MARINE MAMMAL SCIENCE, 11(1):85–93 (January 1995)

STATISTICAL CLASSIFICATION OF DIVING BEHAVIOR

Diving behavior has been studied directly by observing animals, and indirectly by analyzing data that represents the behavior (*e.g.,* dive depths). Most recent studies utilizing data from time-depth recorders (TDRs) have primarily grouped diving behavior subjectively according to perceived similarities in the maximum depth, duration, and general appearance of the dive profile (depth *vs.* time) (Kooyman 1968, Le Boeuf *et al.* 1988, DeLong and Stewart 1991, Goebel *et al.* 1991, Hindell *et al.* 1991, Bengtson and Stewart 1992). The need for more efficient and objective types of analyses has developed due to the enormous amount of data recorded by these devices. Statistical analyses of behavior may be useful in expediting the analysis of large data sets, providing "on board" classification where data compression is necessary (*e.g.,* satellite transmissions), and reducing human subjective bias in interpreting diving behavior.

Weddell seals (*Leptonychotes weddellii*) are a good model on which to test multivariate statistical techniques because large amounts of dive data have been collected and their diving behavior is relatively well known and diverse (Kooyman 1968, 1975, 1981; Kooyman *et al.* 1983; Testa *et al.* 1989; Castellini *et al.* 1992; Testa, in press). Weddell seal dives were originally classified into three types depending on the maximum depth and duration of the dives (Kooyman 1968). In this paper an approach to statistically classify large data sets of Weddell seal dives is described.

Three different types of multivariate techniques used in grouping observations were tested to see which would be the most appropriate for grouping Weddell seal dives: principal component analysis, discriminant function analysis, and cluster analysis. Principal component analysis examines relationships among several quantitative variables and derives linear combinations (principal components) of these variables that retain as much of the information in the original

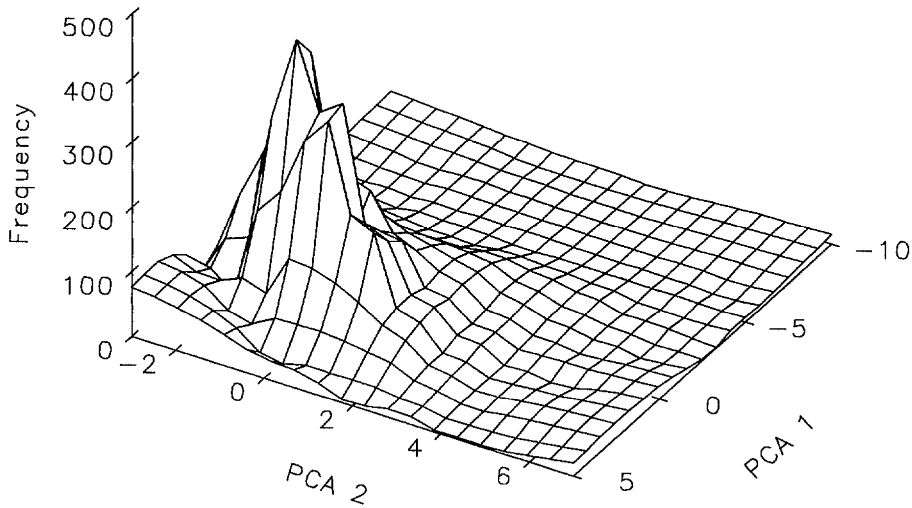

Figure 1. Principal component analysis performed on a subsample of smoothed dive profiles. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

variables as possible (PRINCOMP Procedure, SAS Institute Inc. 1990). Plotting the first two principal components, which may account for much of the variability, can show grouping of observations. Discriminant function analysis computes various discriminant functions (mathematical rules) for classifying observations (DISCRIM Procedure, SAS Institute Inc. 1990), but some observations must be classified previously for the functions to be created. This analysis is unable to classify observations that have not been at least partially grouped. Discriminant function analysis, however, can be used to test the validity of groups created by other techniques (see below). Finally, cluster analysis (specifically disjoint cluster analysis based on Euclidean distances computed from quantitative variables) classifies observations into clusters such that every observation belongs to only one cluster (FASTCLUS Procedure, SAS Institute Inc. 1990). This technique can also suggest the appropriate number of categories for a data set.

