on growth and nutrient digestibility in farmed Atlantic cod, Atlantic salmon, and Atlantic halibut was investigated. Atlantic cod and Atlantic halibut seem unaffected by up to 5% chitin addition in the diet, while in Atlantic salmon the growth and nutrient utilization are affected negatively by adding chitin in the diet at levels higher than 1% (Karlsen, et al., 2017).

On the contrary, another study demonstrated that chitin has protective effects against salmonids, enhance the respiratory burst, improve phagocytic activities in gilthead bream and significantly increases the white blood cell count in chitin-fed groups (Cuesta, et al., 2003). A dietary inclusion of mealworm (tenebrio molitor) meal on European sea bass (Henry, et al., 2018) and yellow catfish (Su, et al., 2017) showed improved immune response and bacterial resistance without any negative growth effects.

Even though the BSF larvae meal is permitted to be used as a fish meal alternative, an efficient infrastructure should be put in place to guarantee the quality and safety of final products. The ideal infrastructure should have a battery of validated analytical techniques to certify not only the quality and quantity of nutrients and antinutrients but also important insect components (e.g. chitin) generally overlooked by the insect-based food industry.

Chitin exists rarely in a pure form in nature but instead is usually in a complex matrix with other compounds. Its determination from insects has been based on acid detergent fiber (ADF) method assuming structural similarity between chitin and cellulose. The alleged structural similarity has resulted in an overestimation of chitin due to the substantial amounts of amino acids contained in the ADF fractions (Finke, 2007). Chitin content is quantified based on the difference from total nitrogen content after subtracting the protein contribution to total nitrogen. A specific nitrogen conversion factor of 14.5 was calculated for chitin, assuming a fully acetylated glucosamine (Caligiania, et al., 2018).

A previous study has reported an acid hydrolysis-based extraction procedure for glucosamine from BSF larvae meal to determine chitin using liquid chromatography-tandem mass spectrometry (LC-MS) (Tefera, 2017). This method was based on the hydrolysis of a commercial purified standard from shrimp shells (~100% purity) by using a concentrated hydrochloric acid. The shrimp shell commercial standard was used for constructing the calibration curve and for quantifying chitin in purified from BSF larvae (~ 80-90%). Unfortunately, there was a remarkable disagreement between the declared (~ 80-90% chitin) and predicted (12%) levels of chitin which could be the result of an incomplete hydrolysis of chitin in the purified sample for BSF larvae.

Accurate detection and quantification of chitin is of paramount. The implementation of acid hydrolysis and LCMS for analysis of chitin (Tefera, 2017) has demonstrated that many

challenges lie ahead in the field of sample treatment and selection of reliable instrumental techniques. In the context of the aquaculture industry and its search for alternative source of proteins (e.g. insects), it is crucial the implementation of fast analytical techniques, able to overcome the limitations of LC-MS. Vibrational spectroscopic techniques (e.g. Fourier transform infrared (FTIR)) in combinations with chemometrical tools might constitute reliable quantitative approaches. To the best of my knowledge, the present thesis is the first scientific report describing the use of FTIR spectroscopy combined with chemometrics for quantification of chitin in an insect-based fish diet. Therefore, the aim of this study is to evaluate the potential of FTIR spectroscopy combined with standard addition and net analytical signal based multivariate calibrations to quantify chitin from the insect-based fish diet.

1.1. Objective of the project

- To characterize chitin from a commercial standard (shrimp) and a purified BSF insect chitin by using FTIR spectroscopy operated in Attenuated Total Reflectance (ATR) mode in order to verify whether the structures are similar or not.
- To optimize the parameters affecting the determination of chitin by FTIR
- To evaluate under optimal conditions the potential of FTIR operated in Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) as a quantitative technique for determining chitin in insect-based fish feed by means of the standard addition method and multivariate calibration models.

1.2. Significance of the project

Black soldier fly (BSF) is one of the insects on an authorized list amended in January 2018 for the European Commission (EC) Regulation No 999/2001, which had been prohibiting the production of proteins derived from farmed insects. Insect-based fish diets are recently attracting the aquaculture industry in many countries as a potential fish feed alternative. Since the cost of cultivation of BSF insect larvae is possible at a lower price, the proteins are chiefly available in it and the cost of fish feed is increasing year to year due to the increasing human consumption of fish, the regulatory bodies for aquaculture in Norway and the European Commission are striving to replace fish feed with BSF insect-based meals to improve the nutritional values of fish. However, the quality requirements regarding the non-nutrient materials composition, for example percentage of chitin tolerated in the feed, have not yet been addressed. The regulatory bodies are seeking fast methods that can easily quantify the percentage of chitin in insect-based fish diets. Hence, this project has contributed a tremendous effort to develop multivariate calibration models that can be easily applied for quantification of chitin by the use of DRIFTS-FTIR spectroscopy.

2. Theoretical background

2.1. Chitin

Chitin, the second most abundant biopolymer after cellulose, is usually present in a mixture with other polysaccharides, lipids and proteins in exoskeletons of arthropods (insects, crustaceans, arachnids and myriapods), cuticles of nematodes, algae and cell walls of fungi (Majtan, et al., 2007; Kaya, et al., 2015).

Chitin is a linear biopolymer of β-(1→4)-N-acetyl-D-glucosamine repeating unit. The structure of chitin is similar to cellulose except for the hydroxy group at C-2 position in cellulose is substituted by acetamido group in chitin (

Figure 1).

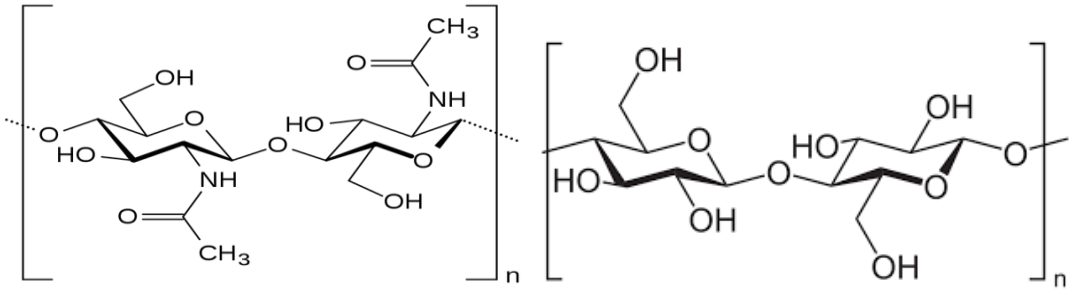


Figure 1. Chemical structure of chitin (left) and cellulose (right).

Chitin exists in three crystalline polymorphic forms designated as α, β and γ (Kaya, et al., 2017). The α-chitin present in anti-parallel arrangement is the most abundant and stable polymorphic form found in shells of crustaceans, the skeletons of insects and cell walls of fungi. Chitin in β-form has a parallel arrangement and is packed with weak intramolecular hydrogen bonds. It is found in squid pens, in the extracellular spines of the euryhaline diatom, and in pogonophore tubes. The γ-crystalline form has a mixture of antiparallel and parallel chains and is found in the cocoons of insects (Gonil & Sajomsang, 2012; Sajomsang & Gonil, 2010) such as moth (*Orgyia dubia*), threads of larvae of the spider beetle, Ptinus tectus, larval and adult peritrophic membranes (PM) of adult locust, cockroach, mantis and dragon fly silkworm larva (*Antheraea pernyi*) and in a sawfly larva (*Phymatocera aterrima*) (Kaya, et al., 2017).

Presence of amine functional groups in chitin and its partially deacetylated derivative chitosan represents a great advantage for modification reactions. Despite of its insolubility in common

organic and inorganic solvents, chitin has excellent properties: biocompatibility, biodegradability, bioactivity, non-toxicity and good adsorption properties means that this biomaterial has drawn a great deal of industrial attention (Islam, et al., 2016). Because of these properties, chitin and its derivatives have plenty of applications in wastewater treatment, pharmaceuticals, biomedical engineering, biotechnology and food science. The surface morphology, acetylation degree and molecular weight are the three main criteria determining the industrial use of chitin and its derivatives (Kaya & Baran, 2015; Nitschke, et al., 2011).

2.2. Structural characterization and quantification techniques for chitin

Several analytical techniques have been developed for detection and quantification of chitin and its derivatives. Gas chromatography-mass spectrometry in selected ion monitoring mode has been used to detect and quantify acid-extracted hydrolysated monomer, D-glucosamine after using epimer-specific derivatization to confirm the presence of chitin in fossils of arthropods (Flannery, et al., 2001). Furthermore, high performance liquid chromatography (HPLC) analysis have been used to separate D-glucosamine after pre-column derivatization with 9-fluorenylmethoxycarbonyl chloride for quantifying and determining the purity of chitin in biological materials and evaluating the quality of chitin products. The sample preparation step used for chromatographic analysis, in general, is acid hydrolysis to cleave the β (1→ 4) glucosidic bonds and removal of the acetyl amide group to produce glucosamine for subsequent quantitation. This HPLC analysis has been developed to overcome the time-consuming pre-column derivatization methods with several stages of vacuum drying and less reproducible methods of spectrophotometry. However, the hydrolysis of chitin with concentrated HCl (8 M) into glucosamine is not ideal as it can possibly breakdown the glucosamine (Zhu, et al., 2005). Nitschke, et al., (2011) has determined chitin and chitosan from mushroom samples in a coloured polyiodide-chitosan insoluble complex obtained as result of specific reaction of polyiodide anions contained in Lugol's solution with chitosan. Lugol's solution in microscopy have been used as a specific dye for chitin in cell walls. A spot assay has been developed on thin layer chromatographic (TLC) plates and measured their optical density with photographic technique. Measurement of optical densities for standards of chitosan have used to prepare calibration curve for determination of chitosan and in turn for chitin in mushroom samples.

Fourier transform infrared (FTIR) and solid-state ^{13}C cross-polarization magic angle spinning nuclear magnetic resonance spectroscopy is commonly used for structural characterisation and determination of the degree of acetylation percentage of chitin (Majtan, et al., 2007; Sajomsang

& Gonil, 2010; Kaya, et al., 2015). Elemental analysis has been also used to characterize chitin and chitosan in which the degree of acetylation and deacetylation were determined by relating with percentage composition of carbon and nitrogen (Soon, et al., 2018).

Scanning electron microscopy is a powerful tool for characterizing surface morphology of chitin and chitosan extracted from orthoptera species (Kaya & Baran, 2015), cicada sloughs (Sajomsang & Gonil, 2010), shrimps, desert locusts, honey bees and beetles (Marei, et al., 2016). Chitins characterised with nanopores surface structures have applied in metal ion adsorption while those with fibrillar surface morphology can be used in textiles.

2.3. Principles of FTIR spectroscopy

The infrared (IR) portion of the electromagnetic spectrum is divided into three regions; the near-infrared (NIR), mid-infrared (MIR) and far-infrared (FIR). The mid-IR region ranges approximately from 4000-400 cm^{-1} and is used to study the fundamental vibrations (signatures) and associated rotational-vibrational structures (Huck, 2015). An infrared spectrum represents a fingerprint of a sample with absorption peaks that correspond to the frequencies of vibrations between the bonds of the atoms making up the material. Because each different material is a unique combination of atoms, no two compounds produce the exact same infrared spectrum. Therefore, infrared spectroscopy can result in a positive identification (qualitative analysis) of many different kinds of material. In addition, the size of the peaks in the spectrum is a direct indication of the amount of material present. With help of modern software algorithms, FTIR spectroscopy is an excellent tool for quantitative analysis (Stuart, 2004).

FTIR spectrometers work basically based on an interferometer to generate an interferogram. All wavelengths are measured simultaneously which results in faster sampling with a better signal-to-noise ratio. An interferometer is an optical device that allows the controlled generation of interference patterns or interferograms. The most commonly used interferometer is Michelson interferometer. The Michelson interferometer consists of a source, a beam splitter, a fixed mirror, and a moving mirror (**Figure 2**) (Subramanian & Rodriguez-Saona, 2009).

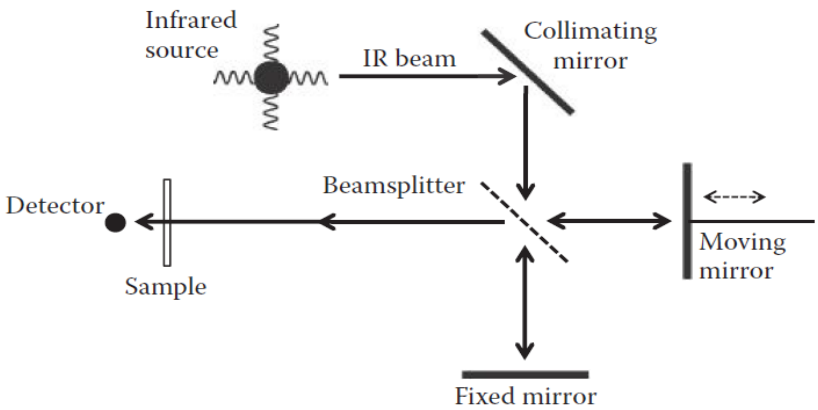


Figure 2. Schematic diagram of Michelson interferometer.

The source emits light in the IR region when electricity is passed through it. The beam splitter serves to split the incident IR light into two. The mirrors are aligned to reflect the light waves in a direction that would allow recombination of the waves at the beam splitter. The movable

mirror is capable of moving along the axis, away from and towards the beam splitter. One half of the light passes through the beam splitter and is reflected by a stationary mirror back to the beam splitter. The other half of the light is reflected on to the moving mirror, which in turn reflects the light back to the beam splitter. The two reflected beams from the mirrors recombine at the beam splitter. The difference in distance travelled by the two light beams is the optical path difference. The recombined light beam passed from the interferometer and it is then directed by mirrors into the sample compartment and detected by the detector (Subramanian & Rodriguez-Saona, 2009). FTIR uses Fourier transform modern mathematical algorithm to convert the raw wavelength data collected by the detector into spectra (Beasley, et al., 2014).

The three conventional testing modes of FTIR techniques for bulk sample analysis are transmission FTIR, ATR-FTIR and DRIFT spectroscopies. The transmission mode is fast and cost-efficient technique applied by mixing solid powder samples with KBr and pressing with the hydraulic press to make pellets. KBr is used for sample preparations as a background matrix because it has a wide spectral range, no significant wavelength in MIR region and produces smooth IR transparent disc upon mixing with powder samples. The sample pellet is placed in the path of the IR beam and the resulting transmitted IR signal is recorded by the detector (Chen, et al., 2015). The main solvent used in the preparation of the sample is water. Signal saturation due to -OH stretching vibrations leads to a non-linear detector response and makes the subtraction difficult. Even though the spacer thickness is reduced to less 1 µm to avoid saturation effect, it is very difficult to ensure the reproducibility of the spacer thickness. So, ATR-FTIR method can be an alternative solution to overcome this difficulty (Grdadolnik, 2002).

Attenuated total reflectance (ATR)- FTIR spectroscopy involves directing an infrared beam at an interface between an internal reflection element (IRE) and a sample placed in contact with the IRE at a certain angle of incidence as shown in **Figure 3**. This IRE is an optically dense crystal made of ZnSe, diamond, Si or Ge with a higher refractive index than the sample or the surrounding medium. Total internal reflection occurs when the angle of incidence of IR beam exceeds the critical angle. This internal reflectance causes a generation of standing waves of radiation called an evanescent wave that extends beyond the surface of the crystal into the sample held in contact with the crystal (Houa, et al., 2018). An ATR spectrum can be obtained by measuring the interactions of the evanescent wave with the sample. This evanescent wave will be absorbed by the sample and its intensity is attenuated in regions of the IR spectrum where the sample absorbs.

In general, the penetration depth is defined by the amplitude decrease of the electromagnetic field in the optically less dense material to the fraction 1/e of its initial level.

$$d_p = \frac{\lambda}{n_1 . \, 2\pi \sqrt{\sin^2\theta - (\frac{n_1}{n_2})^2}} \qquad \text{Eq. 1}$$

The depth of penetration (d_p), is proportional to the wavelength λ and also depends on the angle of incidence θ and the refractive indices of the crystal n_1 and the sample n_2. Upon decreasing the angle of incidence and refractive index of the crystal results in a shallower depth of penetration. The depth of penetration is limited only to a fraction of the wavelength, typically ranges from 0.5 – 2 µm although it should be noted that absorbance information is obtained from the sample beyond this range, as the evanescent wave probes deeper than the depth of penetration. The detection power of the ATR technique is increased by employing a trapezoidal crystal which allows multiple reflections (Schmitt & Flemming, 1998).

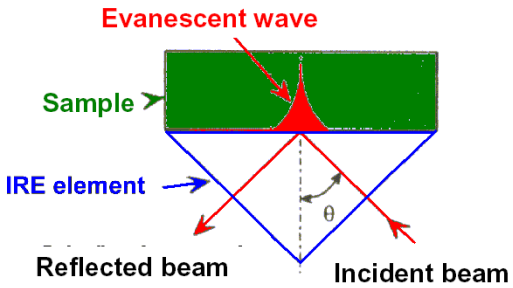


Figure 3. Illustrations of the principles of ATR spectroscopy.

Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy is suitable for analysis of organic and inorganic samples ground into fine powders (< 10 µm size). It is a non-destructive FTIR technique that needs little or no sample preparations, and a fast and easy clean up. The principles of DRIFT-FTIR spectroscopy is demonstrated in **Figure 4**. This technique is designed for analysis of samples with rough surfaces. It operates by directing the IR beam into a sample cup filled with a mixture of sample and an IR transparent matrix (such as KBr and KCl). The IR radiation interacts with the particles and then reflected off their surfaces (specular reflection), causing the light to diffuse (diffuse reflection), or scatter, as it moves throughout the sample (Beasley, et al., 2014). The output mirror then directs this scattered light energy to the IR detector, where quantitative information from the chemical components is measured. The penetration depth into the sample is dependent on the physical and optical properties of the sample and diluting powder. These properties are refractive index, particle dimensions, packing density, homogeneity, concentration and absorption coefficients that influences the

quality of the DRIFT spectra. Samples are generally diluted 10-100 times to guarantee deeper penetration of the IR beam and less specular reflection from the sample surface, thereby increasing the contribution to the spectrum of components which contain data on the absorbance properties (Armaroli, et al., 2004).

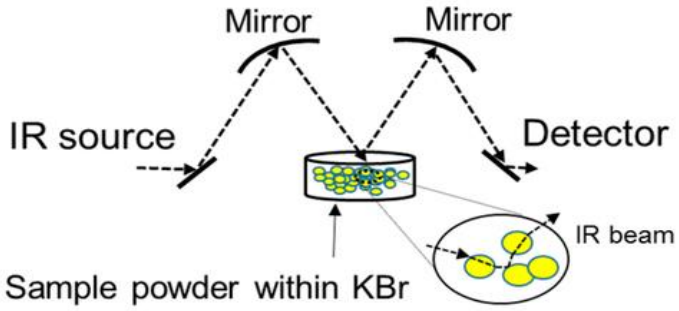


Figure 4. Simplified schematic diagram for diffuse reflectance FTIR spectroscopy (Chen, et al., 2015).

Diffuse reflection spectra are very similar to transmission spectra. However, diffusely reflected light undergoes repeated transmission through the sample, thus the low absorption bands are emphasized as compared to the transmission spectrum of the same sample. Ideally, there is no linear relation between a band intensity and concentration which makes DRIFTS method more complicated for quantification. The Lambert-Beer law applied in the transmission spectrum is not applicable for diffuse reflection (Armaroli, et al., 2004; Soares, 2014). The spectra are largely influenced by several experimental parameters such as, shape and size, refractive index, absorption characteristics, and porosity of the powder particles. An approximation of the Kubelka-Munk (KM) function is widely applied for quantitative analysis though it is questionable of its applicability and accuracy for non-absorbing and highly absorbing samples (Jinda Sirita, 2007). The Kubelka-Munk formula (Eq. 2) relates the absolute reflectance (R_∞) of the scattered radiation (R_∞ = intensity scattered divided by that of the incident radiation) to the sample absorption and scattering coefficients k and s, respectively.

$$f(R) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} \qquad \text{Eq. 2}$$

Where, k is molar absorption coefficient, $k = 2.303\alpha c$, α is absorptivity, c is concentration and s is the diffusion (scattering) coefficient.

2.4. Chemometrics

The term chemometrics was first used by Svante Wold and Bruce Kowalski back in 1972 (Hibbert, 2016). Chemometrics is a discipline concerned with the application of statistical and mathematical methods as well as those methods based on mathematical logic, to chemistry (Brown, et al., 1992). It improves the understanding of chemical information and to correlate quality parameters or physical properties to analytical instrument data. When working with spectra data containing complex chemical composition such as in biological samples, herbal medicines and drug development, it is not easy to apply the commonly used univariate methods. Application of chemometric methods provide a good opportunity for mining more useful chemical information from the original complex data. With the advancement of computer technologies, chemometric methods have become a leading tool among the scientific communities towards faster analysis results and shorter product development time (Bansal, et al., 2014).

Multivariate calibration is one of the frequently used chemometric tools. It refers to the process of relating analyte concentration or the measured value of a physical or chemical property to a measured response such as NIR spectra of multicomponent mixtures (Lavin, 2000). Multivariate calibration has been applied in many different fields, especially in food chemistry, pharmaceutical analysis, agriculture, environment, industrial and clinical chemistry.

The reason of the large interest in multivariate calibration is that the method is fast, cheap and accurate for many real problems (Forina, et al., 2007). It was an immediate success where selectivity and precision increased dramatically in comparison with the traditional scientific approach of using just one or two variables are examined at a time. The reliability of results for unknown samples is higher than the univariate calibration in which it can detect the presence of an uncontrolled interference errors in samples. Moreover, a large number of variables are used for the analytical signal and examining the residuals can expose whether a given sample is different from those used for calibration and the result that it provides is thus unreliable (Romía & Bernàrdez, 2009).

2.4.1. Data pre-processing

Data preprocessing (Zeaiter & Rutledge, 2009) is widely used for spectroscopic applications such as NIR, MIR, UV and NMR to correct the spectral data for baseline drift, nonlinearity, additive and multiplicative effects, and other variations in the spectra not related to the

properties being studied. These methods may be used to improve the predictive ability of the multivariate calibration. In vibrational spectroscopy, extended multiplicative scatter correction (EMSC), standard normal variate (SNV) and spectral derivatives (Afseth & Kohler, 2012; FitzPatrick, et al., 2012) are commonly used pre-processing methods. The first two are model-based methods, which allow quantifying and separating the different types of physical and chemical variations in the spectra. Whereas, the spectral derivative methods transform spectra to a better-smoothed version of the same data by filtering out undesired types of variations. EMSC was introduced to correct problems of signal amplification and offset in spectra due to additive and multiplicative effects. It is done by calculating slope and offset correction coefficients of the regression line between each individual spectrum and the average spectrum of the calibration data set. However, the average spectrum is not always a representative spectrum unless a large number of samples is employed (Rinnan, et al., 2009). SNV and EMSC methods are linearly related and provide similar results (Romía & Bernàrdez, 2009). SNV is done by subtracting the mean from each spectrum and then dividing by the standard deviation of the spectrum (Zeaiter & Rutledge, 2009).

2.4.2. Principal component analysis

Principal component analysis (PCA) (Camacho, et al., 2010) is a method of exploratory data analysis that aims to find the subspace in the space of the variables where data mostly vary. The original variables, commonly correlated, are linearly transformed into a lower number of uncorrelated variables, called principal components (PCs). Thus, the original data set can be simplified by performing data reduction to make it more easily interpretable. Conventional notations have been used throughout this thesis. Boldface capital letters are used for matrices, a boldface lower case for a vector and lower-case italic for a scalar. PCA is used to reduce the number of variables by decomposing the spectral data matrix $\mathbf{X}$ into a product of two matrices, scores (**T**) and loadings (**P**) plus an error matrix (**E**), as shown in Eq. 3.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E} \qquad \text{Eq. 3}$$

Where $\mathbf{X}$ is the $i \times j$ data matrix, T is the $i \times A$ matrix of score vectors, containing A orthogonal score vectors $\mathbf{t}_a$, $\mathbf{P}^T$ is the $j \times A$ transposed matrix of loading vectors and the superscript "T" indicates the transposition, E is the $i \times j$ residual matrix, i is the number of objects (rows), j is the number of variables (columns), and A is the number of calculated principal components retained in the PCA model.

The number of principal components (PCs or A) used in the PCA model should be large enough to capture the systematic part of the variation in the data. This can e.g. be determined by looking at the percentage of the variance explained.

2.4.2. Partial least squares regression (PLSR)

Partial least square (PLS) (Valderrama, et al., 2007) is a well-known multivariate calibration method in chemometrics and its implementation might be valuable for the vibrational spectroscopic data to be generated in the present research for predicting percentage of chitin in insect-based fish diet samples. PLS regression methods can determine analytes in a situation where a selective signal is not possible to obtain. It is a very powerful and fast technique for process modelling and calibration methods where the predictor variables are collinear, measurement data contain noise, variables have high dimensionality and the number of samples is smaller than the variables (Xin Bao, 2009).

PLS calculates latent variables ($a = 1, 2, \ldots A$), which are a linear combinations of independent variables **X** that best predict the dependent variable **y**. This is done by focusing on the covariance between the matrix of **X**-block and **y**-vector. The relationship between the data matrix **X** and **y** is represented as a linear algebraic relation between their scores. The data matrix **X** is formed by the multivariate matrix (e.g., MIR spectra) and the vector **y** contains the reference values (e.g., concentration) (Eq. 4 and Eq. 5). The scores are obtained by decomposing the data matrices into a sum of rank one-component matrices (Valderrama, et al., 2007; Oliveira, et al., 2012).

$$X = \mathbf{T}\mathbf{P}^{T} + \mathbf{E} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{T} + \mathbf{E} \qquad \text{Eq. 4}$$

$$\mathbf{y} = \mathbf{T}\mathbf{q}^{T} + \mathbf{f} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{q}_a^{T} + \mathbf{f} \qquad \text{Eq. 5}$$

where **E** is an error matrix for **X** and **f** ($i \times 1$) is an error vector for **y** ($i \times 1$), that are not explained by the model. $\mathbf{t}_a$ is the column vector that comprise the score matrix **T**; $\mathbf{p}_a$ and $\mathbf{q}_a$ are loadings for **P** ($j \times A$) and **q** ($1 \times A$) loading matrices of **X** and **y** respectively, with A equal to the number of latent variables suitable to explain the variance of the variables.

Estimates for the parameter of interest ($\hat{\mathbf{y}}$), for a set of samples, are obtained by multiplication of the spectra by an appropriate regression vector, **b** ($j \times 1$) calculated as,

$$\hat{\mathbf{y}} = \mathbf{T}\mathbf{Q}^{\mathbf{T}} = \mathbf{X}\mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{Q}^{\mathbf{T}} = \mathbf{X}\mathbf{b} \qquad \text{Eq. 6}$$

Where, W is the weight vector. This is calculated in way to maximize the covariance between **X** and **y**. It gives information about how the variables combine to form the quantitative relation between **X** and **y**, thus providing an interpretation of the scores. Hence, these weights, $\boldsymbol{w}_a$ are essential for the understanding of which X-variables are important (numerically large $\boldsymbol{w}_a$-values), and which X-variables that provide the same information (similar profiles of $\boldsymbol{w}_a$-values). The desired number of latent variables are stored in a scores matrix **T** which is used to iteratively model the variables in **X** and **y** until convergence is reached. Similarly, the variables in **y** can be modelled from those in **X** via the matrix of regression coefficients **b**. The coefficients in **b** can be estimated as a function of the loadings of **X** and **y**, and P and Q, respectively, in addition to **W**.

The PLS scores and loadings differ numerically from PCA scores and loadings are dependent both on experimental measurements (e.g. spectra) and the concentrations. In PCA, the scores and loadings depend only on the spectra.
The multivariate calibration model is first constructed using calibration samples where the predictor and predictand (i.e. concentration) variables are known. The model can be validated by comparing the predictions against measured values for samples that were not used to build the model.

The accuracy of multivariate calibration models is expressed by RMSECV (root mean square error of cross-validation) obtained from the calibration set by internal validation (leave-one-out) and RMSEP (root mean square error of prediction), calculated from an external validation set of samples (Brereton, 2000). RMSECV and RMSEP are expressed by

$$\text{RMSE (CV or P)} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \qquad \text{Eq. 7}$$

where n is the number of samples, $\hat{y}_i$ is predicted value and y_i is the nominal value for i^{th} sample.

2.4.3. Standard addition method

When measurements are carried out in samples with complex matrices such as biological samples, the signal obtained is a contribution from both the analyte and the matrix of the sample, which in turn promotes spectral interferences. Spectral interferences occur when spectroscopic sensors are not completely selective for all analytes in chemical analysis. In composition analysis of an unknown mixtures, the effect of the matrix is a serious problem and

renders chemical analysis invalid (Hemmateenejad & Yousefinejad, 2009). Standard addition method (SAM) is most commonly applied to correct matrix effects, minimize spectral interferences and compensate instrumental drift in multicomponent analysis. The SAM generalized by Saxberg & Kowalski (1978) provides a means of detecting interference effects, quantifying the magnitude of the interferences, and simultaneously determining analyte concentrations. This generalized SAM overcomes the presence of spectral and matrix interference effects occurring in classical SAM. In fact, the strategies in classical univariate SAM is quite straightforward even in the presence of different analytes as far as their responses do not interfere with each other, but, the situation is somewhat different in spectroscopic measurements, which leads to interactions between species and signal overlap makes the method invalid (Martínez, et al., 2018).

So, the application of robust and versatile multivariate calibration models, such as PLS regression is desired to solve the problem of signal overlapping and matrix effects.

2.4.4. Net Analyte Signal Analysis

The Net Analyte Signal (NAS) of a component (Lorber, 1986), is the part of its spectrum which is orthogonal to the spectra of the other components. Whereas, the part of the spectra that is not orthogonal to the data of the other component is a linear combination of the spectrum of the others. The digitized spectrum is referred to as a spectrum vector or simply as a vector, while a spectrum vector of a pure component is called a component vector.

The NAS algorithm splits the data matrix **X** into two matrices: one containing the spectral information pertaining to the analyte k, **X**k, and the other containing the spectral information pertaining to all of the other variability sources, $\mathbf{X}_{-\mathrm{k}}$, including the contribution of the interferents ($\mathbf{X}_{\mathrm{k}} + \mathbf{X}_{-\mathrm{k}}$). The NAS vector for a given sample is calculated by projecting the spectrum vector **r** onto the space defined by the interferents (matrix $\mathbf{X}_{-\mathrm{k}}$), with the NAS being the orthogonal resultant ($\mathbf{r}_{\mathrm{k}}^{*}$) (Eq. 8) (He, et al., 2017).

$$\mathbf{r}_{\mathrm{k}}^{*} = (\mathbf{I} - \mathbf{X}_{-\mathrm{k}}\mathbf{X}_{-\mathrm{k}}^{+})\mathbf{r}_{\mathrm{k}} \qquad \text{Eq. 8}$$

Where, $\mathbf{X}_{-k}$ is the matrix of spectra all components other than the k^{th} analyte, $\mathbf{X}_{-k}^{+}$ represents the Moore–Penrose pseudo inverse matrix of $\mathbf{X}_{-k}$, the superscript ‘‘+’’ denotes the pseudo-inverse of a non-square matrix, **I** is an identity matrix having the same dimension as $\mathbf{X}_{-k}\mathbf{X}_{-k}^{+}$ and the $(\mathbf{I} - \mathbf{X}_{-k}\mathbf{X}_{-k}^{+})$ is a projection matrix that projects $\mathbf{r}_{k}$ onto the null space of the rows of $\mathbf{X}_{-k}$, which is the orthogonal complement of the column space of $\mathbf{X}_{-k}$.

Since we have the same data of **X** and the known concentrations of the k[th] analyte, c_k in **X**, it is possible to calculate the space of the spectra spanned by all constituents except the kth analyte with rank annihilation in the *A*-dimensional subspace (Eq. 9). Usually, the number of variables used to collect spectra in the MIR region exceeds the number of calibration samples. Due to the fact that only the significant latent variables are calculated by PLS regression or PCA when building the data matrix, $\mathbf{X}_{\mathrm{reb}}$ (Lorber, et al., 1997).

$$\mathbf{X}_{-\mathrm{k}} = \mathbf{X}_{\mathrm{reb}} - \alpha\hat{\mathrm{c}}_{\mathrm{k}}\mathbf{r}^{\mathrm{T}} \qquad \text{Eq. 9}$$

Where, $\hat{c}_k$ is the analyte concentration vector projected onto the A-dimensional subspace, as expressed in Eq. 10;

$$\hat{\mathbf{c}}_{\mathrm{k}} = \mathbf{X}_{\mathrm{reb}}\mathbf{X}_{\mathrm{reb}}^{+}\mathbf{c}_{\mathrm{k}} \qquad \text{Eq. 10}$$

The vector "**r**" is a linear combination of the rows of **X**, which should include the contribution of the pure spectrum of the k[th] analyte component. The scalar value, α can be computed as,

$$\alpha = 1 \,/\, (\mathbf{r}^{\mathrm{T}}\mathbf{X}_{\mathrm{reb}}^{+}\hat{\mathbf{c}}_{\mathrm{k}}) \qquad \text{Eq. 11}$$

The standard addition spectra consist of m (1, 2 ..., m) variables (wavenumbers) and n (1, 2.... n) number of calibration standard addition samples, giving rise to $\mathbf{X}_{\mathrm{sa}}$ ($i \times j$) data matrix. A spectrum for the unknown sample but has equal amount in all the spiked series of samples is recorded to make up a row vector, $\mathbf{r}_{\mathrm{u}}$. The concentration of the analyte of interest in each standard addition samples ($\mathbf{c}_{\mathrm{k}}$) is equal to sum of the concentration of the analyte in the unknown ($\mathbf{c}_{uk}$) and the added standard ($\mathbf{c}_{\mathrm{sk}}$), Eq. 12.

$$\mathbf{c}_{\mathrm{k}} = \mathbf{c}_{\mathrm{s_k}} + \mathbf{c}_{\mathrm{u}} \qquad \text{Eq. 12}$$

The calculation of the data matrix $\mathbf{X}_{-\mathrm{k}}$ is impractical since c_{uk} is an unknown term and $\mathbf{s}_k$ has matrix effect. Yousefinejad et al. proposed the calculation of $\mathbf{X}_{-\mathrm{k}}$ based on each row vector of the standard addition data matrix, $\mathbf{X}_{\mathrm{sa}}(\mathrm{i})$ can be taken as the sum of the spectrum of the unknown analyte of interest and rest of standard addition mixture.

$$\mathbf{X}_{\mathrm{sa}}(\mathrm{i}) = \mathbf{r}_{\mathrm{u}} + \mathbf{X}_{\mathrm{sm}}(\mathrm{i}) \qquad \text{Eq. 13}$$

Where $\mathbf{X}_{\mathrm{sm}}$ contains spectra of the calibration set standards in the presence of the matrix effect coming from the unknown. That means, we can subtract the IR spectral signal of the analyte of interest of the unknown, $\mathbf{r}_{\mathrm{u}}$ from the standard addition mixture, $\mathbf{X}_{sa}$ to get only data matrix $\mathbf{X}_{sm}$ of series of standards added. Therefore, the data matrix, $\mathbf{X}_{\mathrm{sm}}$ containing known concentration data of the standards added with matrix effect and their corresponding pure

spectra which is possible to calculate $\mathbf{X}_{-k}$ provided that the pre-processed dataset matrix of $\mathbf{X}_{sm}$ are subjected to PCA or PLS to get the most significant principal components, yielding $\mathbf{X}_{reb}$.

