Master's Thesis

Diet estimation and comparison of fatty acid-based diet modelling methods

Jaakko Litmanen

University of Jyväskylä

Department of Biological and Environmental Science

Environmental Science

19 June 2019

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science
Department of Biological and Environmental Science
Environmental Science

Jaakko Litmanen: Diet estimation and comparison of fatty acid-based diet modelling methods
MSc thesis: 48 p., 3 appendices (9 p.)
Supervisors: Docent Sami Taipale and PhD Tommi Perälä
Reviewers: Docent Silva Uusi-Heikkilä and Docent Sami Taipale
June 2019

Keywords: biological tracers, Daphnia, FASTAR, food web, MixSIR, QFASA, SIAR, zooplankton

Trophic interactions have been a popular research subject among ecologists for decades because understanding the structure of food webs is essential in understanding consumer-resource interactions and complex ecosystems. Consumer diet estimation with biological tracer-based mixing models is an important and recently rapidly developing tool for deciphering aquatic food webs. Bayesian frameworks have been introduced to estimate diets accurately with stable isotope proportions. The low quantity of different stable isotope tracers, however, limits greatly the diet estimation accuracy of complex consumer diets. Therefore, fatty acids have been used as biological tracers to multiply the quantity of tracers in the Bayesian mixing models. Aquatic consumer diet estimation has also been conducted with a numerical optimization method. Daphnia is an important herbivorous zooplankton genus and thus a popular model species to research aquatic food webs. In this thesis I identified the most reliable fatty acid-based diet estimation method by comparing Bayesian methods MixSIR and SIAR conducted with FASTAR, and a numerical method QFASA conducted with QFASAR in R statistical software. These methods were compared with systematic and extensive simulations using an extended version of a previously published Daphnia resource library. MixSIR was the most reliable method. However, the structure of resource library significantly affected the diet estimations, and an upgraded error structure fixing modelling method MixSIAR was published after the testing phase of this thesis. Therefore, I recommend precise construction of the resource library, and the usage of MixSIAR instead of any of the modelling models tested in this thesis.

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta
Bio- ja ympäristötieteiden laitos
Ympäristötiede

Jaakko Litmanen: Ruokavalion arviointi ja rasvahappoperustaisten ruokavaliomallinnusmetodien vertailu
Pro gradu -tutkielma: 48 s., 3 liitettä (9 s.)
Työn ohjaajat: Dosentti Sami Taipale ja FT Tommi Perälä
Tarkastajat: Dosentti Silva Uusi-Heikkilä ja Dosentti Sami Taipale
Kesäkuu 2019

Hakusanat: biomarkkerit, eläinplankton, Daphnia, FASTAR, MixSIR, QFASA, ravintoverkko, SIAR

Trofiatasojen välisten vuorovaikutusten tutkiminen on ollut suosittu aihe ekologien keskuudessa jo vuosikymmeniä, koska ravintoverkkojen ymmärtäminen on oleellista kuluttaja-ruokavalio-suhteiden ja monimutkaisten ekosysteemien tarkastelussa. Kuluttajan ruokavalion arviointi biomarkkereihin perustuvilla sekamalleilla on vesiekosysteemien ravintoverkkojen tärkein ja nopeimmin kehittyvä työkalu. Pysyville isotoopeille on sovellettu Bayesiläisiä metodeja tarkkojen ruokavalioarvioiden selvittämiseksi. Erilaisten pysyvien isotooppien vähäinen määrä on hankaloittanut monimutkaisten ruokavalioiden selvittämistä, joten rasvahappoja on käytetty korvaamaan pysyviä isotooppeja Bayesiläisissä sekamalleissa. Vedessä elävien kuluttajien ruokavalion arviointiin on myös kehitetty numeerinen optimointimetodi. Daphnia on ekologisesti tärkeä ja siten malliorganismina suosittu kasvinsyöjäeläinplanktonsuku. Tässä tutkielmassa selvitin luotettavimman rasvahappopohjaisen ruokavalion mallinnusmetodin vertaamalla FASTAR:lla suoritettuja Bayesiläisiä metodeja, MixSIR:ä ja SIAR:a, sekä QFASAR:lla suoritettua numeerista metodia, QFASA:a, tilasto-ohjelma R:ssä. Vertailussa käytettiin aiemmin julkaistua Daphnia ruokavaliokirjastoa, jota laajensin lisätiedolla, ja jonka avulla loin laajat systemaattiset testit. Testien perusteella MixSIR oli luotettavin mallinnusmetodi, mutta kirjaston rakenne vaikutti merkittävästi testien tuloksiin. Lisäksi uusi, Bayesiläisten mallinnusmetodien virherakenteen korjaava metodi, MixSIAR, julkaistiin tutkielman testivaiheen jälkeen. Tästä syystä suosittelen erityistä huolellisuutta kirjastoa laadittaessa, sekä MixSIAR:n käyttöä tässä tutkielmassa esitettyjen metodien sijaan.

TABLE OF CONTENTS

ABBREVIATIONS

ALA	Alpha-linoleic acid (18:3ω3)
CSIA	Compound-specific Stable Isotope Analysis
EFA	Essential fatty acid
EPA	Eicosapentaenoic acid (20:5ω3)
FASTAR	Fatty Acid Source-Tracing Algorithm in R
JAGS	Just another Gibbs sampler
MCMC	Markov chain Monte Carlo
MDI	Missing diet item
MixSIR	Mixing model utilizing Sample Importance Resampling
PUFA	Poly-unsaturated fatty acid
QFASA	Quantitative Fatty Acid Signature Analysis
QFASAR	Quantitative Fatty Acid Signature Analysis in R
SIAR	Stable Isotope Analysis in R
t-POM	Terrestrial particulate organic matter
t-POMb	t-POM assimilated by bacteria

1 INTRODUCTION

Trophic interactions have been a subject of interest since the 19th century as they explain the transfer of energy and biochemicals in food webs (Müller-Navarra 2008). Understanding the trophic interactions is important because they decipher the interactions and structure of ecological communities and estimate energy, carbon and nutrient flow from primary producers and detritus to higher trophic levels through consumer-resource interactions (Cole et al. 2011, Nielsen et al. 2018). Trophic interactions are also a key element in understanding and predicting the ecological changes of climate change (Harrington et al. 1999, Winder and Schindler 2004).

Estimating consumer diets is an integral part of studying trophic interactions, and multiple different estimation approaches have been used over the past few decades (Nielsen et al. 2018). Visual analysis of stomach content or faecal matter are the most conventional way of analysing diets (e.g. McAtee 1912, Hynes 1950, Katona and Altbäcker 2002) but they don't consider the easily digested parts of the resource (Jobling and Breiby 1986). These unassimilated samples can also be analysed with high throughput DNA sequencing to achieve highly accurate results (Bohmann et al. 2014). The unassimilated samples, however, only show the last few hours or days of the consumers' diet, and therefore analyses based on assimilated biological tracers have been developed to estimate long term diets (Hayden et al. 2014). Observing the stable isotope ratios of carbon or nitrogen in consumer tissues (Boecklen et al. 2011) and observing the fatty acid signature profiles of consumers (Dalsgaard et al. 2003) are currently the most utilized forms of biological tracer analyses (Brett et al. 2016). These studies based on assimilated tracers are based on the "you are what you eat" principle. The small amount of tracers in stable isotope analysis – usually just one or two isotope ratios – makes estimating complex diets biased and potentially unreliable, because potential diet items exceeds the amount of tracers in models (Fry 2013, Brett 2014, Galloway et al. 2015). Bayesian mixing

models have been implemented to accommodate for this issue (e.g. Stock et al. 2018) but another solution is to add more tracers (Fry 2013, Brett 2014). Fatty acid signatures offer tens of markers instead of the few stable isotope ratios, increasing the resolution greatly (Galloway et al. 2015). Fatty acid signature-based diet estimation modelling has been implemented with different statistical approaches (Iverson et al. 2004, Galloway et al. 2015). The analysis of stable isotope ratios of certain biomolecules such as fatty acids or amino acids has also been experimented with recently, but these techniques are so far expensive and do not distinguish between different diets adequately (McMahon et al. 2013, Twining et al. 2016).

Zooplankton are important heterotrophs in aquatic ecosystems linking primary producers and higher trophic level organisms, as they consume microalgae and aquatic bacteria (Galloway et al. 2014a). Zooplankton can also be supported by allochthonous carbon which is synthesized in the catchment area, thus bringing external carbon to the aquatic ecosystem (Cole et al. 2011, Taipale et al. 2016a). *Daphnia* is an ecologically important and well-studied herbivorous zooplankton genus, well suited for diet estimation studies in aquatic ecosystems (Brett et al. 2006).

Here, I review the concepts of trophic markers and biological tracers, the most common diet estimation techniques in aquatic context, and the role of zooplankton in aquatic food webs. I then compare the performance of the most common fatty acid-based diet estimation methods currently used – two Bayesian methods, MixSIR and SIAR, and one method based on numerical optimization, QFASA – with systematic and extensive simulations to evaluate the most reliable diet estimation method. The simulations are constructed with a published *Daphnia* fatty acid signature resource library (Galloway et al. 2014a) supplemented with unpublished data to add data points and new diet items. Lastly, I will discuss the possible challenges and unreliabilities of these modelling methods, and the future possibilities of diet estimation and diet modelling.

2 BACKGROUND

Medium to long term diet analysis is mainly conducted by analysing biological tracers, and choosing the right tracers and analysis methods are an important part of the diet estimation studies (Nielsen et al. 2018). Herbivorous zooplankton are an important microscopic part of the aquatic ecosystem with diets consisting of phytoplankton, bacteria and terrestrial matter (Galloway et al. 2014a). Estimating complex zooplankton diets is unsolvable with stable isotope analysis as there are more diet items than biological tracers, but fatty acid signature analysis with significantly more tracers can yield accurate diet estimates (Galloway et al. 2015).

2.1 Biomarkers, trophic markers and biological tracers

Biomolecules such as lipids, proteins and DNA are essential carbon-based molecules for life. Some of these biomolecules can be used as biological markers, or biomarkers, for the observation of biological changes and states. Good biomarkers are biomolecules that are unique, easily identified, inert, non-harmful to organisms, selectively processed during food uptake and incorporation, metabolically stable, and transferred in the food web (Dalsgaard et al. 2003). Biomarkers can also work as trophic markers to observe carbon, energy and nutrient fluxes throughout trophic levels (Dalsgaard et al. 2003).

While not biomolecules, stable isotopes have been used as biological tracers in ecological studies for decades (Peterson and Fry 1987). They have been used to observe spatial and temporal changes in ecosystems (Altabet and McCarthy 1985) and in diet estimation (Fry 1988, Parnell et al. 2010).

Fatty acids have been observed as robust assimilable biomarkers and trophic markers, as they are essential in all organisms, metabolically stable and transfer between trophic levels (Dalsgaard et al. 2003). Several different other important assimilable biomolecules have been studied such as sterols (Martin-Creuzburg and von Elert 2004, Müller-Navarra 2008, Hiltunen et al. 2017), amino acids (Müller-

Navarra 2008) and pigments (Oechsler-Christensen et al. 2011) but none of them are as good general trophic markers as fatty acids (Galloway et al. 2015). DNA are highly accurate and durable biomolecules, but they are not assimilated, which reduces their usability as a long term biological tracer (Nielsen et al. 2018).

Otoliths, exoskeletons and other unassimilable bodily structures have been used as dietary biomarkers as they are found on digestive tracts and faeces (Jobling and Breiby 1986). While precise, their usage is limited to larger organisms such as fish and mammals, and constrained by the digestion of the structures (Jobling and Breiby 1986).

2.1.1 Stable isotopes

Stable isotopes are stable variants of elements with differing amounts of neutrons in the atoms core. For example, a hydrogen atom normally has no neutrons and only one proton in its core, but there are heavier hydrogen isotopes found in the nature that have a neutron and a proton, called deuterium. Hydrogen also has a third isotope with two neutrons and a proton, called tritium which, however, is not stable, but radioactive. Several other elements such as carbon and nitrogen have stable isotopes. The ratio of these isotopes differs in the nature and this phenomenon is called fractionation (Jobling and Breiby 1986).

The usage of isotopes in natural sciences was discovered in 1930's (Gilfillan 1934). Stable isotopes can be used in ecological science as biological tracers. Most commonly observed isotopes are carbon isotopes, but nitrogen, sulphur, hydrogen and oxygen are used in biological studies as well. All these elements have more than one stable isotope and the ratio of isotopes can be observed with mass spectrometry (Peterson and Fry 1987). Measuring the fractionation of stable isotopes has been utilised since 1950's in geosciences (Dansgaard 1964). The potential of stable isotopes in ecological studies was realized in the 1970's and stable isotopes have been the main empirical method of diet estimation studies since the 1980's (Peterson and Fry 1987).

As carbon and other elements flow through the ecosystems, the isotopes make their way into the tissues of organisms. Because there are notable differences in fractionation in the nature, certain kinds of observed stable isotope ratio profiles can be linked to a certain resource species in a certain part of the aquatic ecosystem (Peterson and Fry 1987). When a consumer consumes an organism with a certain stable isotope ratio it will change the stable isotope ratio of the consumer's tissues (Peterson and Fry 1987). Stable isotopes can be isolated from a number of tissues and studied in bulk (Peterson and Fry 1987) or they can be isolated from certain biomolecules such as fatty acids (Bec et al. 2011) or amino acids (Fantle et al. 1999, McMahon et al. 2010).

2.1.2 Fatty acids

Fatty acids are carboxylic acids that have 4 to 28 carbon atoms, and they can be saturated, unsaturated or branched. Fatty acids are usually incorporated into triglycerides, phospholipids or cholesterol esters, all of which are important dietary sources of energy and structural components for cells.

Fatty acids were discovered in the 19th century (Viénot 2002) and the conservative transfer of fatty acids in aquatic food webs was discovered in the 1930's (Lovern 1935). Like stable isotopes, different resource species have different fatty acid profiles and these profiles carry out to the tissues of the consumer. Since the 1970's, fatty acid signatures have been used as qualitative or semi-quantitative trophic markers and they have since been used to identify some of the key processes in major aquatic ecosystem dynamics (Graeve et al. 1994, Dalsgaard et al. 2003). Since 2004, fatty acid signatures have been used quantitatively in aquatic diet estimation research with different kind of statistical approaches to using the signature data (Iverson et al. 2004, Galloway et al. 2015).

Some poly-unsaturated fatty acids (PUFA) are essential for growth and health in most organisms, and heterotrophs cannot synthesize these essential fatty acids (EFA) themselves (Sargent et al. 1995). By a conservative definition, only omega-6 (ω-6) fatty acid linoleic acid (LIN, 18:2ω-6) and omega-3 (ω-3) fatty acid alpha-

linoleic acid (ALA, 18:3ω-3) (formatted as: (number of carbons):(number of double bonds)ω-(placement of the first double bond)) are essential fatty acids (Parrish 2009). By a more open definition, arachidonic acid (ARA, 20:4ω-6), Eicosapentaenoic acid (EPA, 20:5ω-3) and Docosahexaenoic acid (DHA, 22:6ω-3) are EFA (Parrish 2009, Taipale et al. 2018). The distinction is that ARA, EPA and DHA (essential metabolites) can be theoretically converted from LIN and ALA (essential nutrients) (e.g. EPA can be converted in small quantities from ALA) (Sargent et al. 1995, Parrish 2009, Taipale et al. 2011, Taipale et al. 2018). EFA are synthetized by primary producers and animals must assimilate them from their diet. The imperfect transferring of fatty acids and de novo synthesis of fatty acids must be taken into account when considering fatty acids as biomarkers (Taipale et al. 2011).

There are also essential fatty acids for humans (Das 2006) and the aspect of deficiencies of EFA in human diet and their societal effect brings additional interest into the study of fatty acids (Arts et al. 2001, Lands 2009).

2.1.3 Other biomolecules and their usability as biological tracers

Sterols, or steroid alcohols, are steroids that are synthesized in plants, animals, fungi (eukaryotes) and even some bacteria (prokaryotes) (Wei et al. 2016). For eukaryotes, sterols are an important part of cell membrane construction. Therefore, dietary sterols are important for zooplankton growth (Martin-Creuzburg and von Elert 2004, Hiltunen et al. 2017). Sterol profiles can be used to identify algae (Taipale et al. 2016b). However, plant originated phytosterols do not form such strong unique signatures as fatty acids do in algae (Volkman et al. 1998). Sterols can be self-synthesized, and their weaker signatures make their use as biological tracers species-specific.