A total of 59,394 dives were recorded from 15 adult female Weddell seals during the summer of 1986 and the overwinter periods of 1990 and 1991 using TDRs (Wildlife Computers, Woodinville, WA, USA) in McMurdo Sound, Antarctica. Dives with a maximum depth of less than 50 m and a duration of less than 10 min were excluded because the sampling rate of the TDRs (every 60 sec) provided insufficient representation of the shorter dives. All dive records were pooled across the 15 individuals and the three years, and after the shorter dives were excluded, 39,119 dives were used for the analyses. Depth values for each dive (a series of depths over time starting and ending with depth equal to zero) were interpolated so that each dive was represented by 100 depths, allowing corresponding depths among all dives to be compared. One hundred depths were used because all dive durations were considerably less than 100 min

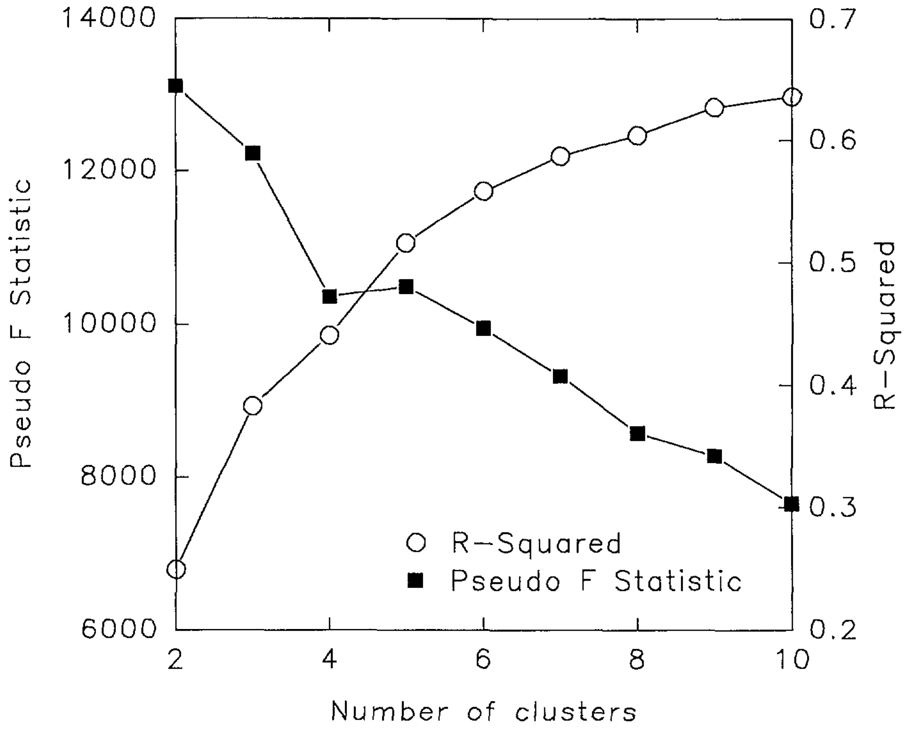

Figure 2. Overall R^2 and Pseudo *F* Statistic values produced by cluster analysis of depth data. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

(recorded every 60 sec), leaving the general dive shape unchanged. Dives were then standardized such that the maximum depth was one, and the rest of the depths scaled less than one. The means for every 10 depths, for a total of 10 means, were calculated for data reduction and to smooth the dive profiles. These 10 means were used in the analyses. Thus, all dives were compared regardless of their maximum depths, enabling comparison of the dive shapes.

Cluster analysis (FASTCLUS Procedure, SAS Institute Inc. 1990) was used to assess the number of dive types and to categorize the dives. Cross-validation error rates produced by discriminant function analysis (DISCRIM Procedure, SAS Institute Inc. 1990) and subjective observations were used to check the validity of the categorization.

