Bayesian estimation of predator diet composition from fatty acids and stable isotopes

Philipp Neubauer*

Dragonfly Science,
PO Box 27535, Wellington 6141, New Zealand

Olaf P. Jensen

Institute of Marine and Coastal Sciences
Rutgers University, New Brunswick, NJ 08901, USA

Dated: June 12, 2014

*Corresponding author electronic address: neubauer.phil@gmail.com

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

Abstract

Quantitative analysis of stable isotopes (SI) and, more recently, fatty acid profiles (FAP) are useful and complementary tools for estimating the relative contribution of different prey items in the diet of a predator. The combination of these two approaches, however, has thus far been limited and qualitative. We propose a mixing model for FAP that follows the Bayesian machinery employed in state-of-the-art mixing models for SI. This framework provides both point estimates and probability distributions for individual and population level diet proportions. Where fat content and conversion coefficients are available, they can be used to improve diet estimates. This model can be explicitly integrated with analogous models for SI to increase resolution and clarify predator-prey relationships. We apply our model to simulated data and an experimental dataset that allows us to illustrate modeling strategies and demonstrate model performance. Our methods are provided as an open source software package for the statistical computing environment R.

Keywords Stable isotope analysis, quantitative fatty acid analysis, QFASA, lipid profile, diet analysis, Bayesian mixing model, fatty acid signature, dietary marker

1 Introduction

Quantitative estimates of an animals diet are a critical component of predator-prey studies, ecosystem models, and ecosystem-based management. Existing methods of estimating diet proportions all have strengths and

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

weaknesses (Bowen & Iverson, 2012). Traditional stomach content and fecal matter analysis represent a brief snapshot of diet at a particularly place and time and can be invasive, time-consuming, and potentially biased by differential rates of digestion of prey or ingestion of identifiable prey parts (Bowen & Iverson, 2012). Chemical markers such as stable isotopes (SI) and fatty acid profiles (FAP) solve some of these problems. For example, both approaches integrate diet composition over an extended time period - typically weeks to months, depending on tissue turnover rates (Tucker, Bowen & Iverson, 2008). These advantages have led to rapid growth in the use of chemical markers in diet studies (Bowen & Iverson, 2012; Elsdon, 2010; Kelly & Scheibling, 2011; Williams & Buck, 2010). However, chemical dietary markers generally lack the specificity of traditional stomach content analysis. In particular, several prey species often have similar isotopic signatures. More recent studies have sought greater dietary resolution through the use of SI of other elements in addition to carbon and nitrogen(Belicka *et al.*, 2012), compound specific SI ratios (Budge *et al.*, 2008; Jack & Wing, 2011), or a combination of stomach content analysis and SI or FAP (Pethybridge *et al.*, 2012). The use of SI and FAP in combination also holds great promise; however the few studies to date that have used both chemical markers have been qualitative (Guest *et al.*, 2009) or based on positive correlation of results from both methods (Tucker, Bowen & Iverson, 2008).

Analysis tools for SI data have become very sophisticated in recent years, starting with the development of general Bayesian analysis tools for estimating diet proportions, and leading to customized (hierarchical) models for

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

individual applications (Hopkins & Ferguson, 2012; Moore & Semmens, 2008; Parnell *et al.*, 2013). The latter models can, for instance, estimate dietary differences of geographically distinct populations (Semmens *et al.*, 2009), accommodate temporal changes in diets or estimate the effect of covariates (e.g., age, size, sex) on diet proportions (Parnell *et al.*, 2013).While these models provide a considerable step towards ecologically relevant models in diet studies, the underlying SI data is limited in the resolution that it can provide. Since typically only 2-3 SI are measured, the contrast that is achievable from such a low number of variables is necessarily limited, especially when the number of potential prey items increases (Phillips & Gregg, 2003; Ward *et al.*, 2011). Optimally aggregating prey items into prey groups may circumvent this problem (Ward *et al.*, 2011), but may also be less satisfactory in complex food webs.

FAP data can, in theory, provide considerably more resolution compared to SI data, due to large number of potential fatty acids that can be measured. Furthermore, Blanchard (2011) developed a Bayesian model for diet inference from fatty acids (furthering the development of Bayesian mixing models for compositional data by Billheimer (2001)), showing that model based inferences of predator diets from fatty acids are achievable. Nevertheless, studies employing FAP remain either qualitative in their estimates of prey proportions in predator diets, or use Quantitative Fatty Acid Signature Analysis (Iverson *et al.*, 2004) to obtain quantitative estimates of diet proportions.

QFASA is the only available (i.e., off the shelf) method thus far for use with FAP data, and, in contrast to recent (Bayesian) SI and FAP mixing models,

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

relies on a distance metric rather than a model based formulation to estimate the most likely diet proportions. This framework provided the first quantitative approach to estimating diet proportions using fatty acids and it has already seen widespread use, particularly in studies of marine mammals (Bowen & Iverson, 2012) and seabirds (Williams & Buck, 2010). Nevertheless, QFASA has a number of limitations. Since it is not based on a probabilistic model, it is difficult to estimate uncertainty associated with estimated diet proportions (but see Steward 2005 as cited in Blanchard, 2011). The absence of an explicit model also makes it impossible to build ecological mechanisms (e.g., covariates of consumed diets) directly into the model. Furthermore, uncertainty about conversion coefficients representing enrichment and preferential uptake of fatty acids cannot be considered, yet small changes in these coefficients can lead to differences in inferred diet proportions (Wang, Hollmen & Iverson, 2010).