Amino acids are the building blocks of proteins and some of the amino acids are essential for animals. These amino acids are synthesized by plants or microorganisms and shift through the food web. Amino acid compositions of tissues are, however, controlled by RNA, thus undermining their usability as

biological tracers, as only a deficiency in certain amino acids could be observed (Sargent et al. 1993).

Pigment composition that determines the colouring of an algae can be species-specific, but most pigments degrade rapidly (Kleppel 1988). The highly specific use-cases and poor durability make pigments mostly obsolete as biological tracers.

DNA is a fundamental building block of the life as we know it. As an important biomolecule, DNA can be used as a highly accurate biological tracer through high throughput sequencing techniques on matter obtained from faeces, stomach content or the environment (Piñol et al. 2014, Bohmann et al. 2014, Nielsen et al. 2018). However, as DNA is not assimilated, its' usability as a long-term biological tracer is weak.

2.2 Diet estimation

Diet estimation has been conducted for decades as trophic interactions have been a topic of interest since the end of 19th century (Müller-Navarra 2008).

Visual analyses or foraging patterns can be useful for larger mammals or birds (Pineda-Munoz and Alroy 2014), and to some extent, harder to observe animals can be analysed with the help of a camera attached to the research subject (Marshall 1998, Iverson et al. 2004).

Visual analysis of stomach or gut content, or faecal matter or scat content are the most intuitive, and for centuries the only method to observe diets (e.g. McAtee 1912, Hynes 1950, Katona and Altbäcker 2002). Although being the most common method, diet analyses based on ingested matter do not consider the easily digested parts of the resource and the estimation is likely biased because of the error caused by digestion (Jobling and Breiby 1986, Baker et al. 2014). These analyses also only show the diet of the last few hours or days (Jobling and Breiby 1986, Hayden et al. 2014). High throughput DNA sequencing can be used to more accurately assess the

contents of these samples, although these samples are still only a short snapshot into the consumer diet (Bohmann et al. 2014).

To achieve more accurate long-term results, diet estimation has shifted to biological tracers such as stable isotopes and fatty acids over the last decades (Nielsen et al. 2018). These methods have at first been used qualitatively but nowadays the increasing computational power has enabled statistical methods to be utilised with relative ease in quantitative diet modelling (e.g. Phillips and Gregg 2003, Iverson et al. 2004, Stock et al. 2018). These methods are based on creating groups of diet items – e.g. one species of fish, or one genus of algae, or terrestrial sources – depending on the used tracer and the species under study.

The basic concept of diet estimation with biological tracers follows the "you are what you eat" principle inversely. In this simple problem the "what you are" can be deduced from "what you eat", but the challenge of the inverse problem in diet estimation is deducing the "what you eat" from "what you are". In a simple example a consumer diet consisting of 80% diet item 1, 10% diet item 2 and 10% diet item 3, results in the consumer tracer profile being a weighed combination of these diet items (Figure 1). In diet estimation the tracer profile of a consumer is known as well as the tracer profiles of the diet items, but the proportions of the diet items are unknown. This moderately simple inverse problem becomes complex when metabolic changes and other biological phenomena affect the transfer of diet item tracers to the consumer tracer profile.

To obtain the most accurate diet estimates, combining different types of biological tracers has been strongly advised (Boschker and Middelburg 2002, Bowen and Iverson 2013, Neubauer and Jensen 2015, Brett et al. 2016, Nielsen et al. 2018).

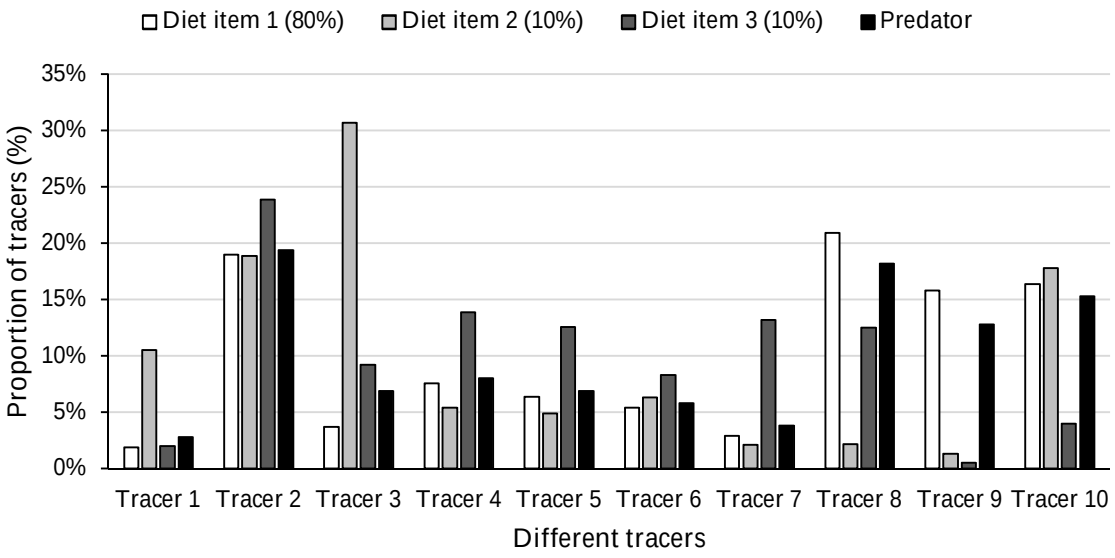


Figure 1. An example of three hypothetical diet item tracer profiles where different amounts of the diet items are consumed and the tracer profile of a consumer. In this example any biological changes to the consumer profile are not accounted for, which makes the consumer profile a weighted average of the diet items.

2.2.1 Stable isotope analysis

Observing the changes in stable isotope fractionations of carbon, nitrogen and sulphur has been the main empirical method of biological tracer-based diet estimation studies since the 1980's (Peterson and Fry 1987). However, most studies nowadays only use carbon and nitrogen isotopes (Boecklen et al. 2011). Although the usage of sulphur, hydrogen or oxygen isotopes can make the estimation more accurate, they are not always useful. Sulphur is practical only for determining some differences, such as detecting the source of the diet item between pelagic and benthic zones or marine and freshwater systems (Connolly et al. 2004, Michener and Kaufman 2007). Supplementing diet estimation data with hydrogen or oxygen isotope analysis data is not very applicable. These isotopes come from both the food and the water the organisms live in, making them less usable dietary tracers, as differentiating between isotopes from the food and the water can be impossible (Vander Zanden et al. 2016).

A mixing model approach for stable isotopes was introduced by Phillips and Gregg (2003) in the form of a program called IsoSource, that was built on top of a

spreadsheet known as IsoError created by Phillips and Gregg (2001). IsoSource, however, had a problem, because with only two or three stable isotope tracers the number of reliable diet estimation is restricted to three to four diet items (Phillips and Gregg 2003). Moore and Semmens (2008) introduced a more advanced Bayesian stable isotope mixing model called MixSIR, which can be utilised in numerical-analysis softwares. MixSIR is an implementation of sample importance resampling (SIR) technique, and it tried to solve the problem with too many diet items in IsoSource by including the possibility of adding prior information (Moore and Semmens 2008). MixSIR does not include a residual error value to account for unknow sources of error. Parnell et al. (2010) introduced an R package called Stable Isotope Analysis in R (SIAR) to add the residual error value to the model. SIAR utilises Markov chain Monte Carlo (MCMC) sampling technique instead of the SIR. SIAR is also the first Bayesian analysis tool to be implemented in the free R Statistical Software (R Core Team 2018).

MixSIR and SIAR, however, have flawed error structures (Stock and Semmens 2016). These flaws are addressed in an R package called MixSIAR, introduced by Stock et al. (2018), which includes several methods to incorporate variance and error to the model. This R-based Bayesian method utilises MCMC algorithm called the Gibbs sampler, and the code is executed with JAGS (Just Another Gibbs Sampler) software (Plummer 2003). The small number of tracers is still, however, a problem that requires meticulous preparation and understanding of the Bayesian method and its assumptions (Stock et al. 2018).

There is also a lesser known Bayesian stable isotope diet estimation software presented by Fernandes et al. (2014) called FRUITS (Food Reconstruction Using Isotopic Transferred Signals). It considers macronutrient dietary routing which has been observed to affect stable isotope dietary estimation (Fernandes et al. 2012) and utilises the Gibbs sampler executed with openBUGS (Bayesian inference Using Gibbs Sampling) (Lunn et al. 2000) instead of the more recent JAGS. FRUITS has not gathered interest after it was published.

Estimating diets with stable isotopes can be highly inaccurate depending on the studied ecosystem, species, trophic level, the occasionally large variances of stable isotope values (Boecklen et al. 2011, Taipale et al. 2011, Brett 2014, Galloway et al. 2015, Brett et al. 2018). The problem of too few tracers is also still present (Fry 2013, Stock et al. 2018). Stable isotope analysis, however, has uses in determining the properties of consumer and resource organisms, the properties of environment, as well as analytical properties, such as acidification (as reviewed by Boecklen et al. 2011).

2.2.2 Fatty acid signature analysis

Diet estimation based on fatty acid analysis is grounded on the resource species leaving their signature profile in to the consumers' fatty acid profile (Dalsgaard et al. 2003), and this profile can be utilised qualitatively (Lovern 1935) or quantitatively (Iverson et al. 2004, Galloway et al. 2015) in diet estimation. The profiles are used like stable isotope ratio profiles, but the possible number of tracers is between 20 and 60 compared to a maximum of five elements and their stable isotopes in stable isotope analysis. This removes the constraint of only few possible diet items, thus making estimating a realistic diet more viable (Galloway et al. 2015). The fatty acid turnover time, a time in which the change in fatty acid profile as a reaction to diet normalises (Graeve et al. 2005), must be considered when conducting fatty acid signature analysis as the change of fatty acid compositions in tissues varies between species (Brett et al. 2009a).

Fatty acid diet modelling took a different approach compared to stable isotopes. The quantitative fatty acid signature analysis (QFASA), a statistical diet estimation method, was first introduced by Iverson et al. (2004). It is "a statistical model that provides quantitative estimates of the proportions of prey species in the diets of individual predators using fatty acid signatures". It is originally based on minimising the Kullback-Leibler divergence measure, which is a measure of how one probability distribution is different from a reference probability distribution, between the observed and predicted fatty acid profiles.

The QFASA model was initially created for marine mammal predators whose diet consist of different fish. A feeding study was carried out in which individuals were fed a homogenous diet of fish for 20 days to ensure fatty acid turnover. The fatty acid composition of the predator was then measured, and the fatty acid composition of the predator represents the fatty acid metabolism for a certain diet. The fatty acid composition of the fish was known as well. The model uses calibration coefficients to account for the predator lipid metabolism calculated from the difference between the prey and predator fatty acid profiles. These calibration coefficients are calculated individually for every fatty acid based on the feeding experiments and then incorporated into the model. The calibration coefficients attempt to capture the fatty acid metabolism by constituting a linear mapping from prey to predator. It is therefore possible to estimate the diet based only on the predator fatty acid analysis and the known prey item fatty acid profiles.

Creating a consumer-resource library is a required step for modelling (Iverson et al. 2004) and if no calibration coefficients are used, the library is constructed of feeding study results with homogenous diet (Happel et al. 2016a). There might be problems with these kind of feeding studies when the consumer cannot sustain itself with certain homogenous diets because of the poor quality of the food (Taipale et al. 2012, Hiltunen et al. 2017). These kinds of diets need a small amount of high-quality supplementary food to keep the consumers alive during the experiments (Taipale et al. 2012). Construction of the resource signature library requires knowledge of all the possible resource items in the habitat of the consumer. The effects of library without the full possible foods has not been studied. If the library is missing resource items, the models are likely to give invalid results because current models cannot estimate unknown diet items but instead fill the void with the ones in the library (Iverson et al. 2004). Sometimes different resource items can be grouped if their fatty acid profiles are close to each other (Galloway et al. 2014a). The differences between these resource items could then be later determined with other biological tracers or prior knowledge of the habitat or consumer-resource interactions. Grouping certain fatty acids can save computing time without affecting the model performance (Neubauer and Jensen 2015, Brett et al. 2016). The

removal of trace amount fatty acids from a library has been tested and debated because of the potential for error (Iverson et al. 2004, Wang et al. 2010, Budge et al. 2012, Neubauer and Jensen 2015) but not removing data points is believed to be the better choice for increased accuracy (Happel et al. 2016a).

In the recent article by Happel et al. (2016a), the calibration coefficients were calculated in various ways to explore the potential problems calculating of these coefficients. First, conventional calculations were conducted with one predator-prey interaction and using those coefficients for all diets. This yielded poor results for anything other than the prey species that the coefficients were based on. Second, the coefficients were calculated for all possible prey items based on individual feeding studies, and the mean of these coefficients was used. This method produced accurate results for all prey items. Third, the coefficients were not calculated at all but the fatty acid signatures of feeding study predators were used as the prey item entries in the library (as in Galloway et al. 2015). This yielded the most accurate results and using this method is advocated, should the time and money budgeted allow it. Considering these aspects, the calibration coefficients have been shown to be highly consumer species dependent (Rosen and Tollit 2012, Happel et al. 2016b). The resource libraries should also not be used interchangeably between different aquatic ecosystems even with similar subject species, as the libraries are highly likely to be ecosystem-specific (Dethier et al. 2013).

An upgraded QFASA-model was introduced by Bromaghin et al. (2017). This new model does not require the estimation of calibration coefficients and thus avoids the need for conducting the potentially expensive and cumbersome controlled feeding studies, as it can calculate the coefficients itself (Bromaghin et al. 2017). This has resulted in more accurate results (Bromaghin et al. 2017). QFASA has also been made into an easy to use R-package QFASAR (Bromaghin 2017) but it does not include the upgraded model for now. QFASAR uses a different scoring method because Bromaghin et al. (2015) found that Aitchison's distance measure gives more accurate results compared to the originally used Kullback-Leibler's divergence measure.

A more recent method to analyse fatty acid signature data is an R-script called FASTAR (Fatty Acid Source-Tracing Algorithm in R), a modelling script introduced by Galloway et al. (2014a). FASTAR uses MixSIR (Moore and Semmens 2008) or SIAR (Parnell et al. 2010), Bayesian mixing models where the posterior distributions for the diet proportions are solved using MCMC Gibbs sampling method implemented in JAGS software (Galloway et al. 2015). Unlike the original MixSIR technique, in FASTAR the SIR technique is replaced with Gibbs sampling (Galloway et al. 2015). Both modelling methods are denounced defective as discussed in previous chapter, but the stable isotope mixing modelling algorithm, MixSIAR (Stock et al. 2018), is not yet been implemented in the context of fatty acid analysis. Unlike QFASA, FASTAR is not designed to incorporate calibration coefficients but instead it requires a library of consumer fatty acid signatures created with the results from homogenous diet feeding tests. The Bayesian modelling methods essentially require all the possible resource species to be analysed in a homogenous diet feeding study, which might be too cumbersome for more complex food webs (Happel et al. 2016a).

Another Bayesian estimation R-package called fastinR is purposely built for fatty acids unlike FASTAR (Neubauer and Jensen 2015). The package allows using fatty acids and stable isotopes at the same time and it is also using JAGS. The fastinR package has, however, gathered little attention.

The fundamental difference between QFASA and FASTAR is in diet estimate structure. The diet estimates of QFASA are just singular values with a standard error derived from a variance matrix. Contrarily the Bayesian diet estimates are posterior probability density distributions of which a point estimate (median in the case of FASTAR) can be calculated. The diet uncertainty should, however, be considered. Therefore, credible intervals should be used in reporting FASTAR diet estimates (Phillips and Gregg 2003).

Studying the fatty acid metabolism of each fatty acid for every consumer species is cumbersome (De Troch et al. 2012). The FASTAR method does not try to estimate metabolism at all (Galloway et al. 2015) but the calibration coefficients of QFASA

are estimated to simulate the metabolism (Iverson et al. 2004). These estimates are, however, just good guesses and only a simple number attempting to describe the complex fatty acid metabolism and biotransformation (Happel et al. 2016a) because of the linear mapping. To add biological realism, existing knowledge of metabolism and biotransformation could be incorporated with a simple non-linear model fitted to the data.