Cluster analysis was more appropriate than principal component analysis for classifying Weddell seal diving behavior. The first two components of the principal component analysis accounted for only 60% of the total variance. Plotting the first two components suggested only two or three groups for Weddell seal dives and the groups did not resolve into distinct clusters (Fig. 1). It is possible that the range of dive types is continuous and that it is difficult or impossible to separate these dives with artificial boundaries. Principal component

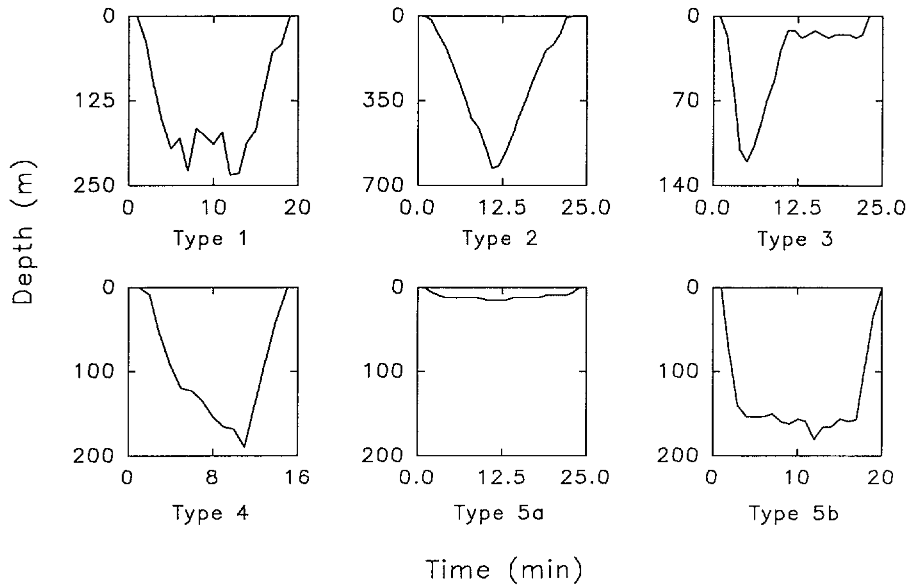

Figure 3. Examples of six types of dive profiles determined by cluster analysis. Dive profiles are plotted from time-depth recorder data recorded every 60 sec. Dive types 5a and 5b were determined from a second cluster analysis performed on the original type 5 dives. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

analysis may have been less successful than cluster analysis in separating groups because it is most sensitive to the direction of the greatest variance through the multidimensional space which may or may not be useful in separating groups.

Disjoint cluster analysis specifically looks for groups by finding cluster seeds that are farthest from each other in space and is more suited for finding clusters. Plotting the overall R^2 and Pseudo F Statistic values from cluster analyses (FASTCLUS Procedure, SAS Institute Inc. 1990) *vs.* number of clusters (Fig. 2) indicated five types of dives (Fig. 3) was an appropriate number for the data set. The R^2 values increased rapidly with increasing number of clusters until five clusters, when the rate of change decreased. The Pseudo F Statistic is an index used as a stopping rule for determining the most appropriate number of clusters for a data set (Calinski and Harabasz 1974, Milligan and Cooper 1985). Pseudo F Statistic values decreased with increasing numbers of clusters, but at five clusters a local increase was seen (Fig. 2). This suggests that five clusters is an appropriate number for the data set. Using these criteria, 37% of the dives were type 1, 21% were type 2, 14% were type 3, 12% were type 4, and 16% were type 5. A plot of the first three canonical variables (a dimension-reduction technique) showed overall clumping for all dives, but also that the groups are discrete (Fig. 4). Canonical variables are linear combinations of quantitative variables related to principal components that summarize between-class variation (CANDISC Procedure, SAS Institute Inc. 1990).

