Should fatty acid signature proportions sum to 1 for diet estimation?

Jeffrey F. Bromaghin, Suzanne M. Budge & Gregory W. Thiemann

Ecological Research

ISSN 0912-3814
Volume 31
Number 4

Ecol Res (2016) 31:597-606
DOI 10.1007/s11284-016-1357-8

Springer

Ecol Res (2016) 31: 597–606
DOI 10.1007/s11284-016-1357-8

TECHNICAL REPORT

Jeffrey F. Bromaghin · Suzanne M. Budge
Gregory W. Thiemann

Should fatty acid signature proportions sum to 1 for diet estimation?

Received: 10 December 2015 / Accepted: 18 March 2016 / Published online: 29 March 2016

Abstract Knowledge of predator diets, including how diets might change through time or differ among predators, provides essential insights into their ecology. Diet estimation therefore remains an active area of research within quantitative ecology. Quantitative fatty acid signature analysis (QFASA) is an increasingly common method of diet estimation. QFASA is based on a data library of prey signatures, which are vectors of proportions summarizing the fatty acid composition of lipids, and diet is estimated as the mixture of prey signatures that most closely approximates a predator's signature. Diets are typically estimated using proportions from a subset of all fatty acids that are known to be solely or largely influenced by diet. Given the subset of fatty acids selected, the current practice is to scale their proportions to sum to 1.0. However, scaling signature proportions has the potential to distort the structural relationships within a prey library and between predators and prey. To investigate that possibility, we compared the practice of scaling proportions with two alternatives and found that the traditional scaling can meaningfully bias diet estimators under some conditions. Two aspects of the prey types that contributed to a predator's diet influenced the magnitude of the bias: the degree to which the sums of unscaled proportions differed among prey types and the identifiability of prey types within the prey library. We caution investigators against the routine scaling of signature proportions in QFASA.

Keywords Bias · Bootstrap sample size · Diet composition · Distance measure · Quantitative fatty acid signature analysis QFASA

J. F. Bromaghin (✉)
U.S. Geological Survey, Alaska Science Center, 4210 University Drive, Anchorage, AK 99508, USA
Tel.: 907-786-7086
E-mail: jbromaghin@usgs.gov
URL: http://alaska.usgs.gov/science/biology/quantitative_ecology/index.php

S. M. Budge
Process Engineering and Applied Science, Dalhousie University, Halifax, NS B3H 4R2, Canada

G. W. Thiemann
Faculty of Environmental Studies, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada

Introduction

Knowledge of animal diets, and how diets differ among individuals or age and sex classes, provides essential insights into ecological processes. Consequently, the development of methods to estimate diets has long been a focus of research in quantitative ecology and several methods remain in common use (e.g., Bowles et al. 2011; Chambellant et al. 2013; Haynes et al. 2015; Roberts and Lalas 2015; Zeppelin et al. 2015). Quantitative fatty acid signature analysis (QFASA; Iverson et al. 2004) is one such method, being implemented most widely for diet estimation of marine species (e.g., Bowen and Iverson 2013).

QFASA is based upon a data library of fatty acid signatures, which are vectors of proportions that represent the fatty acid composition of lipids, compiled from samples of prey types potentially consumed by predators. Calibration coefficients to adjust for the differential metabolism of fatty acids by predators are obtained in controlled feeding trials (e.g., Nordstrom et al. 2008; Wang et al. 2010), and are used to either transform a prey signature to the predator space or a predator signature to the prey space (Bromaghin et al. 2015). Given those data inputs, a predator signature is modeled as a linear mixture of prey signatures and diet is estimated as the mixture that minimizes a measure of distance between the modeled and observed predator signatures (Iverson et al. 2004). Several distance measures have been used in the literature (e.g., Iverson et al. 2004; Stewart et al. 2014), although the Kullback–Leibler measure originally rec-

ommended by Iverson et al. (2004) is currently used most frequently.

Although marine lipids can be comprised of more than 70 fatty acids (Budge et al. 2006), investigators typically base diet estimation on a subset of fatty acids that are primarily derived from diet. Some investigators routinely use what have been termed the "dietary" or "extended dietary" suites of fatty acids (e.g., Thiemann et al. 2008; Wang et al. 2010; Haynes et al. 2015), though additional criteria may be applied (e.g., Bromaghin et al. 2013). Given the suite of fatty acids selected, their proportions are scaled to sum to 1.0 so that vectors of proportions are compositional. If the sums of the unscaled proportions, hereafter termed a "partial sum", differ between individuals, each individual's proportions are scaled by a different constant. Although the Aitchison distances between unscaled partial prey signatures whose proportions have different partial sums are not changed when the signatures are individually scaled to sum to 1.0, the Kullback–Leibler and Chi square distances are not generally unaffected. More importantly for diet estimation, scaling-induced magnitude changes in prey proportions are propagated to a modeled predator signature irrespective of the distance measure in use. For these reasons, the customary scaling has the potential to distort the structural relationships within a prey library and between predator and prey, and thereby bias diet estimation.

To investigate the potential for scaling to bias diet estimation, we compared the influence of three approaches to scaling on the bias and variance of diet estimators based on the Aitchison (Stewart et al. 2014), Kullback–Leibler (Iverson et al. 2004), and Chi square (Stewart et al. 2014) distance measures using computer simulation. Independent simulations were conducted using two libraries of prey signatures having substantially different characteristics to enhance the general applicability of our conclusions.

Methods

Prey libraries

Simulations were conducted using two libraries of prey signature data, one with 357 signatures from seven marine mammal species (Rode et al. 2014) and the second with 954 signatures from 28 fish and shellfish species (hereafter referred to as the fish library; Budge et al. 2002). Both prey libraries have previously been used to investigate the performance of QFASA estimators (Bromaghin et al. 2015, 2016), although we here excluded the monkfish (*Lophius americanus*) prey type from the fish library due to its small sample size ($n = 3$). Bromaghin et al. (2016) provides links to both prey libraries.