Studies have been conducted with QFASA on a wide variety of vertebrates like mammals (e.g. Iverson et al. 2004, Rosen and Tollit 2012), fishes (Happel et al. 2016a) and birds (Wang et al. 2010). FASTAR has so far been used on studies with zooplankton (Galloway et al. 2014a, Taipale et al. 2016a), isopods (Galloway et al. 2014b) and algae (Strandberg et al. 2015).

2.2.3 Novel methods of diet estimation

Compound-specific stable isotope analysis (CSIA) has been experimented with lately to increase accuracy in stable isotope analysis (Boecklen et al. 2011). Fatty acid -specific stable isotope analysis of carbon has been tested to get more information of fatty acid metabolism and to differentiate between the sources of the fatty acids with studying the stable isotopes of PUFAs (Bec et al. 2011, De Troch et al. 2012, Taipale et al. 2015) and branched fatty acids (Bec et al. 2011, Taipale et al. 2015). CSIA is, however, more expensive and significantly less available compared to bulk stable isotope and fatty acid analyses (Twining et al. 2016). Amino acid-specific carbon and nitrogen isotope analysis has also been experimented with to increase the resolution of stable isotope analysis and to get more information about amino acid metabolism (Fantle et al. 1999, McMahon et al. 2013). Even though these CSIA methods could be used to improve the accuracy of future food web studies, there are problems with them. Only carbon, nitrogen and hydrogen techniques are verifiably reliable, whereas sulphur and oxygen stable isotopes have been problematic. The precision of these measurements varies highly between compounds and instruments, and the derivatization of the compounds makes corrections of the results required, adding more measurement error (Boecklen et al. 2011).

DNA sequencing could be a next step in increasing the accuracy of diet estimation as high throughput sequencing techniques are getting better and less expensive. However, DNA shows only a snapshot of the consumers' diet as DNA does not assimilate and must be collected from stomach, faeces or the environment (Bohmann et al. 2014). For certain species, the DNA tests could find such resource species that are never assimilated into the consumer and thus only disrupt the resource library, thus potentially causing biased results.

2.3 Zooplankton

Zooplankton are a heterogenous group of heterotrophic plankton encountered drifting, floating or slowly swimming freely in almost every aquatic habitat excluding streams and rivers. Jellyfish are among the largest zooplankton but mostly zooplankton are microscopic. Some organisms spend only their infancy as zooplankton (mesoplankton), but most zooplankton are permanent (holoplankton). Some of the important zooplankton in freshwater systems are rotifers, copepods and cladocerans. These herbivorous zooplankton transfer energy and essential biomolecules up in the trophic pyramid of aquatic systems. The role of zooplankton in aquatic food webs has been studied since the beginning of the 20th century (Müller-Navarra 2008).

Zooplankton are an important link between primary production and zooplanktivorous organisms such as fish (Brett and Müller-Navarra 1997). As heterotrophs, zooplankton must obtain essential fatty acids from algae that biosynthesize these fatty acids themselves (Harwood and Guschina 2009, Cagliari et al. 2011). Different consumer organisms, such as zooplankton, can synthesize a limited amount of essential metabolite fatty acids from essential nutrient fatty acids at very low rates but getting these fatty acids from food is required for sustainable growth (Parrish 2009, Taipale et al. 2011).

Cladocerans are a group of small grazing herbivore crustaceans. One of the most common cladocerans are Daphnia, commonly called water fleas. They are an integral

part of practically every standing water aquatic food web as they are an abundant zooplankton genus, thus making them a common study species in aquatic ecosystems. Daphnia has been observed to be selective in its feeding (Hartmann and Kunkel 1991). The fatty acid turnover time has been observed to be six days, thus making Daphnia an excellent model organism (Taipale et al. 2011).

Copepods are another abundant zooplankton subclass. They are a group of small crustaceans of which some are planktonic, and some are benthic. Planktonic copepods are selective feeders (Meunier et al. 2016) and can contain seasonally high amount of lipids due to their overwintering strategies (Kattner and Hagen 2009).

The diet of Daphnia and copepods consists mostly of phytoplankton, but also bacteria and allochthonous sources like terrestrial particulate organic matter (t-POM) or terrestrial dissolved organic matter (t-DOC) (Galloway et al. 2014a). The fatty acid signatures of zooplankton diet items are distinct enough to differentiate between resources on class level (Taipale et al. 2013). Allochthonous carbon can also enter the food web via zooplankters eating bacteria that has consumed terrestrial matter (Hiltunen et al. 2015, Hiltunen et al. 2017).

2.3.1 Zooplankton diet estimation

The importance of allochthonous carbon in aquatic food webs is debated as there is evidence that t-POM is a poor quality resource compared to phytoplankton and it should not therefore support zooplankton communities (Brett et al. 2009b). There is, however, evidence that t-POM would support zooplankton communities based on studies conducted with stable isotopes (Jones et al. 1999, Pace et al. 2004, Doucett et al. 2007, Cole et al. 2011), fatty acids (Masclaux et al. 2013, Taipale et al. 2016a, Hiltunen et al. 2017), and sterols (Hiltunen et al. 2017). Mass carbon flux calculations show that terrestrial inputs are likely overstated (Brett et al. 2012). Additionally, using stable isotopes of hydrogen is observed to provide overstated results of the importance of t-POM (Brett et al. 2018). There is also evidence that when high quality resources, such as phytoplankton, are present zooplankton do not assimilate terrestrial carbon (Galloway et al. 2014a, Taipale et al. 2016a, Hiltunen et al. 2017).

It is highly likely that terrestrial carbon is an important part of the food web in certain instances, but it is believed that the stable isotope studies overestimate the importance of terrestrial carbon in aquatic food webs (Brett et al. 2017).

The quality of zooplankton food can be specified by the availability of certain sterols, and ω-3 and ω-6 PUFA in the diet, as deficiencies in these biomolecules have been linked to poor growth (Brett and Müller-Navarra 1997, Taipale et al. 2014, Hiltunen et al. 2017). Daphnia can convert some EPA from short chain ω-3 PUFA (ALA) at trace levels (<1%), but only with high presence of energy (Goulden and Place 1990, von Elert 2002, Taipale et al. 2011). Deficiency in dietary phosphorous has been observed to restrain growth and egg production (Becker and Boersma 2005). The quality of food has been considered in at least one study by adding quality data as prior information to the FASTAR model (Galloway et al. 2014a).

Temperature and starvation has been found to affect the results of zooplankton fatty acid studies as the copepod fatty acid composition adaptation to cold temperatures has been established over the past decades (Brett et al. 2009a). A study showed that a starved Daphnia pulex can adjust and retain its PUFA to survive colder temperatures (11 °C) and that the mortality was considerably higher at a higher temperature (22 °C) (Schlechtriem et al. 2006). Few of notable observations were concluded in a recent study feeding study focusing on the effects of temperature on Eudiaptomus gracilis, the most common calanoid copepod species in European lakes (Koussoroplis et al. 2014). Copepods can lose half of their body mass when starved and use their PUFA reserves unevenly during that time. If these copepods are fed after starving, they partially recovered only in higher temperatures (8 °C or 12 °C) contrarily to low temperature (4 °C). The quality of food also affected recovery rate and success of these copepods. Light intensity, phosphorous supply and temperature have been observed to affect algal fatty acid compositions (Piepho et al. 2012). These concepts have yet to be considered in the fatty acid signature analysis modelling studies.

The bioconversion of ω-3 and ω-6 PUFA has been studied on Daphnia (von Elert 2002, Taipale et al. 2011, Strandberg et al. 2014) and copepods (De Troch et al. 2012). Zooplankton has also been observed to selectively incorporate the most physiologically important fatty acids, although the mechanism is not yet determined (Burns et al. 2011). These observations have not been incorporated in modelling studies either.

DNA sequencing assist in the coming diet estimation studies, but it does not account for diet assimilation. Nonetheless, DNA sequencing could be used to determine which food items to choose for the model as it can present a comprehensive list of all possible algae and bacteria in the system.

In the context of zooplankton and the debate over the role of terrestrial matter in aquatic food webs, getting the quality of diet items right in diet estimation is more important than the exact diet. There is a substantial amount of data about food quality and fatty acid metabolism in different environments and they should be somehow incorporated into the models to increase estimation accuracy and validity.

3 DATA AND METHODS

3.1 Daphnia fatty acid signature diet library

The Daphnia fatty acid signature resource library used in this study was based on a library presented by Galloway et al. (2014a) which was expanded by adding several new previously unpublished datapoints for existing resource items and new resource items. The resource items in this resource library are formed from the results of homogenous diet feeding tests conducted on Daphnia and presented as quantitative fatty acid proportions with 33 fatty acid tracers. The newly constructed library (N = 151) consists of algae (cryptomonads (N = 16), diatoms (N = 22), dinoflagellate (N = 11), euglenoids (N = 12), golden algae (N = 8) and green algae

(N = 33)), bacteria (actinobacteria (N = 5), cyanobacteria (N = 20) and methane oxidizing bacteria (MOB) (N = 6)), terrestrial particulate organic matter (t-POM) (N = 11) and bacteria which has consumed t-POM (t-POMb) (N = 7). To illustrate the diversity of the diet item groups a principal component analysis was conducted on the library.

Several algae-specific fatty acid groups were created as it should not affect the results (Wang et al. 2010, Budge et al. 2012, Neubauer and Jensen 2015, Happel et al. 2016a), but make the library more concentrated to save computing time (Neubauer and Jensen 2015). No fatty acids were removed from the library but some harder to distinguish fatty acids were grouped to mitigate error.

Generally, the consumer library is structured as follows: $x_1 = (x_{11}, x_{12}, ..., x_{1M})$, …, $x_N = (x_{N1}, x_{N2}, ..., x_{NM})$, where the fatty acid profiles of N resource items x_i consist of M relative fatty acid proportions x_{ij} that sum up to 100%. In the library of this study there are 151 resource items and 33 fatty acids. The consumer fatty acid profiles are constructed similarly: $y = (y_1, y_2, ..., y_M)$, where the consumer fatty acid profile y consist of M relative fatty acid proportions y_j that sum up to 100%.

Gathering the fatty acid signature data was not a part of this thesis. However, the data which forms the consumer profiles and diet libraries in this study are constructed trough analysing the samples by extracting fatty acids from the samples and then derivatising them to fatty acid methyl esters. These methyl esters are then analysed in a gas-chromatograph and identified by comparing retention times to known standards and quantified based on standard dilution series (as described by Brett et al. 2009b).

3.2 Modelling methods

Three modelling methods – namely Bayesian methods MixSIR (Moore and Semmens 2008) and SIAR (Parnell et al. 2010) performed with FASTAR R-script version 4.1 (Galloway et al. 2015) and a numerical method QFASA (Iverson et al. 2004) performed with QFASAR package (Bromaghin 2017) – were examined in this

thesis using the free R Statistical Software version 3.5 (R Core Team 2018). To reach the most objective results, no prior knowledge about the relative proportions of the diet items in the consumer's diet based on food quality were used, as QFASA cannot utilize such information.

Although MixSIAR (Stock et al. 2018) – the new modelling method replacing MixSIR and SIAR – had already been published, the bulk of the analyses in this thesis was already performed at the time of its publication. Furthermore, the new model was not implemented on fatty acids. Thus, MixSIAR was not included in this thesis.

The model is similar for all methods:

$$y = \sum_i \pi_i x_i \,, \tag{1}$$

where the consumer fatty acid profile y is known as well as all the resource item signatures x_i. The relative proportion π_i of each diet item i in the consumer diet are, however, unknown. The modelling methods estimate the relative proportion π_i of each fatty acid in each diet item x_{ij} for each of the consumers' fatty acids y_j.

3.2.1 FASTAR

The FASTAR R-script can utilize two almost identical Bayesian modelling methods to fit models to data; MixSIR and SIAR. Each diet item group is used in the FASTAR models as a single diet item, constructed by calculating the fatty acid proportion mean $\bar{x}_{ij}$ and variance σ_{ij} within the diet item group. The equations presented here are derived from their respective articles (MixSIR: Moore and Semmens 2008, SIAR: Parnell et al. 2010) with the stable isotope-specific parts excluded.

The MixSIR model conducts the conditional probability density functions of y_j given π is modelled as a normal distribution:

$$p(y_j|\pi) = N\left(y_j; \sum_i \pi_i \bar{x}_{ij}, \sum_i \pi_i^2 \sigma_{ij}^2\right), \quad (2)$$

where the mean of the normal distribution is $\sum_i \pi_i \bar{x}_{ij}$ and variance is $\sum_i \pi_i^2 \sigma_{ij}^2$. In this study no prior knowledge about the relative proportions of the diet items in the consumer's diet were used, so π was given an uninformative Dirichlet prior probability density function:

$$p(\pi) = Dir(\pi; \alpha_i, \ldots, \alpha_N), \quad (3)$$

where a_i = 1 for all i. The posterior probability density function can be written:

$$p(\pi|y) \propto p(y_1, \ldots, y_M|\pi)\, p(\pi), \quad (4)$$

as per the Bayes rule. The model assumes that y_j are mutually conditionally independent and thus:

$$p(y_1, \ldots, y_M|\pi) = \prod_j p(y_j|\pi). \quad (5)$$

Therefore, the probability density function of π given the fatty acid profile of y can then be written:

$$p(\pi|y) \propto p(\pi) \prod_j p(y_j|\pi), \quad (6)$$

which was computed in JAGS. MCMC chains were run for 100 000 iterations with a 50 000 iteration burn-in and a thinning rate of 50 (Galloway et al. 2015). The density plots were then produced of the MCMC posterior sample.

To add inter-observation variance, a residual error term ε_j, constructed from the gamma distribution, was introduced in SIAR:

$$p(y_j|\pi, \varepsilon_j) = N\left(y_j; \sum_i \pi_i \bar{x}_{ij}, \sum_i \pi_i^2 \sigma_{ij}^2 + \varepsilon_j\right), \quad (7)$$

$$p(\varepsilon_j) = Gamma(\varepsilon_j, 0.001, 0.001). \quad (8)$$

Otherwise the operation of the model is identical to MixSIR.

3.2.2 QFASA

The numerical approach of QFASA differs significantly from the Bayesian modelling methods. Here the diet item groups are represented only as the mean $\bar{x}_{ij}$ of the diet items within each diet item group. Because calibration coefficients were not used they were excluded from the equations presented by Bromaghin et al. (2015).

QFASA numerically minimises the Aitchison distance measure:

$$Q(\pi; y, x_1, \ldots, x_M) = \sum_j \left(\log\left[\frac{y_j}{g(y)}\right] - \log\left[\frac{\sum_i \pi_i \bar{x}_i}{g(\sum_i \pi_i \bar{x}_i)}\right]\right)^2. \quad (9)$$

The estimate $\hat{\pi}_i$ of the diet item proportion π_i is then obtained as the solution to the following optimization problem:

$$\hat{\pi}_i = \min_{\pi} Q(\pi; y, x_1, \ldots, x_M), \quad (10)$$

when $\sum_i \pi_i = 1$, $\pi_i \geq 0$ and $\pi_i \leq 1$. The objective functions are minimised using a reduced gradient algorithm with a golden ratio line search (Bromaghin et al. 2015).

3.3 Simulation tests

The modelling methods were compared with synthetic systematic and extensive (N = 11132) simulations to find out about the models' ability to find diet items that are not in the consumer's diet, i.e. missing diet items (MDI). The synthetic data was

constructed with systematic matrices to simulate between one to nine MDIs per diet and designed in a way where one diet item was the primary source (Appendix 1; Table 1). Pseudo-consumers were created with weighing every resource fatty acid in the Daphnia library with 33 fatty acid tracers based on the systematic matrices. The simulations were then carried out using the pseudo-consumers and the Daphnia library.

The overall performance of the models was assessed in an equivalent way with an attempt at a relatively more realistic systematic test with no MDIs. The test matrix was created by creating a weighing matrix to generate random deviates from the Dirichlet distribution in R (rdirichlet()-function) to achieve more realistic diet proportions (N = 2761). The result of the rdirichlet()-function is a matrix of which each row contains weighed randomly generated values that sum to 1, which then can be used as diet proportions for simulations (Appendix 1; Table 2).

