# Reply: Fatty acid signatures and classification trees: new tools for investigating the foraging ecology of seals[1]

Stephen J. Smith, Sara J. Iverson, and W.D. Bowen

It is unfortunate that the comment by Dr. Grahl-Nielsen is based on a misinterpretation of our use of classification and regression trees (CART), a misinterpretation of our discussion of their merits, and an erroneous application of principle components analysis (PCA) to our example data. Herein, we address the issues brought up in his comment.

CART has been advanced as a useful way of classifying data based on a number of variables. Indeed, as Grahl-Nielsen's comment acknowledges, CART's classification power has been demonstrated previously and has been found to perform as well as discriminant analysis and PCA, but with the added advantages of simplicity and ease of interpretation. The purpose of our paper (Smith et al. 1997) was to illustrate the features and value of CART in the analysis of fatty acid data and to compare these results with other types of multivariate analyses. For this purpose, we used milk samples collected throughout the lactation period from female harbour seals. As stated in our paper, since female harbour seals fast during the first 4–7 days of lactation, relying solely on blubber stores for milk production, followed by feeding trips (Boness et al. 1994), we hypothesized that we might see changes in milk fatty acids over time and therefore could use fatty acids as a tool to indicate changes in foraging patterns (e.g., Iverson 1993; Iverson and Oftedal 1995). We chose classification trees because unlike some of the standard multivariate methods (linear discriminant analysis, PCA, etc.), one can screen all 65 measured fatty acids with this method, even when their number exceeds the number of seals sampled. Grahl-Neilsen has taken issue with our approach and has proposed that PCA is applicable in this situation, does not suffer from needing more observations than variables, and provides a more meaningful analysis.

In our study, a PCA would proceed as follows. For each female seal in the study, we sample its milk and measure $p$ fatty acids. Given a sample of $n$ seals, we would have an $n \times p$ matrix of observations. A PCA would aim to reduce the $p$ dimension of this matrix to $q$ dimensions ($q << p$) via linear combinations (usually) such that combination $i$ explains the $i$th largest portion of the total response variance of the observations. In order to do this, $n \geq p$ and the covariance or correlation matrix of the observations is used. The general result is that we would have a small number of principal components that explain most of the variability in the data. Plotting the scores for the first principal component (which will be a linear combination of all fatty acids for each seal) against the second principal component for each animal, we would look for some natural clustering of points that may reflect postparturition periods. This method is exploratory at best and requires that we have more seals than fatty acids.

Grahl-Nielsen states that PCA can be used even when the number of fatty acids exceeds the number of observations. Legendre and Legendre (1998) listed this situation in their "Misuses of principal components" subsection. Application of PCA to the covariance or correlation matrix assumes that we have a statistically valid estimate of this matrix, which is only possible when $n \geq p$ (also see Ripley 1996, p. 289).

Grahl-Nielsen chooses to show that PCA would work for our data by first using the data in our table 1 (Smith et al. 1997) where mean fatty acid compositions for all seals within a time period postparturition are presented for each fatty acid. By using mean compositions, Grahl-Nielson ignores individual variation in fatty acid composition. Next, he applies a PCA to these data where it appears, given that individual fatty acids are labelled on his fig. 1, that the fatty acids are treated as observations and time periods are treated as variables (hence, in this case, $n > p$). Our original problem was, given a fatty acid signature at any one time, could we tell how many days postparturition the female's milk was collected. However, Grahl-Nielsen has twisted the problem to be that, given a vector of compositions over the time periods for a given fatty acid, we might be able to tell what fatty acid it was or what group it may belong to. That is, we do not seem to be working on the same problem. Furthermore, the problem analyzed by Grahl-Neilsen is of doubtful biological interest.