Given the discrepancy in methods applied to SI and FAP data, it is perhaps not surprising that their joint application has commonly relied on qualitative comparisons. Because both markers integrate diet composition over often comparable time-scales, however, an explicit integration of these data types could provide substantial benefits. While FAP data could mitigate the resolution problem in SI data, SI data could provide increased resolution and clarify predator-prey relationships, the knowledge of which is usually a pre-requisite for FAP data. For example, for many non-modified fatty acids, FAP alone cannot discriminate between the case of two species which share a common diet and the situation in which one of these species eats the other. In

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

either case, the two species may have similar FAP. The addition of a stable isotope with trophic fractionation (e.g., ^{15}N), however, can readily distinguish predation from dietary overlap.

Here, we develop a mixing model for FAP data based on a probabilistic model whose parameters are estimated using Bayesian methods. Using both simulated and published data, we demonstrate the suitability of this model for FAP analysis and highlight the potential benefit of explicit integration with SI data to increase the precision of diet estimates.

2 Methods

2.1 A Bayesian mixing model for FAP

Bayesian models for SI data are commonly based on the assumption that SI ratios are normally distributed. This assumption cannot be made for FAP data, since for most methods of analysis, the concentration of individual fatty acids is normalized to the total lipid content of the sample. Thus, the FAP are a collection of proportions (referred to as a composition), which lie between 0 and 1, and are constrained to sum to 1. A common solution to this problem, however, is to consider transformations that make the data approximately normal (Budge, Iverson & Koopman, 2006). To construct our model, we considered the additive log ratio transformation Aitchison & Bacon-Shone (1999), also called alr transformation, such that

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

$$y_{i,s} = alr(\phi_{i,s}) = log\left(\frac{\phi_{i,s,1\ldots p-1}}{\phi_{i,s,p}}\right) \quad (1)$$

where $\phi_{i,s}$ is the p-variate fatty acid composition of individual i of prey species s, with a total of n potential prey species considered. Note that in the following we often drop the subscript for fatty acids, e.g., $\phi_{i,s}$ and $y_{i,s}$ are thus p and $p-1$ dimensional vectors, respectively. We assumed that the distribution of y is multivariate normal, with species specific mean μ_s and covariance matrix Σ_S, or $y_{i,s} \sim N(\mu_s, \Sigma_s)$. A vaguely informative prior on μ_s and Σ_s allows for uncertainty in prey distributions (Ward, Semmens & Schindler, 2010) to propagate to estimates of diet proportions.

Each predator j consumes a proportion π_j of each prey source, and analogous to stable isotope mixing models, predator FAP are then a linear combination of prey FAPs, normalised to sum to one. Since predators may selectively assimilate or metabolize fatty acids (Budge, Iverson & Koopman, 2006; Iverson *et al.*, 2004; Rosen & Tollit, 2012), we specify prey-specific conversion coefficients $\kappa_s = \kappa_{s,1}\ldots\kappa_{s,p}$ for each of the p fatty acids (Rosen & Tollit, 2012). Furthermore, the n prey species likely have different fat content Φ that will affect the total amount of fatty acids assimilated from each prey species by the predator. The expected FAP of predator τ_j is then a linear combination of the prey FAP, modified by conversion coefficients for each fatty acid p and fat content for each prey i:

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

$$t_j \sim N(alr(\tau_j), \Sigma_\tau) \quad (2)$$

$$\tau_j = C\left\{\sum_s^n (\pi_{j,s}\Phi_s)(\kappa_s \otimes \phi_{j,s})\right\} \quad (3)$$

Here, C is the closure operation which normalizes the FAP to sum to one and $\otimes$ is the outer (element wise) product. $\phi_{s,j}$ is the FAP of prey items of species s consumed by predator j. Similarly to Parnell *et al.* (2013), we thus assumed that individual predators do not necessarily feed on 'average' prey items, but rather consume prey items with signatures drawn from the estimated prey distribution. We again formulate predator signatures t as draws from a normal distribution after transformation. We further assumed that Φ and κ are log-normally and gamma distributed, respectively, around known mean and variance values (estimated or calculated from controlled diet experiments, see below). The closure operation in Equation 2 (i.e., the sum-to-one constraint on the FAP) leads to κ being determined in terms of relative uptake of fatty acids (i.e., up to a multiplicative constant), and implicitly makes the multivariate distribution over all κ a Dirichlet distribution. The same logic applies to Φ, and in both cases we opted for formulations that can be readily parametrised from priors studies or published values (e.g., sample means and variances from experiments).

The diet proportions π of predators are the main focus of investigation in diet studies. These may be modeled at the (statistical) population level (thus dropping the subscript j in Equation 2) or at the individual level, as suggested

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

in Equation 2. In the latter case individual predator FAP can be modeled as draws from a population level distribution of predator diet proportions. Recent approaches to stable isotope mixing have focused on transformations of the diet proportion vector π to get around the problems associated with the compositional nature of the diet proportions in such a hierarchical setup, and we follow this approach in our model. The diet proportions are transformed using clr transformations (Semmens *et al.*, 2009), such that the support of is the real line rather than the interval [0;1], and we then assume that $clr(\pi_j) \sim N(\Pi, \Sigma_\Pi)$, where Π is the vector of mean (population level) diet proportions. It is then possible to model diet proportions as function of covariates, such as size, sex, or region (i.e., in a regression formulation). While this approach is appealing, it adds to computation time employed to estimate model parameters, and generally slower convergence. We therefore use a vague Dirichlet prior on the proportions when convenient (e.g., when we estimate only population level parameters).