The effect of concentrating libraries by grouping and removing tracers was studied with constructing libraries from the Daphnia data library with 33 and 50 tracers, and testing the models' performance with systematic synthetic consumer data with one MDI generated with the rdirichlet()-function in R (Appendix 1; Table 3) (N = 495).

Three methods of comparison were used to explore the differences between the modelling methods. The absolute errors, i.e. the absolute distance from the real value, of MixSIR and SIAR diet estimate medians and QFASA diet estimates were used to calculate the average absolute error of the modelling methods. Credible intervals for FASTAR and confidence intervals for QFASA of 68th, 95th and 99.7th percentiles were assessed for each model to calculate the average of how often the real diet value was on the diet estimation distribution. Confidence interval is an interval centred around the mean ± standard deviation, normal distribution is expected (Figure 2). In this study, credible interval is a density plot produced of the MCMC posterior sample with the centre being the median of which the percentages are calculated. Hereafter the credible interval and confidence interval are noted as confidence interval. 95th percentile is the most commonly used confidence interval, but 68th percentile identifies the accuracy of the estimates, and 99.7th percentile tells

whether the diet estimate has the real value on its' distribution at all. In an equivalent manner with the confidence intervals, the models' capability of detecting MDIs was calculated. The confidence interval comparisons were also calculated on individual diet items to uncover potential problems with certain diet items on certain models. The mean absolute errors and the mean confidence interval errors were calculated with their errors of mean to assess the performance of the models.

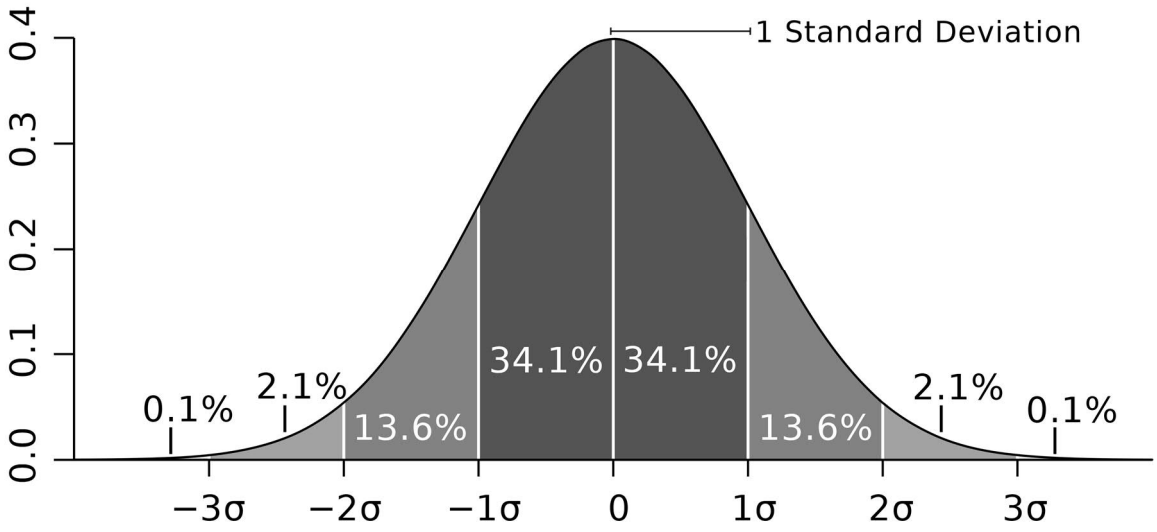


Figure 2. Confidence intervals are based on standard deviation (SD, σ). Here is illustrated the confidence intervals of 68% with ± 1 SD, 95% with ± 2 SD and 99.7% (~100%) with ± 3 SD.

The most problematic diet items were examined by summing the instances of overestimating and underestimating the real value and calculating their proportions to determine the direction of their bias.

3.4 Tests with natural consumer data

The new *Daphnia* library was tested on a set of natural unpublished consumer data to estimate *Daphnia* diet compositions, and whether there were any significant differences between the estimates of the models. The original library (Galloway et al. 2014a) was also compared to the new one to see differences between the two libraries, and whether the original library had deficiencies in diet items. The old

library is constructed in a way that QFASA cannot utilize it, thus QFASA diet estimations could not be conducted with the old library. The average of four unpublished Daphnia profiles obtained from an eutrophicated lake (a lake enriched with minerals and nutrients resulting in excessive growth of algae) were used to conduct diet estimations to assess the differences of the models and libraries. Estimates of single consumers were utilized in pointing out potential problems with the modelling methods.

4 RESULTS

The diet items of the resource library were diverse (e.g. the MOB group consists of two distinct clusters) and overlapped with each other (e.g. the golden algae group is completely within the cryptomonad group) according to the two first principal components of the principal component analysis (Figure 3). The PC1 explained 36.9% of the variability and the vector consisted of the following fatty acids: 16:1ω7 (0.5), 18:3ω3 (ALA) (-0.8) and c20:5ω3 (EPA) (0.3). The PC2 explained 17.0% of the variability and the vector consisted of the following fatty acids: 16:1ω7 (-0.7), 18:3ω3 (ALA) (-0.2), 18:4ω3 (0.4) and c20:5ω3 (EPA) (0.5).

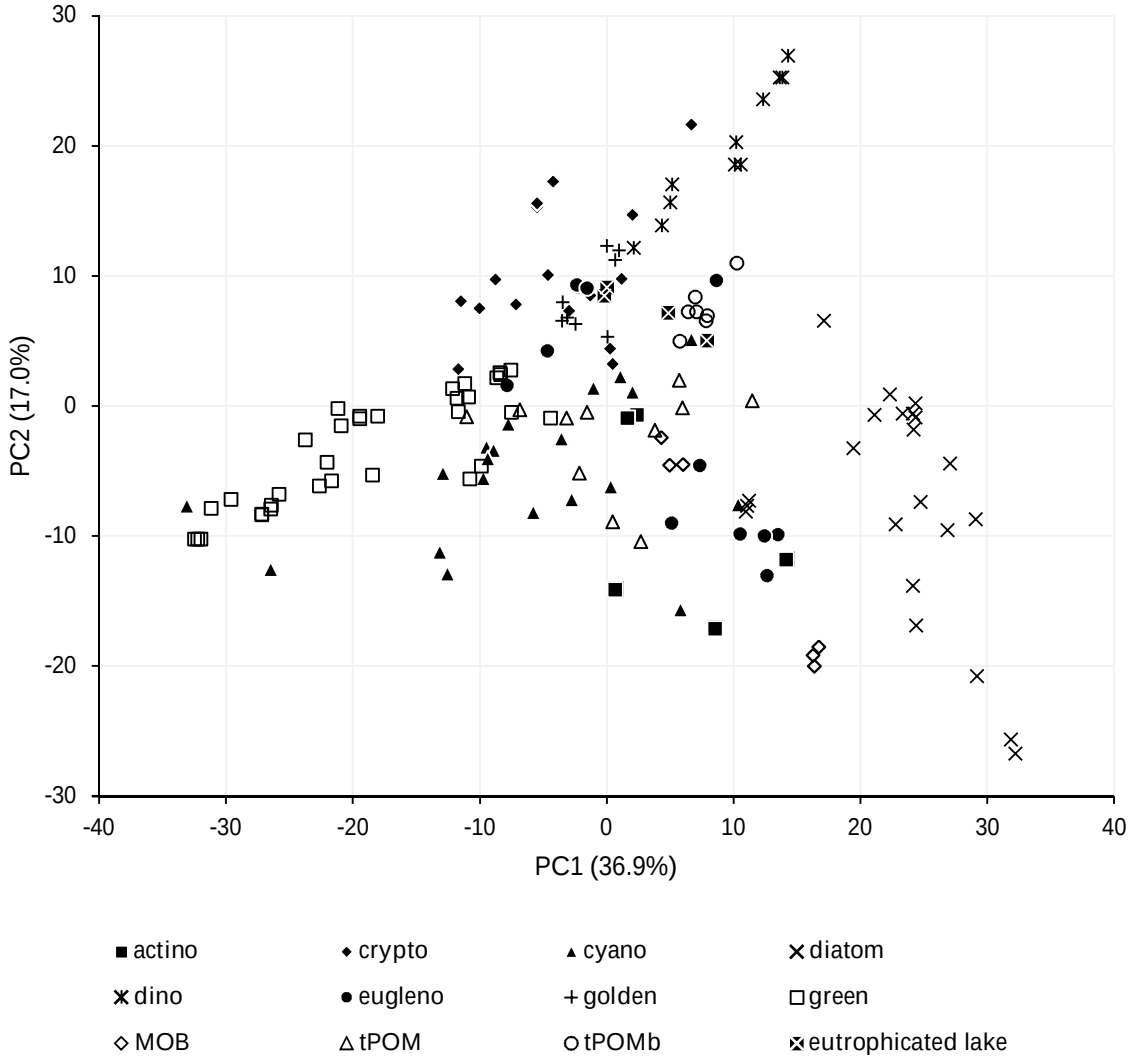


Figure 3. The principal component analysis plot of the used extended Daphnia resource library and the eutrophicated lake samples, where PC1 explains 36.9% and PC2 explains 17.0%.

4.1 Simulation tests

The average absolute error was less than 6% in all models as MixSIR (2.8 ± 0.1%) performed the best followed by SIAR (4.4 ± 0.1%) and QFASA (5.4 ± 0.2%) (Table 1). MixSIR was the most accurate method according to the confidence interval comparison with 20.0 ± 0.3% average error at the 95th percentile, while SIAR (44.6 ± 0.2%) and QFASA (47.1 ± 0.4%) performed inferiorly (Table 1). At the 95th percentile of finding MDIs from the confidence interval, QFASA (32.8 ± 0.5%) was the most accurate, followed by MixSIR (41.1 ± 0.6%) and SIAR (89.0 ± 0.3%) (Table 1).

Table 1. The mean absolute errors and confidence interval errors for all cases and finding MDIs with the errors of the means. The most important values are bolded.

N = 11132			MixSIR	SIAR	QFASA
Average absolute error (%)			2.8 ± 0.1	4.4 ± 0.1	5.4 ± 0.2
Average confidence interval error (%)	All cases	68th	41.0 ± 0.4	58.2 ± 0.2	59.9 ± 0.4
		95th	20.0 ± 0.3	44.6 ± 0.2	47.1 ± 0.4
		99.7th	1.9 ± 0.1	2.5 ± 0.1	38.6 ± 0.4
	Finding MDIs	68th	59.1 ± 0.6	100.0 ± 0.0	40.3 ± 0.5
		95th	41.1 ± 0.6	89.0 ± 0.3	32.8 ± 0.5
		99.7th	2.0 ± 0.2	0.8 ± 0.1	27.3 ± 0.5

SIAR had the highest probable error of the models (75% at lowest) whereas QFASA is stable at around 33% (Figure 4). MixSIR could identify MDIs with lowest error if the consumer diet consists of more than 6 MDIs (29% to 12%) while QFASA has the lowest error with the number of MDIs lower than 7 (36% to 32%) (Figure 4).

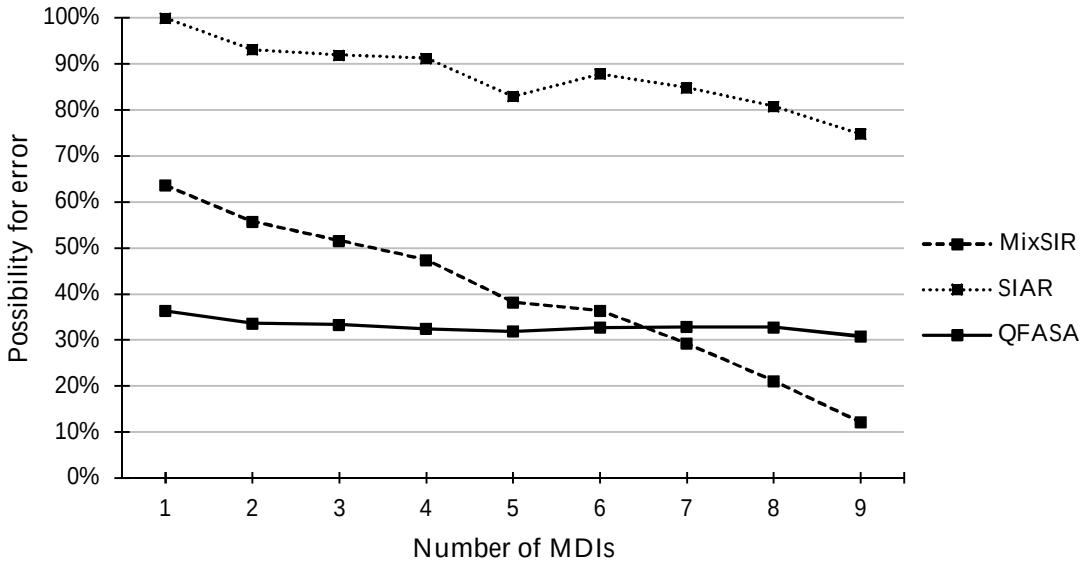


Figure 4. The average 95th percentile confidence interval errors of different number of MDIs in the consumer diet for the different models. SIAR is poor at estimating diets with MDIs (75% to 100% average error). MixSIR is the most accurate with more than six MDIs (12% to 29%). QFASA is the most consistent at finding the MDIs with less than seven MDIs (32% to 36%).

The average diet specific 95th percentile confidence interval errors range with MixSIR from 0.4% to 41.0%, SIAR from 11.2% to 54.1%, and QFASA from 26.4% to 81.4% (Appendix 2; Table 1). The differences were more moderate at the 99.7th percentile with MixSIR ranging from 0.0% to 6.9%, SIAR from 0.2% to 6.6%, and QFASA from 20.1% to 47.5% (Appendix 2; Table 1). The average diet specific 95th percentile confidence interval errors of finding MDIs range with MixSIR from 0.5% to 63.5%, SIAR from 41.7% to 99.6%, and QFASA from 6.3% to 74.6% (Appendix 2; Table 2). The differences were, again, more moderate at the 99.7th percentile with MixSIR ranging from 0.0% to 5.5%, SIAR from 0.0% to 2.4%, and QFASA from 5.0% to 67.1% (Appendix 2; Table 2).

The average absolute errors of the more realistic synthetic diet test were higher than in the test with MDIs as the average errors were for MixSIR (6.5 ± 0.2%), for SIAR (6.0 ± 0.1%) and for QFASA (8.4 ± 0.2%) (Table 2). At the 95th confidence interval MixSIR was the most accurate (7.1 ± 0.2%) while closely followed by SIAR (10.8 ± 0.2%) while QFASA was struggling (68.8 ± 0.3%) (Table 2).

Table 2. The mean absolute errors and confidence interval errors with the errors of the means of the realistic diet test. The most important values are bolded.

N = 2761		MixSIR	SIAR	QFASA
Average absolute error (%)		6.5 ± 0.2	6.0 ± 0.1	8.4 ± 0.2
Average confidence interval error (%)	68th	34.0 ± 0.3	41.2 ± 0.2	81.9 ± 0.2
	95th	7.1 ± 0.2	10.8 ± 0.2	68.8 ± 0.3
	99.7th	1.2 ± 0.1	1.6 ± 0.1	59.8 ± 0.3

The average diet specific 95th percentile confidence interval errors range with MixSIR from 0.0% to 15.5%, SIAR from 6.5% to 17.1%, and QFASA from 50.7% to 91.2% (Appendix 2; Table 3). The differences were more moderate but reaching its highest value slightly higher than at the test with MDIs at the 99.7th percentile with MixSIR ranging from 0.0% to 8.3%, SIAR from 0.0% to 7.5%, and QFASA from 42.9% to 88.7%, with the highest errors of the Bayesian methods originating from t-POM (Appendix 2; Table 3).

The diet estimates for t-POM were distinctly underestimated. The underestimated t-POM proportions were 76% for MixSIR, 87% for SIAR and 69% for QFASA.

Grouping tracers had a positive effect on the performance of QFASA based on the 95th percentile confidence interval error, which decreased from 71.0% to 59.0% with a more condensed library, while MixSIR and SIAR perform slightly better with the larger library (Table 3).

Table 3. The mean 95th percentile confidence interval errors (%) and the errors of the mean of 50 fatty acid resource library and the same library with grouped fatty acids, resulting in a 33 fatty acid library. Smaller error implies better model performance.