The Chi square distance measure is defined for signature proportions of zero (Stewart et al. 2014), but the Kullback–Leibler and Aitchison distance measures are not. Therefore, to prepare the prey libraries for simulations, we first replaced any proportions equal to zero with a small constant (0.0001) and the full suite of all fatty acid proportions in each library (66 in the mammal library and 68 in the fish library) were scaled to sum to 1.0 using the multiplicative method (Martín-Fernández et al. 2011). Because the proportions of the full suite of fatty acids sum to 1, this replacement of zeros produces virtually no distortion in the signatures but allows all three distance measures we considered to be computed. All individuals in the mammal library had proportions of 0 for fatty acid 16:3n-1, so that fatty acid was deleted from the mammal library prior to the replacement of zeros.

We evaluated each prey library with respect to prey-type identifiability using a leave-one-out procedure, modified from the subsampling approach of Iverson et al. (2004). An individual prey signature was temporarily excluded from a library, the mean signature of its prey type was recomputed, and the "diet" of the excluded specimen was estimated as if it were a predator, after which the excluded prey signature was returned to the library. The estimated diet proportion for the prey type to which the excluded signature belonged measured the identifiability of the prey type in relation to the prey library as a whole. We excluded each prey signature in turn and computed the average diet proportion correctly attributed to each prey type. An average proportion close to 1.0 indicated a prey type that was highly identifiable within a library, while values less than 1.0 indicated some degree of confounding with other prey types. Because our simulations included three distance measures, we replicated this procedure for each distance measure and averaged the results.

Simulation design

We developed an algorithm to construct a regular grid of diet proportions, allowing us to compare estimator performance with diets systematically located and evenly distributed throughout the span of all possible diet compositions. We initiated the algorithm with a diet consisting wholly of the first prey type. The algorithm then constructed other diets by systematically shifting an increment of diet to the adjacent prey type in stepwise fashion until all of the diet was associated with the last prey type. We then excluded diets comprised of a single prey type so that at least two prey types contributed to each diet. A small example of such a diet grid for three prey types and a diet increment of 0.25 is presented in Table 1. We used a diet increment of 0.10 for the mammal library, resulting in a total of 8001 diets for that library. The larger number of prey types in the fish library necessitated a larger increment of 0.25, to limit execution time of the simulations, resulting in 31,437 diets.

Table 1 An example diet grid for three prey types and a diet increment of 0.25

Diet	Prey 1	Prey 2	Prey 3
1	0.75	0.25	0.00
2	0.75	0.00	0.25
3	0.50	0.50	0.00
4	0.50	0.25	0.25
5	0.50	0.00	0.50
6	0.25	0.75	0.00
7	0.25	0.50	0.25
8	0.25	0.25	0.50
9	0.25	0.00	0.75
10	0.00	0.75	0.25
11	0.00	0.50	0.50
12	0.00	0.25	0.75

For each diet, we randomly generated signatures for 30 pseudo-predators using all fatty acids available in each library. For each predator signature, a bootstrap sample of prey signatures was drawn from each prey type in the library, the mean signature was computed from each prey type sample, and the predator signature was computed as the sum of the diet proportion for each prey type multiplied by the corresponding sample mean fatty acid proportion (Bromaghin et al. 2015). We used two levels of bootstrap sample size. In the small sample size case, each bootstrap sample was of size two, resulting in predator signatures with a relatively high degree of variability around the mean signature. In the large sample size case, bootstrap sample sizes equaled the number of individuals of each prey type in the library, producing signatures with a relatively low level of variability.

We based diet estimation on the suite of 33 dietary fatty acids (Iverson et al. 2004). However, the elimination of 16:3n-1 from mammal library, noted above, resulted in diet estimation with the mammal library being based on the remaining 32 dietary fatty acids. We estimated the diet for each predator signature in the prey space (Bromaghin et al. 2015), consistent with how the predator signatures were constructed, using nine estimators formed by the combination of three distance measures and three scaling options. The distance measures were the Kullback–Leibler (Iverson et al. 2004), the Aitchison (Stewart et al. 2014), and a Chi square measure with the power parameter γ equal to 1 (Stewart et al. 2014; value of 1 recommended by C. Stewart, personal communication). In the "traditional" scaling option, proportions for the dietary suite of fatty acids were scaled to sum to 1.0 as is common practice (Iverson et al. 2004). The "unscaled" option consisted of using the dietary fatty acid proportions without change, so their partial sums were less than 1.0 and varied among individuals. In the "augmented" option, the subset of fatty acid proportions were also used without change, but each signature was augmented with an additional proportion equal to 1.0 minus the partial sum. The augmented option therefore preserved the compositional (sum to 1.0) characteristic of a signature. For each diet and estimator, we computed a measure of bias as the sum of the absolute values of the differences between the true diet and the average diet estimate across the 30 pseudo-predators. All computations were performed using R version 3.2.2 (R Core Team 2015) and diets were estimated by minimizing distance measures using the "solnp" function of the R package Rsolnp (Ghalanos and Theussl 2014).

We expected bias to depend on the degree to which the prey types contributing to each diet were identifiable within the prey library. In addition, the selection of a scaling option could reasonably be expected to be most consequential when the partial sums of prey types contributing to a diet differed most greatly. We therefore computed measures of both factors with the expectation that they would aid in the interpretation of results. The measure of "diet identifiability" for diet d was computed as

$$I_d = \sum_k \pi_k p_k$$

where π_k is the diet proportion for prey type k, p_k is the average proportion of prey type k correctly attributed to prey type k in the leave-one-out analysis of prey group identifiability, and the sum is over all k prey types. The measure of the dissimilarity of partial sums among prey types, hereafter termed "sum dissimilarity", for diet d was computed as

$$D_d = \sqrt{\sum_{k \ni \pi_k > 0} \pi_k (s_k - \bar{s}_d)^2}$$

where s_k is the average partial sum for prey type k, $\bar{s}_d$ is the average value of s_k among prey types contributing to diet d, and the sum is over the prey types contributing to diet d. For purposes of graphical summarization, the observed ranges of both measures were divided into approximately 100 equally-sized bins and the average bias of each estimator and the average difference in bias for each pair of estimators were computed for each combination of bins.