An R (R Core Team, 2014) package (called fastinR) implementing methods outlined here, along with simulated examples and the analysis of experimental data described further below, is available on the open source repository github.com/philipp-neubauer/fastinR. Models implemented in the package include the above-mentioned formulations for individual diet estimates, population level estimates or both as well as linear model (regression and ANOVA) formulations for diet proportions, all available for SI and FAP individually or as combined models (see below). Model parameters were estimated using Markov Chain Monte Carlo methods implemented in JAGS

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

(Plummer, 2003), called from R through higher level functions in the fastinR package that allow for data input, inspection and manipulation.

Depending on the amount of samples for prey and predators, it may be necessary to use informative priors for Σ_s and Σ_τ. Both were given inverse-Wishart priors, and since both are co-variances of transformed data, it is not straightforward to formulate default priors for these parameters. We have found that in practice manual adjustment of these priors is often needed to be able to achieve convergence and mixing (efficient exploration of the posterior distribution by the sampling algorithm) of the MCMC algorithms employed by JAGS. This is especially true when there are few source and/or predator samples. The package allows for high level adjustment of these parameters through the specification of the order of magnitude of the diagonal of each covariance matrix.

2.2 Joint diet estimation from FAP and SI

There are at least three potential benefits of integrating FAP and SI data: i) increased information to discriminate among sources, ii) the potential of SI to resolve predator prey relationships due to trophic enrichment of SI, and iii) the potential reduction in estimation error due to a larger body of research on fractionation coefficients for stable isotopes as opposed to conversion coefficients in FAP. Integrating the two complimentary types of data in a single model to estimate diet proportions may thus considerably improve estimates of diet proportions over estimation from wither data-source alone.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

Our model for FAP is conceptually similar to recent models proposed for SI data, and integration of FAP and SI data into a single model is straightforward in the present setting. We again assume that the vector of SI signatures of prey items q follow a multivariate normal distribution, such that $y_{q,s}^{SI} \sim N(\mu_s^{SI}, \Sigma_s^{SI})$, where the superscript SI denotes that these are stable isotope signatures. Predator SI signatures are again a linear combination of prey SI, this time modified by additive fractionation coefficients γ. Fractionation may, in turn, depend on prey isotope concentrations (Caut, Angulo & Courchamp, 2009; Hussey *et al.*, 2014). In our model, we assume additive fractionation, and suggest that concentration dependence is taken into account when specifying distributions for prey and SI specific fractionation coefficients γ_S (see examples below). The expected SI signature for predator r is then

$$t_r^{SI} = \sum_s^n \pi_{r,s} \left(y_{q,r} + \gamma_s \right) \tag{4}$$

$$clr(\pi_r) \sim N(\Pi, \Sigma_\Pi) \tag{5}$$

$$\gamma_{s,SI} \sim N(\nu_{SI}, \sigma_{SI}) \tag{6}$$

Note that the different subscripts to the FAP model imply that there is no need to have SI and FAP from the same prey or predator samples, as long as we can assume that the prey samples are drawn from the same statistical population as those for FAP, and that individual diet proportions of predators

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

are drawn from the same population distribution of diet proportions.

The exact formulation of the integration of SI and FAP depends on the assumptions that one is comfortable with in a given setting: identical dietary proportions may be appropriate if diets (and hence SI and FAP) are thought to be stable, or if both chemical tracers are thought to integrate over similar time-scales. If the time scales of these two elements are thought to be different (e.g., for different tissue types), individual diet proportions may be more appropriate, and may be drawn from an overall population distribution of diet proportions. Any of these options can be implemented in the fastinR package.

2.3 Simulation studies

We initially explored the feasibility and performance of our model setup in a range of simulations, which are illustrated (including code) in supplemental information S1. Simulations were also used to explore sensitivities of inferred diet proportions to the source configuration and diet evenness in a series of simulation experiments. We hypothesized that estimated diet proportions are sensitive to diet source separation in FAP space, co-linearity in FAP space (Blanchard, 2011) and diet makeup (e.g., specialist versus generalist diets). Further details and simulation results can be found in supplemental information S2.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

2.4 Selecting fatty acids for analysis: an ordination approach

A potentially large number of FAs are available from analysis methods such as gas-chromatography. A common practice is to simply set a threshold and keep the most abundant FA for analysis. This practice may, however, discard potential useful information, and a more judicious approach is to retain FAs based on the among diet source variability that they explain. Wang, Hollmen & Iverson (2010) used a method by which they tested the QFASA method on a series of subsets to determine the subset that gave the best accuracy. Although feasible, such a method is prohibitive with fully Bayesian models, which can take a long time to run with a realistic dataset.

Here, we propose a variable selection method based on constrained ordination, which considers the contribution of individual fatty acids to axes separating diet sources. Based on this contribution relative to the overall separation, the user can choose fatty acids that contribute most to source separation. This procedure is intended to reduce computation time (and dimensionality) of the models, retaining accuracy in diet estimates. Further details about the procedure are given in supplemental information S3.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

2.5 Application: estimating predator diets in a controlled experiment

To illustrate the potential of the models presented above, we analysed data from an experimental study by Stowasser *et al.* (2006), which investigated changes in squid FAP and SI as a function of diet treatments. The treatments consisted of exclusive fish and crustacean diets, as well as switched and mixed diets, with the former switching diets from fish (henceforth SF, n=4) to crustacean (SC, n=5) after 15 days of the 30 day experiment.

In order to apply our model, we first estimated conversion coefficients of FAP and fractionation in SI, using squid from the 30 day diet treatments feeding exclusively crustacean and fish diets. The model for estimation of SI fractionation followed the model in Hussey *et al.* (2014), thus accounting for diet $\delta^{15}N$ and $\delta^{13}C$, and used their results as priors for fractionation parameters for $\delta^{15}N$, and results from Caut, Angulo & Courchamp (2009) to construct priors for $\delta^{13}C$. Estimation of FA conversion coefficients used (2) with proportions assumed known from feeding trials. Details on the estimation of conversion coefficients and fractionation are given in supplemental information S4.