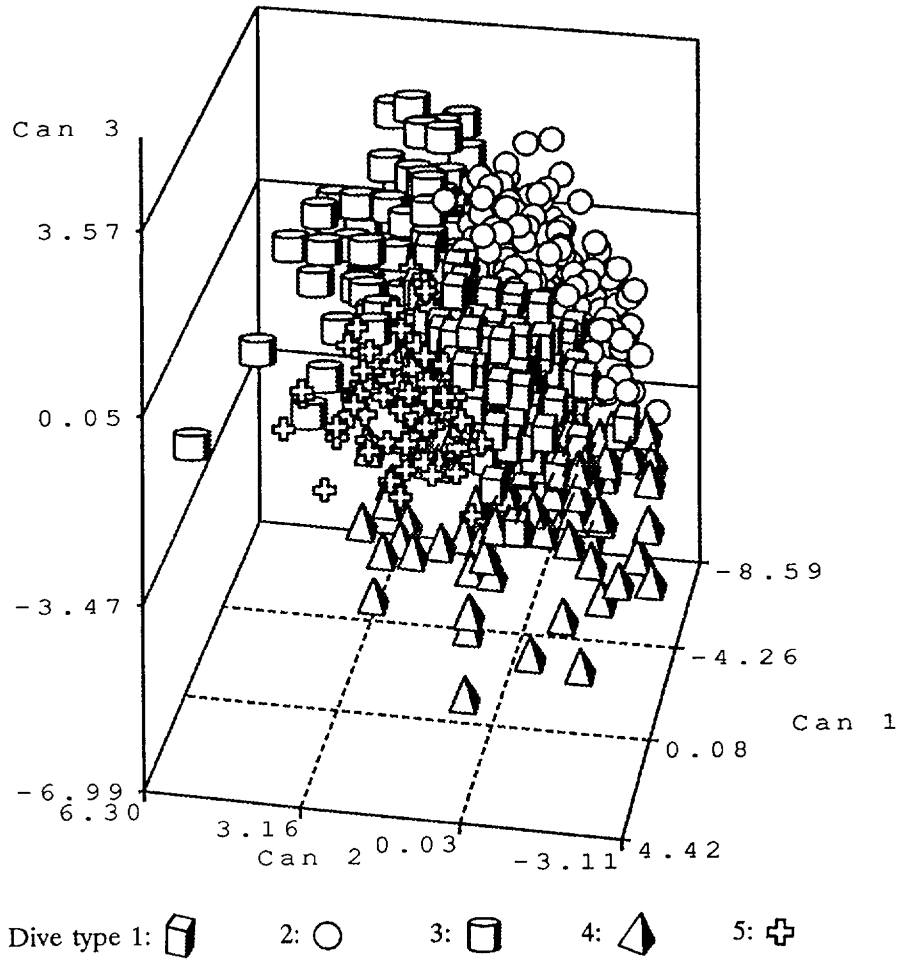

Figure 4. Canonical discriminant analysis performed on a subsample of smoothed dive profiles. The five dive types shown were obtained from cluster analysis. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

Cross-validation error rates produced by discriminant function analysis (two nearest neighbors) showed that 90% of the dives were classified correctly suggesting that five dive types is an adequate number. Cross-validation error rates are the percentage of observations misclassified using discriminant functions created while excluding the observation being classified (DISCRIM Procedure, SAS Institute Inc. 1990). A misclassified dive was usually placed in a similar group that was nearest to it in the 10 dimensional space (10 means). A plot of cross-validation error rates *vs.* number of clusters (Fig. 5) indicated that generally, as the number of groups increased, error rate also increased. At five