Results

The partial sums differed among prey types in both libraries (Table 2). Maximum differences between partial sums were 0.090 for the mammal library and 0.227 for the fish library. The leave-one-out analysis of prey type identifiability revealed substantial differences in the complexity of the two prey libraries (Table 3), as previously reported (Bromaghin et al. 2016). Prey types in the mammal library were relatively identifiable, with measures of prey-type identifiability ranging from 0.526 to 0.973 and averaging 0.782. The measures of identifiability for prey types in the fish library were smaller on average and more variable, ranging from 0.023 to 0.928

Table 2 Statistics summarizing the partial sums of the dietary fatty acids among prey types in each prey library

Statistic	Prey library	
	Mammal	Fish
Minimum	0.301	0.367
Maximum	0.391	0.594
Average	0.336	0.509
SD	0.034	0.050

SD standard deviation

Table 3 Statistics summarizing the mean proportion of prey attributed to the correct prey type in the leave-one-out analysis to measure prey-type identifiability

Statistic	Prey library	
	Mammal	Fish
Minimum	0.526	0.023
Maximum	0.973	0.928
Average	0.782	0.531
SD	0.153	0.258

SD standard deviation

and averaging 0.531, reflecting the greater degree of complexity in that library.

The time required to complete the simulations ranged from 2 days for the small sample size mammal simulation to 26 days for the large sample size fish simulation on a desktop computer with a 3.4 GHz Intel i7-2600K processor. The larger number of prey types and the greater complexity of the fish library substantially slowed the optimization routine.

The mean and standard deviation of estimator bias computed across all diets in the grid for each library varied as a function of all aspects of simulation design (Table 4). Bias varied substantially between the two prey libraries, with the greater complexity and reduced prey-type identifiability of the fish library leading to increased bias. Bias also differed substantially between the two bootstrap sample sizes. With the mammal library, bias in the small sample size case ranged from approximately 2 (0.23 versus 0.12 for the traditional Kullback–Leibler estimator) to nearly 4 (0.23 versus 0.06 for the unscaled Kullback–Leibler estimator) times greater than the bias in the large sample size case among estimators. For the fish library, bias in the small sample size case was consistently about 3 times greater than the bias in the large sample size case (e.g., 0.72 versus 0.22 for the unscaled Aitchison estimator). Bias differed to a lesser degree between the scaling options. Biases of the unscaled and augmented estimators were consistently similar and tended to approximately equal or be less than the bias of the traditional estimators. The greatest differences were observed with large bootstrap sample sizes and the mammal library, where the bias of the traditional Kullback–Leibler and Chi square estimators was approximately double that of the unscaled and augmented estimators (0.12 versus 0.06 or 0.07), and the bias of the traditional Aitchison estimator was 2.6 times greater than the bias of unscaled and augmented estimators (0.13 versus 0.05). Differences in the bias among distance measures were relatively small and did not display a strong pattern, though the biases of the Kullback–Leibler and Chi square estimators were consistently similar.

Although the magnitude of estimator bias did not display recognizable patterns with respect to diet identifiability or sum dissimilarity (results not shown), strong and consistent patterns in the difference between estimator biases were observed (Figs. 1, 2, 3, 4). When diet identifiability and sum dissimilarity were both moderate to high, the biases of the unscaled and augmented estimators were consistently less than the biases of the traditional estimators. Large bootstrap sample sizes increased the size of the region in which the unscaled and augmented estimators had the least bias (Figs. 1 versus 2; Figs. 3 versus 4). As sum dissimilarity decreased, the consistency with which the unscaled and augmented estimators had the lowest bias decreased, especially for the small sample size cases (Figs. 1, 3). A similar effect was observed as diet identifiability decreased. The relationship between the unscaled and augmented estima-

Table 4 The mean (standard deviation) of bias by scaling option and distance measure (distance) for small and large bootstrap sample sizes (sample) computed across all diets considered with each prey library

Library	Sample	Distance	Scaling option		
			Traditional	Unscaled	Augmented
Mammal	Small	Aitchison	0.23 (0.07)	0.18 (0.08)	0.18 (0.08)
Mammal	Small	Kullback–Leibler	0.23 (0.08)	0.23 (0.10)	0.24 (0.11)
Mammal	Small	Chi square	0.23 (0.08)	0.23 (0.10)	0.24 (0.11)
Mammal	Large	Aitchison	0.13 (0.04)	0.05 (0.02)	0.05 (0.02)
Mammal	Large	Kullback–Leibler	0.12 (0.03)	0.06 (0.03)	0.07 (0.03)
Mammal	Large	Chi square	0.12 (0.03)	0.06 (0.03)	0.07 (0.03)
Fish	Small	Aitchison	0.73 (0.23)	0.72 (0.24)	0.71 (0.24)
Fish	Small	Kullback–Leibler	0.69 (0.21)	0.67 (0.22)	0.68 (0.22)
Fish	Small	Chi square	0.69 (0.21)	0.67 (0.22)	0.68 (0.22)
Fish	Large	Aitchison	0.24 (0.08)	0.22 (0.08)	0.22 (0.08)
Fish	Large	Kullback–Leibler	0.26 (0.08)	0.22 (0.08)	0.23 (0.08)
Fish	Large	Chi square	0.25 (0.08)	0.22 (0.08)	0.22 (0.08)