In our diet analysis, we analyzed samples from the switched diet treatments, and used both SI and FAP to investigate if our models allow us to infer diet proportions in either treatments. We subset the data to use only switched diet squid that were analysed for FAP and SI after at least 10 days under the respective treatment. We only had overlapping SI and FAP for the SC

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

treatment squid, and we therefore started by analyzing this treatment in isolation to demonstrate that both SI and FAP can resolve diet proportions, and to demonstrate the benefit of using the two tracers in a joint model. We then analyzed the SF squid, for which we only had 3 specimen with FAP and 1 specimen with SI. The markers available for this treatment did not overlap for any of the sampled squid.

We lastly estimated individual diet proportions in the SC treatment. To demonstrate how the model based approach to diet estimation can be use to answer ecologically relevant questions about predator diets, we also analyzed SF and SC treatment squid together in a linear model setup that investigated treatment differences explicitly. The linear model used treatment dummy variables to estimate individual intercepts for each treatment and prey combination, and allows us to estimate, conditional on the data and priors, whether squid in either one treatment group consumed significantly more of any one prey type.

FAP analyses used data obtained by analyzing digestive glad tissue, which is thought to rapidly assimilate dietary fatty acids in relatively unmodified proportions relative to the original diet. SI were analyzed from muscle tissue since we had more individuals sampled for SI from this tissue, which may be more prone to fractionation and slower turnover than digestive glad tissue. In the original study, a total of 25 FAs were reported. Here, we selected FAs using ordination methods described above. For estimation of model parameters, priors for prey and predator specific variances were adjusted manually to give reasonable behaviour in the MCMC algorithm. The analyses

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

are detailed in supplemental information S5.

3 Results

3.1 Simulation studies

Simulated test cases suggested that our model can estimate diet proportions from both SI and FAP (supplemental information S1), with accuracy depending mainly on source separation and diet evenness (supplemental information S2). For very uneven diet proportions, such as in the feeding trials analyzed in the squid example, we found the choice of posterior means as point estimate for diet proportions inevitably introduced error at the margins of the 0-1 interval when compared to true simulated diet proportions.

Models with low accuracy conversion coefficients (with prior mean for all FA set to 1 and large prior variance) also performed substantially worse than models with accurately specified coefficients when comparing point estimates of diet proportions to simulated diet proportions(supplemental information S2), showing decreasing accuracy with increasing variance among simulated convergence coefficients.

3.2 Squid diet experiments

Dimension reduction by NMDS on FAP of squid and their potential prey suggested that crustacean diets were readily distinguishable from fish diets (Figure 1a). For fish diet items, however, no single fish species could be clearly

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

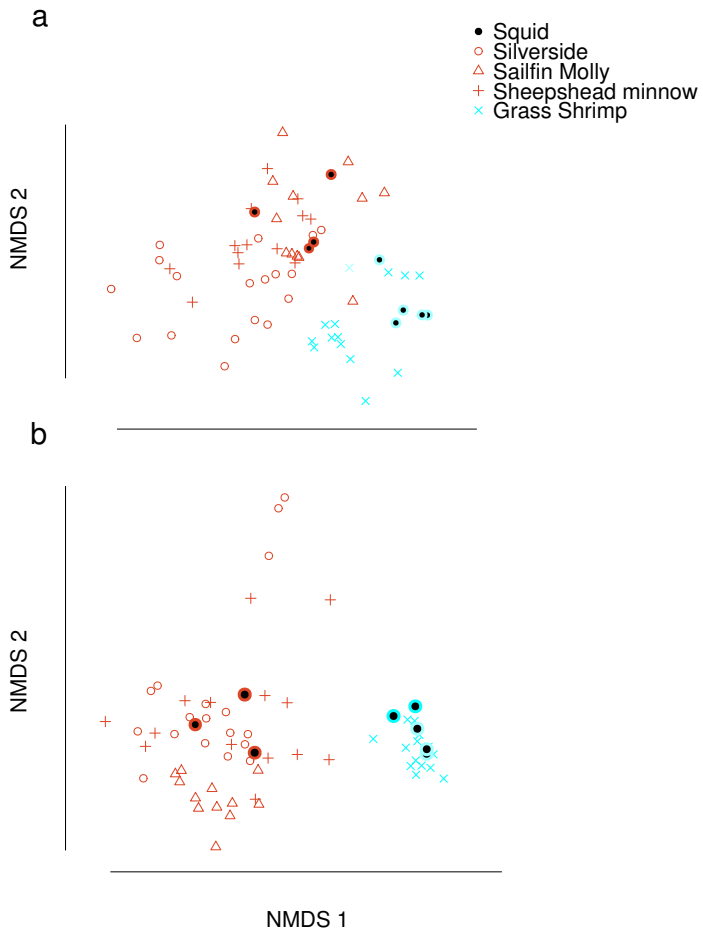

Figure 1: Non-metric Multi-Dimensional Scaling (NMDS) plots of FAP for squid and their potential prey a) before and b) after variable selection.

distinguished from any other fish species. Predator signatures of switched diet squid aligned with their respective diets after correcting by posterior means of estimated conversion coefficients. The latter were different from expected $(1/p)$ for many FA in the analysis (supplemental information S4).