Finally, the NAS vector of the k[th] analyte component in the standard addition samples can be calculated by replacing x_k in Eq. 8 by $R_{sa,reb}$,

$$\mathbf{X}^{*}_{\mathrm{sa,k}} = (\mathbf{I} - \mathbf{X}_{-\mathrm{k}}\mathbf{X}^{+}_{-\mathrm{k}})\mathbf{X}_{\mathrm{sa,reb}} \qquad \text{Eq. 14}$$

Where, $\mathbf{X}_{sa,reb}$ is a standard addition data matrix with only the significant number of principal components built from $\mathbf{X}_{\mathrm{sa}}$. So, the main idea here is the row vectors of $\mathbf{X}^{*}_{\mathrm{sa,k}}$ have a direct relation with the concentration of the kth analyte of the added calibration standards. An extrapolated standard addition graph is obtained by plotting the Euclidean norm of the row vectors, $\left\|\mathbf{X}^{*}_{\mathrm{sa,k}}\right\|$ against series of added calibration standards as shown in the hypothetical graph in **Figure 5**.

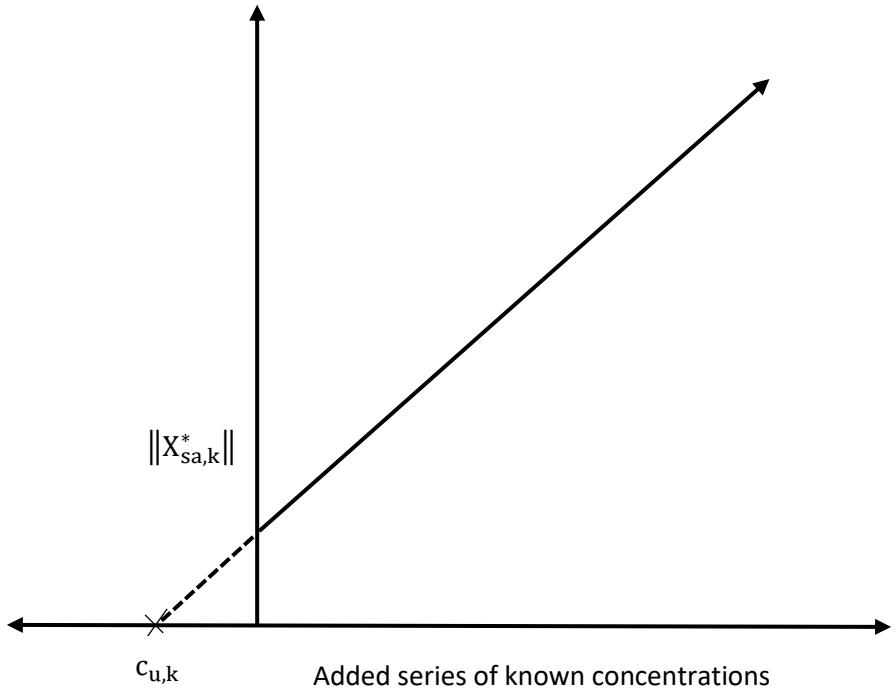


Figure 5. hypothetical plot of standard addition method based on NAS multivariate calibration method

3. Experimental

3.1. Reagents and materials

Hexane (HPLC grade, 99% purity), Methanol (HPLC grade, 99.8% purity), HCl (37% w/w), NaOH (powder), H_2O_2 (30% w/w) and KBr powder for IR spectroscopy were from Merck KGaA, Darmstadt, Germany. The chitin powder purified from shrimp shells (~100% purity) were from Sigma-Aldrich Norway. The chitin standards from cocoon of black soldier fly (BSF) were designated as raw (21-22%), purified (80-90%) BSF Powders and were provided by Dr. Thomas Hahn, from the Fraunhofer Institute for Interfacial Engineering and Biotechnology (IGB, Stuttgart, Germany). An insect raw material (IRM) from BSF, three insect-based fish feed samples containing unknown chitin content and one fish feed without added insect meal were received from the Institute of Marine Research, Bergen, Norway.

3.2. Sample preparation

Chitin from shrimp shell, raw and purified BSF powders were used as received without sample preparation for structure characterization since the ATR sampling mode does not require any form of sample cleaning, e.g. extraction; all it needs is the intimate contact between sample and ATR crystal. However, a minimal sample preparation like grinding samples to a lower particle size and clean-up of oils contained in the fish feed sample was required for DRIFT-FTIR spectroscopy.

The IRM material was subjected to dilute HCl acid (1 M) and NaOH (1 M) treatment to remove minerals and excess of proteins. Hence, the IRM was washed with distilled water, final residues were oven dried (60 ^{o}C) overnight and ground with ball mill machine model MM200 from Retsch company to get a fine powder (<10 µm size). The fine powder (1 g) was demineralized with 1 M HCl for 20 min in a hot oil bath (100^{o}C) and neutralized (~ pH=7) by washing several times with distilled water. This demineralization step was followed by deproteinization, treatment with dilute NaOH (1 M) for 24 h in hot oil bath (85^{O}C). Afterwards, the sample was filtered using a filter paper and neutralized with distilled water by washing for several times. Trace pigments responsible for the brown colour of the chitin product were removed by mild oxidizing treatment with H_2O_2/33%HCl (9/1, v/v) yielding a light brown product.

The purified, raw and demineralized IRM product of BSF dried samples were finely ground (<10 µm particle sizes) to be used for quantification of chitin with DRIFT-FTIR spectroscopy. The feed samples contained fish oil, insect meal and fish feed itself in a mixture. Before extracting a sample, the feeds were homogenized. The oily part of the sample was also difficult

to fill into the sampling cup accessory and to record its spectra. Thus, a 1.0 g from each homogenized sample was transferred to the extraction tube and subjected to ultrasonication followed by centrifugation for 10 minutes to get rid of the oils from the feed. The supernatant was discarded, and residues left on the extraction tube were oven dried overnight. Dried samples were ground to a fine powder (< 10 μm size).

Sample homogeneity test: when working with powder samples in DRIFT-FTIR spectroscopy, sample inhomogeneity is a major source of unwanted variation. To detect whether the variations were random or systematic, two samples (0.5 g) with 1.5 and 1.3% of chitin (w/w) purified from shrimp shells each containing 3.0% purified chitin from BSF insect were prepared with KBr salt diluent. Samples were mixed well with the help of Specamill (Specac, UK) mixing machine.

Calibration sets of samples used was chitin purified from shrimp shells and was prepared in six calibration levels (0.29, 0.50, 1.15, 1.90, 3.20 and 4.76 % chitin w/w) by mixing with the IR transparent matrix (i.e. KBr powder) for reflection and dilution purposes. Prediction sets of samples were prepared from purified BSF (0.94, 1.78, 2.8, 4.0 and 5.0 % chitin w/w), raw BSF (0.54 and 1.1 % chitin w/w), insect-based fish feed samples and fish feed without added insectmeal (3.0 % sample w/w), and demineralized IRM obtained from BSF (5.0 % sample w/w) by mixing with KBr powder. The % of chitin prepared as prediction sets from purified and raw BSF powders were based on the chitin content claimed by IGB, Germany that is 80-90 and 21-22 %, respectively.

The chitin purified from shrimp shells were also used to develop a standard addition calibration model for quantification of chitin in insect-based fish feed, raw and purified BSF samples. A standard addition calibration samples containing 3.0 % (w/w) insect-based diet samples spiked in each of the known concentration (0.0, 0.5, 0.8, 1.1, 1.4, 1.75 and 2.0 % w/w) of chitin standard powders from shrimps were prepared. KBr powder were used to dilute the samples to an appropriate composition. All prepared samples in the calibration were homogenized using Specamill mixing device prior to collection of DRIFT spectra.

3.3. FTIR instrumentation

3.3.1. **ATR-**FTIR characterisation

ATR spectra of BSF powder samples and chitin from purified shrimp shells were collected using Nicolet iS50 FTIR spectrometer (Thermo Scientific, USA) equipped with diamond internal reflection element (IRE), IR source, KBr beamsplitter and deuterated triglycine sulfate (DTGS) room temperature detector. All spectra were recorded at 4000-400 cm^{-1} wavelength range, 4 cm^{-1} resolution, and 32 scans.

3.3.2. DRIFT-FTIR analysis

DRIFT spectra in MIR range (4000-400 cm^{-1}) of samples in six replicates were recorded with Nicolet iS50 FTIR spectrometer (Thermo Fisher Scientific, USA) equipped with DTGS detector at 4 cm-1 resolution, the data points were collected at each of 0.48 cm^{-1} and 128 scans per sample using DRIFT sampling cup accessory. The 128 scans were done by manually rotating the DRIFT sampling cup accessory up-to 8 orientations (16 scans each) to get representative spectra. Spectra collection were performed using OMNIC version 9.8.286 statistical software package (Thermo Fischer Scientific Inc.). Dried spectroscopic grade KBr powder was employed as a reference material to collect background spectra at every start of each day.

For testing sample homogeneity, six spectra at each of the four orientations with refilling and packing of the sampling cup for each of the six replicates were collected. The light beam coming from the IR source cannot fully hit the area of the sample crystals. Since the sample refilling is carried out manually, the orientation of particles in each replicate is different, which leads to high variations in the intensity of the spectra for the same sample. Thus, the spectrum was recorded in each of the orientation of the sampling cup by manually rotating it to the direction of IR light as shown in **Figure 6.** DRIFT sampling cup accessory filled with sample powder mixed with KBr..

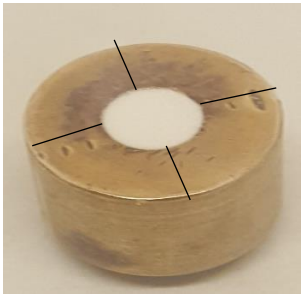


Figure 6. DRIFT sampling cup accessory filled with sample powder mixed with KBr.

3.4. Multivariate data analysis

In order to construct multivariate calibration models, DRIFT spectral datasets were subjected to principal component analysis (PCA) and partial least square (PLS) regression by using Sirius v11.0 (PRS, Bergen, Norway). All spectra were baseline corrected and data pre-processing method was carried out via Savitzky-Golay smoothing polynomial (FitzPatrick, et al., 2012) (15-point window size). Outlying samples were identified and discarded by applying PCA score plot, residuals and leverage points. Calibration datasets (six levels with six replicates) and the variable wavenumbers under the peak areas of N-H stretch (3350-3050 cm^{-1}), amide I and II (1700-1490 cm^{-1}) and amide III (1347-1280 cm^{-1}) IR vibrational bands were selected. Multivariate calibration method based on PLS regression was applied on mean centered data at 0.05 significance level and cross-validation method using four groups (i.e. default setting) were applied. Prediction performances of the model were tested using RMSECV and RMSEP to estimate chitin percentage on samples.

Identifying outlier samples that can affect the quality of a multivariate calibration model is an important procedure. Outliers contained in the calibration data may occur because of many reasons such as instrument and sampling errors, and then, the outliers may have a significant effect on the quality of the model (ZhiChao, et al., 2008). Such abnormal samples show some type of departure from the majority of the dataset. Identification of outliers were based on detection of samples with extreme leverage points, large residuals in **X**-block (spectra data) and **y**-block (predicted concentrations) (Silva, et al., 2012). Leverage (h_i) is a measure of a distance of an object from center of the data and explains that how much the individual object has an influence on the model. It can be defined as;

$$\mathrm{h_i} = \mathrm{t_{A,i}^T}(\mathrm{T_A^T T})^{-1}\mathrm{t_{A,i}} \qquad \text{Eq. 15}$$

Where, **T** is the score matrix of calibration samples, $\mathbf{t}_i$ is the score vector of sample i, and A is the number of significant principal componets. The samples with h_i larger than the $\mathrm{h_{limit}}$ (EqEq. 24) are detected as outliers.

$$\mathrm{h_{limit}} = \frac{3(\mathrm{A}+1)}{\mathrm{I_{cal}}} \qquad \text{Eq. 16}$$

Where $\mathrm{I_{cal}}$ is the number of calibration samples.

Detection of outliers based on unmodeled spectral data residuals is carried out by comparison between the total standard deviation of the residuals (RSD) of the whole calibration set with the standard deviation of an individual sample replicates (RSD_i) as defined in 17 and 18.

$$\mathrm{RSD} = \sqrt{\frac{1}{\mathrm{v}}\sum_{\mathrm{i=1}}^{\mathrm{I_{cal}}}\left(\sum_{\mathrm{j=1}}^{\mathrm{J}}(\mathrm{X_{i,j}} - \widehat{\mathrm{X}}_{\mathrm{i,j}})^2\right)} \quad 17$$

$$\mathrm{RSD_i} = \sqrt{\frac{1}{\mathrm{v}}\sum_{\mathrm{j=1}}^{\mathrm{J}}(\mathrm{X_{i,j}} - \widehat{\mathrm{X}}_{\mathrm{i,j}})^2} \quad 18$$

where J is the number of spectral variables, $\mathbf{X}_{i,j}$ is the Kubelka Munk (K.M) value of sample i at wavenumber j and this absorbance in K.M unit is similar to the Lambert-Beer relation applied in DRIFT-FTIR spectroscopy situation. $\widehat{\mathbf{X}}_{i,j}$ is the absorbance value in K.M unit estimated with A latent variables and the number of degrees of freedom v is given by $I_{cal} \times J - J - A(\max(I_{cal}, J)$. If a sample has $\mathrm{RSD_i} >$ nRSD (n = 2 or 3 for ~95 or 99% confidence level). The other method of detecting outlying samples is through the comparison of the RMSEC of the model with the absolute residual error of individual samples. Generally, samples are discarded when the difference between the reference and predicted value is larger than three times RMSEC.

Precision was calculated based on the calibration model (Gontijo, et al., 2014) in six concentration levels (0.29, 0.50, 1.15, 1.9, 3.2 and 4.76%) of chitin standards in six replicate measurements. The standard deviation (δ_i) of predictions at each level of sample is calculated as,

$$\delta_{\mathrm{i}} = \sqrt{\frac{\sum_{\mathrm{r=1}}^{\mathrm{R_i}}(\hat{\mathrm{y}}_{\mathrm{i,r}} - \bar{\mathrm{y}}_{\mathrm{i}})^2}{\mathrm{R_i} - 1}} \quad \text{Eq. 19}$$

Where, $\hat{y}_{i,r}$ is predicted values in each level, $\bar{y}_i$ is the mean value of the predicted values, R_i is the number of replicate measurements per each level of sample i. The coefficient of variation (CV) was calculated by dividing the standard deviation of individual sample's replicates with the mean predicted value and taking their averages as shown in Eq. 20

$$\mathrm{CV} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}}\frac{\delta_{\mathrm{i}}}{\bar{\mathrm{y}}_{\mathrm{i}}} \times 100}{\mathrm{p}} \quad \text{Eq. 20}$$

Where, p is the number of concentration levels in the calibration model.

According to the recommendations of the International Union of Pure and Applied Chemistry (IUPAC), the limit of detection (LOD) estimation in multivariate calibration was based on the theory of hypothesis testing, considering the probabilities of false-positive and false-negative

decisions and the different sources of errors both in calibration and prediction steps which could affect the result. The International Standardization Organization defined LOD as "the lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit (Allegrini & Olivieri, 2014). Thus, LOD and limit of quantification (LOQ) were calculated from the standard deviation of predicted values for six replicate spectra of blank (without added BSF insect diet into fish feed sample) shown in Eq. 21 and Eq. 22 respectively.

$$\mathrm{LOD} = 3.3 \times \delta_{\mathrm{blank}} \qquad \text{Eq. 21}$$

$$\mathrm{LOQ} = 10 \times \delta_{\mathrm{blank}} \qquad \text{Eq. 22}$$

In the case of standard addition based net analyte signal analysis procedure, the spectra dataset contained one pure chitin standard, one from the sample without the added standard and six samples with added standards each containing six replicates. Each standard addition spectra dataset was imported to Sirius and raw spectra were visually inspected. Afterwards, pre-processing of spectra data was performed using second-derivative Savitzky-Golay third order polynomial with window size 15 to remove the unwanted variations. Potential outlier samples were identified on PCA score plots and the determination of their leverage points. The number of variables selected and the IR region of higher interest (3350-3050, 1700-1490 and 1347-1280 cm^{-1}) were the same as used in the PLS calibration model. Since Sirius does not contains procedures for the standard addition method, the transformed datasets were exported to MATLAB R2018a, version 9.4 (The MathWorks, Inc., MA, USA) and a program were created to build two-dimensional SANAS plot in order to calculate the norm of NAS for unknown BSF insect-based fish feed samples.