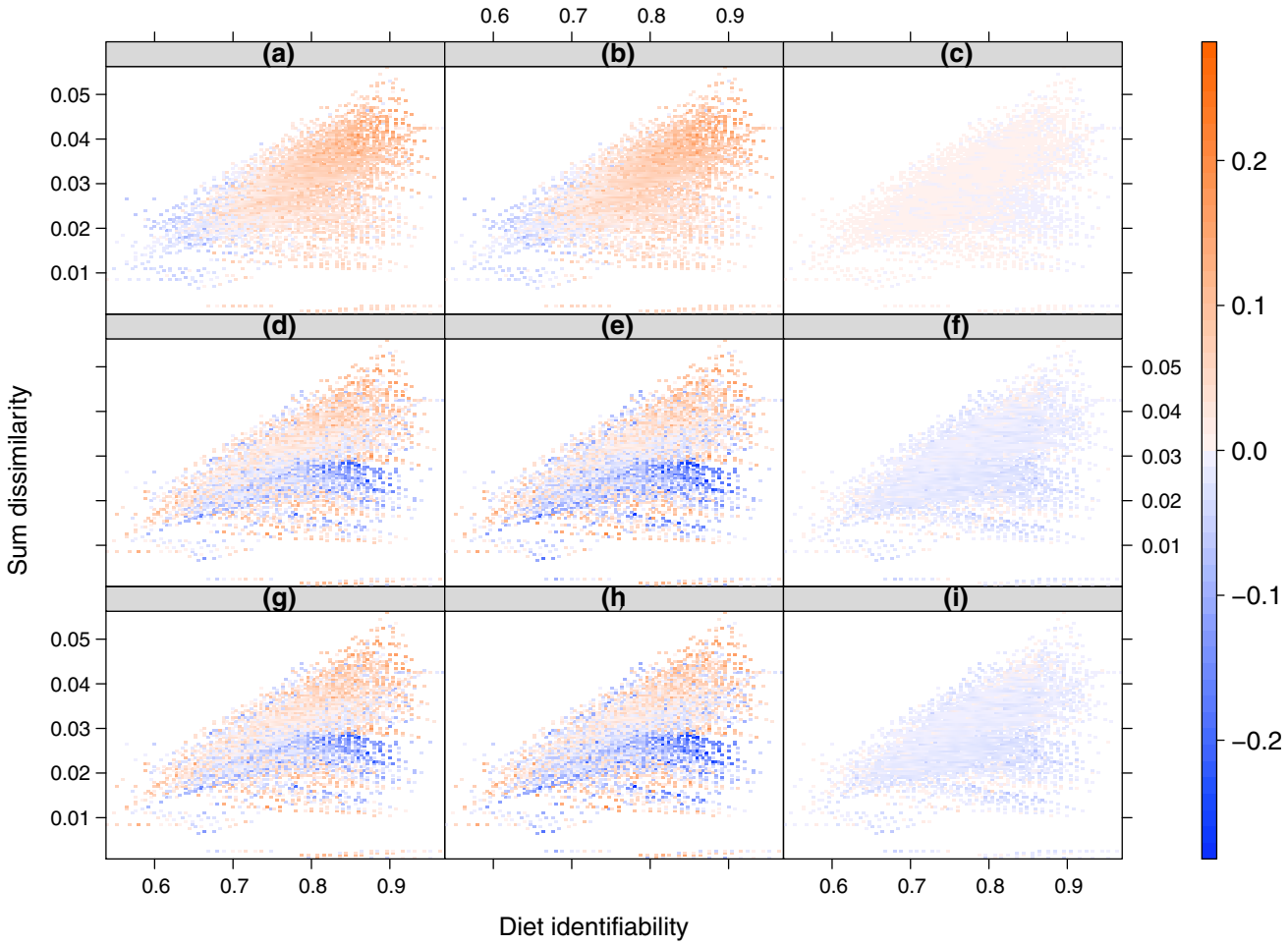

Fig. 1 Differences in mean bias between pairs of estimators (positive difference indicates the first estimator had greater bias) as a function of measures of differences in partial sums (*sum dissimilarity*) and the identifiability of a diet (*diet identifiability*) for the marine mammal prey library: **a** traditional minus unscaled, Aitchison measure; **b** traditional minus augmented, Aitchison measure; **c** unscaled minus augmented, Aitchison measure, **d** traditional minus unscaled, Kullback–Leibler measure, **e** traditional minus augmented, Kullback–Leibler measure; **f** unscaled minus augmented, Kullback–Leibler measure, **g** traditional minus unscaled, Chi square measure, **h** traditional minus augmented, Chi square measure; **i** unscaled minus augmented, Chi square measure. Bootstrap sample sizes for generating pseudo-predator signatures were equal to 2 from each prey type

tors differed between distance measures, with the unscaled estimator having slightly greater bias with the Aitchison measure and slightly less bias with the Kullback–Leibler and Chi square measures.

Discussion

We found that the selection of a scaling option can meaningfully influence QFASA diet estimation and therefore merits thoughtful consideration in QFASA applications. If partial sums are similar among prey types, the selection of a scaling option should be relatively inconsequential and other aspects of the model, such as the distance measure, predator diet, and properties of the prey library, are likely to more strongly influence diet estimation. However, as the disparity among partial sums increases, the traditional scaling of fatty acid proportions appears to progressively distort the structure among signatures and induce a bias in diet estimation that can be substantial.

Our findings suggest that the degree of confounding among prey within the library is the most important determinant of model performance. An obvious source of prey confounding is similarities among the mean prey-type signatures used to estimate diet. However, the degree to which the signatures of individual prey sampled to simulate predator signatures differ from the mean signature of their prey type, and perhaps resemble other prey types, can also be viewed as a source of prey confounding. When these sources of confounding were low, i.e., diet identifiability was high and predator variation was low (large bootstrap sample sizes), the model performed well and the influence of scaling was most apparent. Conversely, reduced model performance caused by low diet identifiability or high predator variation largely masked or overwhelmed differences attributable to the scaling options. The influence of