Selection of FAs using constrained ordination lead to four FAs, 22.6n.3, 20.5n.3, 20.4n.6 and 18.1n.9 being retained for analysis (Figure 2), accounting

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

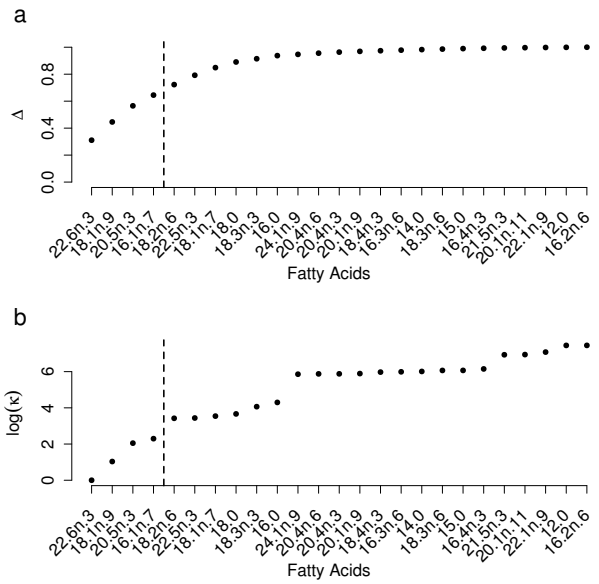

Figure 2: a) Cumulative proportion of between prey variance along CAP axes explained by individual fatty acids being added to the datasets, ordered by the contribution of each fatty acid to the total variance. b) Prey matrix condition number as a function of individual fatty acids being added as in a).

for a total of 74% of total among source variation on ordination axes while maintaining a low prey matrix condition number ($\kappa = 15.67$), suggesting limited co-linearity. The matrix condition number nearly doubled for the next most important fatty acid ($\kappa = 29.17$) and increased exponentially thereafter with addition of other fatty acids. The resulting NMDS plot suggested that the reduction from 22 to four FA did not significantly alter the configuration of predators and prey items in FAP space, despite the drastically lowered number of input dimensions (Figure 1b). Retaining a larger subset of FAs (8 FAs) did not qualitatively alter the results, but did lead to lower uncertainty in diet proportion estimates, suggesting that we lost some relevant information by retaining four of

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

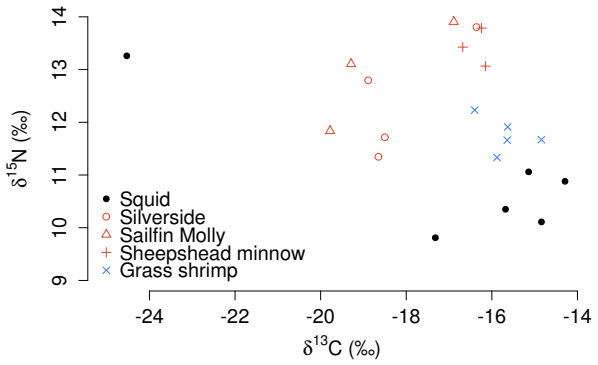

Figure 3: Stable isotope signatures of squid and their potential prey.

SI also showed clear separation between crustacean and fish prey (Figure 3), but showed two groups for fish prey items, both consisting of specimen from more than one fish species. Squid $\delta^{15}N$ was also substantially lower than any of the prey species analysed even after correcting for estimated fractionation coefficients.

FAP were able to resolve population level SC treatment squid diets, suggesting a diet predominantly based on crustaceans (Figure 4). While uncertainty about the exact diet proportions remained for both crustaceans and fish, most of the posterior density for crustacean diet proportions was clearly concentrated towards high proportions of squid. For fish, posteriors were peaked near zero, however, all fish species posteriors had long tails that spanned nearly the whole interval of possible diet contributions. An analyses based on SI alone gave very similar results, despite different tissue types examined (Figure 4).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

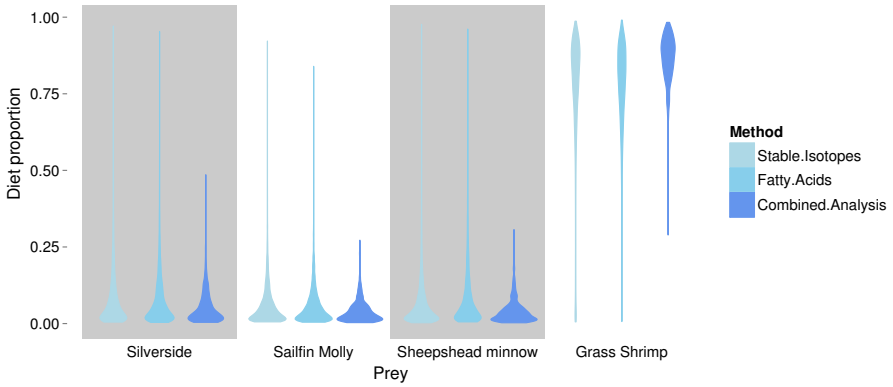

Figure 4: Posterior densities for diet proportion estimates of SC (crustacean only diet) treatment squid based on FAP, SI and a combined (FAP & SI) analysis.

Combining the two markers lead to a substantial reduction in the uncertainty of estimated diet proportions (Figure 4), and suggested a clear dominance of crustaceans in the diet. For the combined analysis the spread of the posterior distribution for crustaceans in the squid diet was reduced by approximately 30%, and most of the probability density was shifted closer to one, and the reductions in the spread of posterior distributions for fish diet items were as high as 70%. Lastly, estimates of individual diet proportions closely mirrored population level estimates (Figure 5).