The replicate spectra of each standard addition sample are rows in the $\mathbf{X_{sa}}$ data matrix and from which the first spectra were only the unknown without added standard but exist in every row of $\mathbf{X_{sa}}$. The mean spectrum from each group of replicates was calculated. The unknown sample spectrum vector, $\mathbf{r_u}$ was subtracted from every row of $\mathbf{X_{sa}}$ and gives $\mathbf{X_{sm}}$ matrix. After making the average of replicates of spectra, the dimensions of matrices became; 6×163 ($\mathbf{X}_{\mathrm{sa}}$), 1×163 ($\mathbf{r_u}$) and 6×163 ($\mathbf{X_{sa}}$). Six different standard addition samples were followed the same approach for NAS calculation of chitin in their unknowns.

Figures of merits such as selectivity, sensitivity, limit of detection (LOD) and quantification (LOQ) were calculated to study the quality of the analytical method using the NAS concept (Hemmateenejad & Yousefinejad, 2009). The sensitivity can be computed as the norm of the

NAS vector of the kth component in the unknown sample, $\mathbf{s}_{\mathrm{k}}^{*}$ using the significant number of latent variables, (Eq. 23) (Sarraguca & Lopes, 2009).

$$\mathrm{SEN_k} = \|\mathbf{s}_{\mathrm{k}}^{*}\| \qquad \text{Eq. 23}$$

Selectivity ($\mathrm{SEL_k}$) is a measure of degree of overlap aimed to indicate what part of the total signal is lost due to overlap (Lorber, et al., 1997). It was computed by dividing the norm of the k^{th} component of the unknown spectrum ($\mathbf{s}_{k}^{*}$) by the norm of the original spectrum of the unknown sample as shown in Eq. 24.

$$\mathrm{SEL_k} = \frac{\|\mathbf{s}_{\mathrm{k}}^{*}\|}{\|\mathbf{s}_{\mathrm{k}}\|} \qquad \text{Eq. 24}$$

The LOD and LOQ are the important figures of merit and can be estimated based on the following equations.

$$\mathrm{LOD} = 3 \times \frac{\|\delta\|}{\|\mathrm{s}_{\mathrm{k}}^{*}\|} \qquad \text{Eq. 25}$$

$$\mathrm{LOQ} = 10 \times \frac{\|\delta\|}{\|\mathrm{s}_{\mathrm{k}}^{*}\|} \qquad \text{Eq. 26}$$

Where, δ is the standard deviation at the selected variable wavenumber. The norm of δ was estimated for six replicate spectra of fish feed sample without added BSF insect diet which is treated as the blank sample in our case and s_k^* is its corresponding net sensitivity vector.

4. Result and Discussion

4.1. ATR-FTIR characterization

The aim of this characterization was to check similarity of spectra in each band and to evaluate whether it was possible to use the chitin from purified shrimp shells as a calibration standard for quantification of chitin in BSF cocoons and BSF insect-fish feed samples. The ATR-FTIR spectrum of chitin from purified shrimp shells and purified BSF are shown in **Figure 7**. **Figure 8** represent the spectrum of raw BSF and the demineralized IRM of BSF samples. The spectrum of chitin purified from shrimp shells (**Figure 7**, red) was similar with the purified chitin from BSF (**Figure 7**, pink) and demineralized IRM (**Figure 8**, green) but looks different from raw BSF spectrum (**Figure 8**, red). The raw BSF spectrum was characterised with a single broad peak at amide I and II bands. Presence of amide II band at 1540 cm-1 was due to an excess of proteins (Wasko, et al., 2016). Chitin occurs in three polymorphic structures (α, β and γ forms). The spectral bands of α-chitin from different sources such as cicada sloughs (Sajomsang & Gonil, 2010), nymphs and adults of grasshoppers (Erdogan & Kaya, 2016), Hermetia illucens (Wasko, et al., 2016), shrimp shells and bumblebees (Majtan, et al., 2007) are characterized by their double peak at amide I band.

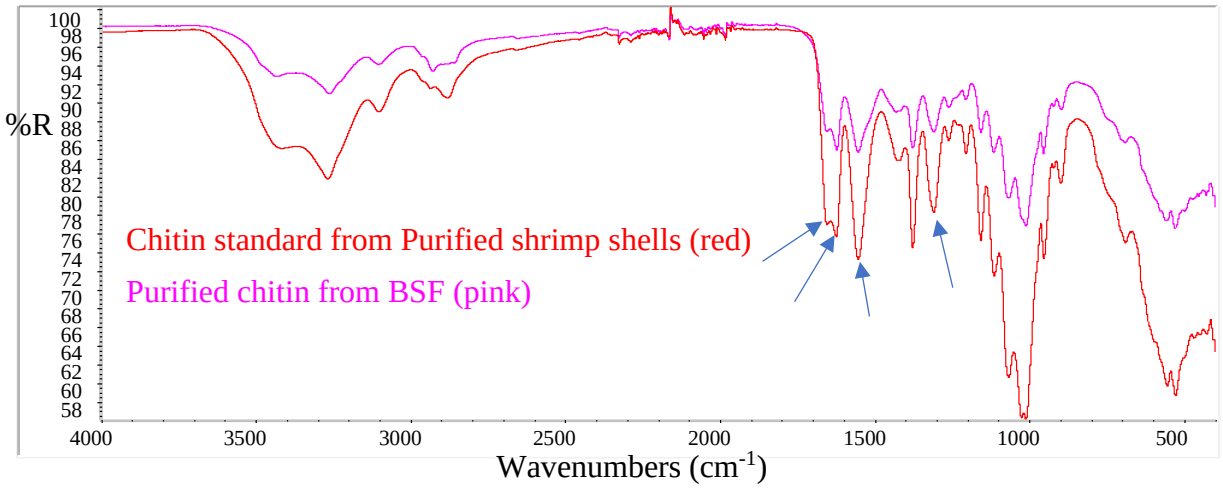


Figure 7. ATR-FTIR spectra of chitin purified from shrimp shells (red) and purified chitin from BSF (pink).

The spectrum of purified chitin from BSF cocoons was characterized by amide I, II and III bands at 1653, 1620 1552 and 1307 cm^{-1}. The amide I band was ascribed to the carbonyl secondary amide stretch. The α-chitin is capable of forming two intermolecular and intramolecular hydrogen bonds. So that, a well-defined peak at 1620 cm^{-1} and a weaker intensity at 1653 cm^{-1} of the amide I band were ascribed to the intermolecular hydrogen bonds

(CO-N-H) and intramolecular hydrogen bond (CO-H-OCH_2) stretches respectively (Erdogan & Kaya, 2016). Moreover, the amide II band was observed at 1552 cm^{-1} and amide III band was also displayed at 1307 cm^{-1}, which were assigned to N-H bending and C-N stretching vibrations of amides respectively.

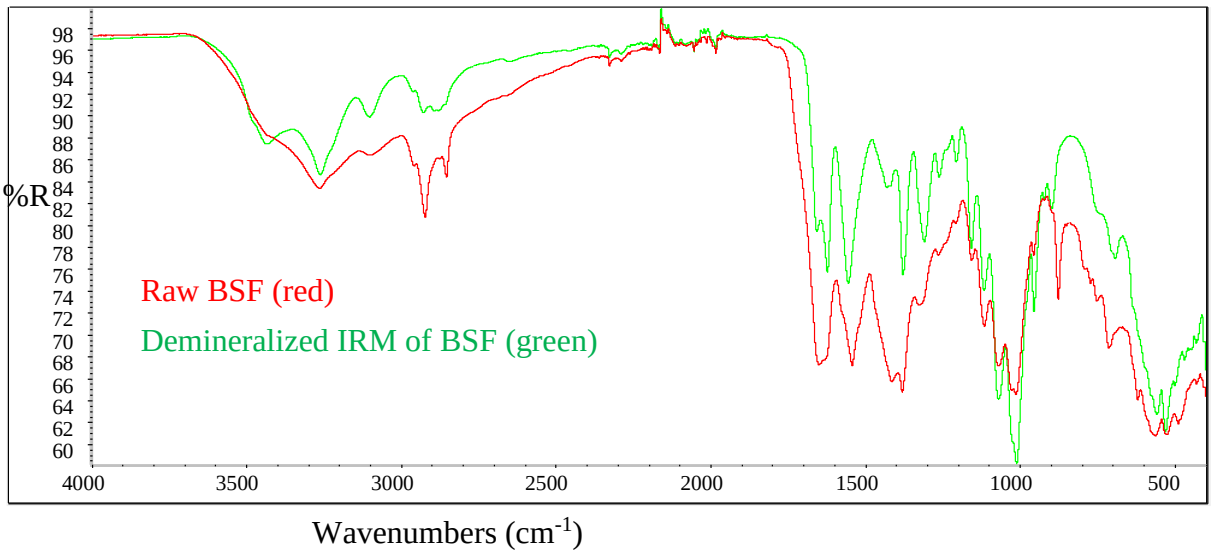


Figure 8. ATR-FTIR spectra for raw BSF (red) and demineralized IRM of BSF (green).

The other vibrational bands observed were at 3430 cm^{-1} (O-H stretching), 3261 and 3102 cm^{-1} (asymmetric and symmetric N-H stretching), 1154 cm^{-1} (C-O-C asymmetric stretch) and 1113, 1067 and 1009 cm^{-1} (C-O-C symmetric stretching). The band at 1375 cm^{-1} is also attributed to C-H bend of the rocking methyl group (Soon, et al., 2018; Kaya & Baran, 2015; Majtan, et al., 2007). The assignment of significant bands of chitin from shrimp shells, purified BSF cocoons, raw BSF cocoons and demineralized IRM of BSF cocoons are shown in **Table 1**. Therefore, high similarity of the vibrational bands and shape of peaks of chitin among samples of BSF cocoons and shrimp shells were observed.

Table 1. Assignment of relevant bands of chitin from shrimp, and purified, raw and demineralized IRM of BSF cocoon samples.

Assignments	**Wavenumbers, cm^{-1}**			
	Shrimp	Purified BSF	Raw BSF	Demineralized IRM of BSF
***v*(O-H)**	3415	3430	-	3430
***v*(N-H)**	3264, 3100	3261, 3102	3261, 3097	3258, 3100
***v*(C-H)**	2933, 2877	2925, 2856	2920, 2850	2925, 2877
***v*(C=O)**	1652	1653	1648	1654
***v*(C=O)**	1622	1620	-	1619
***δ*(N-H)**	1551	1552	1540	1552
***δ*(C-H)**	1417, 1375	1429, 1375	1412, 1378	1428, 1375
***v*(C-N)**	1307	1307	1325	1307
***v*(C-O-C)**	1155, 1112, 1065, 1023 and 1009	1154, 1113, 1067, 1009	1153, 1113, 1068, 1010	1154, 1113, 1067, 1008

4.2. Multivariate calibration methods for quantification of chitin

4.2.1. Spectral preprocessing

The data matrix (107×1203) comprised of spectra of chitin from purified shrimp shells as a calibration set, and the spectra of purified, raw, the demineralized product of BSF, and insect-based fishfeed samples as prediction sets. The DRIFT-FTIR raw spectra data is illustrated in **Figure 9**, that clearly shows that the presence of high variability within samples. PCA method using two principal components was applied to visually inspect the score plot of raw data, However, the variability between different classes (such as different concentration and samples) were not visible enough. That means a mixing up of information were observed. The raw spectra contain useful information related to the sample of interest and the instrumental variations as well. Besides, particle size, sample packing/thickness, and particle size distribution are basic factors which result on distortions of signals and unexpected higher intensities due to light scattering effects and instrumental variations (Rinnan, et al., 2009; Christy, et al., 1993). As a result of these noises and light scattering effects, extraction of the useful chemical information of interest become a problem. For example, biological samples

are very soft and causes light scattering effects. So that, to reduce the unwanted spectral contributions (i.e., random noises and systematic variations due to scattering effects) and to simultaneously enhance the spectral features of the analyte of interest; several data preprocessing methods were investigated prior to building the PLS regression model.

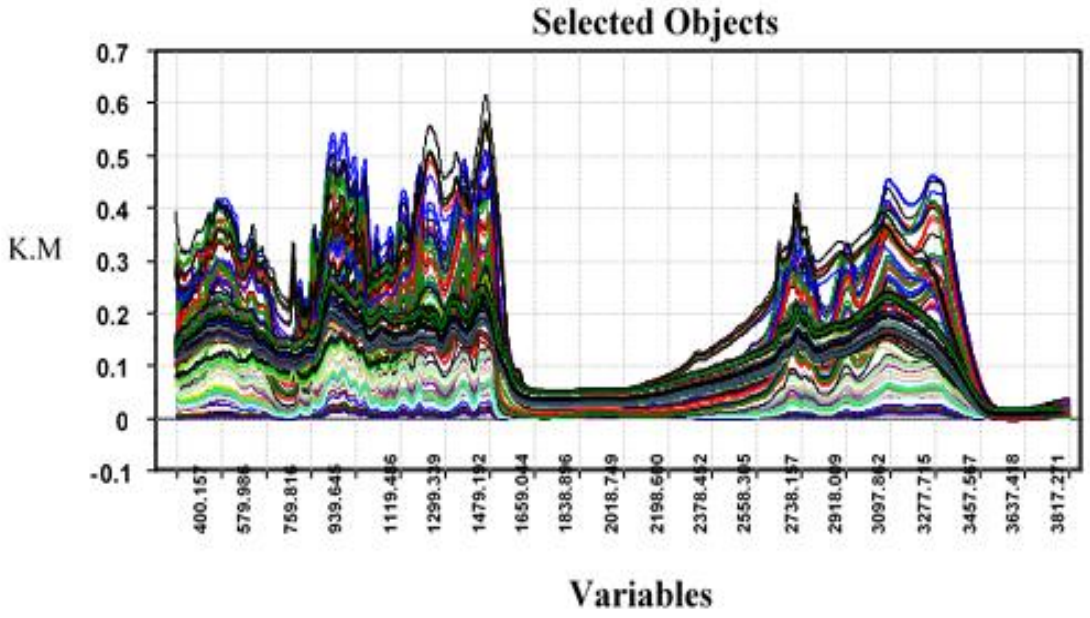


Figure 9. Raw DRIFT-FTIR spectra of chitin calibration standards and BSF cocoons samples.

Extended multiplicative scatter correction (EMSC) and standard normal variate (SNV) transformed spectra are displayed in **Figure 10** and **Figure 11**, one can see from the spectra that both methods have reduced the spectral dispersion of samples. PLS regression calibration method was applied on EMSC and SNV transformed data sets to see the quality of the model and their PCA score plot on chitin calibration set and samples of purified chitin from BSF insect, raw BSF insect and dilute acid-based treated BSF insect material results in better visualization of each group. However, non-linear variations between calibration levels and the residuals among replicates were very high. These data preprocessing methods were not fitting with the data characteristics and applying spectral derivatives was another good option.

Second-degree derivative Savitzky–Golay was chosen with third order polynomial and 15-points window size to smooth the data as displayed in **Figure 12**. It corrects baseline offset and linear baseline shift and smoothes the data by fitting a polynomial expression to the data within the moving window selected for derivation and results in new values by replacing the original dataset with no loss of useful information. The portions of the transformed data showing linear

variation upon increasing concentration of the calibration set were selected to build PCA and PLS regression models.

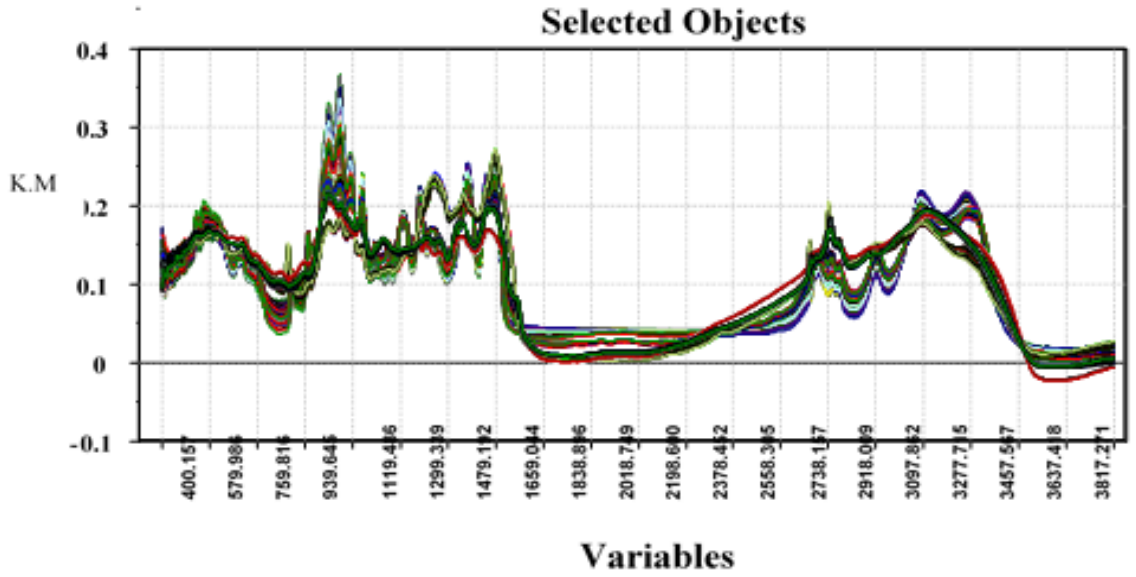


Figure 10. EMSC transformed DIRFT spectra of calibration and prediction sets.