Grahl-Neilsen states that the classification tree method gives no information of where misclassified samples may actually belong. Given that we have to know the original classification of the seals to build the tree, then we do know where the misclassified samples actually belong. However, if his point was that because we only show the predicted class for each seal, we do not know how reliable the prediction is, then, again, he is incorrect because there is information available in CART that is analogous to distances in the

Received July 27, 1999. Accepted September 10, 1999.
J15266

**S.J. Smith[2] and W.D. Bowen.** Department of Fisheries and Oceans, P.O. Box 1006, Bedford Institute of Oceanography, Dartmouth, NS B2Y 4A2, Canada.
**S.J. Iverson.** Department of Biology, Dalhousie University, Halifax, NS B3H 4J1, Canada.

[1]Reply to comment by O. Grahl-Nielsen. 1999. Can. J. Fish. Aquat. Sci. **56**: 2219–2223.

[2]Author to whom all correspondence should be addressed.
e-mail: smithsj@mar.dfo-mpo.gc.ca

**Table 1.** Posterior probabilities from a pruned classification tree (Smith et al. 1997, fig. 4) for misclassified specimens.

| | Posterior probability | | | | Class | |
|---|---|---|---|---|---|---|
| Specimen | Day 0 | Days 4–7 | Days 12–14 | Days 19–21 | Predicted | Actual |
| M6MK14 | 0.00 | 0.00 | 0.40 | 0.60 | Days 19–21 | Days 12–14 |
| M7MK14 | 0.00 | 0.00 | 0.40 | 0.60 | Days 19–21 | Days 12–14 |
| M8MK07 | 0.00 | 0.18 | 0.82 | 0.00 | Days 12–14 | Days 4–7 |
| M9MKNB | 0.07 | 0.86 | 0.07 | 0.00 | Days 4–7 | Day 0 |
| M10MK04 | 0.93 | 0.07 | 0.00 | 0.00 | Day 0 | Days 4–7 |
| M12MK07 | 0.00 | 0.18 | 0.82 | 0.00 | Days 12–14 | Days 4–7 |
| M13MK14 | 0.07 | 0.86 | 0.07 | 0.00 | Days 4–7 | Days 12–14 |

**Note**: Predicted class is given by the highest posterior probability.

multivariate techniques listed by Grahl-Neilson. For each seal, the classification tree model predicts the probability that the seal is a member of each of the classes (or terminal nodes). The class with the largest probability is assigned as the predicted class. Consider these probabilities for the seven misclassified samples from the harbour seal milk data using the pruned tree in fig. 4 of Smith et al. (1997). The model prediction for the first two specimens appears to be less certain than for the remaining five and may suggest that the former group of specimens may be transitional in some sense (Table 1). While the actual class is known for these specimens, this kind of information would be useful for judging the certainty of predictions even when the class is unknown.

The rest of Grahl-Nielsen's comment dwells on how much better his analysis is and on how we claimed things that were completely unfounded. He states that we tried to demonstrate the superiority of CART over PCA. We did no such thing. We only suggested that CART was an additional tool with some attractive features for these kinds of data. He then goes on to state that CART is not an appropriate tool in studies of foraging ecology because it cannot be used to predict the composition of diets. Unfortunately, Grahl-Nielsen seems not to appreciate that there are several ways of studying foraging ecology. One way is to ask if fatty acid signatures of one group of animals differ from those of another in time or space. CART seems to be quite a valuable tool in answering this type of question (Iverson et al. 1997*a*, 1997*b*; Smith et al. 1997). However, contrary to Grahl-Nielsen's claim, we have never suggested that a CART analysis of predator fatty acids could be used to estimate the composition of a predator's diet. Neither CART nor PCA could be used for such a purpose. To suggest this seems quite naive. The problem of quantitatively relating predator fatty acids to diet is a far more complex issue. Currently, research is underway at Dalhousie University on the development of statistical models that address this more complex problem (e.g., Iverson et al. 1999).