Due to overlap of fish species in FAP and SI space, similar models for SF treatment fish were unclear about the contribution of individual fish species (Figure 6), but suggested that crustaceans were a small part in the diet of these squid. SI and FAP combined (i.e., adding one squid with SI but no FAP data) did not provide much improvement for individual fish species, however, combining fish species post-hoc as the sum of individual posterior distributions clearly shows a fish based diet (Figure 7).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

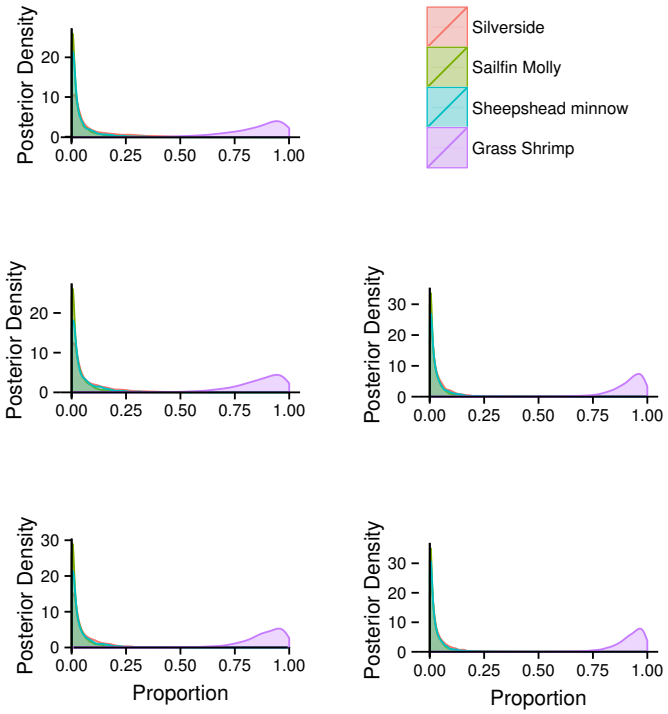

Figure 5: Posterior densities for individual diet proportion estimates of SC squid based on a hierarchical model for diet proportions using both FAP and SI.

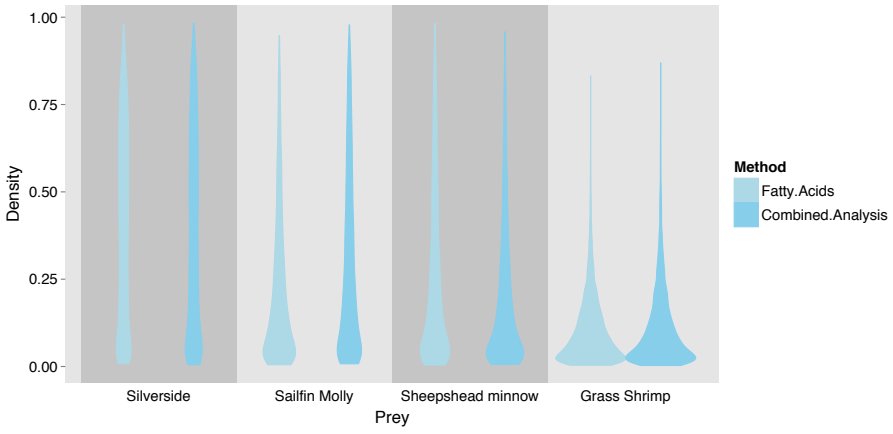

Figure 6: Posterior densities for diet proportion estimates of SF (fish only diet) treatment squid based on FAP and a combined (FAP & SI) analysis. Note that no separate analysis using SI only was run.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

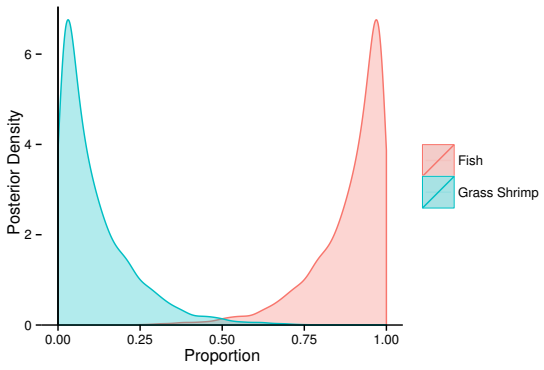

Figure 7: Posterior densities for diet proportion estimates of SF (fish only diet) treatment squid using both SI and FAP, combining all fish species into a fish prey group.

4 Discussion

We presented here a general way to analyse FAP in a Bayesian mixing model, and demonstrated that the method can estimate diet proportions in feeding trials while accounting for fatty acid conversion and diet fat content. The Bayesian framework allows explicit representation of uncertainty about mixing proportions as a function of uncertainty about prey distributions, conversion coefficients and fat content, which represents a substantial improvement over QFASA, the only other currently available method to analyse diet proportions from fatty acids.

The general mixing model framework also allowed us to integrate SI and FAP into a joint model for diet estimation. Both approaches have their own limits, and the application to squid feeding trials suggests that their combination can help to overcome each tracers shortcomings to substantially reduce uncertainty

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

in diet estimates. As an increasing number of studies combine these two tracers (Bank *et al.*, 2011; Guest *et al.*, 2008; Guest *et al.*, 2009; Jaschinski, Brepohl & Sommer, 2008; Stowasser *et al.*, 2006; Tucker, Bowen & Iverson, 2008), we expect that a quantitative method to explicitly compare and combine markers will allow practitioners to make more robust inference and explicitly highlight discrepancies among methods that may warrant future research.

Simulation experiments and sensitivity tests suggested that the mixing model for FAP can achieve high accuracy of estimated diet proportions in idealised settings, and the application to squid feeding trials demonstrated the applicability of the model in a practical setting. Our results in the squid study further confirm many of the points made by Stowasser *et al.* (2006), thereby giving further credibility to our results. In particular, our analysis of discrimination coefficients showed that FA in the digestive glad may undergo significant modification and our analysis of switched diet treatments suggested that despite the short acclimation time (10-15 days) we can detect dominant proportions of the switched diet treatments from both SI and FA. While a complete discussion of these findings is beyond the scope of this manuscript, these results suggest that the time frame over which FAP and SI integrate diet proportions in squid is on the order of weeks rather than month.