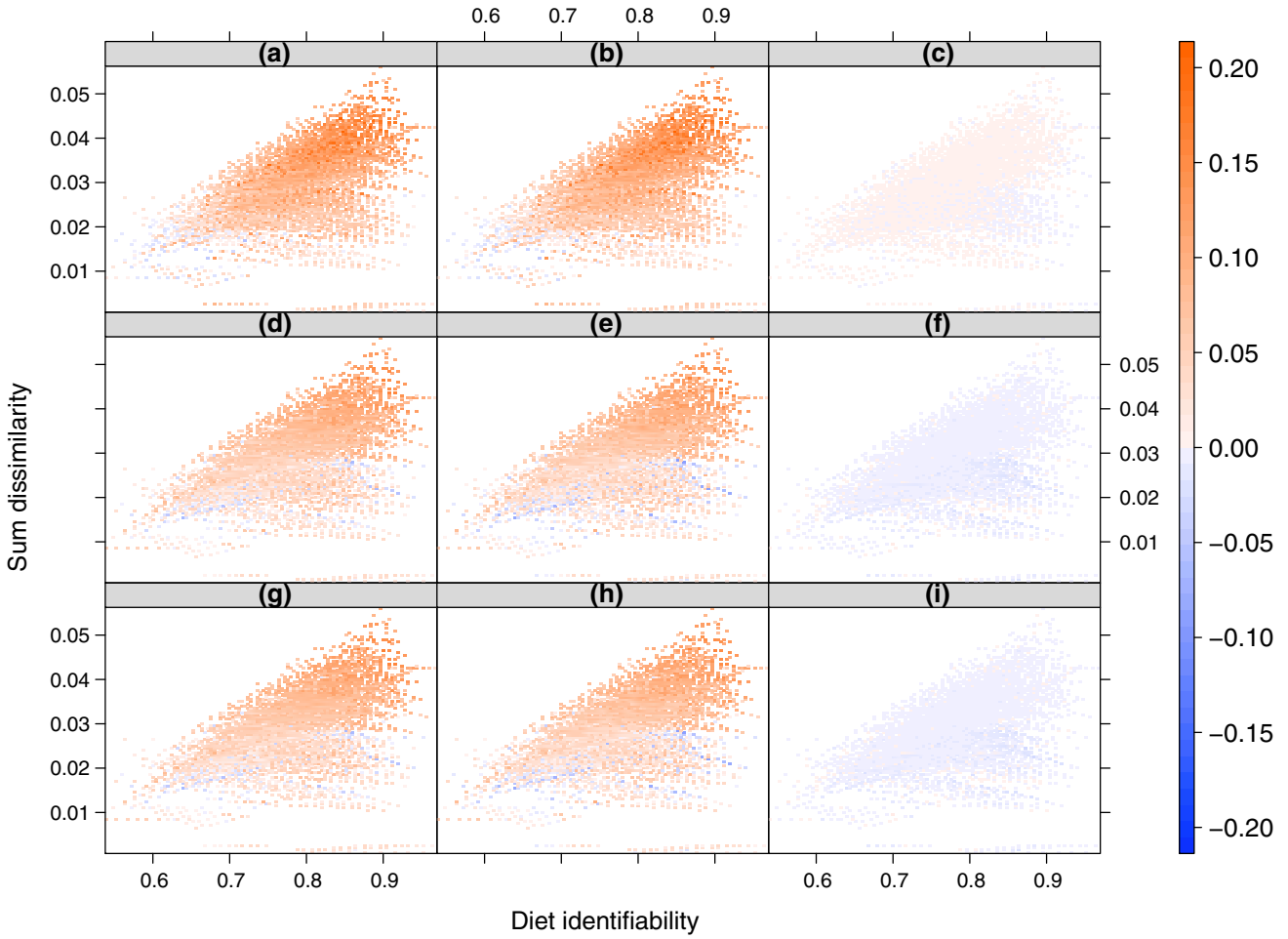

Fig. 2 Differences in mean bias between pairs of estimators (positive difference indicates the first estimator had greater bias) as a function of measures of differences in partial sums (*sum dissimilarity*) and the identifiability of a diet (*diet identifiability*) for the marine mammal prey library: **a** traditional minus unscaled, Aitchison measure; **b** traditional minus augmented, Aitchison measure; **c** unscaled minus augmented, Aitchison measure, **d** traditional minus unscaled, Kullback–Leibler measure, **e** traditional minus augmented, Kullback–Leibler measure; **f** unscaled minus augmented, Kullback–Leibler measure, **g** traditional minus unscaled, Chi square measure, **h** traditional minus augmented, Chi square measure; **i** unscaled minus augmented, Chi square measure. Bootstrap sample sizes for generating pseudo-predator signatures were equal to the number of individuals in each prey type

bootstrap sample size that we observed suggests that variation among individual signatures within prey types may be an underappreciated determinant of model performance.

We were surprised by the extent to which bootstrap sample size influenced bias. For a group of predators with a common diet, the random selection of individual prey in the bootstrap samples used to simulate their signatures is the only source of variation, so bootstrap sample size directly controls the degree of variation among predator signatures. Our initial expectation was that the degree of variation in predator signatures would primarily influence the degree of variation in their diet estimates. However, our results clearly demonstrate that variation in predator signatures strongly influences both bias and variance in diet estimation. We hypothesize that when bootstrap sample sizes are large, the mean signature of prey in the bootstrap sample, which contributes to a predator signature, is very similar to the mean prey-type signature in the library that is used in diet estimation, maximizing the ability of the model to accurately identify the correct mixture of prey types contributing to diet. Conversely, when bootstrap sample sizes are small, the mean signature of prey in the bootstrap sample may be quite different from the mean signature of that prey type in the library, or even more closely resemble the mean signature of another prey type, thereby increasing both bias and variation. The degree to which variation in predator signatures influences bias is undoubtedly dependent on the interaction of the prey types contributing to a diet and the structure within the prey library.

The strong influence of bootstrap sample size on both bias and variance suggests that care is needed when conducting simulation studies to evaluate the capabilities of QFASA in particular applications. Using bootstrap sample sizes that are either too large or too small may produce misleading results that are overly opti-