N= 495	95% confidence interval error (%)		
Number of fatty acids	MixSIR	SIAR	QFASA
50	7.2 ± 0.4	9.0 ± 0.2	71.0 ± 0.7
33	7.9 ± 0.3	10.8 ± 0.2	59.0 ± 0.8

4.2 Tests with natural consumer data

Comparing the diet estimates conducted on the new library to the estimates conducted on the one presented in the study by Galloway et al. (2014a) show that there can be large differences between libraries (e.g. actinobacteria, cryptomonads, diatoms, dinoflagellates and MOB), and between models (e.g. cryptomonads, diatoms, dinoflagellates and t-POMb) (Table 4). Density plots of these results illustrate the distribution of the estimates (Appendix 3; Figure 1). The density plots produced by the Bayesian methods can be bimodal or distorted (Appendix 3; Figure 2).

Table 4. Average diet estimates as proportions (%) and the errors of means of different modelling methods and the old library (Galloway et al. 2014a) and the new library extended in this thesis.

N = 4	Average diet estimates (%)				
	MixSIR		SIAR		QFASA
resource	new	old	new	old	new
actino	14.8 ± 2.3	4.4 ± 0.6	8.1 ± 1.1	4.4 ± 0.7	6.4 ± 1.8
crypto	14.1 ± 5.5	35.4 ± 7.7	21.4 ± 5.5	36.4 ± 4.9	24.2 ± 8.3
cyano	1.1 ± 0.1	1.7 ± 0.6	3.4 ± 0.3	4.6 ± 0.7	4.0 ± 2.6
diatom	22.6 ± 4.1	8.6 ± 1.1	10.2 ± 3.8	20.0 ± 5.0	6.1 ± 2.5
dino	25.6 ± 1.6	10.9 ± 6.4	12.7 ± 0.6	4.4 ± 0.6	31.6 ± 0.2
eugleno	2.0 ± 0.4		7.7 ± 1.4		9.2 ± 5.1
golden	1.4 ± 0.3	0.4 ± 0.0	6.6 ± 1.5	4.5 ± 1.7	7.2 ± 3.7
green	7.2 ± 3.2	8.2 ± 2.3	5.0 ± 1.3	9.9 ± 2.7	0.0 ± 0.0
MOB	3.9 ± 2.2	29.4 ± 2.8	2.1 ± 0.2	0.8 ± 0.2	1.0 ± 1.0
tPOM	0.9 ± 0.1	0.4 ± 0.1	4.0 ± 0.6	5.2 ± 0.8	0.0 ± 0.0
tPOMb	1.4 ± 0.2		7.6 ± 1.9		10.3 ± 3.6

5 DISCUSSION

In this thesis I compared three fatty acid profile-based diet estimation modelling methods. Based on the results, the Bayesian methods exceeds QFASA in diet estimation with MixSIR being the most accurate modelling method before SIAR. Not only were the Bayesian methods more accurate in absolute terms but the nature of Bayesian diet estimates with distributions instead of single values makes the estimates profoundly more realistic.

The resource library seems to have a profound effect on the results on all modelling methods. Having MDIs in the resource library might distort the results greatly and none of the methods can reliably identify MDIs. Being certain of having all the possible diet items in the resource library is paramount and the grouping should be carefully created. More realistic models for fatty acid metabolism and selective

assimilation could be incorporated into the models. The grouping of fatty acids to save computing time is not justifiable.

In a recent study, FASTAR and QFASAR were compared with the diets of Beluga whales (*Delphinapterus leucas*) (Choy et al. 2019). The study used calibration coefficients with MixSIR and QFASA, and in the study QFASA performed better. For MixSIR the calibration coefficients were accounted for in the resource diet profiles and the uninformative Dirichlet priors were used. The researchers point out that the QFASA model is more robust to errors in calibration coefficients, which is likely to be true. Implementing such information into a Bayesian model should be done by an informative prior rather than transforming the data.

During the writing phase of this thesis a more advanced Bayesian modelling method, MixSIAR, was published to replace MixSIR and SIAR (Stock et al. 2018). Thus, future fatty acid research should be conducted with the new method.

5.1 Simulation tests

MixSIR was the most accurate model on almost all accounts followed closely by SIAR. Contrarily QFASA performed poorly even though the library was formatted to benefit QFASA. The overall performance based on the most common metric, 95th percentile confidence interval, on both finding the real diet proportion and finding MDIs shows that none of the models are highly accurate. Finding the real diet proportions with MixSIR had an error rate of 20%, whereas SIAR and QFASA had error rates over 40%. At best, these models can find the real diet proportion only four times out of five. Finding MDIs is even more inaccurate as at least third of the MDIs are not recognized. Furthermore, these average error rates are biased by the differing performance with different quantities of MDIs in the synthetic consumer diets (Figure 4).

5.1.1 Finding the real diet proportions

MixSIR was the most accurate model in the simulation test with MDIs (Table 1). The average absolute error was smallest, as were the confidence interval errors for all cases on all three percentiles. QFASA suffers from consistently high error rates at all confidence interval percentiles, and SIAR performs at the same level as QFASA on finding real proportions on 68th and 95th percentiles. However, the accuracy increases greatly at the 99.7th percentile for the Bayesian methods, meaning that only 1.9% (MixSIR) or 2.5% (SIAR) of the time they do not have the real diet on their distribution, unlike the error rate of QFASA at over 38%. Similar pattern can be noted with the more realistic pseudo-consumers (Table 2). Here the Bayesian methods are fairly accurate at even the 95th percentile with average errors of 7.1% (MixSIR) and 10.8% (SIAR), while QFASA is very inaccurate at even the 99.7th percentile with almost a 60% average error.

The diet estimation errors can be diet item-specific with different models (Appendix 2; Tables 1 and 3). Diet specific analyses of confidence interval errors shows that diet estimation is diet item specific particularly at the 95th percentile but the differences are moderate at the 99.7th percentile for the Bayesian methods. MixSIR was adequately accurate – under 5% mean error at 95th percentile confidence interval – with diatoms, euglenoids and MOB whereas neither SIAR nor QFASA could identify any diet item with under a 10% error (Appendix 2; Table 1). At the same limit of 5% mean error at the 99.7th percentile had MixSIR only struggling with t-POM and SIAR with cyanobacteria and t-POM, while QFASA's lowest error was 20% with actinobacteria (Appendix 2; Table 1). The estimates of the Bayesian methods were more accurate at the realistic pseudo-consumer test at the 95th percentile but the 99.7th percentile errors were similar, while QFASA was more accurate with the MDI-dataset compared to the realistic pseudo-consumers (Appendix 2; Tables 1 and 3).

These observations highlight the inferior performance of QFASA on a resource library that has diversity within diet item groups but not enough distinction between diet item groups (Figure 3). With low-diversity diet item groups the

estimates are highly accurate (Happel et al. 2016b). The amount of diet items should not be the reason for this as there is evidence that the model can handle more than eleven diet items accurately (e.g. Iverson et al. 2004, Bromaghin et al. 2017). Improving the resource library is recommended for more accurate results with QFASA (Bromaghin et al. 2016). The results of FASTAR models are in line with more concise simulations conducted in a previous study (Galloway et al. 2015). In the realistic pseudo-consumer test the Bayesian methods' vague Dirichlet distribution prior, that expects a priori that every diet item has been consumed at the same level, might be beneficial for identifying trace amounts diet items.

The most interesting diet item in the current zooplankton diet debate is t-POM (Brett et al. 2017) which also happened to be the most problematic for the models to identify; it was most commonly not identified even in the 99.7th confidence interval by the Bayesian models (Appendix 2; Tables 1 and 3). The diet item group was regularly underestimated by the models as the group was underestimated 69% to 87% of the time. Thus, the role of t-POM might be moderately underestimated by the methods. This might be due to the diverse samples of the t-POM diet item group (Figure 3).

5.1.2 Finding the missing diet items

The more there are MDIs in the consumer diet, the more likely it is that the models can identify them in their 95th percentile confidence intervals (Figure 4). Finding MDIs from the diet is generally not reliable on any model – except finding missing MOB with MixSIR – based on the average 95th percentile confidence interval errors. However, the Bayesian methods can identify the MDIs on their 99.7th percentile distribution (error less than 5.6% in all diets), unlike the struggling QFASA (ranging from 5.0% to 67.1%) (Appendix 2; Table 2).

Having an MDI in the consumer diet is somewhat likely to not be found and this phantom diet item could change the results particularly if the MDI was not found. Thus, it is important for the models to be able to recognize MDIs because all diet items might not be present in all habitats. With current tools, it should be heavily

considered whether there could be diet items in the resource library that are not in the diet of the consumer and remove those diet items or downweigh their role greatly with the use of informative prior distributions in the Bayesian methods.

5.1.3 Grouping tracers

Grouping fatty acid tracers seems to have a slightly negative effect on SIAR and MixSIR while it has a positive effect on QFASA. This shows that grouping fatty acids is not so straightforward. QFASA makes use of the grouping while the Bayesian methods are not greatly impacted by it. Here MixSIR is again the least erroneous model closely followed by SIAR, while QFASA performs very poorly with either library. The condensed library was used based on this test for the rest of the tests as it had such a small impact on the Bayesian methods and clearly a positive effect on QFASA, and it would save computing time.

However, the computing time is only a real problem on large simulation studies like the tests in this thesis. In the normal use of the Bayesian modelling methods the increased time needed for computing is tens of minutes at most. Thus, saving relatively little time should not justify losing accuracy with the Bayesian methods.

5.2 Tests with natural consumer data

There are significant differences in the diet estimates conducted on Daphnia from nature (Table 4). The differences are between the old and the new libraries as well as between the different models.

There is a chance that the Daphnia diets include euglenoids and t-POMb as all the models estimate at least some amount of these diet items. This highlights the importance of the diet library as based on just these tests it is impossible to assess whether euglenoids or t-POMb are MDIs or not. However, using prior knowledge of the quality of these resources could affect the results and even conclude on these diet items being MDIs.

A good example of the impact of the diet item group diversity is the amount of MOB in the estimates. Although MOB can be a valuable resource for Daphnia (Taipale et al. 2008), it should not compose 29 to 39 percent of the diet even at an eutrophicated lake (Taipale et al. 2016a) as MixSIR estimates with the old library (Table 4). These kinds of anomalies could be corrected with the usage of an informative a priori or with constructing a more uniform diet item groups in the resource library.

The reporting of density plots instead of medians of the diet estimate is crucial in avoiding losing information of a bimodal distribution, a distorted distribution (Appendix 3; Figure 2) or the overall reliability of a single value estimate.

Based on my experiments the most accurate diet estimates are produced by MixSIR when using the new library, as there are substantial anomalies with MOB using the old library. The potentially considerable effects on diet estimates stemming from the resource library variance should, however, be acknowledged with all modelling methods.

5.3 Criticism of the modelling methods

MixSIR and SIAR have deficiencies in their error structure as reported by Stock and Semmens (2016). The residual error of SIAR can be problematic (Semmens et al. 2009). Adding a residual error to the model makes the model generalist, resulting in poor recognition of MDIs, potentially distorting the estimates greatly. The problems with the error structure have been observed and repaired in the MixSIAR method by Stock et al. (2018). Additionally, MixSIR can have problems with MDIs where the MCMC algorithm can get stuck making the density plot produced from the posterior MCMC sample distorted which did not, however, seem to affect the estimate accuracy in my tests (Appendix 3; Figure 2).

Grouping and removing fatty acid tracers have different effects in different models. Furthermore, the different libraries can give highly differing estimates on certain diet items with the same consumer data. The convention of the FASTAR resource library consisting of mean and standard deviance instead of the individual diet item

samples is a potential source of error. A problem with MixSIR, the old Daphnia library and the issue with MOB, is a good example of this kind of error. Condensing the resource library in this fashion eliminates the possibility to remove anomalies, and to reassess the groupings to achieve better diet items groups. Although the QFASA library format is complete, the method itself only calculates the means of individual diet items within the diet item group. Omitting variance data from the model is worse than the approach in FASTAR.

5.4 Future insights of diet estimation

The fatty acid profiles in higher trophic levels originate from the primary producers through the zooplankton further into the higher trophic levels. This means that e.g. all fish and mammals get their fatty acid compositions fundamentally from primary producers and detritus apart from a few fatty acids they can produce themselves. The importance of lake-specific (or any ecosystem-specific) feeding trials has been noted before (Budge et al. 2002, Galloway et al. 2012, Happel et al. 2016a, Choy et al. 2019). There are, however, preliminary results indicating that some parts of a resource library could be used between adequately similar ecosystems (Happel et al. 2019). The individual lake specific trials are expensive and being more holistic will probably make diet estimation of higher trophic level organisms even more expensive. It must be considered, however, whether the current models are serviceable at all if they are not holistic enough.

Understanding the food web interactions of different ecosystems is critical in comprehending the effects of climate change (Harrington et al. 1999, Winder and Schindler 2004, Blanchard 2015). Accurate information of food webs can also clarify the source of phenomena such as trophic cascades, as all perturbations in ecosystems are not caused by effects of climate change, but other human action (Pace et al. 1999). Long term research and monitoring of food web interactions is important in identifying ongoing changes in ecosystems (Hays et al. 2005, Brown et al. 2010). This information could be utilized in planning nature conservation

projects and legislation, and the aspiration for a more holistic approach to diet estimation is an essential piece in the future.

5.4.1 The future of fatty acid-based modelling

There is movement to get rid of MCMC algorithms and JAGS (e.g. Ward et al. 2011, Neubauer and Jensen 2015) but so far they mostly bring more complexity and longer computing times (Stock et al. 2018). Lengthy computing times of MCMC-algorithms could be more efficiently conducted with STAN (Carpenter et al. 2017) instead of JAGS with a potential for more accurate results (Monnahan et al. 2017).

Temperature, light, food quality and fatty acid metabolism should be more meticulously considered in the models. To make the Bayesian framework more relevant with larger organisms such as mammals the proper introduction of calibration coefficients into the model should be considered. Tools for assessing the robustness of the fatty acid tracer-based resource library could be developed to ease the usage of the modelling method. The new toolset should also take into account the poor consideration or ignoring of Bayesian model assumptions that has been noted as a problem in ecological studies (Stock et al. 2018).

5.5 Test deficiencies

MixSIAR was published during the writing stage of this thesis, and it would have had a more robust error structure and would have made the experimental part of the thesis more relevant. However, analyses would have had to been done again shifting the publication of the thesis by at least a month.

QFASA deficiencies should have been examined more clearly as it has worked in other papers adequately. More diet item entries or different grouping of diet items could help with the assessing of the problems. It is also possible that the library is poorly constructed as the relative diversity within diet item groups and relative similarity between diet item groups is high (Figure 3).

Prior information about the diet proportions were not used in the tests in this study to reach objectivity, because unlike FASTAR, QFASA cannot utilize such information. Adding that information would, however, have had a significant effect on the Bayesian estimates of the natural consumer diet estimates. The role of food quality was dismissed in constructing the systematic studies. Thus, the tests include a high number of highly improbable diet compositions potentially distorting the results.

This thesis' experiments were ultimately constrained by time. The deficiencies in the experiments were noticed during the writing phase of the thesis and to get more satisfactory results at least an extra month of testing would have been required.

6 CONCLUSIONS

The feasibility of diet estimation must be considered in terms of cost and accuracy. Based on the literary review and my experiments, future diet estimation should thrive for a more holistic approach with multiple diet estimation methods particularly in higher trophic levels.

Although the Bayesian methods MixSIR and SIAR proved more accurate than QFASA, they have been replaced with MixSIAR (Stock et al. 2018). Thus, using MixSIAR-based analysis is advised in the future. Adding prior information of food quality to the model is the first step to get more accurate results, but other parameters such as temperature and fatty acid bioconversion data are the next steps in fatty acid-based diet estimation. The role of the resource library is highly crucial in achieving accurate results. Thus, particular care should be taken to construct an ecosystem-specific resource library with sufficiently distinct diet item groups. The next generation fatty acid-based diet estimation method could have tools to deal with these challenges.

ACKNOWLEDGEMENTS

I thank Sami Taipale and Tommi Perälä for guiding and encouraging me through the thesis process. I thank Laura Piiroinen for understanding the pain of authoring a thesis and providing feedback on the thesis.