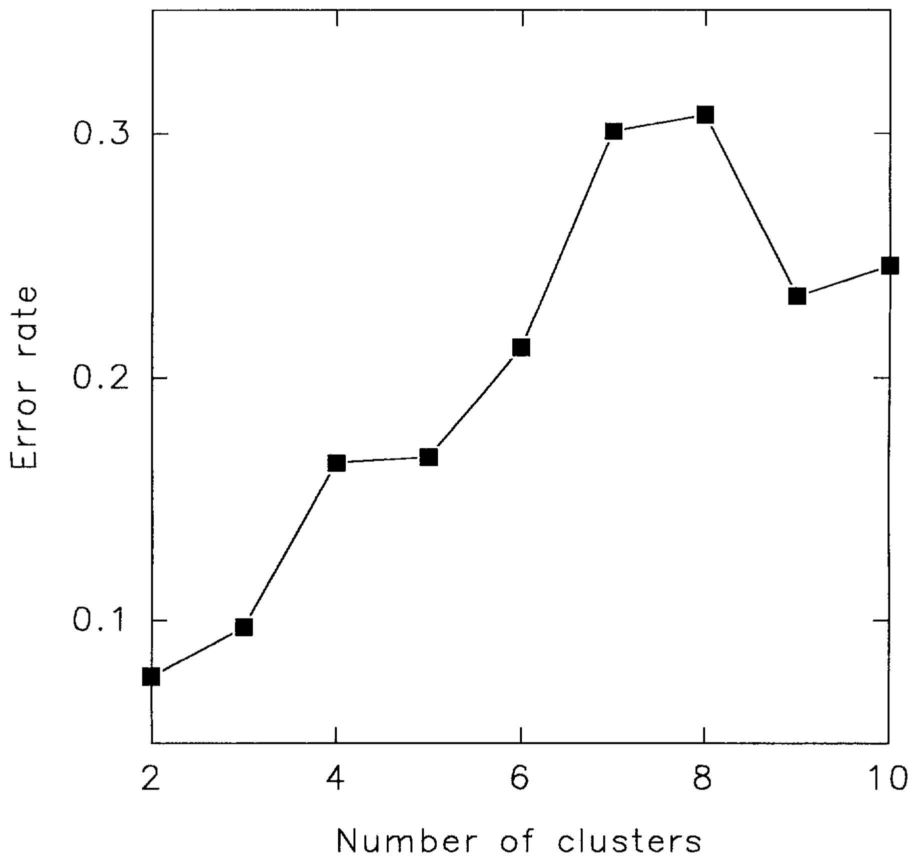

Figure 5. Error rates produced by discriminant function analysis with cross-validation performed on a subsample of smoothed dive profiles. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

groups, however, almost no change occurred indicating that five groups caused no increase in error rate and fit the data well.

Histograms for maximum depth, duration, and the quotient of maximum depth/duration were calculated for each dive type to determine if any further categorization could be accomplished. This also enabled information (maximum depth and duration) lost during the standardization to be reintroduced to the analyses. The histograms for dive types 1 through 4 all showed unimodal distributions indicating that no further categorization was warranted using these variables. The histogram of maximum depth/duration for dive type 5, however, showed two peaks indicating that the data could be divided into two groups (Fig. 6). The quotient of maximum depth/duration enabled the average depths of the dives to be compared. Cluster analysis was performed on the type 5 dives using the quotient of maximum depth/duration to split the dives into two groups. The first group (type 5a) included dives with long durations relative to