Our results from the squid experimental data also showcased the model sensitivities found using simulated data. Fish species within treatments could not be discriminated using FAP (and/or SI), and estimated diet proportions corresponding to fish species in the SF treatment remained very uncertain. These correlations suggest insufficient prey separation at the species level,

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

which is a major determinant of accuracy as shown by simulation experiments. Despite the uncertainty in estimated diet proportions for individual fish species, the estimate for the group of all fish species as opposed to crustacean diets reveals a clear dominance of fish in the diets Figure 7. This example thus illustrates another important benefit of a fully Bayesian treatment: rather than giving potentially erroneous point estimates in such situations, the wide 95% intervals suggest that there is insufficient signal in the data to discriminate among diets at the species level.

The decrease in accuracy with decreasing source separation and increasing co-linearity reported from simulations and shown in the squid experiments is thus due to choosing a point estimate within a large interval rather than the model suggesting erroneous point estimates of diet proportions. Similarly, for unknown conversion coefficients, posterior distributions of diet estimates are generally wide, provided that the prior for conversion coefficients reflects uncertainty. Even when uncertainty about diet proportions is relatively low, posterior distributions of diet proportions close to 0 or 1 were generally skewed rather than symmetric due to the constrained nature of the diet proportions, meaning the posterior mode (the highest posterior probability) is often not located at the mean of the posterior distribution. In this case, as for very wide and/or flat posterior distributions, any point estimate chosen for diet proportions is somewhat arbitrary. Overall estimation error from (posterior mean) point estimates thus scales with the evenness of the diet proportions as well as overall uncertainty in diet proportions, and, rather than relying on point estimates of diet proportions in that case, it becomes increasingly

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

important to acknowledge uncertainty in the posterior distributions.

We opted for a fully Bayesian analysis that estimates prey and predator distributions, as well as individual proportions. However, the Bayesian approach for FAP comes at a relatively high computational cost: we found that there are limits to the dimensionality that the estimation procedure (as we formulated it) can deal with. When working with fully Bayesian methods in high dimensional applications such as FAP, where the number of measured variables can be large (>20 FAs is common), there is an inevitable trade-off between computational feasibility and model dimensionality. Since the model dimensionality depends at once on the number of prey items, predators and fatty acids in the analysis, we have found it to be useful to initially use predator FAP (geometric) means or relatively few predator signatures to estimate a single population distribution. Once one has determined that the model can effectively estimate diet proportions given the data at hand and knowledge of conversion coefficients, the model can be re-run with a larger number of predators and/or FAs and, although time consuming, may provide additional insights. The squid diet example illustrates this strategy: we first estimated population level parameters for predators (although we used all predator signatures rather than their geometric mean), and then proceeded to more complex analyses of individual diet proportions.

To further address the issue of computational complexity, we presented an approach to variable selection for FAPs. An optimal subset of variables is usually one that explains the bulk of among prey variance (represented by CAP axes), but eliminates FAs that only contribute minimally to separation

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

among sources, and thus only add noise. In our squid application, we found that retaining only 4 FA was enough to explain over 95% of among source variance, and adding additional FA only added a small amount of signal for rapidly increasing co-linearity in prey signatures. While a limited number of FA may often be diagnostic of a particular prey type, it may not generally be the case that a small number of FA account for the bulk of the signal. The computational cost of high dimensional models in the Bayesian framework can be limiting in such instances, and the practical trade-off between model run-time and accuracy of estimated diet proportions will have to be considered. Our aim is to further develop the fastinR package to include empirical Bayes options (as described in (Parnell *et al.*, 2013) that would likely speed up the models considerably. However, the empirical Bayes approach comes at the cost of considering prey distribution parameters as known quantities, which may not be desirable with a small number of prey samples.

Recent developments in SI mixing models have led to increasingly realistic models in terms of their error structure (Hopkins & Ferguson, 2012) and incorporation of relevant biology, such as time dependent diet proportions and SI signatures (Parnell *et al.*, 2013). Given that our FAP and combined FAP and SI models employ the same general structure as these models, such developments are readily achievable within this framework. It should be noted that they present the practitioner with requirements for substantial amounts of data of various kinds (i.e., measurement error estimates, collection of SI and FAP through time, respectively), and may substantially increase computational requirements. Nevertheless, we suggest that the methods

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

presented here provides a basis to use and combine the two most powerful markers for diet estimation available in a single framework to produce more robust and comparable.

References

Aitchison, J. & Bacon-Shone, J. (1999). Convex linear combinations of compositions. *Biometrika*, 86, 351–364.

Bank, M. G. van der, Utne-Palm, A. C., Pittman, K., Sweetman, A. K., Richoux, N. B., Brchert, V. & Gibbons, M. J. (2011). Dietary success of a *new* key fish in an overfished ecosystem: evidence from fatty acid and stable isotope signatures. *Marine Ecology Progress Series*, 428, 219–233.

Belicka, L. L., Burkholder, D., Fourqurean, J. W., Heithaus, M. R., Macko, S. A. & Jaff, R. (2012). Stable isotope and fatty acid biomarkers of seagrass, epiphytic, and algal organic matter to consumers in a pristine seagrass ecosystem. *Marine and Freshwater Research*, 63, 1085–1097.

Billheimer, D. (2001). Compositional receptor modeling. *Environmetrics*, 12, 451–467.