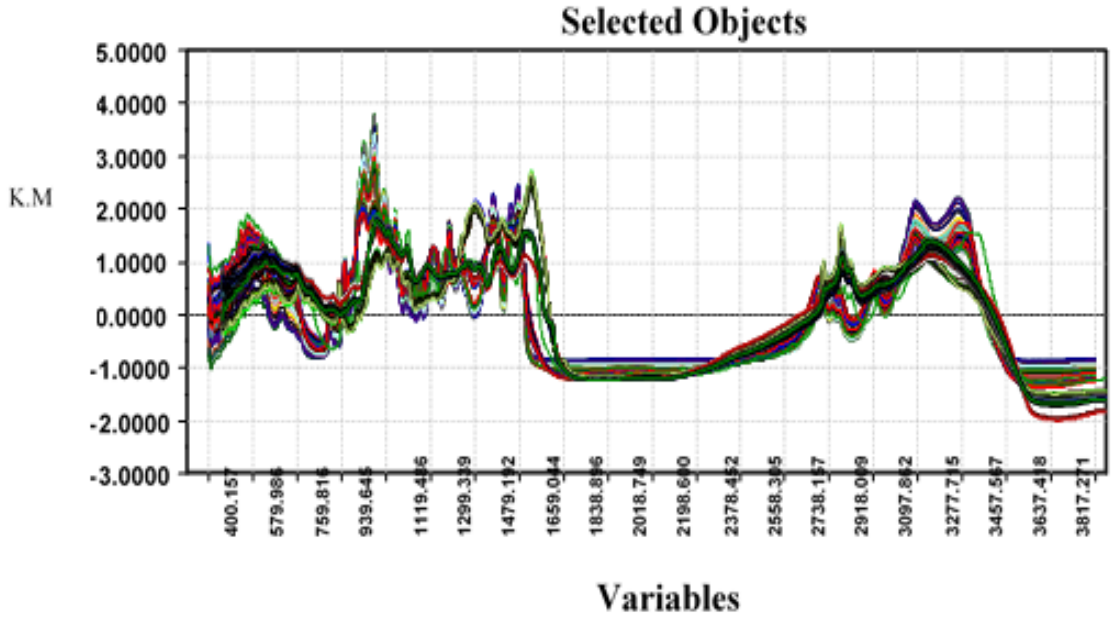


Figure 11. SNV transformed DIRFT spectra of calibration and prediction sets.

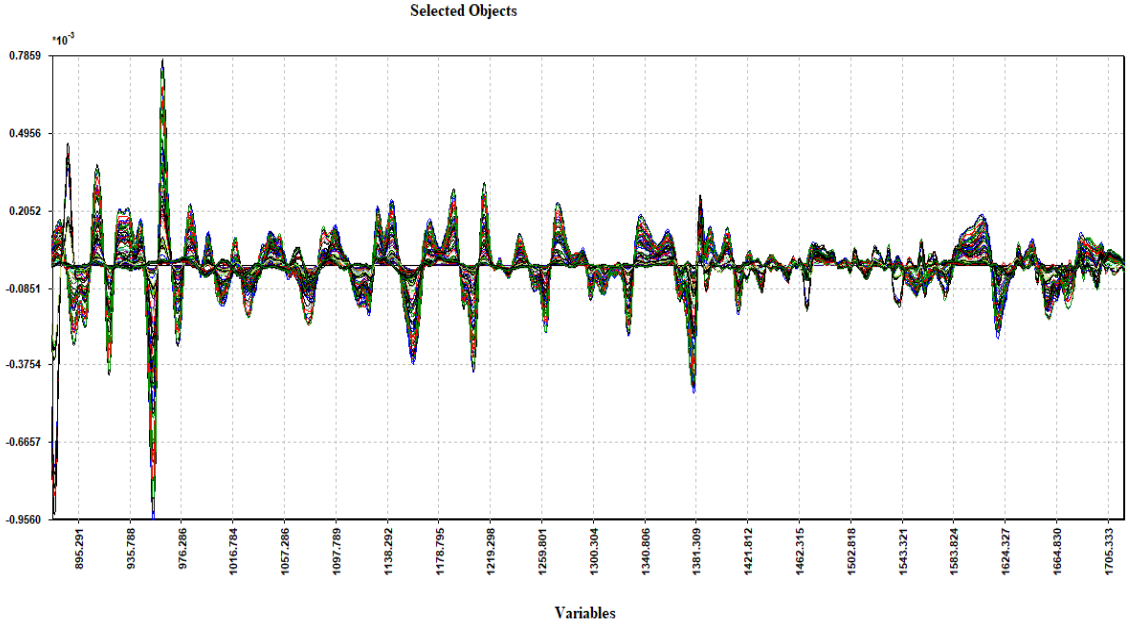


Figure 12. Second-derivative Savitzky-Golay pre-processed DIRFT spectra for calibration and prediction sets.

4.2.2. Principal component analysis (PCA)

The data matrix contains 107 objects and 1203 variables . The chemical of interest, chitin, contains an acetylamide group in each of the monomer units of the biopolymer. This group is not present in the other constituents of the samples. Thus, the signals of higher interest were mainly for those peaks of the acetylamide group (N-H stretch, amide I, II & III bands) excluding the methyl group which has common IR band with aliphatics result from the matrix of samples. Direct observation of data distributions of datasets containing several hundred to thousands of variables is tedious and an impossible task. However, using multivariate calibration tools, such as PCA , is quite straightforward to visualize the distribution of data in the latent subspace since the number of latent variables are much lower than the number variables. PCA is commonly employed to reduce the dimensionality of spectral data and to obtain preliminary information about data distribution. After data preprocessing, variables of higher interest with IR vibrational bands from 3350-3050 cm^{-1}, 1700-1490 cm^{-1}, and 1347-1280 cm^{-1} which corrospond to N-H stretching, amide I and II, and III, respectively containing 107×1203 spectra data matrix were subjected to PCA to estimate and visualize variations between-replicate as well as variations between samples. The resulting score plot (**Figure 13**) clearly visualizes the variations in the sample set. The first and second principal components have explained much of the total explained variance - about 83.2 and 7.9 %, respectively.

The score plot for the chitin standard purified from shrimp shells (blue), purified (green), raw (red), demineralized IRM (black) of BSF, three insect-based fishfeed samples and one fishfeed sample without added insect diet (orange) are shown in **Figure 13**.

The variations of demineralized BSF insect raw material (black), purified chitin of BSF cocoon (green) and chitin calibration standard (blue) were not clearly visible because they are projected in the same direction. But samples from fish feed (orange) and raw BSF insect (red) were clearly distinguished from the others. Variations between levels of concentration were apparently differentiated and but difficult when the concentration range was very close to each other.

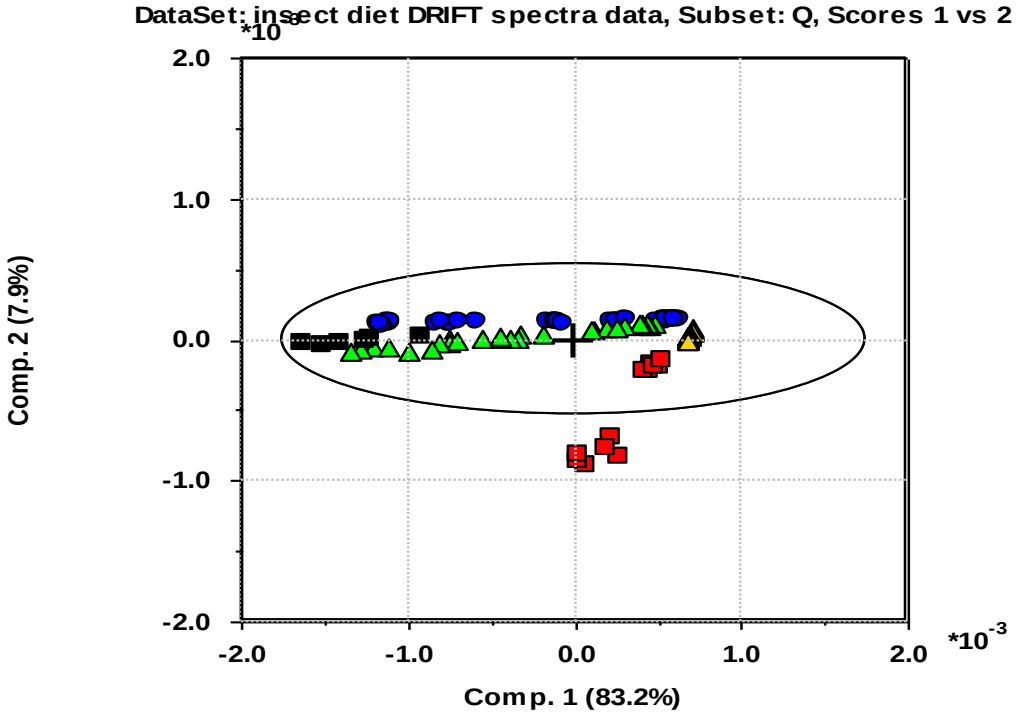


Figure 13. Score plot of transformed spectra data of chitin calibration standard from shrimp (blue), purified chitin from BSF cocoons (green), raw BSF insect chitin (red), demineralized BSF insect raw material (black) and fish feed samples (orange).

A score plot for the calibration sets of samples were obtained after mean centering of datasets. **Figure 14** (top) shows RSD vs leverage plot at 95% confidence level where the horizontal line tells us the threshold value of residuals to reject the objects beyond it. The leverage point falls between 0 and 1. It is found that the objects with extreme leverage points detected have a sizable impact on the regression coefficients. The maximum limit of leverage value (h_{limit}) to reject outlying samples is beyond 0.25. So that, only one sample (object 55) was found furthest

from the rest that is detected as an outlier. In addition, calibration samples found in the higher concentration levels were obtained with higher residuals than the lower concentrations. That means increasing concentration has an impact for dispersion of results. Our suspicion was confirmed by using a normal plot of the scores (**Figure 14**, bottom) and detected the same object as an outlier. Similarly, the score plot was also done for each sample to estimate within-sample variation. In this case, only one sample (object 2) was detected as an outlier. These two outlier objects were not included while building the model.

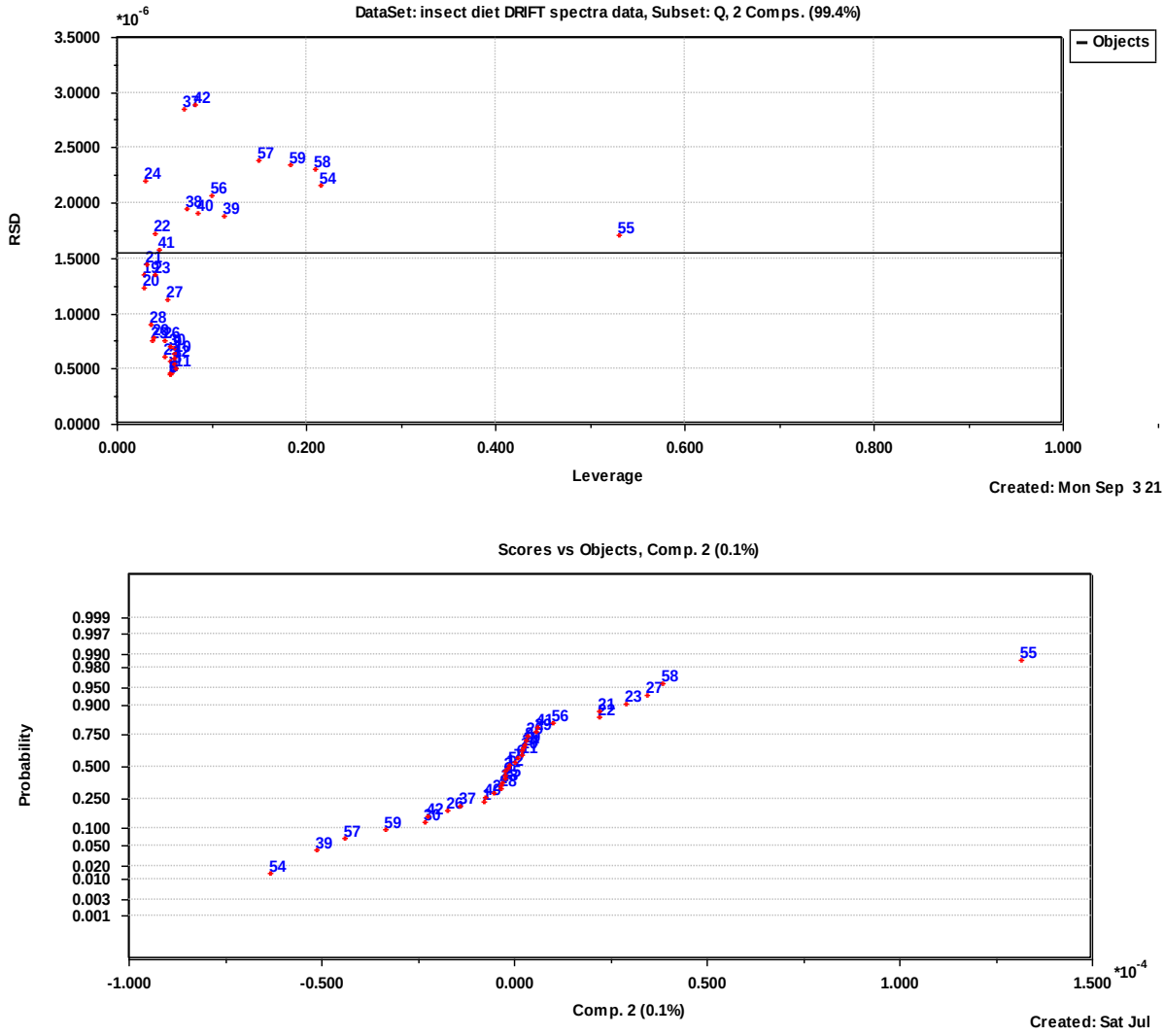


Figure 14. Plots for outlier detection: RSD vs leverage (top) and normal plot of the scores (bottom).

The score plot displayed in **Figure 15**a present the spectra of 1.5% chitin standard purified from shrimp which is spiked with 3% purified of BSF insect sample recorded in six replicates, each with six refilling and at each of the four orientations of the sampling cup, 144 spectra in total. The variations did not show a specific pattern between the same orientations, refills and replicates rather it seems randomly distributed. This can have assured us that there were no

systematic variations and the samples prepared were well homogenized. The random variations were resulted mostly from differences in packing and refilling of the sampling cup. In a similar way, a closely related concentration of chitin (1.3%) from shrimp shells containing 3.0% purified BSF were visualized in PCA score plot (**Figure 15**b) to compare whether they can be clearly distinguished among two concentrations. Since both concentrations were very close to each other, the variations between concentrations were unclear or less pronounced.

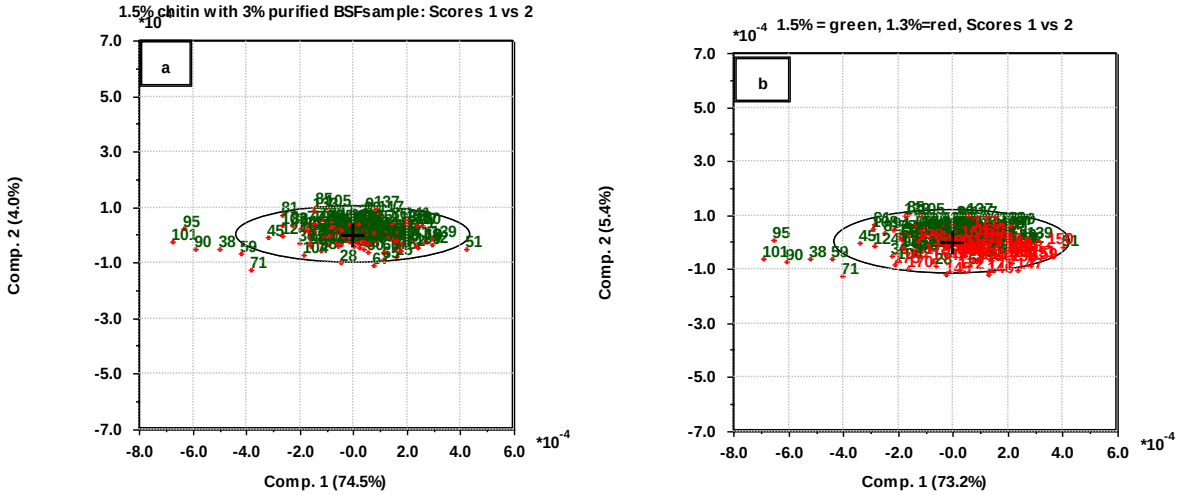


Figure 15. (a) PCA score plot for the 1.5% chitin replicate spectra used for sample homogeneity test and **(b)** PCA score plot to show the concentration variation both for 1.5% (green) and 1.3% (red).

4.2.3. Partial least square regression (PLSR)

The calibration model based on PLS regression were developed with the selected IR band between 3350-3050, 1700-1490 and 1347-1280 cm^{-1} (total 1203 variables). Hence, the variables under the area of peaks of amides vibrational bands of chitin were preliminarily selected. Second order Savitzky-Golay diffentiation smoothing using window size 15 was performed for all IR bands of the spectra prior to the regression modelling. In many cases, most of the variables have little or no contribution to the response to be modelled. While decreasing the numbers of such variables drastically, it only leads to a minor or no loss of the useful information. A robust variable selection method retains a small set of variables capable of providing better, or at least, equivalent model performance to those obtained by the original set of variables (Sarathjith, et al., 2016).