REFERENCES

Altabet M.A. & McCarthy J.J. 1985. Temporal and spatial variations in the natural abundance of 15N in PON from a warm-core ring. Deep Sea Res A 32: 755-772.

Arts M.T., Ackman R.G. & Holub B.J. 2001. " Essential fatty acids" in aquatic ecosystems: a crucial link between diet and human health and evolution. Can J Fish Aquat Sci 58: 122-137.

Baker R., Buckland A. & Sheaves M. 2014. Fish gut content analysis: robust measures of diet composition. Fish Fish 15: 170-177.

Bec A., Perga M., Koussoroplis A., Bardoux G., Desvilettes C., Bourdier G. & Mariotti A. 2011. Assessing the reliability of fatty acid–specific stable isotope analysis for trophic studies. Methods Ecol Evol 2: 651-659.

Becker C. & Boersma M. 2005. Differential effects of phosphorus and fatty acids on Daphnia magna growth and reproduction. Limnol Oceanogr 50: 388-397.

Blanchard J.L. 2015. Climate change: A rewired food web. Nature 527: 173.

Boecklen W.J., Yarnes C.T., Cook B.A. & James A.C. 2011. On the use of stable isotopes in trophic ecology. Annu Rev Ecol Evol Syst 42: 411-440.

Bohmann K., Evans A., Gilbert M.T.P., Carvalho G.R., Creer S., Knapp M., Douglas W.Y. & De Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol Evol 29: 358-367.

Boschker H. & Middelburg J.J. 2002. Stable isotopes and biomarkers in microbial ecology. FEMS Microbiol Ecol 40: 85-95.

Bowen W.D. & Iverson S.J. 2013. Methods of estimating marine mammal diets: a review of validation experiments and sources of bias and uncertainty. Mar Mamm Sci 29: 719-754.

Brett M.T. 2014. Resource polygon geometry predicts Bayesian stable isotope mixing model bias. Mar Ecol Prog Ser 514: 1-12.

Brett M.T., Arhonditsis G.B., Chandra S. & Kainz M.J. 2012. Mass flux calculations show strong allochthonous support of freshwater zooplankton production is unlikely. PLoS One 7: e39508.

Brett M.T., Bunn S.E., Chandra S., Galloway A.W., Guo F., Kainz M.J., Kankaala P., Lau D.C., Moulton T.P. & Power M.E. 2017. How important are terrestrial organic carbon inputs for secondary production in freshwater ecosystems? Freshwat Biol 62: 833-853.

Brett M.T., Eisenlord M.E. & Galloway A. 2016. Using multiple tracers and directly accounting for trophic modification improves dietary mixing-model performance. Ecosphere 7.

Brett M.T., Holtgrieve G.W. & Schindler D.E. 2018. An assessment of assumptions and uncertainty in deuterium-based estimates of terrestrial subsidies to aquatic consumers. Ecology 99: 1073-1088.

Brett M.T., Kainz M.J., Taipale S.J. & Seshan H. 2009b. Phytoplankton, not allochthonous carbon, sustains herbivorous zooplankton production. PNAS 106: 21197-21201.

Brett M.T., Möller-Navarra D.C., Ballantyne A.P., Ravet J.L. & Goldman C.R. 2006. Daphnia fatty acid composition reflects that of their diet. Limnol Oceanogr 51: 2428-2437.

Brett M.T. & Müller-Navarra D. 1997. The role of highly unsaturated fatty acids in aquatic foodweb processes. Freshwat Biol 38: 483-499.

Brett M.T., Müller-Navarra D.C. & Persson J. 2009a. Crustacean zooplankton fatty acid composition. In: Anonymous Lipids in aquatic ecosystems, Springer, pp. 115-146.

Bromaghin J.F. 2017. qfasar: Quantitative fatty acid signature analysis with R. Methods Ecol Evol 8: 1158-1162.

Bromaghin J.F., Budge S.M., Thiemann G.W. & Rode K.D. 2016. Assessing the robustness of quantitative fatty acid signature analysis to assumption violations. Methods Ecol Evol 7: 51-59.

Bromaghin J.F., Budge S.M., Thiemann G.W. & Rode K.D. 2017. Simultaneous estimation of diet composition and calibration coefficients with fatty acid signature data. Ecol Evol 7: 6103-6113.

Bromaghin J.F., Rode K.D., Budge S.M. & Thiemann G.W. 2015. Distance measures and optimization spaces in quantitative fatty acid signature analysis. Ecology and evolution 5: 1249-1262.

Brown C.J., Fulton E.A., Hobday A.J., Matear R.J., Possingham H.P., Bulman C., Christensen V., Forrest R.E., Gehrke P.C. & Gribble N.A. 2010. Effects of climate-driven primary production change on marine food webs: implications for fisheries and conservation. Glob Chang Biol 16: 1194-1212.

Budge S.M., Iverson S.J., Bowen W.D. & Ackman R.G. 2002. Among-and within-species variability in fatty acid signatures of marine fish and invertebrates on the Scotian Shelf, Georges Bank, and southern Gulf of St. Lawrence. Can J Fish Aquat Sci 59: 886-898.

Budge S.M., Penney S.N. & Lall S.P. 2012. Estimating diets of Atlantic salmon (Salmo salar) using fatty acid signature analyses; validation with controlled feeding studies. Can J Fish Aquat Sci 69: 1033-1046.

Burns C.W., Brett M.T. & Schallenberg M. 2011. A comparison of the trophic transfer of fatty acids in freshwater plankton by cladocerans and calanoid copepods. Freshwat Biol 56: 889-903.

Cagliari A., Margis R., dos Santos Maraschin F., Turchetto-Zolet A.C., Loss G. & Margis-Pinheiro M. 2011. Biosynthesis of triacylglycerols (TAGs) in plants and algae. Int J Plant Sci 2: e10.

Carpenter B., Gelman A., Hoffman M.D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P. & Riddell A. 2017. Stan: A probabilistic programming language. J Stat Softw 76.

Choy E.S., Sheehan B., Haulena M., Rosenberg B., Roth J.D. & Loseto L.L. 2019. A comparison of diet estimates of captive beluga whales using fatty acid mixing models with their true diets. J Exp Mar Biol Ecol 516: 132-139.

Cole J.J., Carpenter S.R., Kitchell J., Pace M.L., Solomon C.T. & Weidel B. 2011. Strong evidence for terrestrial support of zooplankton in small lakes based on stable isotopes of carbon, nitrogen, and hydrogen. PNAS 108: 1975-1980.

Connolly R.M., Guest M.A., Melville A.J. & Oakes J.M. 2004. Sulfur stable isotopes separate producers in marine food-web analysis. Oecologia 138: 161-167.

Dalsgaard J., St John M., Kattner G., Müller-Navarra D. & Hagen W. 2003. Fatty acid trophic markers in the pelagic marine environment. Adv Mar Biol 46: 225-340.

Dansgaard W. 1964. Stable isotopes in precipitation. Tellus 16: 436-468.

Das U.N. 2006. Essential fatty acids: biochemistry, physiology and pathology. Biotechnol J 1: 420-439.

De Troch M., Boeckx P., Cnudde C., Van Gansbeke D., Vanreusel A., Vincx M. & Caramujo M.J. 2012. Bioconversion of fatty acids at the basis of marine food webs: insights from a compound-specific stable isotope analysis. Mar Ecol Prog Ser 465: 53-67.

Dethier M.N., Sosik E., Galloway A.W., Duggins D.O. & Simenstad C.A. 2013. Addressing assumptions: variation in stable isotopes and fatty acids of marine macrophytes can confound conclusions of food web studies. Mar Ecol Prog Ser 478: 1-14.

Doucett R.R., Marks J.C., Blinn D.W., Caron M. & Hungate B.A. 2007. Measuring terrestrial subsidies to aquatic food webs using stable isotopes of hydrogen. Ecology 88: 1587-1592.

Fantle M.S., Dittel A.I., Schwalm S.M., Epifanio C.E. & Fogel M.L. 1999. A food web analysis of the juvenile blue crab, Callinectes sapidus, using stable isotopes in whole animals and individual amino acids. Oecologia 120: 416-426.

Fernandes R., Millard A.R., Brabec M., Nadeau M. & Grootes P. 2014. Food reconstruction using isotopic transferred signals (FRUITS): a Bayesian model for diet reconstruction. PloS one 9: e87436.

Fernandes R., Nadeau M. & Grootes P.M. 2012. Macronutrient-based model for dietary carbon routing in bone collagen and bioapatite. Archaeol Anthrop Sci 4: 291-301.

Fry B. 1988. Food web structure on Georges Bank from stable C, N, and S isotopic compositions. Limnol Oceanogr 33: 1182-1190.

Fry B. 2013. Alternative approaches for solving underdetermined isotope mixing problems. Mar Ecol Prog Ser 472: 1-13.

Galloway A.W., Brett M.T., Holtgrieve G.W., Ward E.J., Ballantyne A.P., Burns C.W., Kainz M.J., Müller-Navarra D.C., Persson J. & Ravet J.L. 2015. A fatty acid based Bayesian approach for inferring diet in aquatic consumers. PloS one 10: e0129723.

Galloway A.W., Britton-Simmons K.H., Duggins D.O., Gabrielson P.W. & Brett M.T. 2012. Fatty acid signatures differentiate marine macrophytes at ordinal and family ranks. J Phycol 48: 956-965.

Galloway A.W., Eisenlord M.E., Dethier M.N., Holtgrieve G.W. & Brett M.T. 2014b. Quantitative estimates of isopod resource utilization using a Bayesian fatty acid mixing model. Mar Ecol Prog Ser 507: 219-232.

Galloway A.W., Taipale S.J., Hiltunen M., Peltomaa E., Strandberg U., Brett M.T. & Kankaala P. 2014a. Diet-specific biomarkers show that high-quality phytoplankton fuels herbivorous zooplankton in large boreal lakes. Freshwat Biol 59: 1902-1915.

Gilfillan E.S.J. 1934. The isotopic composition of sea water. J Am Chem Soc 56: 406-408.

Goulden C.E. & Place A.R. 1990. Fatty acid synthesis and accumulation rates in daphniids. J Exp Zool A Ecol Genet Physiol 256: 168-178.

Graeve M., Albers C. & Kattner G. 2005. Assimilation and biosynthesis of lipids in Arctic Calanus species based on feeding experiments with a 13C labelled diatom. J Exp Mar Biol Ecol 317: 109-125.

Graeve M., Kattner G. & Hagen W. 1994. Diet-induced changes in the fatty acid composition of Arctic herbivorous copepods: experimental evidence of trophic markers. J Exp Mar Biol Ecol 182: 97-110.

Happel A., Maier C., Farese N., Czesny S. & Rinchard J. 2019. Fatty acids differentiate consumers despite variation within prey fatty acid profiles. Freshwat Biol .

Happel A., Stratton L., Kolb C., Hays C., Rinchard J. & Czesny S. 2016a. Evaluating quantitative fatty acid signature analysis (QFASA) in fish using controlled feeding experiments. Can J Fish Aquat Sci 73: 1222-1229.

Happel A., Stratton L., Pattridge R., Rinchard J. & Czesny S. 2016b. Fatty-acid profiles of juvenile lake trout reflect experimental diets consisting of natural prey. Freshwat Biol 61: 1466-1476.

Harrington R., Woiwod I. & Sparks T. 1999. Climate change and trophic interactions. Trends Ecol Evol 14: 146-150.

Hartmann H.J. & Kunkel D.D. 1991. Mechanisms of food selection in Daphnia. In: Anonymous Biology of Cladocera, Springer, pp. 129-154.

Harwood J.L. & Guschina I.A. 2009. The versatility of algae and their lipid metabolism. Biochimie 91: 679-684.

Hayden B., Harrod C. & Kahilainen K.K. 2014. Dual fuels: intra-annual variation in the relative importance of benthic and pelagic resources to maintenance, growth and reproduction in a generalist salmonid fish. J Anim Ecol 83: 1501-1512.

Hays G.C., Richardson A.J. & Robinson C. 2005. Climate change and marine plankton. Trends Ecol Evol 20: 337-344.

Hiltunen M., Honkanen M., Taipale S., Strandberg U. & Kankaala P. 2017. Trophic upgrading via the microbial food web may link terrestrial dissolved organic matter to Daphnia. J Plankton Res 39: 861-869.

Hiltunen M., Strandberg U., Taipale S.J. & Kankaala P. 2015. Taxonomic identity and phytoplankton diet affect fatty acid composition of zooplankton in large lakes with differing dissolved organic carbon concentration. Limnol Oceanogr 60: 303-317.

Hynes H. 1950. The food of fresh-water sticklebacks (Gasterosteus aculeatus and Pygosteus pungitius), with a review of methods used in studies of the food of fishes. J Anim Ecol 19: 36-58.

Iverson S.J., Field C., Don Bowen W. & Blanchard W. 2004. Quantitative fatty acid signature analysis: a new method of estimating predator diets. Ecol Monogr 74: 211-235.

Jobling M. & Breiby A. 1986. The use and abuse of fish otoliths in studies of feeding habits of marine piscivores. Sarsia 71: 265-274.

Jones R.I., Grey J., Sleep D. & Arvola L. 1999. Stable isotope analysis of zooplankton carbon nutrition in humic lakes. Oikos : 97-104.

Katona K. & Altbäcker V. 2002. Diet estimation by faeces analysis: sampling optimisation for the European hare. Folia Zool 51: 11-16.

Kattner G. & Hagen W. 2009. Lipids in marine copepods: latitudinal characteristics and perspective to global warming. In: Anonymous Lipids in aquatic ecosystems, Springer, pp. 257-280.

Kleppel G.S. 1988. Plant and animal pigments as trophodynamic indicators. In: Anonymous Marine organisms as indicators, Springer, pp. 73-90.

Koussoroplis A., Nussbaumer J., Arts M.T., Guschina I.A. & Kainz M.J. 2014. Famine and feast in a common freshwater calanoid: Effects of diet and temperature on fatty acid dynamics of Eudiaptomus gracilis. Limnol Oceanogr 59: 947-958.

Lands W.E. 2009. Human life: caught in the food web. In: Anonymous Lipids in aquatic ecosystems, Springer, pp. 327-354.

Lovern J.A. 1935. Fat metabolism in fishes: The depot fats of certain fish fed on known diets. Biochem J 29: 1894.

Lunn D.J., Thomas A., Best N. & Spiegelhalter D. 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10: 325-337.

Marshall G.J. 1998. Crittercam: an animal-borne imaging and data logging system. Mar Technol Soc J 32: 11.

Martin-Creuzburg D. & von Elert E. 2004. Impact of 10 dietary sterols on growth and reproduction of Daphnia galeata. J Chem Ecol 30: 483-500.

Masclaux H., Perga M., Kagami M., Desvilettes C., Bourdier G. & Bec A. 2013. How pollen organic matter enters freshwater food webs. Limnol Oceanogr 58: 1185-1195.

McAtee W.L. 1912. Methods of estimating the contents of bird stomachs. Auk 29: 449-464.

McMahon K.W., Fogel M.L., Elsdon T.S. & Thorrold S.R. 2010. Carbon isotope fractionation of amino acids in fish muscle reflects biosynthesis and isotopic routing from dietary protein. J Anim Ecol 79: 1132-1141.

McMahon K.W., Hamady L.L. & Thorrold S.R. 2013. Ocean ecogeochemistry: a review. Oceanogr Mar Biol 51: 327-373.

Meunier C.L., Boersma M., Wiltshire K.H. & Malzahn A.M. 2016. Zooplankton eat what they need: copepod selective feeding and potential consequences for marine systems. Oikos 125: 50-58.

Michener R.H. & Kaufman L. 2007. Stable isotope ratios as tracers in marine food webs: an update. Stable isotopes in ecology and environmental science 2: 238-282.

Monnahan C.C., Thorson J.T. & Branch T.A. 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods Ecol Evol 8: 339-348.

Moore J.W. & Semmens B.X. 2008. Incorporating uncertainty and prior information into stable isotope mixing models. Ecol Lett 11: 470-480.

Müller-Navarra D.C. 2008. Food web paradigms: the biochemical view on trophic interactions. Int Rev Hydrobiol 93: 489-505.

Neubauer P. & Jensen O.P. 2015. Bayesian estimation of predator diet composition from fatty acids and stable isotopes. PeerJ 3: e920.