Blanchard, J. W. (2011). Inference on the diet composition of predators using fatty acid signatures: an application of Bayesian inference on lienar mixing models. PhD thesis. Dalhousie University.

Bowen, W. D. & Iverson, S. J. (2012). Methods of estimating marine mammal diets: A review of validation experiments and sources of bias and uncertainty. *Marine Mammal Science*, 29, 719–754.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

Budge, S. M., Iverson, S. J. & Koopman, H. N. (2006). Studying trophic ecology in marine ecosystems using fatty acids: A Primer on analysis and interpretation. *Marine Mammal Science*, 22, 759–801.

Budge, S. M., Wooller, M. J., Springer, A. M., Iverson, S. J., McRoy, C. P. & Divoky, G. J. (2008). Tracing carbon flow in an arctic marine food web using fatty acid-stable isotope analysis. *Oecologia*, 157, 117–129.

Caut, S., Angulo, E. & Courchamp, F. (2009). Variation in discrimination factors (δ15N and δ13C): the effect of diet isotopic values and applications for diet reconstruction. *Journal of Applied Ecology*, 46, 443–453.

Elsdon, T. S. (2010). Unraveling diet and feeding histories of fish using fatty acids as natural tracers. *Journal of Experimental Marine Biology and Ecology*, 386, 61–68.

Guest, M. A., Nichols, P. D., Frusher, S. D. & Hirst, A. J. (2008). Evidence of abalone (*Haliotis rubra*) diet from combined fatty acid and stable isotope analyses. *Marine Biology*, 153, 579–588.

Guest, M. A., Frusher, S. D., Nichols, P. D., Johnson, C. & Wheatley, K. (2009). Trophic effects of fishing southern rock lobster *Jasus edwardsii* shown by combined fatty acid and stable isotope analyses. *Marine Ecology Progress Series*, 388, 169–184.

Hopkins, J. B. & Ferguson, J. M. (2012). Estimating the diets of animals using stable isotopes and a comprehensive Bayesian mixing model. *PLoS ONE*, 7, e28478.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

Hussey, N. E., MacNeil, M. A., McMeans, B. C., Olin, J. A., Dudley, S. F., Cliff, G., Wintner, S. P., Fennessy, S. T. & Fisk, A. T. (2014). Rescaling the trophic structure of marine food webs. *Ecology Letters*, 17, 239–250.

Iverson, S. J., Field, C., Don Bowen, W. & Blanchard, W. (2004). Quantitative fatty acid signature analysis: a new method of estimating predator diets. *Ecological Monographs*, 74, 211–235.

Jack, L. & Wing, S. R. (2011). Individual variability in trophic position and diet of a marine omnivore is linked to kelp bed habitat. *Marine Ecology Progress Series*, 443, 129–139.

Jaschinski, S., Brepohl, D. C. & Sommer, U. (2008). Carbon sources and trophic structure in an eelgrass *Zostera marina* bed, based on stable isotope and fatty acid analyses. *Marine Ecology Progress Series*, 358, 103–114.

Kelly, J. R. & Scheibling, R. E. (2011). Fatty acids as dietary tracers in benthic food webs. *Marine Ecology Progress Series*, 446, 1–22.

Moore, J. W. & Semmens, B. X. (2008). Incorporating uncertainty and prior information into stable isotope mixing models. *Ecology Letters*, 11, 470–480.

Parnell, A. C., Phillips, D. L., Bearhop, S., Semmens, B. X., Ward, E. J., Moore, J. W., Jackson, A. L., Grey, J., Kelly, D. J. & Inger, R. (2013). Bayesian stable isotope mixing models. *Environmetrics*, 24, 387–399.

Pethybridge, H., Virtue, P., Casper, R., Yoshida, T., Green, C., Jackson, G. & Nichols, P. (2012). Seasonal variations in diet of arrow squid (*Nototodarus gouldi*): stomach content and signature fatty acid analysis. *Journal of the Marine Biological Association of the United Kingdom*, 92, 187–196.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

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

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pp. 20–22.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rosen, D. & Tollit, D. (2012). Effects of phylogeny and prey type on fatty acid calibration coefficients in three pinniped species: implications for the QFASA dietary quantification technique. *Marine Ecology Progress Series*, 467, 263–276.

Semmens, B. X., Ward, E. J., Moore, J. W. & Darimont, C. T. (2009). Quantifying inter- and intra-population niche variability using hierarchical bayesian stable isotope mixing models. *PLoS ONE*, 4, e6187.

Stowasser, G., Pierce, G. J., Moffat, C. F., Collins, M. A. & Forsythe, J. W. (2006). Experimental study on the effect of diet on fatty acid and stable isotope profiles of the squid *Lolliguncula brevis*. *Journal of Experimental Marine Biology and Ecology*, 333, 97–114.

Tucker, S., Bowen, W. D. & Iverson, S. J. (2008). Convergence of diet estimates derived from fatty acids and stable isotopes within individual grey seals. *Marine Ecology Progress Series*, 354, 267–276.

Wang, S. W., Hollmen, T. E. & Iverson, S. J. (2010). Validating quantitative fatty acid signature analysis to estimate diets of spectacled and Stellers

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

eiders (*Somateria fischeri* and *Polysticta stelleri)*. *Journal of Comparative Physiology B*, 180, 125–139.

Ward, E. J., Semmens, B. X. & Schindler, D. E. (2010). Including source uncertainty and prior information in the analysis of stable isotope mixing models. *Environmental science and technology*, 44, 4645–4650.

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, art19.

Williams, C. T. & Buck, C. L. (2010). Using fatty acids as dietary tracers in seabird trophic ecology: theory, application and limitations. *Journal of Ornithology*, 151, 531–543.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.413v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014