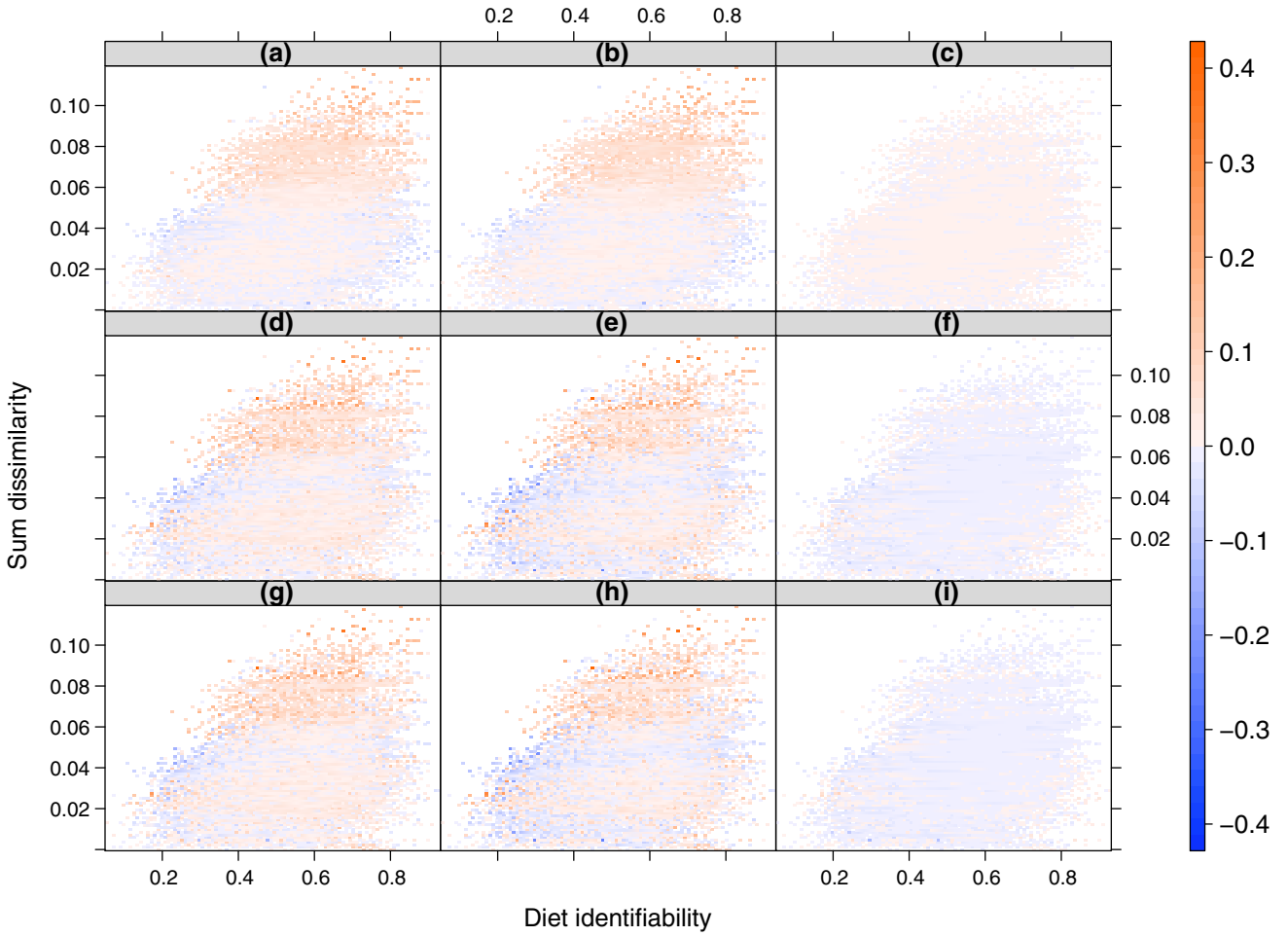

Fig. 3 Differences in mean bias between pairs of estimators (positive difference indicates the first estimator had greater bias) as a function of measures of differences in partial sums (*sum dissimilarity*) and the identifiability of a diet (*diet identifiability*) for the fish library: **a** traditional minus unscaled, Aitchison measure; **b** traditional minus augmented, Aitchison measure; **c** unscaled minus augmented, Aitchison measure, **d** traditional minus unscaled, Kullback–Leibler measure, **e** traditional minus augmented, Kullback–Leibler measure; **f** unscaled minus augmented, Kullback–Leibler measure, **g** traditional minus unscaled, Chi square measure, **h** traditional minus augmented, Chi square measure; **i** unscaled minus augmented, Chi square measure. Bootstrap sample sizes for generating pseudo-predator signatures were equal to 2 from each prey type

mistic or pessimistic, respectively, with respect to model performance. A further complication is that our results, particularly the small and large sample size cases with the mammal library, demonstrate that even relative performance among estimators can change as a function of bootstrap sample size. The empirical algorithm to establish bootstrap sample sizes of Bromaghin (2015) provides the only viable solution to this problem that we are aware of.

Although the selection of bootstrap samples size may seem to have importance only with respect to computer simulation studies, in fact it relates directly to the structure within a prey library and is therefore pertinent to QFASA applications. Predators in the wild could have identical diets in terms of prey types, but they are not generally consuming the same individual prey animals, and it is that selection of individual prey that bootstrap sampling mimics. Both logic and our results suggest that the structure within a prey library is a crucial determinant of QFASA performance, with confounding among prey increasing both bias and variance in diet estimation. The ideal library would therefore contain only those prey types actually consumed by predators, and be characterized by minimal signature variation within prey types and maximal signature variation between prey types. Unfortunately, that ideal seems unlikely to be achievable in most QFASA applications. One potential remedy might be to estimate diet on the basis of individual prey animals (e.g., Meynier et al. 2010) and then sum the estimates with respect to the desired prey types. That approach seems likely to improve diet estimation, but its computational burden is likely to quickly exceed practical limits as the size of the prey library increases. An intermediate strategy, in which the desired prey types are partitioned into a larger number of similar groups for estimation, might be worth