Nielsen J.M., Clare E.L., Hayden B., Brett M.T. & Kratina P. 2018. Diet tracing in ecology: method comparison and selection. Methods Ecol Evol 9: 278-291.

Oechsler-Christensen B., Jónasdóttir S.H., Henriksen P. & Hansen P.J. 2011. Use of phytoplankton pigments in estimating food selection of three marine copepods. J Plankton Res 34: 161-172.

Pace M.L., Cole J.J., Carpenter S.R. & Kitchell J.F. 1999. Trophic cascades revealed in diverse ecosystems. Trends Ecol Evol 14: 483-488.

Pace M.L., Cole J.J., Carpenter S.R., Kitchell J.F., Hodgson J.R., Van de Bogert, Matthew C, Bade D.L., Kritzberg E.S. & Bastviken D. 2004. Whole-lake carbon-13 additions reveal terrestrial support of aquatic food webs. Nature 427: 240.

Parnell A.C., Inger R., Bearhop S. & Jackson A.L. 2010. Source partitioning using stable isotopes: coping with too much variation. PloS one 5: e9672.

Parrish C.C. 2009. Essential fatty acids in aquatic food webs. In: Anonymous Lipids in aquatic ecosystems, Springer, pp. 309-326.

Peterson B.J. & Fry B. 1987. Stable isotopes in ecosystem studies. Annu Rev Ecol Syst 18: 293-320.

Phillips D.L. & Gregg J.W. 2001. Uncertainty in source partitioning using stable isotopes. Oecologia 127: 171-179.

Phillips D.L. & Gregg J.W. 2003. Source partitioning using stable isotopes: coping with too many sources. Oecologia 136: 261-269.

Piepho M., Arts M.T. & Wacker A. 2012. Species-specific variation in fatty acid concentrations of four phytoplankton species: does phosphorus supply influence the effect of light intensity or temperature? J Phycol 48: 64-73.

Pineda-Munoz S. & Alroy J. 2014. Dietary characterization of terrestrial mammals. Proc R Soc Lond B Biol Sci 281: 20141173.

Piñol J., San Andrés V., Clare E.L., Mir G. & Symondson W. 2014. A pragmatic approach to the analysis of diets of generalist predators: The use of next-generation sequencing with no blocking probes. Mol Ecol Resour 14: 18-26.

Plummer M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. 124.

R Core Team. 2018. R: A Language and Environment for Statistical Computing.

Rosen D.A. & Tollit D.J. 2012. Effects of phylogeny and prey type on fatty acid calibration coefficients in three pinniped species: implications for the QFASA dietary quantification technique. Mar Ecol Prog Ser 467: 263-276.

Sargent J.R., Bell J.G., Bell M.V., Henderson R.J. & Tocher D.R. 1993. The metabolism of phospholipids and polyunsaturated fatty acids in fish. Aquaculture 43: 103-124.

Sargent J.R., Bell J.G., Bell M.V., Henderson R.J. & Tocher D.R. 1995. Requirement criteria for essential fatty acids. J Appl Ichthyol 11: 183-198.

Schlechtriem C., Arts M.T. & Zellmer I.D. 2006. Effect of temperature on the fatty acid composition and temporal trajectories of fatty acids in fasting Daphnia pulex (Crustacea, Cladocera). Lipids 41: 397-400.

Semmens B.X., Moore J.W. & Ward E.J. 2009. Improving Bayesian isotope mixing models: a response to Jackson et al.(2009). Ecol Lett 12: E8.

Stock B.C., Jackson A.L., Ward E.J., Parnell A.C., Phillips D.L. & Semmens B.X. 2018. Analyzing mixing systems using a new generation of Bayesian tracer mixing models. PeerJ 6: e5096.

Stock B.C. & Semmens B.X. 2016. Unifying error structures in commonly used biotracer mixing models. Ecology 97: 2562-2569.

Strandberg U., Taipale S.J., Kainz M.J. & Brett M.T. 2014. Retroconversion of docosapentaenoic acid (n-6): an alternative pathway for biosynthesis of arachidonic acid in Daphnia magna. Lipids 49: 591-595.

Strandberg U., Taipale S.J., Hiltunen M., Galloway A.W., Brett M.T. & Kankaala P. 2015. Inferring phytoplankton community composition with a fatty acid mixing model. Ecosphere 6: 1-18.

Taipale S.J., Brett M.T., Hahn M.W., Martin-Creuzburg D., Yeung S., Hiltunen M., Strandberg U. & Kankaala P. 2014. Differing Daphnia magna assimilation efficiencies for terrestrial, bacterial, and algal carbon and fatty acids. Ecology 95: 563-576.

Taipale S.J., Brett M.T., Pulkkinen K. & Kainz M.J. 2012. The influence of bacteria-dominated diets on Daphnia magna somatic growth, reproduction, and lipid composition. FEMS Microbiol Ecol 82: 50-62.

Taipale S.J., Galloway A.W., Aalto S.L., Kahilainen K.K., Strandberg U. & Kankaala P. 2016a. Terrestrial carbohydrates support freshwater zooplankton during phytoplankton deficiency. Sci Rep 6: 30897.

Taipale S.J., Hiltunen M., Vuorio K. & Peltomaa E. 2016b. Suitability of phytosterols alongside fatty acids as chemotaxonomic biomarkers for phytoplankton. Front Plant Sci 7: 212.

Taipale S.J., Kahilainen K.K., Holtgrieve G.W. & Peltomaa E.T. 2018. Simulated eutrophication and browning alters zooplankton nutritional quality and determines juvenile fish growth and survival. Ecol Evol 8: 2671-2687.

Taipale S.J., Kainz M.J. & Brett M.T. 2011. Diet-switching experiments show rapid accumulation and preferential retention of highly unsaturated fatty acids in Daphnia. Oikos 120: 1674-1682.

Taipale S.J., Kankaala P., Tiirola M. & Jones R.I. 2008. Whole-lake dissolved inorganic 13C additions reveal seasonal shifts in zooplankton diet. Ecology 89: 463-474.

Taipale S.J., Peltomaa E., Hiltunen M., Jones R.I., Hahn M.W., Biasi C. & Brett M.T. 2015. Inferring Phytoplankton, Terrestrial Plant and Bacteria Bulk $\delta^{13}C$ Values from Compound Specific Analyses of Lipids and Fatty Acids. PloS one 10: e0133974.

Taipale S.J., Strandberg U., Peltomaa E., Galloway A.W., Ojala A. & Brett M.T. 2013. Fatty acid composition as biomarkers of freshwater microalgae: analysis of 37 strains of microalgae in 22 genera and in seven classes. Aquat Microb Ecol 71: 165-178.

Twining C.W., Brenna J.T., Hairston N.G. & Flecker A.S. 2016. Highly unsaturated fatty acids in nature: what we know and what we need to learn. Oikos 125: 749-760.

Vander Zanden H.B., Soto D.X., Bowen G.J. & Hobson K.A. 2016. Expanding the isotopic toolbox: applications of hydrogen and oxygen stable isotope ratios to food web studies. Front Ecol Evol 4: 20.

Viénot F. 2002. Michel-Eugène Chevreul: From laws and principles to the production of colour plates. Color Res Appl 27: 4-14.

Volkman J.K., Barrett S.M., Blackburn S.I., Mansour M.P., Sikes E.L. & Gelin F. 1998. Microalgal biomarkers: a review of recent research developments. Org Geochem 29: 1163-1179.

von Elert E. 2002. Determination of limiting polyunsaturated fatty acids in Daphnia galeata using a new method to enrich food algae with single fatty acids. Limnol Oceanogr 47: 1764-1773.

Wang S.W., Hollmén T.E. & Iverson S.J. 2010. Validating quantitative fatty acid signature analysis to estimate diets of spectacled and Steller's eiders (Somateria fischeri and Polysticta stelleri). J Comp Physiol 180: 125-139.

Ward E.J., Semmens B.X., Phillips D.L., Moore J.W. & Bouwes N. 2011. A quantitative approach to combine sources in stable isotope mixing models. Ecosphere 2: 1-11.

Wei J.H., Yin X. & Welander P.V. 2016. Sterol synthesis in diverse bacteria. Front Microbiol 7: 990.

Winder M. & Schindler D.E. 2004. Climate change uncouples trophic interactions in an aquatic ecosystem. Ecology 85: 2100-2106.

APPENDIX 1. Systematic simulation matrices

Table 1. The two first rows and the last row of the simulation matrices to illustrate the patterns of how the matrices were constructed (N = 11132).

	consumer	actino	crypto	cyano	diatom	dino	eugleno	golden	green	MOB	tPOM	tPOMb
9 MDIs	1	0.9	0.1	0	0	0	0	0	0	0	0	0
	2	0.9	0	0.1	0	0	0	0	0	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	110	0	0	0	0	0	0	0	0	0	0.1	0.9
8 MDIs	1	0.8	0.1	0.1	0	0	0	0	0	0	0	0
	2	0.8	0.1	0	0.1	0	0	0	0	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	495	0	0	0	0	0	0	0	0	0.1	0.1	0.8
7 MDIs	1	0.7	0.1	0.1	0.1	0	0	0	0	0	0	0
	2	0.7	0.1	0.1	0	0.1	0	0	0	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	1320	0	0	0	0	0	0	0	0.1	0.1	0.1	0.7
6 MDIs	1	0.6	0.1	0.1	0.1	0.1	0	0	0	0	0	0
	2	0.6	0.1	0.1	0.1	0	0.1	0	0	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	2310	0	0	0	0	0	0	0.1	0.1	0.1	0.1	0.6
5 MDIs	1	0.5	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
	2	0.5	0.1	0.1	0.1	0.1	0	0.1	0	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	2761	0	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.5
4 MDIs	1	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0
	2	0.4	0.1	0.1	0.1	0.1	0.1	0	0.1	0	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	2310	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.4
3 MDIs	1	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0
	2	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0	0.1	0	0
	...	...	...	...	...	...	...	...	...	...	...	...
	1320	0	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3
2 MDIs	1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0
	2	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0.1	0
	...	...	...	...	...	...	...	...	...	...	...	...
	495	0	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
1 MDI	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0
	2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0.1
	...	...	...	...	...	...	...	...	...	...	...	...
	11	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Table 2. The two first rows and the last row of the realistic pseudo-consumer simulation weighing matrix for the rdirichlet()-function and the generated relation matrix used in diet estimation simulation (N = 2761).

	consumer	actino	crypto	cyano	diatom	dino	eugleno	golden	green	MOB	tPOM	tPOMb
Weights for rdirichlet()	1	5	1	1	1	1	1	0.1	0.1	0.1	0.1	0.1
	2	5	1	1	1	1	0.1	1	0.1	0.1	0.1	0.1
	...	...	...	...	...	...	...	...	...	...	...	...
	2761	0.1	0.1	0.1	0.1	0.1	1	1	1	1	1	5
Relations for diet estimation	1	0.5250	0.0584	0.0749	0.0903	0.0903	0.1353	0.0003	0.0026	0.0074	0.0095	0.0061
	2	0.4452	0.0962	0.1421	0.1013	0.1008	0.0001	0.0762	0.0023	0.0267	0.0047	0.0045
	...	...	...	...	...	...	...	...	...	...	...	...
	2761	0.0141	0.0008	0.0081	0.0059	0.0071	0.1326	0.1061	0.0546	0.1012	0.1069	0.4627

Table 3. The two first rows and the last row of the concentrated library simulation weighing matrix for the rdirichlet()-function and the generated relation matrix used in diet estimation simulation (N = 495).

	consumer	actino	crypto	cyano	diatom	dino	eugleno	golden	green	MOB	tPOM	tPOMb
Weights for rdirichlet()	1	0	25	25	5	5	5	5	5	5	5	5
	2	0	25	5	25	5	5	5	5	5	5	5
	...	...	...	...	...	...	...	...	...	...	...	...
	495	5	5	5	5	5	5	5	5	25	25	0
Relations for diet estimation	1	0	0.27	0.27	0.06	0.06	0.06	0.05	0.07	0.05	0.05	0.06
	2	0	0.30	0.05	0.28	0.06	0.04	0.05	0.05	0.05	0.07	0.05
	...	...	...	...	...	...	...	...	...	...	...	...
	495	0.05	0.07	0.06	0.05	0.08	0.04	0.06	0.06	0.24	0.28	0

APPENDIX 2. Diet specific simulation test results

Table 1. Average probability for errors in finding the real diet proportions at 95th and 99.7th confidence interval percentiles for each model, and their errors of means (%).

N = 11132		Average confidence interval error (%)		
	resource	MixSIR	SIAR	QFASA
95th percentile	actino	19.0 ± 1.9	37.2 ± 4.6	26.4 ± 1.4
	crypto	26.6 ± 2.3	53.3 ± 9.4	46.6 ± 8.0
	cyano	29.2 ± 2.7	53.5 ± 8.6	49.1 ± 4.1
	diatom	3.0 ± 0.6	47.4 ± 7.6	57.4 ± 1.5
	dino	35.6 ± 3.7	46.3 ± 8.8	40.9 ± 3.9
	eugleno	0.3 ± 0.1	50.9 ± 8.7	38.0 ± 1.5
	golden	41.0 ± 5.1	46.0 ± 8.8	81.4 ± 2.5
	green	15.5 ± 1.5	46.2 ± 7.6	41.1 ± 5.3
	MOB	0.4 ± 0.2	11.2 ± 1.3	50.4 ± 3.7
	tPOM	12.0 ± 1.8	54.1 ± 9.3	30.1 ± 6.9
	tPOMb	37.8 ± 4.3	44.7 ± 7.3	57.3 ± 1.9
99.7th percentile	actino	0.0 ± 0.0	1.5 ± 1.3	20.1 ± 2.3
	crypto	3.1 ± 1.4	4.8 ± 1.5	42.0 ± 8.2
	cyano	2.8 ± 0.7	6.5 ± 1.5	37.5 ± 3.6
	diatom	0.0 ± 0.0	1.3 ± 1.3	47.5 ± 1.1
	dino	0.3 ± 0.2	0.4 ± 0.9	32.9 ± 2.5
	eugleno	0.0 ± 0.0	3.2 ± 1.3	30.1 ± 1.5
	golden	3.0 ± 0.7	0.5 ± 0.9	70.8 ± 2.6
	green	1.5 ± 0.3	1.9 ± 1.3	31.9 ± 4.5
	MOB	0.0 ± 0.0	0.2 ± 0.8	41.3 ± 4.0
	tPOM	6.9 ± 1.6	6.6 ± 1.5	25.3 ± 5.0
	tPOMb	3.1 ± 0.4	0.9 ± 1.2	45.2 ± 1.3

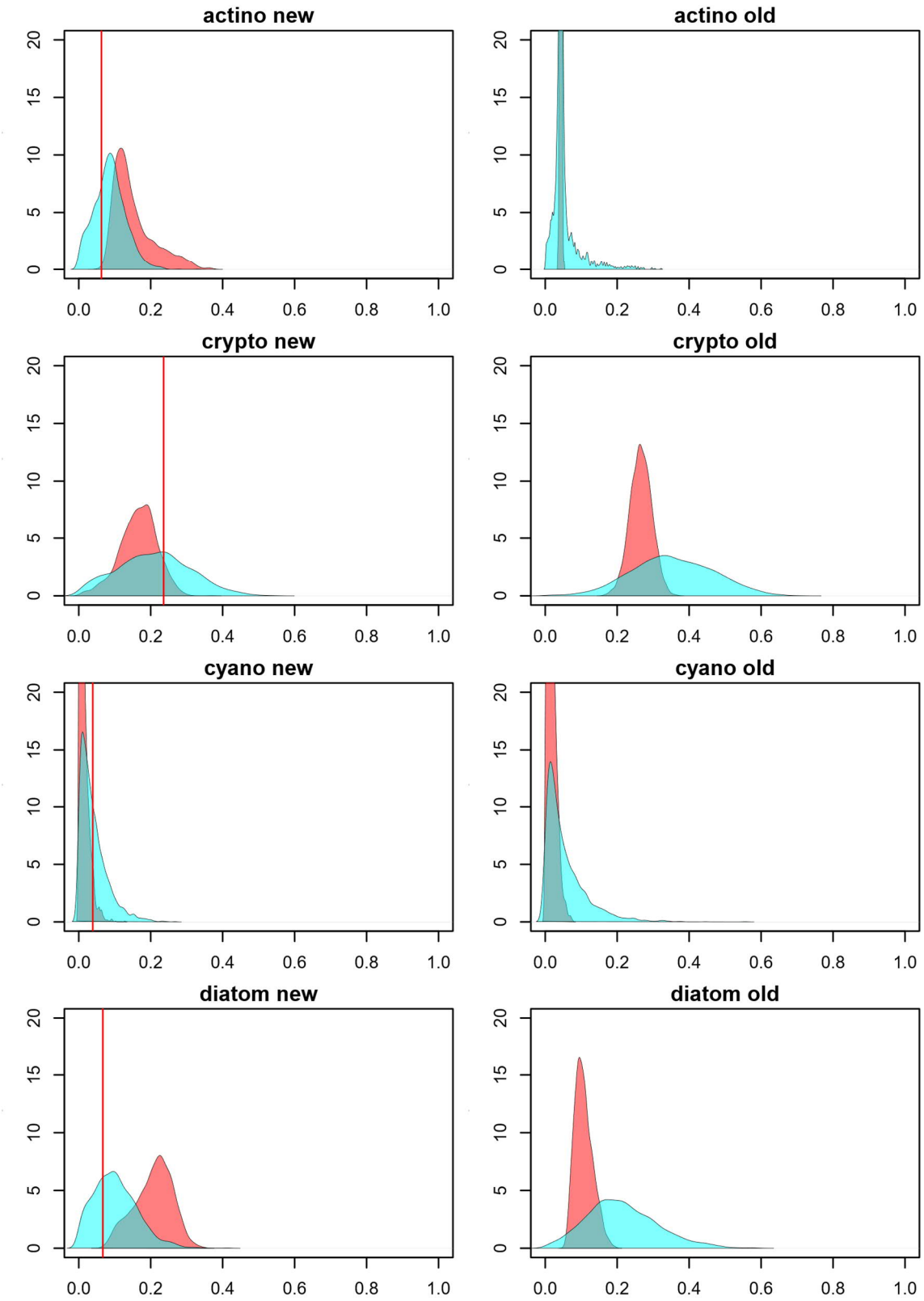
Table 2. Average probability for errors in finding the MDIs at 95th and 99.7th confidence interval percentiles for each model, and their errors of means (%).