Variable selection methods such as, selectivity ratio, variable importance in prediction (VIP), X-weights and weighted regression coefficients were investigated to retain a smaller set of predicting variables which are better correlated to the response variable. The variable selection graph for variable selectivity ratio, VIP, weighted regression coefficient and X-weights are plotted in y-response axes against the variable wavenumbers (x-axes) are shown in **Figure 16**.

Selectivity ratio and VIP are the most commonly used variable selection methods for FTIR spectroscopy data (Farrésa, et al., 2015). Selectivity ratio is estimated by using the ratio of explained to residual variance of a variable. The VIP score of a predictor is a summary of the importance for the projections to find A, latent variables. VIP values can be calculated by summing variable influence (VIN) over all model dimensions. For a given PLS dimension a, $(VIN)_{a,j}^2$ is equal to the squared PLS weight $(w_{a,j})^2$ of that term multiplied by the percent explained of residual sum of squares by that PLS dimension. The accumulated (overall PLS dimensions) value, $VIP_{a,j} = \sum\left((VIN)_{a,j}^2\right)$ is then divided by the total percent explained of residual sum of squares by the PLS model and multiplied by the number of terms in the model.

The VIP, weighted regression coefficients and X-weights variable selection methods were better to reduce most of the variables with little advantage than the selectivity ratio. Selectivity ratio reduce only smaller number of variables and was needed much higher ratio to get rid of the variables with little importance. For these four variable selection methods, the variables found with the values nearer to zero indicates that the variable is less important and should not be included during building the model. As shown in **Figure 16**, there are large positive and negative values for X-weights, weighted regression coefficient and selectivity ratio plots. These large positive values showed that intensity of the variable increases as the percentage of chitin increases. The large negative values also indicate that the variable is high when the percentage is getting lower.

The PLS model finally used were based on how much the RMSECV and RMSEP get lower with an increasing the regression parameters (R^2 and slope approaches 1) and lowest intercept for better prediction. In this case, high selectivity ratio (level=300) was applied to select the important predictor variables.

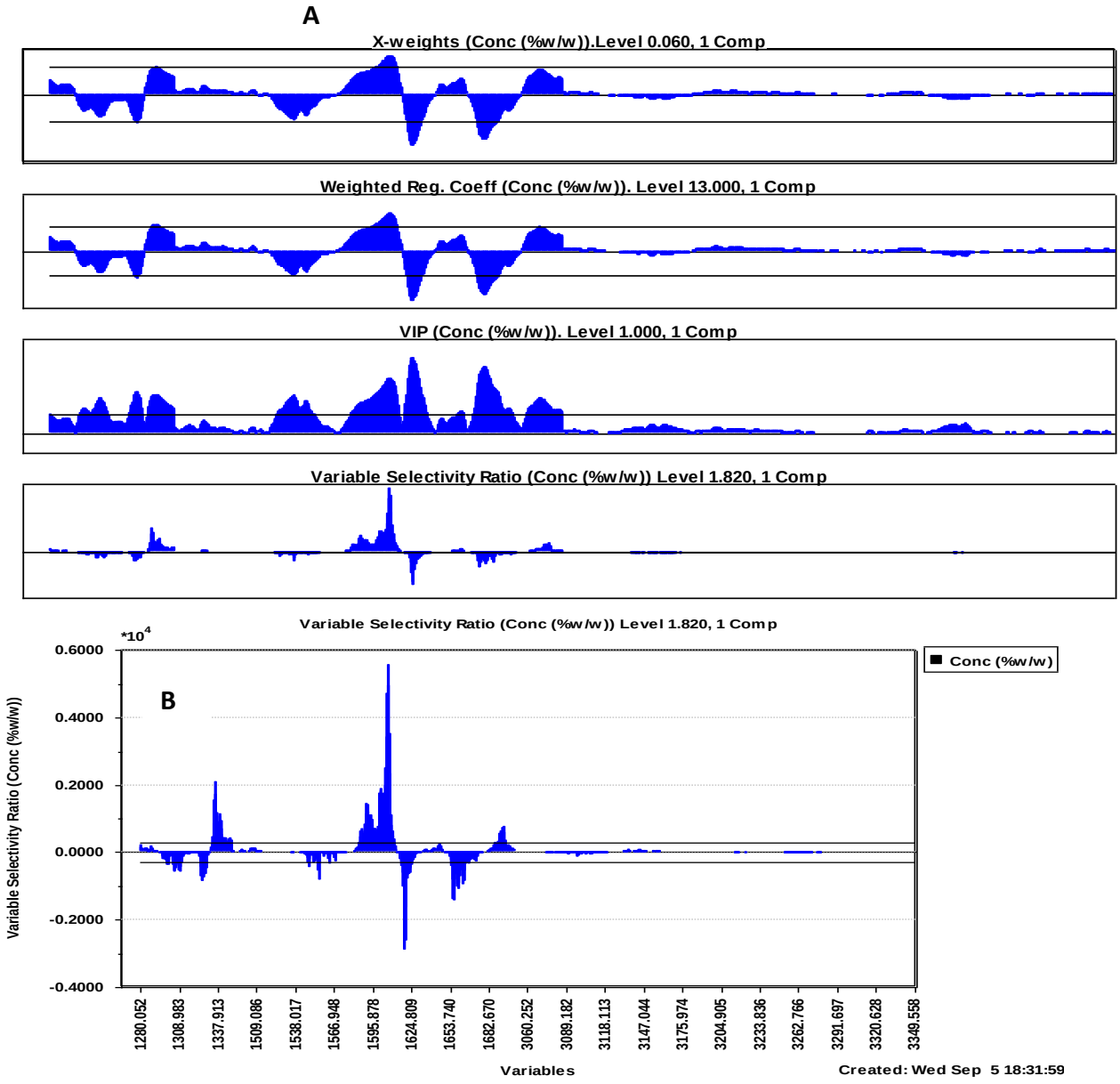


Figure 16. **(A)** Variable selection parameters (y-axes) against wavenumbers (x-axes) applied for reducing variables with little or no contribution to the PLS model and **(B)** variable selectivity ratio plot

The PLS response residuals plot was used to determine the linearity of the model (**Figure 17**). The residual errors of samples were randomly distributed. The fit of the PLS calibration model built for DRIFT-FTIR spectra of chitin standards are shown in **Figure 18**. Only one component was used to build the model. Validation of the fit of the model was done by comparing the true concentration levels of chitin to their predicted values. The number of calibration samples were small and was not realistic to divide it into a training and validation set. So, the validation was

based on deleting objects at a time and substituting the deleted one and discarding another to check its influence on the performance of the PLS model. The RMSECV for the calibration set obtained was 0.220 with a correlation coefficient of 0.982. Once the PLS regression model is established, prediction for a new set of samples were done. The fit of test samples to the built PLS subset model were done at 0.05 significance level as shown from the plot of predicted versus measured for the fitted test samples in **Figure 19**. The percent prediction error for each raw and purified BSF powder samples were greater than 15%. This is because of the nominal percentage of chitin said by the Fraunhofer IGB, Germany is maximum as compared to the predicted value in this method. Since the degree of matrix effect in purified and raw BSF insect itself is not the same, calculations of RMSEPs were done separately for each sample and are presented in **Table 2**.

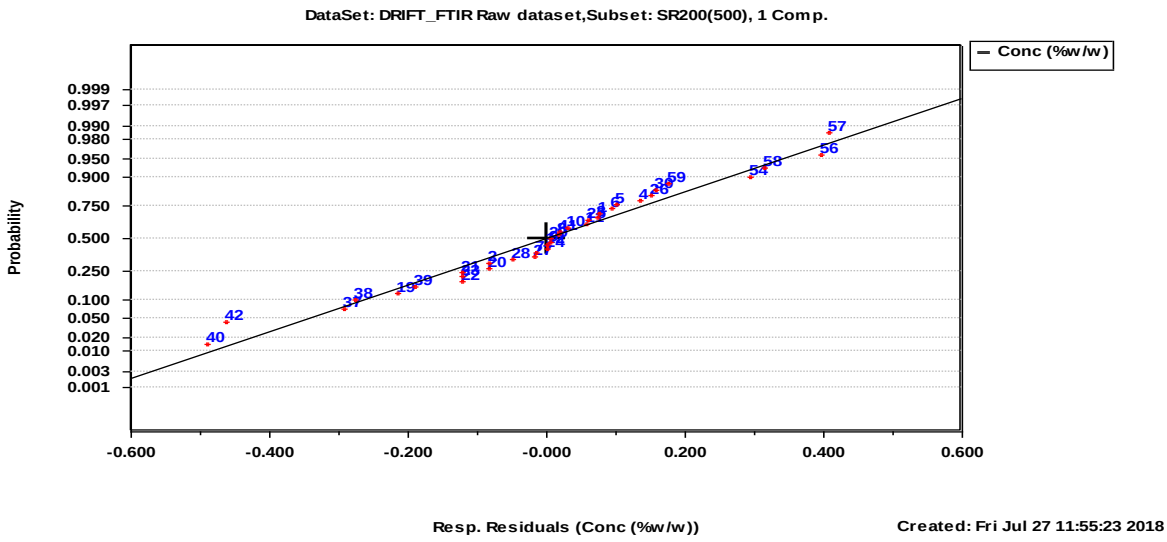


Figure 17. Response residuals plot of calibration sets

Precision is one of the important figures of merits during validation of PLS calibration model, which is defined as a closeness of agreement between the results of a series of measurements for the same homogeneous sample under specified conditions (Gontijo, et al., 2014). The coefficient of variation (CV) calculated for the six-level calibration model was 5.64%. LOD and LOQ calculated from six blank replicate spectra of fish feed sample were found to be 0.026 and 0.084 % chitin (w/w) respectively.

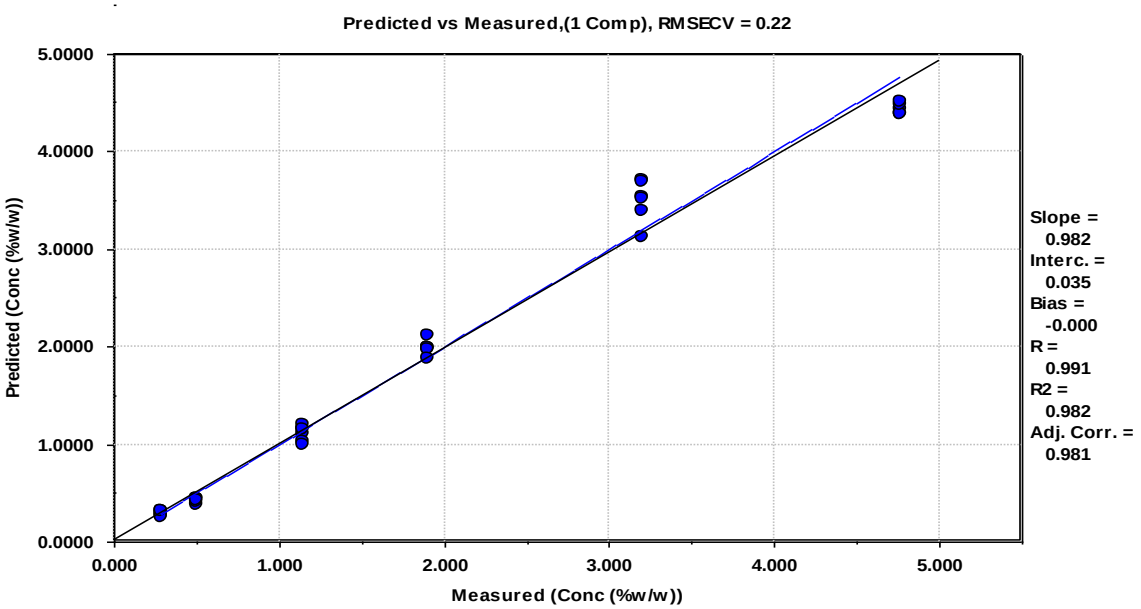


Figure 18. Correlation plot of calibration datasets for quantification of % chitin (w/w) in purified chitin, raw, demineralized product of BSF cocoons and BSF insect-based fish feed using PLS model.

Whereas, the RMSEP % chitin (w/w) calculated for the individual samples, purified and raw BSF cocoon powders were 20.66 and 11.19 respectively. The predicted percentages of chitin in purified chitin from BSF cocoons and raw BSF cocoons were found about 66.82 and 31.95 % chitin (w/w) respectively. The claimed percentage of chitin by the company for the purified chitin and raw BSF cocoons were 80-90 and 21-22 % w/w respectively. In this case, the purified chitin from BSF cocoons was obtained lower than the percentages of chitin claimed by the company. On contrary, the percent chitin predicted for raw BSF cocoons in this study was overestimated. Predictions of chitin performed for three different insect-based fish feed test samples were 8.90, 8.60 and 10.51 % chitin (w/w).

Table 2. Summary of prediction % of chitin (w/w) in purified and raw BSF powder samples, the coefficient of variations (CV, %) and RMSEP values in % chitin (w/w).

Samples	Number of samples and replicates	Nominal	Predicted	RMSEP	CV
Purified chitin obtained from BSF insect	5 samples, 6 replicates	80-90	66.82	19.36	10.74
Raw BSF insect	2 samples, 6 replicates	21-22	31.95	11.19	13.54

To figure out the problem for the overestimated results obtained, the prediction was made on demineralized raw BSF insect materials and found 16.92 % chitin with RMSEP of 1.53%. The spectra of this demineralized raw BSF insect raw material have a well-defined peak in the same position as the chitin purified from shrimp shells but the raw BSF samples spectra were strongly pronounced by other component mixtures, mostly proteins. In such too overlapped spectra, the contribution for the resulting high signal will be from all components. The presence of the other interfering components in test samples were noticed by comparing the magnitude of RMSEC and RMSEP. The error RMSEC was lower than the RMSEP because the physical properties of test samples were different from calibration samples. As a result, a further consideration of a multivariate calibration model based on the standard addition method and net analyte signal calculation was proposed.

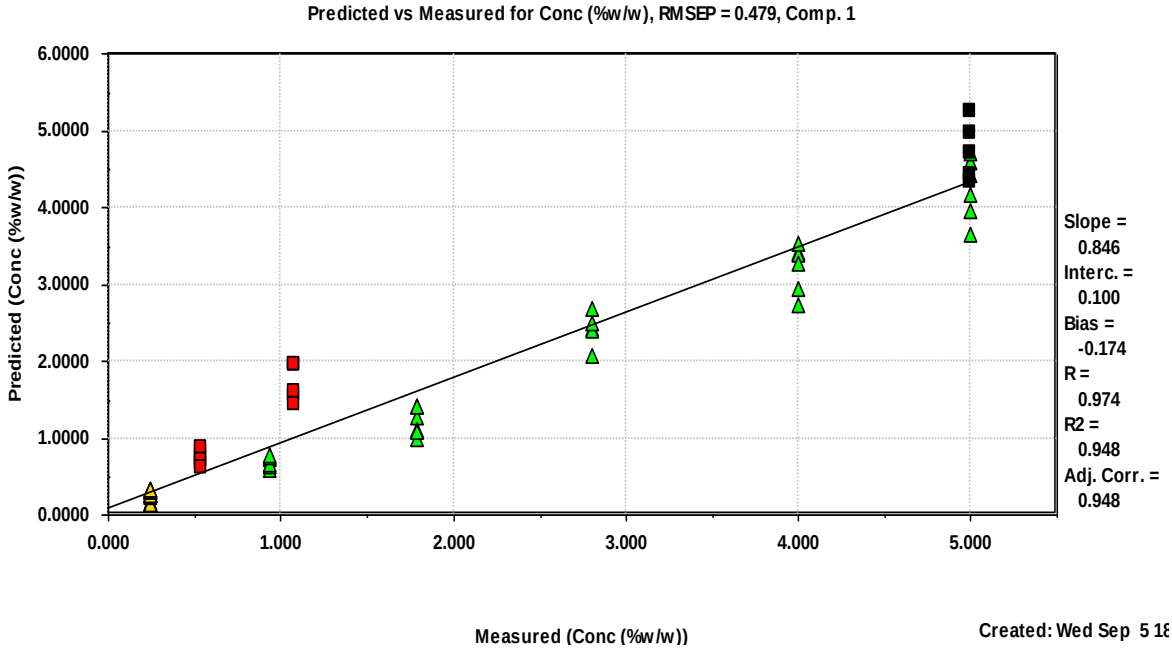


Figure 19. Correlation plot of prediction samples for purified chitin (green), raw BSF cocoons (red), demineralized IRM of BSF cocoons (black), and fish feed (orange).

4.2.4. Standard addition net analyte signal (SANAS)

Five different unknown samples containing chitin and one blank sample (fish feed without added insect diet) were examined by SANAS based multivariate calibration. Each sample has its own standard addition spectra dataset**. Figure 20** shows the standard addition spectra recorded from raw BSF insect sample spiked with a series of known concentration of chitin standard obtained from purified shrimp shells. A_1 is the spectrum of chitin standard, B_1 is the spectrum of raw BSF insect containing unknown chitin percentage and C_1 – H_1 are standard addition spectra each spiked with the same percentage of B_1 in a series of chitin standards. The spectrum of B_1 was assumed to contain highly overlapped spectra because of the signal obtained has contributions from multiple components besides the analyte of interest. As observed from the spectra in **Figure 20**, the raw BSF insect spectrum (B_1) has not similar peak shapes and well-defined peaks in the same IR band position as in the spectrum of chitin standard (A_1) but after subsequent addition of chitin standard, the matrix effects were getting pronounced. For example, the unknown's sample peaks for amide II band (1540 cm^{-1}) has more interferences from proteins which is different from the chitin standard spectra IR band (1551 cm^{-1}) and the peak for amide III band (1325 cm^{-1}) possess too small intensity which makes it difficult to visualize. Peaks for the chitin purified from

shrimp, which were used as the standard spectra, were sharper than the series of calibration standard addition spectra spiked with the sample.