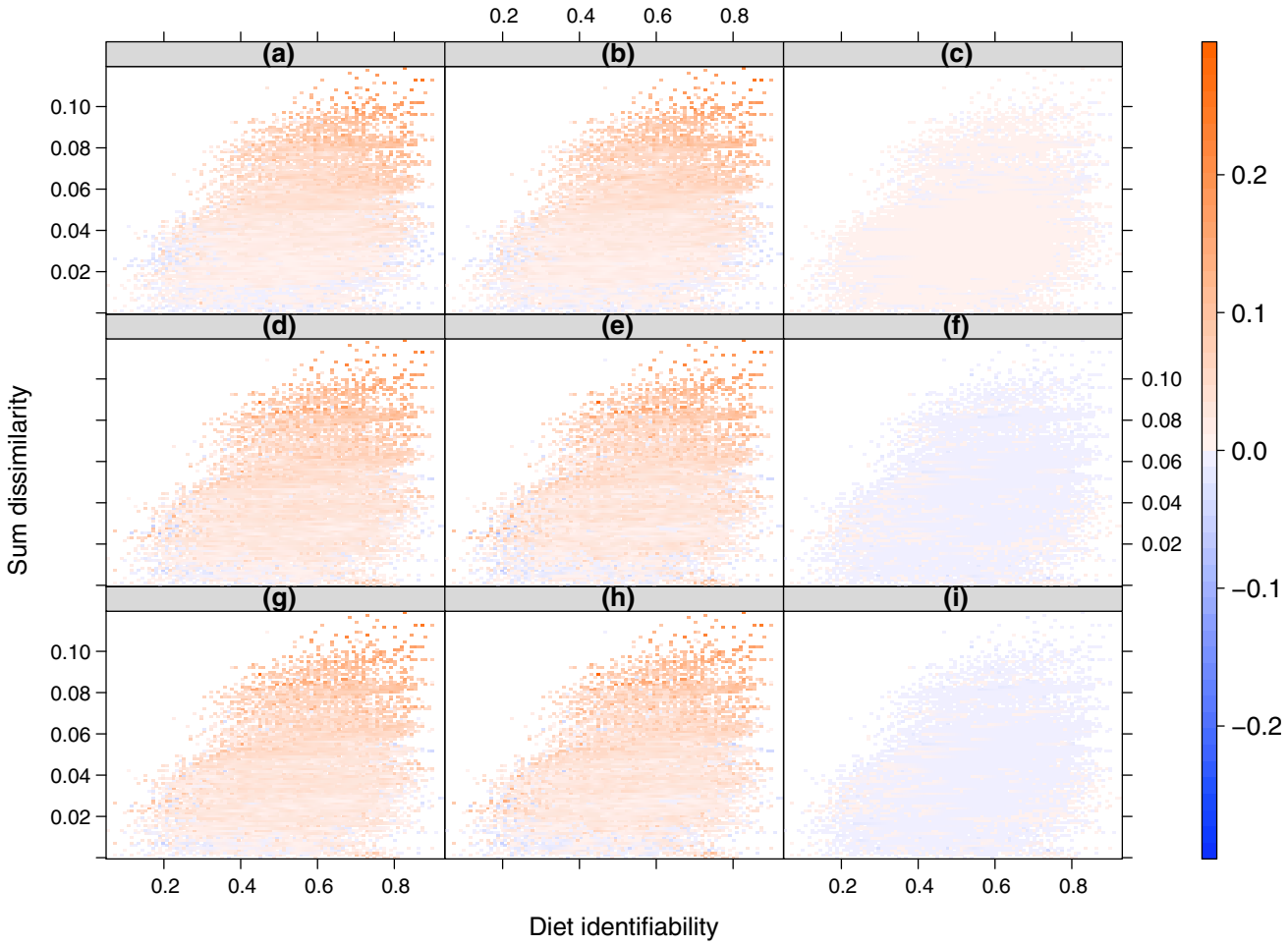

Fig. 4 Differences in mean bias between pairs of estimators (positive difference indicates the first estimator had greater bias) as a function of measures of differences in partial sums (*sum dissimilarity*) and the identifiability of a diet (*diet identifiability*) for the fish library: **a** traditional minus unscaled, Aitchison measure; **b** traditional minus augmented, Aitchison measure; **c** unscaled minus augmented, Aitchison measure, **d** traditional minus unscaled, Kullback–Leibler measure, **e** traditional minus augmented, Kullback–Leibler measure; **f** unscaled minus augmented, Kullback–Leibler measure, **g** traditional minus unscaled, Chi square measure, **h** traditional minus augmented, Chi square measure; **i** unscaled minus augmented, Chi square measure. Bootstrap sample sizes for generating pseudo-predator signatures were equal to the number of individuals in each prey type

investigating in some applications. The "allocate-sum" approach, in which estimation is based on a large number of groups after which estimates are summed to a smaller number of groups for reporting, is routine in some other disciplines in which mixture models are used (e.g., Wood et al. 1987).

We did not observe strong and consistent differences among the distance measures, although they can produce substantially different diet estimates for some individual predator signatures (e.g., Bromaghin et al. 2015) and the Kullback–Leibler and Aitchison measures respond differently to violations of model assumptions (Bromaghin et al. 2016). One aspect of our results that we found intriguing was the apparent similarity in the results for the Kullback–Leibler and Chi square distance measures, because Stewart et al. (2014) report that the Chi square measure converges to the Aitchison measure as the γ parameter tends to zero. It would be interesting to explore how the performance of estimators based on this distance measure might vary over a range of values for γ.

By generating pseudo-predator signatures and estimating diets in the prey space, we avoided the need to use calibration coefficients in our simulations. However, calibration coefficients are necessary to transform predator signatures to the prey space or prey signatures to the predator space (Bromaghin et al. 2015) in QFASA applications, and transformed signatures must be scaled to sum to 1.0. Consequently, use of either the unscaled or augmented approach will require investigators to first apply calibration coefficients to all fatty acids, not just the subset on which estimation is to be based. Once the transformed signatures have been computed, one would then restrict attention to the desired suite of fatty acids and augment the signatures if implementing that approach.

We intentionally used the relatively small suite of 33 (32 for the mammal data) dietary fatty acids to investigate differences in scaling approaches because we expected use of a smaller number of fatty acids would make any patterns more readily apparent. While not guaranteed, using fewer fatty acids should tend to create greater opportunity for the partial sums to vary more greatly among prey types. Conversely, using more fatty acids, such as the 41 fatty acids in the extended dietary suite, would tend to reduce the influence of the scaling option used. While we did not explicitly explore this question, it seems to be a reasonable expectation.