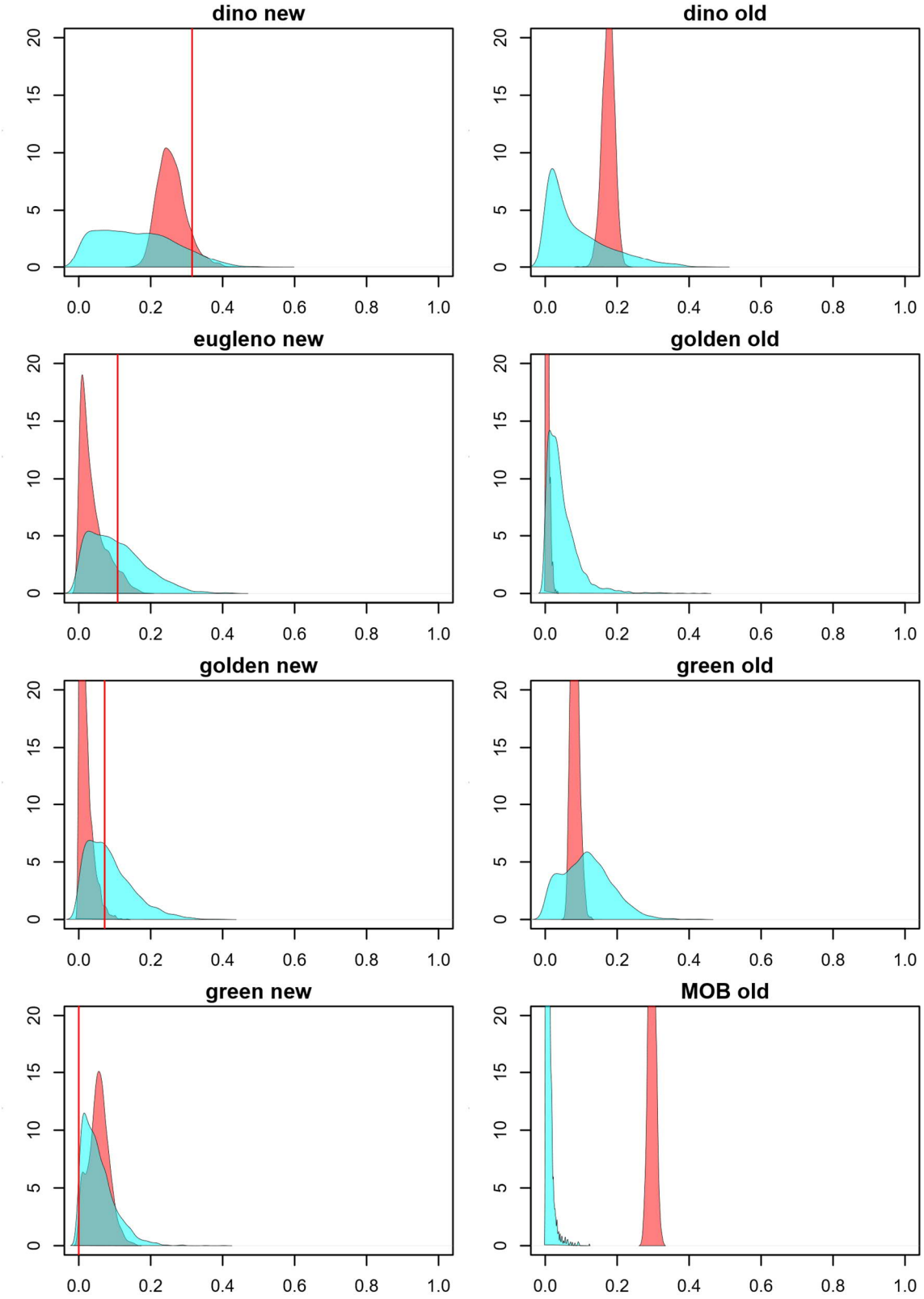
N = 11132		Average confidence interval error (%)		
	resource	MixSIR	SIAR	QFASA
95th percentile	actino	43.8 ± 8.8	80.0 ± 7.1	6.3 ± 1.3
	crypto	47.3 ± 10.0	99.5 ± 0.4	14.8 ± 5.0
	cyano	45.8 ± 10.5	97.6 ± 1.5	44.7 ± 7.7
	diatom	25.1 ± 8.6	95.2 ± 2.9	30.3 ± 5.2
	dino	60.5 ± 10.8	99.0 ± 0.8	28.3 ± 5.4
	eugleno	23.2 ± 9.9	99.1 ± 1.6	41.8 ± 9.9
	golden	77.9 ± 7.7	98.7 ± 0.6	74.6 ± 7.2
	green	29.6 ± 11.8	76.9 ± 7.7	33.7 ± 4.5
	MOB	0.5 ± 0.2	41.7 ± 11.5	30.0 ± 8.7
	tPOM	23.3 ± 9.9	99.6 ± 0.2	14.1 ± 5.5
	tPOMb	63.5 ± 12.2	72.8 ± 10.3	39.9 ± 8.2
99.7th percentile	actino	0.0 ± 0.0	0.0 ± 0.0	5.0 ± 1.1
	crypto	0.9 ± 0.4	0.3 ± 0.1	11.4 ± 3.9
	cyano	2.7 ± 0.8	0.4 ± 0.1	34.9 ± 5.4
	diatom	0.1 ± 0.1	0.1 ± 0.1	24.7 ± 4.3
	dino	0.3 ± 0.3	0.6 ± 0.2	22.6 ± 4.3
	eugleno	1.7 ± 0.7	2.4 ± 0.9	35.2 ± 10.1
	golden	5.4 ± 0.8	1.1 ± 0.2	67.1 ± 7.7
	green	3.2 ± 1.3	1.6 ± 0.3	28.1 ± 3.8
	MOB	0.0 ± 0.0	0.1 ± 0.0	27.6 ± 7.9
	tPOM	1.7 ± 0.7	0.7 ± 0.2	11.3 ± 4.4
	tPOMb	5.5 ± 1.1	1.3 ± 0.3	31.1 ± 8.2

Table 3. Average probability for errors in finding the real diet proportions of the realistic pseudo-consumers at 95th and 99.7th confidence interval percentiles for each model, and their errors of means (%).

N = 2761		Average confidence interval error (%)		
	resource	MixSIR	SIAR	QFASA
95th percentile	actino	2.9 ± 0.3	8.9 ± 0.5	50.7 ± 1.0
	crypto	11.0 ± 0.6	15.8 ± 0.7	91.2 ± 0.5
	cyano	15.5 ± 0.7	17.1 ± 0.7	68.3 ± 0.9
	diatom	0.4 ± 0.1	8.6 ± 0.5	77.0 ± 0.8
	dino	5.1 ± 0.4	6.5 ± 0.5	68.0 ± 0.9
	eugleno	0.3 ± 0.1	15.7 ± 0.7	55.0 ± 0.9
	golden	11.6 ± 0.6	9.1 ± 0.5	79.4 ± 0.8
	green	9.6 ± 0.6	11.0 ± 0.6	54.3 ± 0.9
	MOB	0.0 ± 0.0	1.3 ± 0.2	83.6 ± 0.7
	tPOM	11.6 ± 0.6	16.6 ± 0.7	57.8 ± 0.9
	tPOMb	10.0 ± 0.6	8.1 ± 0.5	71.1 ± 0.9
99.7th percentile	actino	0.0 ± 0.0	0.0 ± 0.0	42.9 ± 0.9
	crypto	2.4 ± 0.3	3.0 ± 0.3	88.7 ± 0.6
	cyano	1.8 ± 0.3	7.0 ± 0.5	54.0 ± 0.9
	diatom	0.0 ± 0.0	0.1 ± 0.1	71.5 ± 0.9
	dino	0.0 ± 0.0	0.0 ± 0.0	57.3 ± 0.9
	eugleno	0.0 ± 0.0	0.3 ± 0.1	46.8 ± 0.9
	golden	0.4 ± 0.1	0.0 ± 0.0	68.6 ± 0.9
	green	0.2 ± 0.1	0.0 ± 0.0	44.3 ± 0.9
	MOB	0.0 ± 0.0	0.0 ± 0.0	74.6 ± 0.8
	tPOM	8.3 ± 0.5	7.5 ± 0.5	52.6 ± 1.0
	tPOMb	0.3 ± 0.1	0.1 ± 0.1	56.9 ± 0.9

APPENDIX 3. Diet specific estimation density plots of natural data




dino new
dino old
eugleno new
golden old
golden new
green old
green new
MOB old
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0

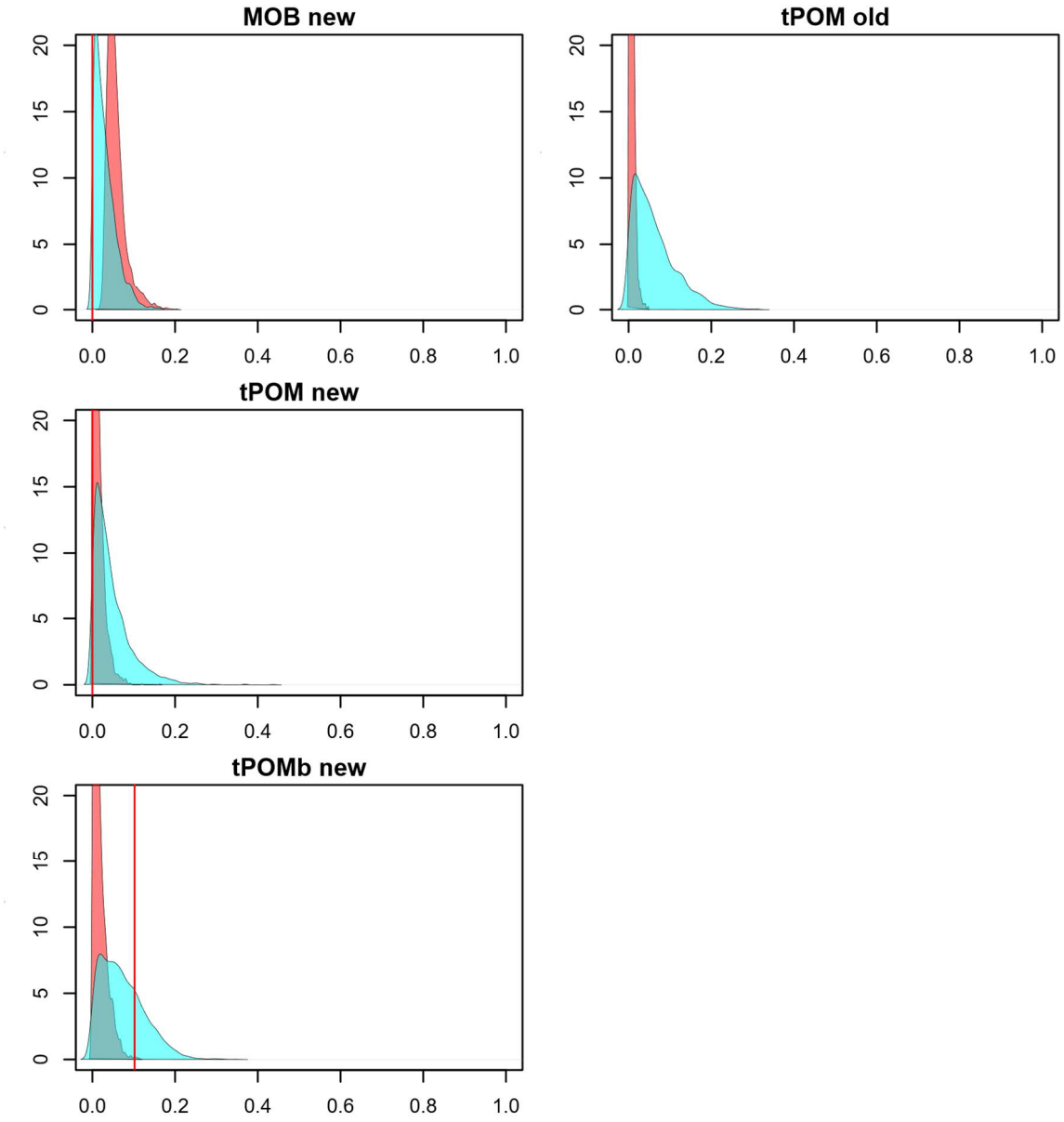


Figure 1. The diet estimation density plots of each diet item group with the newly constructed library and the old library (Galloway et al. 2014a) used to model data of Daphnia from an eutrophicated lake. The red density plot represents the diet estimate of MixSIR, the cyan plot represents the diet estimate of SIAR, and the red vertical line represents the diet estimate of QFASA.

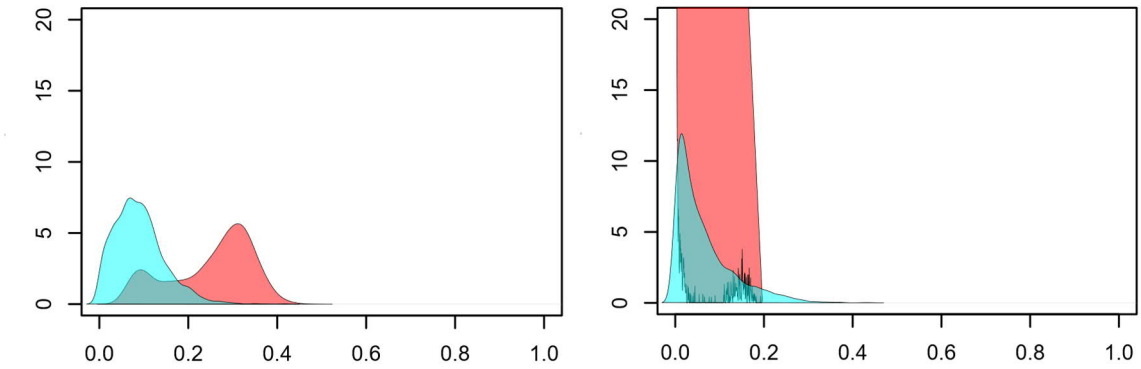


Figure 2. Examples of a bimodal density distribution, and a distorted density distribution.