First, it was of interest to view and inspect the spectra to find how the different pre-processing methods affect the raw spectra. After data transformations with different degrees of differentiation and window sizes, the better smoothed data were found when the 15-point window size second-derivative Savitzky-Golay third order-polynomial expression were used. The outlier samples within each replicate of the calibration points were removed when data have high leverage points. Furthermore, variable selection tools in the PLS regression model were important to get rid of unwanted and noisy variable intervals. All the necessary data treatments and identifying outliers were with Sirius software and finally the norm of NAS were calculated using the MATLAB program.

The matrices present in the unknown sample are contained in the standard addition samples as well. That means, the unknown spectrum vector, $\mathbf{r_u}$ and the spectra data matrix of the calibration standards, $\mathbf{X_s}$ are combined to give the standard addition data matrix, $\mathbf{X_{sa}}$. Average of replicates from each level of the standard addition data were taken to build the SANAS model. The same procedure was followed for the rest of the five samples.

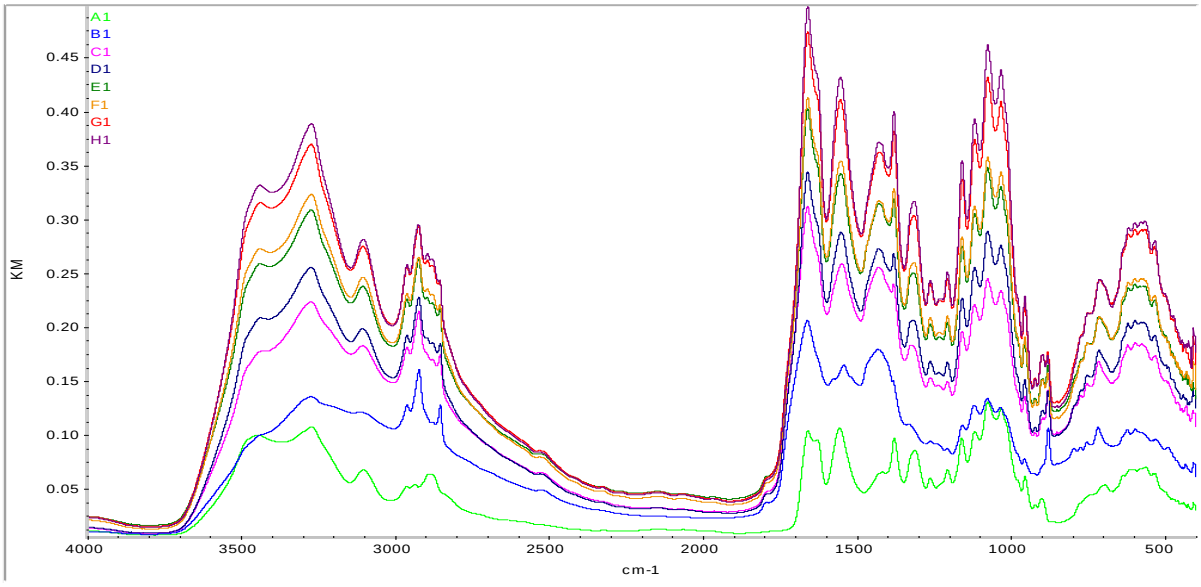


Figure 20. Spectra of the standard addition samples: chitin standard (A_1), raw BSF cocoon containing unknown chitin (B_1) and chitin calibration standards spiked with raw BSF cocoon sample (C_1-H_1).

The standard addition plot obtained by the SANAS method for the six samples are represented in **Figure 21** (**A-F**). Only one significant principal component was needed to calculate the $\mathbf{X}_{\mathrm{sa,reb}}$ from $\mathbf{X}_{\mathbf{sa}}$. The Euclidean norm of row vectors of analyte component in $\mathbf{X}_{\mathrm{sa}}$ and added chitin calibration standards are directly related to get a plot similar to a univariate calibration method as shown in **Figure 21** (**A-F**).

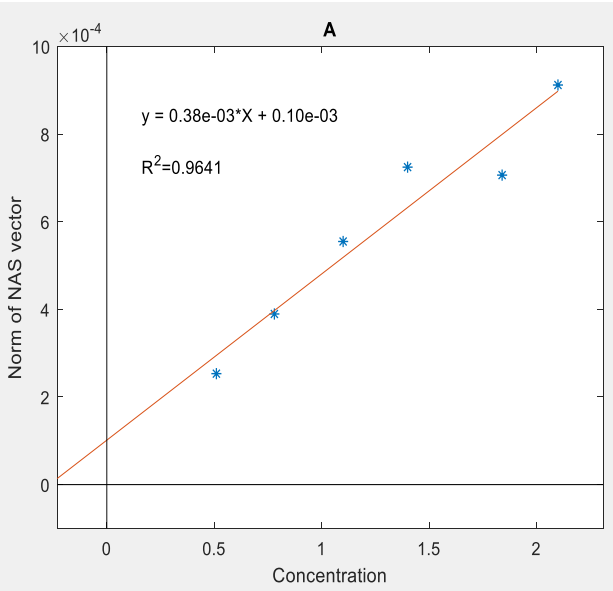
Analytical figures of merit, such as sensitivity and selectivity were estimated from the Eucleadian norm of the NAS as described in section A different selectivity value was acquired for each unknown sample determined with SANAS method since the degree of interferences are different. Lower value of selectivity was obtained for samples having high matrix effects and vice versa as shown in **Table 3**. In this case, the selectivity value for raw BSF sample was almost a third of the value for the corresponding purified chitin of BSF sample. Sensitivity values were measured as the norm of the net analyte signals. The estimates of LOD and LOQ were 1.0 and 3.34 % chitin (w/w) respectively. The estimate was done from the fish feed without added BSF insect diet and showed maximum value of LOD and LOQ higher than expected.

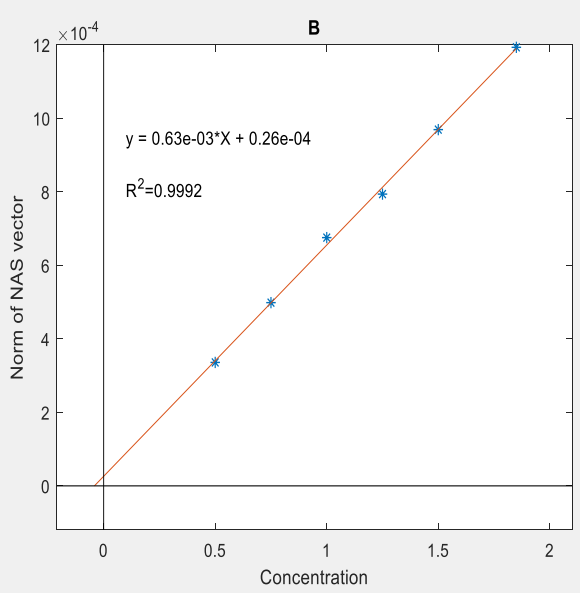
Table 3. Figures of merit for quantification of chitin with SANAS method

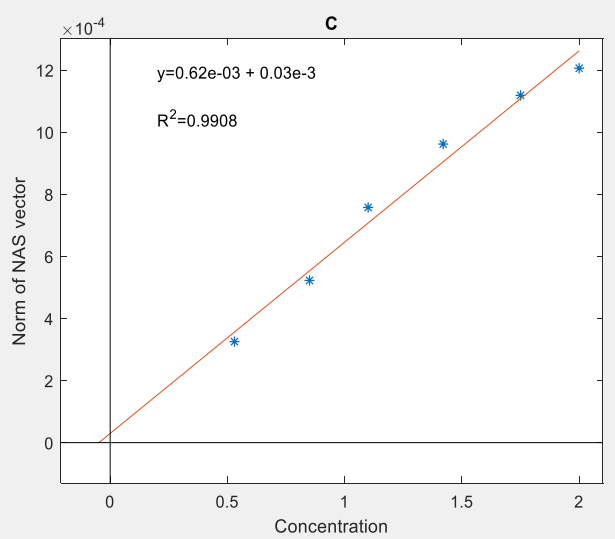
Samples	Selectivity	Sensitivity ($x10^{-4}$)
Purified chitin	0.49	2.03
Raw BSF	0.18	0.52
Fish feed sample-2	0.43	0.72
Fish feed sample-3	0.89	1.40
Fish feed sample-4	0.88	1.80

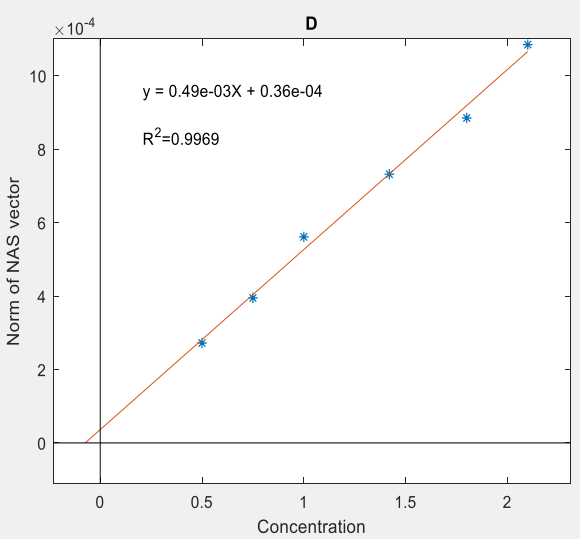
The prediction results for the % of chitin in unknown insect-based fish diets are shown in **Table 4.** The relation between spectra of chitin standards in the presence of matrix effect ($\mathbf{X}_{\mathrm{sm}}$) and their concentrations ($\mathbf{c}_k$) were used to set up a PLS regression model. Multiplying the regression coefficients with the spectra data matrix of standard additions samples ($\mathbf{X}_{sa}$) gives the sum of concentrations contained in the standard and unknown sample. So, subtracting the concentrations of standards from the sum (c_{sa}) provides the unknown concentration of analyte. Based on this calculation, the SANAS predictions obtained for purified chitin of BSF insect and its raw sample were 9.32 and 1.66 % chitin (w/w). These prediction results were lower than those obtained from PLS multivariate calibration, which estimated between 57.8-73.5 and

29.2-34.7% w/w chitin. Furthermore, the estimation of chitin content in four fish feed samples were 1.07, 1.65, 3.12 and 3.83 % w/w. The first sample was expected to contain no chitin.









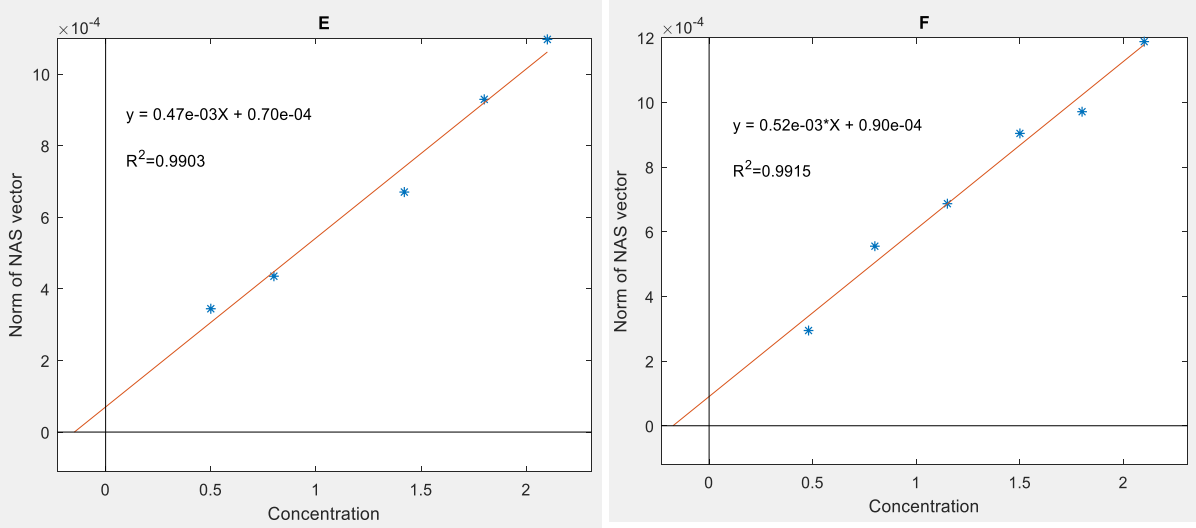


Figure 21. SANAS plots for unknown samples of purified chitin from BSF insect (A), raw BSF insect (B) fish feed sample-1 (blank) (C) and fish feed samples-2-4 (D-F).

Table 4. Predicted percentages of chitin in unknown samples obtained by SANAS analysis method.

sample	Predicted % chitin (w/w)	Correlation coefficient (R^2)	Slope ($x10^{-3}$)
Purified chitin	8.91	0.9641	0.38
Raw BSF	1.66	0.9992	0.63
Fishfeed_1(blank)	1.07	0.9908	0.62
Fishfeed_2	1.65	0.9969	0.49
Fishfeed_3	3.12	0.9903	0.47
Fishfeed_4	3.82	0.9915	0.52

4.2.4. The PLS multivariate calibration method used for predictions of chitin in the unknown samples Partial least square regression (PLSR)

has given significantly higher estimation than the results of SANAS method. This might be due to the degree of removing interferences in NAS method which was better than using the PLS regression method. The interferences caused by chemical components present in the sample mixture might not be removed with the use of spectral pre-processing. Pre-processing methods remove some of the variability associated with physical effects, such as baseline drifts, particle size differences or unwanted light scattering (Sarraguca & Lopes, 2009). However, the regression coefficients of the SANAS plot are so sensitive to the random errors or the presence of outliers (Hemmateenejad & Yousefinejad, 2009) and leads to change the unknown values.

In all of the standard addition methods discussed previously, the chitin purified from shrimp shells were used as a calibration standard to develop the multivariate calibration models. However, the differences in matrices of the shrimp and BSF samples may cause difficulties in quantification of chitin with the SANAS method. **Figure 22** shows the SANAS plot using a series of concentrations (% w/w) from purified BSF powder as a standard spiked into an insect-based fish feed sample. Unfortunately, the unknown concentration of chitin obtained was below zero. Linear SANAS plot (R^2=0.9948) were obtained after rejecting one outlying sample containing six replicate spectra. Insect-based fish feed contains complex matrices from both the BSF insect and fish feed itself which may hinder the quantification of chitin.

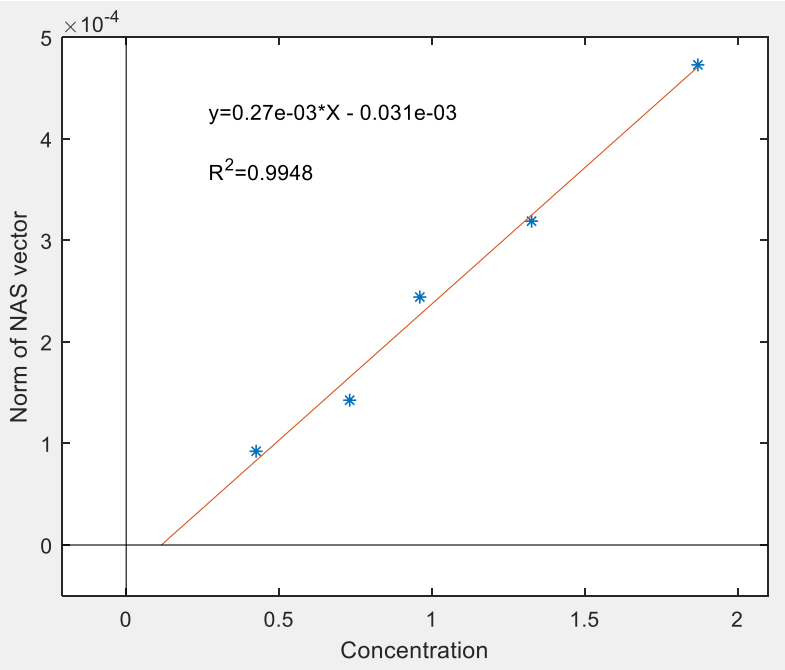


Figure 22. SANAS plot for quantification of chitin from insect-based fish feed using purified BSF as a calibration standard

5. Conclusion

In this work, a rapid and non-destructive DRIFTS-FTIR method has been presented for predictions of chitin % (w/w) in unknown samples of BSF insect-based fish feeds using PLS and SANAS multivariate calibration methods. Samples were ground to smaller particles (<10 µm size) to improve homogeneity in particle distributions and give minimum variations for within sample replicates. Replicate samples spectra collected at the same and one orientation of the sampling cup have showed high variations whereas the spectra collected by manually rotating the DRIFT sampling cup and scanning in the eight orientations for each replicate sample have given more representative spectrum of the sample. Second-derivative of the Savitzky-Golay polynomial data transformation method has given strongest prediction results as compared to EMSC and SNV methods. A PLS regression model have been built by selecting the peaks for amide I, II, and III bands of chitin spectrum and refined further to include only the significant variables with high selectivity ratio. Seemingly irrelevant variables can provide a significant improvement in performance of the model when used in combination with other important variables. So, the reduction of the important variables was done without affecting the model linearity, RMSECV and RMSEP values.