Conclusions

The current practice of scaling proportions of the subset of fatty acids used in QFASA diet estimation so they sum to 1.0 can introduce a meaningful structural bias under certain conditions. Unfortunately, one might have difficulty in determining the degree to which such conditions exist in a particular investigation. At a minimum, we recommend that practitioners inspect the partial sums of unscaled proportions to determine how greatly they differ among prey types as one simple indicator of whether scaling might introduce a bias. Determining whether diet identifiability is high requires a priori knowledge of diets and is therefore difficult to accomplish with confidence. Implementing the leave-one-out procedure with a prey library might be somewhat informative, but that procedure is independent of diet composition. Caution might dictate use of either the unscaled or augmented approach. Our results do not reveal a strong or consistent difference between them. The two approaches use the exact same proportions for each fatty acid in the analysis, the difference being that the augmented approach adds a proportion so that all sum to 1.0. The augmented approach therefore preserves the compositional feature of signatures, which might be viewed as a conceptual advantage. In either case, these two new approaches appear to avoid a potential source of bias whose magnitude could be meaningful in some applications.

Acknowledgments We thank M. Whalen (USGS Alaska Science Center) for assistance with the figures. Shiway Wang (Sedna Ecological Inc.) and two anonymous referees made helpful suggestions that improved the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not constitute endorsement by the U.S. Government.

References

Bowen WD, Iverson SJ (2013) Methods of estimating marine mammal diets: a review of validation experiments and sources of bias and uncertainty. Mar Mammal Sci 29:719–754. doi:10.1111/j.1748-7692.2012.00604.x

Bowles E, Schulte PM, Tollit DJ, Deagle BE, Trites AW (2011) Proportion of prey consumed can be determined from faecal DNA using real-time PCR. Mol Ecol Resour 11:530–540. doi:10.1111/j.1755-0998.2010.02974.x

Bromaghin JF (2015) Simulating realistic predator signatures in quantitative fatty acid signature analysis. Ecol Inform 30:68–71. doi:10.1016/j.ecoinf.2015.09.011

Bromaghin JF, Lance MM, Elliott EW, Jeffries SJ, Acevedo-Gutiérrez A, Kennish JM (2013) New insights into the diets of harbor seals (*Phoca vitulina*) in the Salish Sea revealed by analysis of fatty acid signatures. Fish B 111:13–26. doi:10.7755/FB.111.1.2

Bromaghin JF, Rode KD, Budge SM, Thiemann GW (2015) Distance measures and optimization spaces in quantitative fatty acid signature analysis. Ecol Evol 5:1249–1262. doi:10.1002/ece3.1429

Bromaghin JF, Budge SM, Thiemann GW, Rode KD (2016) Assessing the robustness of quantitative fatty acid signature analysis to assumption violations. Method Ecol Evol 7:51–59. doi:10.1111/2041-210X.12456

Budge SM, Iverson SJ, Bowen WD, Ackman RG (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. doi:10.1139/F02-062

Budge SM, Iverson SJ, Koopman HN (2006) Studying trophic ecology in marine ecosystems using fatty acids: a primer on analysis and interpretation. Mar Mammal Sci 22:759–801. doi:10.1111/j.1748-7692.2006.00079.x

Chambellant M, Stirling I, Ferguson SH (2013) Temporal variation in western Hudson Bay ringed seal *Phoca hispida* diet in relation to environment. Mar Ecol Prog Ser 481:269–287. doi:10.3354/meps10134

Ghalanos A, Theussl S (2014) Rsolnp: general non-linear optimization using augmented Lagrange multiplier method. R package version 1:15

Haynes TB, Schmutz J, Bromaghin JF, Iverson SJ, Padula VM, Rosenberger AE (2015) Diet of yellow-billed loons (*Gavia adamsii*) in Arctic lakes during the nesting season inferred from fatty acid analysis. Polar Biol 38:1239–1247. doi:10.1007/s00300-015-1690-3

Iverson SJ, Field C, Bowen WD, Blanchard W (2004) Quantitative fatty acid signature analysis: a new method of estimating predator diets. Ecol Monogr 74:211–235. doi:10.1890/02-4105

Martín-Fernández JA, Palarea-Albaladejo J, Olea RA (2011) Dealing with zeros. In: Pawlowsky-Glahn V, Buccianto A (eds) Compositional data analysis: theory and application. John Wiley, Chichester, pp 43–58

Meynier L, Morel PCH, Chilvers BL, Mackenzie DDS, Duignan P (2010) Quantitative fatty acid signature analysis on New Zealand sea lions: model sensitivity and diet estimates. J Mammal 91:1484–1495. doi:10.1644/09-MAMM-A-299.1

Nordstrom CA, Wilson LJ, Iverson SJ, Tollit DJ (2008) Evaluating quantitative fatty acid signature analysis (QFASA) using harbour seals *Phoca vitulina richardsi* in captive feeding studies. Mar Ecol Prog Ser 360:245–263. doi:10.3354/meps07378

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roberts J, Lalas C (2015) Diet of New Zealand sea lions (*Phocarctos hookeri*) at their southern breeding limits. Polar Biol 38:1483–1491. doi:10.1007/s00300-015-1710-3

Rode KD, Regehr EV, Douglas DC, Durner G, Derocher AE, Thiemann GW, Budge SM (2014) Variation in the response of an arctic top predator experiencing habitat loss: feeding and reproductive ecology of two polar bear populations. Glob Change Biol 20:76–88. doi:10.1111/gcb.12339

Stewart C, Iverson S, Field C (2014) Testing for a change in diet using fatty acid signatures. Environ Ecol Stat 21:775–792. doi:10.1007/s10651-014-0280-9

Thiemann GW, Iverson SJ, Stirling I (2008) Polar bear diets and arctic marine food webs: insights from fatty acid analysis. Ecol Monogr 78:591–613. doi:10.1890/07-1050.1

Wang SW, Hollmén TE, Iverson SJ (2010) Validating quantitative fatty acid signature analysis to estimate diets of spectacled and Steller's eiders (*Somateria fischeri* and *Polysticta stelleri*). J Comp Physiol B 180:125–139. doi:10.1007/s00360-009-0393-x

Wood CC, McKinnell S, Mulligan TJ, Fournier DA (1987) Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Can J Fish Aquat Sci 44:866–881. doi:10.1139/f87-105

Zeppelin TK, Johnson DS, Kuhn CE, Iverson SJ, Ream RR (2015) Stable isotope models predict foraging habitat of northern fur seals (*Callorhinus ursinus*) in Alaska. PLoS One 10:e0127615. doi:10.1371/journal.pone.0127615