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: July 22, 2014

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

Abstract

Quantitative analysis of stable isotopes (SI) and, more recently, fatty acid (FA) profiles 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 FA profiles 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, demonstrating feasibility and model performance, and re-analyse an experimental dataset to illustrate modeling strategies and applications to real fatty acid profiles. 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 weaknesses (**bowen˙methods˙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˙methods˙2012**). Chemical markers such as stable isotopes (SI) and fatty acids (FA; often called fatty acid signatures or profiles) 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˙convergence˙2008**). These advantages have led to rapid growth in the use of chemical markers in diet studies (**elsdon˙unraveling˙2010**; **williams˙using˙2010**; **kelly˙fatty˙2011**; **bowen˙methods˙2012**). 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˙stable˙2012**), compound specific SI ratios (**budge˙tracing˙2008**; **jack˙individual˙2011**), or a combination of stomach content analysis and SI or FA (**pethybridge˙seasonal˙2012**). The use of SI and FA in combination also holds great promise; however, studies that have used both chemical markers have been qualitative ([**e.g.**;][]**guest˙trophic˙2009**) or based on positive correlation of results from both methods (**tucker˙convergence˙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 individual applications (**moore˙incorporating˙2008**; **hopkins˙estimating˙2012**; **parnell˙bayesian˙2012**). The latter models can, for instance, estimate dietary differences of geographically distinct populations (**semmens˙quantifying˙2009**), accommodate temporal changes in diets or estimate the effect of covariates (e.g., age, size, sex) on diet proportions (**parnell˙bayesian˙2012**).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˙source˙2003**; **ward˙quantitative˙2011**). Optimally aggregating prey items into prey groups may circumvent this problem (**ward˙quantitative˙2011**), but may also be less satisfactory in complex food webs.

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

Analysis (**iverson˙quantitative˙2004**) to obtain quantitative estimates of diet proportions.

QFASA is the only available (i.e., off the shelf) method thus far for use with FA data, and, in contrast to recent (Bayesian) SI and FA mixing models, 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 FA and it has already seen widespread use, particularly in studies of marine mammals (**bowen˙methods˙2012**) and seabirds (**williams˙using˙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 (**blanchard˙inference˙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 FA cannot be considered, yet small changes in these coefficients can lead to differences in inferred diet proportions (**wang˙validating˙2010**).

Given the discrepancy in methods applied to SI and FA 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 FA 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 FA data. For example, for many non-modified fatty acids, FA 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 either case, the two species may have similar FA. 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 FA data based on a probabilistic model whose parameters are estimated using Bayesian methods, and explicitly integrate SI in the estimation of diet proportions. Using both simulated and published data, we demonstrate the suitability of this model for FA 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 fatty acid profiles

Bayesian models for SI data are commonly based on the assumption that SI ratios are normally distributed. This assumption cannot be made for FA data, since for most methods of analysis, the concentration of individual FA is normalized to the total lipid content of the sample. Thus, the FA 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˙studying˙2006**). To construct our model, we considered the additive log ratio transformation (**aitchison˙convex˙1999**), also called alr transformation, such that

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

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 FA, 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 to propagate to estimates of diet proportions (**ward˙including˙2010**).

Each predator j consumes a proportion π_j of each prey source, and analogous to stable isotope mixing models, predator FA are then a linear combination of prey FA, normalized to sum to one. Since predators may selectively assimilate or metabolize FA (**iverson˙quantitative˙2004**; **budge˙studying˙2006**; **rosen˙effects˙2012**), we specify prey-specific conversion coefficients $\kappa_s = \kappa_{s,1}...\kappa_{s,p}$ for each of the p FA (**rosen˙effects˙2012**). Furthermore, the n prey species likely have different fat content Φ that will affect the total amount of FA assimilated from each prey species by the predator. The

expected FA of predator τ_j is then a linear combination of the prey FA, modified by conversion coefficients for each fatty acid p and fat content for each prey i:

$$t_j \sim N(alr(\tau_j), \Sigma_\tau) \tag{2}$$

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

Here, C is the closure operation which normalizes the FA to sum to one and $\otimes$ is the outer (element wise) product. $\phi_{s,j}$ is the FA of prey items of species s consumed by predator j. Similarly to **parnell˙bayesian˙2012**, 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 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 FA) leads to κ being determined in terms of relative uptake of FA (i.e., up to a multiplicative constant), and implicitly makes the multivariate prior 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 parametrized 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 in Equation 2. In the latter case, individual predator FA 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˙quantifying˙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 (**parnell˙bayesian˙2012**). While this approach is appealing, it adds to computation time needed to estimate model parameters, and correlates with generally slower convergence. We therefore use a vague Dirichlet prior on the proportions when convenient (i.e., when we estimate only population level parameters).

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 Markov Chain Monte Carlo (MCMC) employed to estimate model parameters. 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 FA and SI

There are at least three potential benefits of integrating FA 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 FA. 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 either data-source alone.

Our model for FA is conceptually similar to recent models proposed for SI data, and integration of FA and SI data into a single model is straightforward in the present setting. We assume that the vector of SI signatures of sampled 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

(**hussey˙rescaling˙2014**; **caut˙variation˙2009**). 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 FA model imply that there is no need to have SI and FA 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 FA, and that individual diet proportions of predators are drawn from the same population distribution of diet proportions.

The exact formulation of the integration of SI and FA 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 FA) 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.

An R (**R˙core˙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 population level diet estimates, individual diet estimates as well as linear model (regression and ANOVA) formulations for diet proportions, all available for SI and FA individually or as combined models (see below). Model parameters were estimated using MCMC methods implemented in JAGS (**plummer˙jags˙2003**), called from R through higher level functions in the fastinR package that allow for data input, inspection and manipulation.

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 FA space, co-linearity in FA space (**blanchard˙inference˙2011**) and diet makeup (e.g., specialist versus generalist diets). Further details and simulation results can be found in supplemental information S2.

2.4 Selecting fatty acids for analysis: an ordination approach

A potentially large number of FA 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 FA based on the among diet source variability that they explain. **wang˙validating˙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 FA that contribute most to source separation. This procedure is intended to reduce computation time (and dimensionality) of the models, while retaining as much accuracy in diet estimates as possible. Further details about the procedure are given in supplemental information S3.

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˙experimental˙2006**, which investigated changes in squid FA 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 FA 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˙rescaling˙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˙variation˙2009** to construct priors for $\delta^{13}C$. Estimation of FA conversion coefficients used (2) with proportions assumed known from feeding trials. Computational 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 FA 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 FA and SI after at least 10 days under the respective treatment. We only had overlapping SI and FA for the SC

treatment squid, and we therefore started by analyzing this treatment in isolation to demonstrate that both SI and FA can resolve diet proportions, and to demonstrate the benefit of using the two tracers in a joint model. We then analyzed the SF treatment squid, for which we only had 3 specimen with FA 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.

FA analyses used data obtained by analyzing digestive glad tissue, which is thought to rapidly assimilate dietary FA 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 FA were reported. Here, we selected FA 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 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 FA (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 FA of squid and their potential prey suggested that crustacean diets were readily distinguishable from fish diets (**??**a). For fish diet items, however, no single fish species could be clearly