The predictions of chitin (% w/w) in insect-based fish diet samples were also investigated by the application of NAS standard addition multivariate calibration method. The standard addition spectra are both from the contribution of the analyte and matrix effects which gives highly overlapping spectra. The NAS is calculated by projecting the analyte spectra data to a new subspace that is orthogonal to the interferents. The two-dimensional array of DRIFT spectra data is converted to the one-dimensional data similar to the simple univariate calibration method that makes easy for understanding and interpretation of results. The SANAS plot allowed us to calculate the concentration of chitin in the unknown samples in presence of significant matrix effects.

The use of SANAS method has given lower prediction of chitin compared to PLS regression model. These predicted percentage of chitin obtained in purified and raw BSF powders were too low as compared to the nominal values estimated with ADF approach by Fraunhofer IGB, Germany. As the chitin calibration standard used in this study is from purified shrimp shells to quantify chitin in insect-based fish diets, the differences in matrix of both sources might be the problem for giving the unexpected results. It would be ideal if the chitin calibration standard from purified BSF powder are used to quantify chitin in insectmeal provided that the nature of matrix of both calibration standard and test samples are similar.

References

Afseth, N. K. & Kohler, A., 2012. Extended multiplicative signal correction in vibrational spectroscopy. *Chemometrics and Intelligent Laboratory Systems,* Volume 117, pp. 92–99.

Allegrini, F. & Olivieri, A. C., 2014. IUPAC-consistent approach to the limit of detection in partial least squares calibration. *Analytical Chemistry,* Volume 86, pp. 7858−7866.

Armaroli, T., Bécue, T. & Gautier, S., 2004. Diffuse reflection infrared spectroscopy (DRIFTS): application to the in situ analysis of catalysts. *Oil & Gas Science and Technology,* 59(2), pp. 215-237.

Ayoola, A. A., 2015. *Replacement of fishmeal with alternative protein source in aquaculture diets,* Raleigh: MSc thesis, NC State University.

Bansal, A., Chhabra, V., K.Rawal, R. & Sharma, S., 2014. Chemometrics: A new scenario in herbal drug standardization. *Journal of Pharmaceutical Analysis,* 4(4), pp. 223-233.

Barroso, F. G. et al., 2014. The potential of various insect species for use as food for fish. *Aquaculture,* Volume 422–423, pp. 193–201.

Beasley, M. M., Bartelink, E. J., Taylor, L. & Miller, R. M., 2014. Comparison of transmission FTIR, ATR, and DRIFT spectra:implications for assessment of bone bioapatite diagenesis. *Journal of Archaeological Science,* Volume 46, pp. 16-22.

Belghit, I. et al., 2018. Potential of insect-based diets for Atlantic salmon (Salmo salar). *Aquaculture,* Volume 491, pp. 72-81.

Boulesteix, A.-L. & Strimmer, K., 2006. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefing in Bioinformatics,* 8(I), pp. 32-44.

Brereton, R. G., 2000. Introduction to multivariate calibration in analytical chemistry. *The Analst,* Volume 125, pp. 2125-2154.

Brown, S. D., Bear, R. S. & Blank, T. B., 1992. Chemometrics. *Analytical Chemistry,* 64(12), pp. 22-49.

Bu, D., 2007. *Chemometric Analysis for Spectroscopy,* s.l.: CAMO Software Inc..

Caligiania, A. et al., 2018. Composition of black soldier fly prepupae and systematic approaches for extraction and fractionation of proteins, lipids and chitin. *Food Research International,* Volume 105, pp. 812-820.

Caligiani, A. et al., 2018. Composition of black soldier fly prepupae and systematic approaches for extraction and fractionation of proteins, lipids and chitin. *Food Research International,* Issue 105, pp. 812-820.

Camacho, J., Picó, J. & Ferrer, A., 2010. Data understanding with PCA: Structural and variance information plots. *Chemometrics and Intelligent Laboratory Systems,* Volume 100, pp. 48–56.

Chen, Y. et al., 2015. Applications of Micro-Fourier Transform Infrared Spectroscopy (FTIR) in the Geological Sciences—A Review. *International Journal of Molecular Science,* 16(12), pp. 30223-30250.

Christy, A. A., Liang, Y.-Z., Hui, C. & Kvalheim, O. M., 1993. Effect of particle size on diffuse reflectance infrared spectra of polystyrene spheres. *vibrational spectroscopy,* Volume 5, pp. 233-244.

Cuesta, A., Esteban, M. A. n. & Meseguer, J., 2003. In vitro effect of chitin particles on the innate cellular immune system of gilthead seabream (Sparus aurata L.). *Fish and Shellfish Immunology,* Volume 15, pp. 1–11.

Cunha, C. L. et al., 2017. Predicting the properties of biodiesel and its blends using mid-FT-IR spectroscopy and first-order multivariate calibration. *Fuel,* Volume 204, pp. 185-194.

Erdogan, S. & Kaya, M., 2016. High similarity in physicochemical properties of chitin and chitosan from nymphs and adults of a grasshopper. *International Journal of Biological Macromolecules,* Volume 89, pp. 118-126.

Farrésa, M., Platikanova, S., Tsakovski, S. & Taulera, R., 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics,* Volume 29, pp. 528–53.

Finke, M. D., 2007. Estimate of chitin in raw whole insects. *Zoo Biology,* Volume 26, pp. 105-115.

FitzPatrick, M., Champagne, P. & Cunningham, M. F., 2012. Quantitative determination of cellulose dissolved in 1-ethyl-3-methylimidazolium acetate using partial least squares regression on FTIR spectra. *Carbohydrate Polymers,* Volume 87, pp. 1124– 1130.

Flannery, M. B., Sttot, A. W., Briggs, D. E. & Evershed, R. P., 2001. Chitin in the fossil record: identification and quantification of D-glucosamine. *Organic Geochemistry,* Volume 32, pp. 745-754.

Forina, M., Lanteri, S. & Casale, M., 2007. Multivariate calibration. *Journal of Chromatography A,* Volume 1158, pp. 61-93.

Gonil, P. & Sajomsang, W., 2012. Applications of magnetic resonance spectroscopy to chitin from insect cuticles. *International Journal of Biological Macromolecules,* Volume 51, pp. 514– 522.

Gontijo, L. C. et al., 2014. Development and validation of PLS models for quantification of biodiesels content from waste frying oil in diesel by HATR-MIR spectroscopy. *Revista Virtual de Quimica,* 6(5), pp. 1517-1528.

Grdadolnik, J., 2002. ATR-FTIR spectroscopy: its advantages and limitations. *Acta Chimica Slovenica,* Volume 49, pp. 631−642.

Guo, F. et al., 2017. Construction of different calibration models by FTIR/ATR spectra and their application in screening of phenylketonuria. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* Issue 177, pp. 33-40.

Hemmateenejad, B. & Yousefinejad, S., 2009. Multivariate standard addition method solved by net analyte signal calculation and rank annihilation factor analysis. *Analytical and Bioanalytical Chemistry,* Volume 394, pp. 1965–1975.

Henry, M., Gasco, L., Chatzifotis, S. & Piccolo, G., 2018. Does dietary insect meal affect the fish immune system? The case of mealworm, Tenebrio molitor on European sea bass, Dicentrarchus labrax. *Developmental and Comparative Immunology,* Volume 81, pp. 204-209.

Henry, M., Gasco, L., Piccolo, G. & Fountoulaki, E., 2015. Review on the use of insects in the diet of farmed fish: past and future. *Animal Feed Science and Technology,* Volume 203, pp. 1-22.

He, Z., Liu, L., Li, M. & Ma, Z., 2017. Improving specific interval accuracy in multivariate calibration using a net analyte signal-based sample selection method. *Vibrational Spectroscopy* , Volume 92 , pp. 1–8.

Hibbert, D. B., 2016. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations). *Pure and Applied Chemistry,* 88 (4), pp. 407–443.

Houa, X., Lv, S., chen, Z. & xiao, F., 2018. Applications of Fourier transform infrared spectroscopy technologies on asphalt materials. *Measurement,* Volume 121, pp. 304-316.

Huck, C. W., 2015. Advances of infrared spectroscopy in natural product research. *Phytochemistry Letters* , Volume 11 , pp. 384–393.

Islam, S., Bhuiyan, M. A. R. & Islam, M. N., 2016. Chitin and Chitosan: structure, properties and applications in biomedical engineering. *Journal of Polymers and the Environment,* 25(3), pp. 854–866.

Jinda Sirita, S. P. F. C. M., 2007 . Quantitative Analysis of Adsorbate Concentrations by Diffuse Reflectance FT-IR. *Analytical Chemistry* , 79 (10), pp. 3912-3918 .

Karlsen, Ø., Amlund, H., Berg, A. & Olsen, R. E., 2017. The effect of dietary chitin on growth and nutrient digestibility in farmed Atlantic cod, Atlantic salmon and Atlantic halibut. *Aquaculture Research,* Volume 48, pp. 123–133.

Kaya, M. & Baran, T., 2015. Description of a new surface morphology for chitin extracted fromwings of cockroach (Periplaneta americana). *International Journal of Biological Macromolecules,* Volume 75, pp. 7-12.

Kaya, M., Erdogan, S., Molb, A. & Baran, T., 2015. Comparison of chitin structures isolated from seven Orthoptera species. *International Journal of Biological Macromolecules,* Volume 72, pp. 797–805.

Kaya, M. et al., 2017. On chemistry of γ-chitin. *Carbohydrate Polymers,* Volume 176, pp. 177-186.

Khoshhesab, Z. M., 2012. Reflectance IR Spectroscopy. In: *Infrared spectroscopy.* Rijeka: InTech-Open Science, pp. 235.

Kroeckel, S. et al., 2012. When a turbot catches a fly: evaluation of a pre-pupae meal of the black soldier fly (Hermetia illucens) as fish meal substitute — growth performance and chitin degradation in juvenile turbot (Psetta maxima). *Aquaculture,* Volume 364–365, pp. 345–352.

Lavin, B. K., 2000. Chemometrics. *Analytical Chemistry*, 72(12), pp. 91-97.

Li, S. et al., 2017. Defatted black soldier fly (Hermetia illucens) larvae meal in diets for juvenile Jian carp (Cyprinus carpio var. Jian): Growth performance, antioxidant enzyme activities, digestive enzyme activities, intestine and hepatopancreas histological structure. *Aquaculture,* Volume 477, pp. 62-70.

Lorber, A., 1986. Error propagation and figures of merit for quantification by solving matrix equations. *Analytical Chemistry,* 58(6), pp. 1167-1172.

Lorber, A., Faber, K. & Kowalski, B. R., 1997. Net analyte signal calculation in multivariate calibration. *Analytical Chemistry,* 69(8), pp. 1620-1626.

Magalhães, R. et al., 2017. Black soldier fly (Hermetia illucens) pre-pupae meal as a fish meal replacement in diets for European seabass (Dicentrarchus labrax). *Aquaculture,* Volume 476, pp. 79-85.

Majtan, J., Bılikova, K., Markovic, O. & Jan Grof, 2007. Isolation and characterization of chitin from bumblebee (Bombus terrestris). *International Journal of Biological Macromolecules,* Volume 40, pp. 237-241.

Makkar, H. P., Tran, G., Heuzé, V. & Ankers, P., 2014. State-of-the-art on use of insects as animal feed. *Animal Feed Science and Technology,* Volume 197, pp. 1-33.

Mandrile, L. et al., 2018. Detection of insect's meal in compound feed by near infrared spectral imaging. *Food Chemistry,* Volume 267, pp. 240-245.

Marei, N. H. et al., 2016. Isolation and characterization of chitosan from different local insects in Egypt. *International Journal of Biological Macromolecules,* Volume 82, pp. 871–877.

Martínez, K. et al., 2018. Multivariate standard addition for the analysis of overlapping voltammetric signals in the presence of matrix effects: application to the simultaneous determination of hydroquinone and catechol. *Chemometrics and Intelligent Laboratory Systems,* Volume 178, pp. 32-38.

Nitschke, J., Altenbach, H.-J., Malolepszy, T. & Mölleken, H., 2011. A new method for the quantification of chitin and chitosan in edible mushrooms. *Carbohydrate Research,* Volume 346, pp. 1307–1310.

Oliveira, R. R. d., Neves, L. S. d. & Lima, K. M. G. d., 2012. Experimental design, near-infrared spectroscopy, and multivariate calibration: an advanced project in a chemometrics course. *Journal of Chemical Education,* 89(12), pp. 1566−1571.

Ottinger, M., Clauss, K. & Kuenzer, C., 2016. Aquaculture: relevance, distribution, impacts and spatial assessments. *Ocean and Coastal Management,* Volume 119, pp. 244-266.

Rinnan, A., Berg, F. v. d. & Engelsen, S. B., 2009. Review of the Most Common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry,* 28(10), pp. 1201-1222.

Rodriguez-Saona, A. S. a. L., 2009. Fourier Transform Infrared (FTIR) Spectroscopy . In: *Infrared Spectroscopy for Food Quality Analysis and Control* . s.l.:Elsevier Inc., pp. 156-160.

Rodriguez-Saona, A. S. a. L., 2009. Fourier Transform Infrared (FTIR) Spectroscopy. In: *Infrared Spectroscopy for Food Quality Analysis and Control.* s.l.:Elsevier Inc., pp. 156-160.

Rohman, A., Riyanto, S., Sasi, A. M. & Mohd.Yusof, F., 2014. The use of FTIR spectroscopy in combination with chemometrics for the authentication of red fruit (Pandanus conoideus Lam) oil from sunflower and palm oils. *Food Bioscience,* Volume 7, pp. 64-70.

Romía, M. B. & Bernàrdez, M. A., 2009. Multivariate culibration for quantitative analysis. In: D. Sun, ed. *Infrared spectroscopy for food quality analysis and control.* Dublin: Elsevier Inc., pp. 51-82.

Rumpold, B. A. & Schlüter, K. O., 2013. Nutritional composition and safety aspects of edible insects. *Molecular Nutrition and Food Research,* Volume 57, pp. 802–823.

Sajomsang, W. & Gonil, P., 2010. Preparation and characterization of α-chitin from cicada sloughs. *Materials Science and Engineering ,* Volume 30, pp. 357-363.

Sarathjith, M., Das, B. S., Wani, S. P. & Sahrawat, K. L., 2016. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma,* Volume 267, pp. 1-9.

Sarraguca, M. C. & Lopes, J. A., 2009. The use of net analyte signal (NAS) in near infrared spectroscopy pharmaceutical applications: Interpretability and figures of merit. *Analytica Chimica Acta,* Volume 642, pp. 179-185.

Saxberg, B. E. H. & Kowalski, B. R., 1978. The Generalized Standard Addition Method. *Analytical Chemistry,* 51(7), pp. 1031-1038.

Schmitt, J. & Flemming, H.-C., 1998. FTIR-spectroscopy in microbial and material analysis. *International Biodeterioration & Biodegradation,* Volume 41, pp. 1-11.

Silva, M. A., Ferreira, M. H., Braga, J. W. & Sena, M. M., 2012. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. *Talanta,* Volume 89, pp. 342– 351.

Soares, J. A., 2014. Introduction to optical characterization of materials. In: M. Sardela, ed. *Practical Materials Characterization.* New York: Springer, pp. 63-75.

Soon, C. Y. et al., 2018. Extraction and physicochemical characterization of chitin and chitosan from Zophobas morio larvae in varying sodium hydroxide concentration. *International Journal of Biological Macromolecules,* Volume 108, pp. 135-142.

Soon, C. Y. et al., 2018. Extraction and physicochemical characterization of chitin andchitosan from Zophobas morio larvae in varying sodium hydroxide concentration. *International Journal of Biological Macromolecules,* Volume 108, pp. 135–142.

Stuart, B., 2004. *Infrared Spectroscopy: Fundamentals and Applications.* Chichester, United Kingdom : John Wiley & Sons, Ltd.

Subramanian, A. & Rodriguez-Saona, L., 2009. Fourier Transform Infrared (FTIR) Spectroscopy. In: D. Sun, ed. *Infrared spectroscopy for food quality analysis and control.* Dublin: Elsevier Inc., pp. 146-174.

Su, J. et al., 2017. Effects of dietary tenebrio molitor meal on the growth performance, immune response and disease resistance of yellow catfish (Pelteobagrus fulvidraco). *Fish and Shellfish Immunology,* Volume 69, pp. 59-66.

Tefera, T., 2017. *Development of an extraction method for the analysis of chitin in insect-based fish feed by liquid chromtography tandem mass spectrometry*, Bergen: Masters thesis, University of Bergen.

Valderrama, P., Braga, J. W. B. & Poppi, R. J., 2007. Variable selection, outlier detection, and figures of merit estimation in a partial least squares regression multivariate calibration model. a case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry,* 55(21), pp. 8331–8338.

Wasko, A. et al., 2016. The first report of the physicochemical structure of chitin isolatedfrom Hermetia illucens. *International Journal of Biological Macromolecules,* Volume 92, pp. 316-320.

Xin Bao, L. D., 2009. Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties. *Fuel,* Volume 88, pp. 1216–1222.

Zeaiter, M. & Rutledge, D., 2009. Preprocessing methods. In: *Comprehensive Chemometrics.* Hertfordshire: Elsevier, pp. 121-231.

ZhiChao, L., WenSheng, C. & XueGuang, S., 2008. Outlier detection in near-infrared spectroscopic analysis by using Monte Carlo cross-validation. *Science in China Series B: Chemistry,* 51(8), pp. 751-759.

Zhu, X., Cai, J., Yang, J. & Su, Q., 2005. Determination of glucosamine in impure chitin samples by high-performance liquid chromatography. *Carbohydrate Research,* 340(10), pp. 1732–1738.