Grahl-Nielsen goes on to quite erroneously imply that we concluded based on selected fatty acids that the seals used in our test data were eating northern sand lance. We made no such conclusion. We did point out that northern sand lance is the major prey of lactating females at Sable Island. This conclusion is based on lavage samples of the females' stomachs during the breeding season, as we stated in the paper. We then only observed, as a point of discussion, that the increase in several fatty acids known to be abundant in northern sand lance "may reflect...northern sandlance" in the diet. Perhaps Grahl-Nielsen did not understand this point in our paper. The same was true in discussing various aspects of fatty acid changes in Antarctic fur seals (Iverson et al. 1997*a*).

Finally, Grahl-Nielsen ends by pointing out that controlled feeding studies on captive seals will be necessary to validate the use of fatty acid signatures in estimating the species composition of diets. Certainly this is true. Perhaps this statement is with the recognition that his captive study (Grahl-Nielsen and Mjaavatten 1991) was inconclusive. This study reported on sampling of the outermost 1–2 cm (an area least affected by diet) of blubber from only one grey seal pup and one harbour seal fed unknown rations and analyzing only four homogenates of herring (12 individuals) and four individual mackerel throughout the course of the 7-month study, thus precluding any assessment of true diet effects or variability in seals or their prey. Thus, well-designed captive feeding studies are clearly needed. Such experiments are currently underway in our laboratory (S.J. Iverson and W.D. Bowen) as well as in others. After all of this, classification trees and PCA remain exactly as they were: different and complementary means for classifying and differentiating samples based on numerous variables. Beyond this, it is difficult to understand the point of Dr. Grahl-Nielsen's comment.

## References

Boness, D.J., Bowen, W.D., and Oftedal, O.T. 1994. Evidence of a maternal foraging cycle resembling that of otariid seals in a small phocid, the harbor seal. Sociobiology, **34**: 95–104.

Grahl-Nielsen, O., and Mjaavatten, O. 1991. Dietary influence on fatty acid composition of blubber fat of seals as determined by biopsy: a multivariate approach. Mar. Biol. **110**: 59–64.

Iverson, S.J. 1993. Milk secretion in marine mammals in relation to foraging: can milk fatty acids predict diet? Symp. Zool. Soc. Lond. **66**: 263–291.

Iverson, S.J., and Oftedal, O.T. 1995. Phylogenetic and ecological variation in the fatty acid composition of milks. *In* Handbook of milk composition. Academic Press, San Diego, Calif. pp. 789–827.

Iverson, S.J., Arnould, J.P.Y., and Boyd, I.L. 1997*a*. Milk fatty acid signatures indicate both major and minor shifts in the diet of lactating Antarctic fur seals. Can. J. Zool. **75**: 188–197.

Iverson, S.J., Frost, K.J., and Lowry, L.F. 1997*b*. Fatty acids signatures reveal fine scale structure of foraging distribution of harbor seals and their prey in Prince William Sound, Alaska. Mar. Ecol. Prog. Ser. **151**: 255–271.

Iverson, S.J., Frost, K., and Lang, S. 1999. The use of fatty acid signatures to investigate foraging ecology and food webs in Prince Wil-

liam Sound, Alaska: harbor seals and their prey. *In* Monitoring, habitat use, and trophic interactions of harbor seals in Prince William Sound, Alaska. Annual Report to the Exxon Valdez Oil Spill Trustee Council. Restoration Study 98064. Alaska Department of Fish and Game, Division of Wildlife Conservation, Fairbanks, Alaska. pp. 39–99.

Legendre, P., and Legendre, L. 1998. Numerical ecology. 2nd English ed.. Elsevier, Amsterdam, The Netherlands.

Ripley, B.D. 1996. Pattern recognition and neural networks. Cambridge University Press, Cambridge, U.K.

Smith, S.J., Iverson, S.J., and Bowen, W.D. 1997. Fatty acid signatures and classification trees: new tools for investigating the foraging ecology of seals. Can. J. Fish. Aquat. Sci. **54**: 1377–1386.