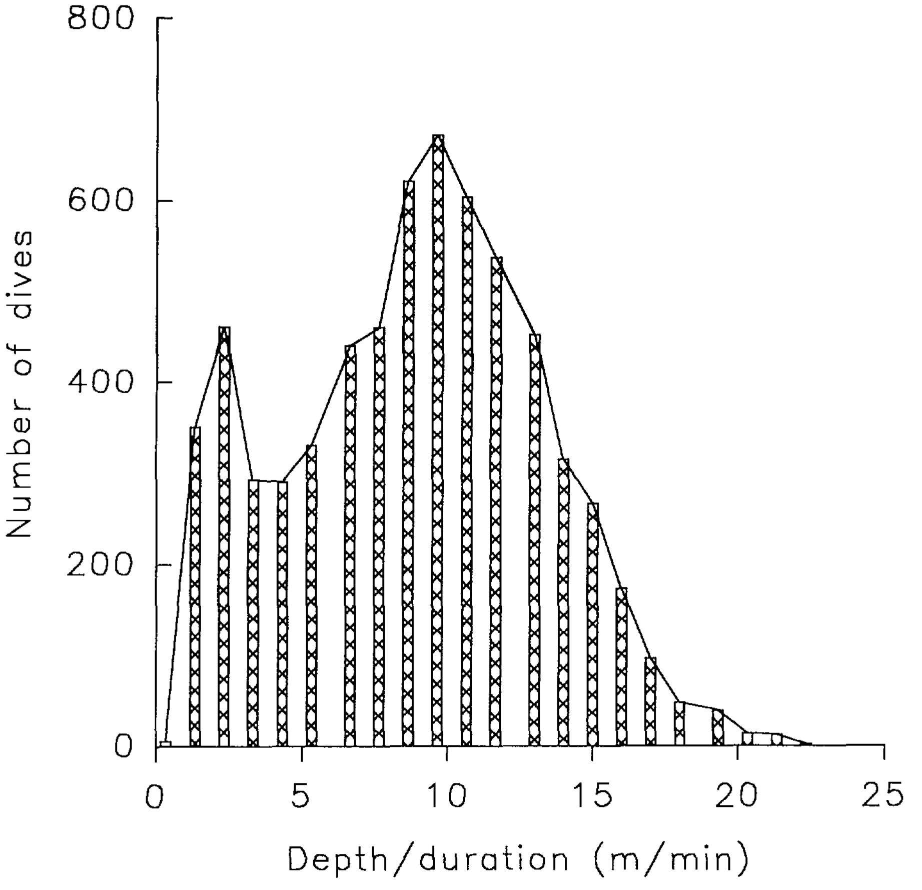

Figure 6. Frequency histogram of maximum depth/duration for type 5 dives. Time-depth recorder data collected from 15 adult female Weddell seals in McMurdo Sound, Antarctica during 1986, 1990, and 1991.

maximum depth and the second group (type 5b) included dives with short durations relative to maximum depth. Subjective comparison of dive categories produced by these statistical tests to actual dive profiles indicated dives were split into legitimate groups.

The results from the analyses of Weddell seal dive data indicate that it is possible to categorize objectively diving behavior using multivariate statistical tests. These analyses, readily available on statistical software packages, could be used on a range of different behaviors in various marine animals and certainly not limited to Weddell seals. The categories produced by the first cluster analysis seemed valid, but other categories that may represent important behavior were missing (*e.g.*, shallow dives with long durations). These types of dives may have been lost when the data were standardized. To find more of the possible categories of dives, additional variables (*e.g.*, maximum depth and duration) were introduced to help separate the dives. This type of multistage cluster analysis first categorizes diving behavior with one set of variables and then further splits those categories with another set of variables. The categories created by the analyses, however, were determined by indirect observations of behavior (*e.g.*, depths). Future studies may entail comparing the dive types determined by statistical analyses to directly observed behavior (*e.g.*, using video cameras) or conducting similar analyses with additional variables (*e.g.*, velocity and jaw movements). The classification of diving behavior is not an end in itself, but rather a tool that can be used to better understand this behavior. The possible functions of the dive types determined by these analyses will be presented elsewhere (Schreer and Testa, in review). In conclusion, the results demonstrate that it is possible to categorize objectively diving behavior using statistical techniques, but also that new ideas and information are needed to verify these categories, to look for new ones, and to understand the role of these behaviors in seal ecology.

Acknowledgments

This work was supported by grants DPP-8816567 and DPP-9119885 from the National Science Foundation. Thanks to the U.S. Antarctic program for their logistic support and to the many field members for their technical support. Special thanks to Ronald Barry for his vast statistical knowledge and assistance. The manuscript also benefitted from reviews by Tania Zenteno-Savin, Michael Castellini, Brian Fadely, Lorrie Rea, Erich Follmann, Ronald Barry, Kristen Schreer, and two anonymous reviewers. Thanks XOQ.

Literature Cited

Bengtson, J. L., and B. S. Stewart. 1992. Diving and haulout behavior of crabeater seals in the Weddell Sea, Antarctica, during March 1986. Polar Biology 12:635–644.

Calinski, T., and J. Harabasz. 1974. A dendrite method for cluster analysis. Communications in Statistics 3:1–27.

Castellini, M. A., R. W. Davis and G. L. Kooyman. 1992. Annual cycles of diving behavior and ecology of the Weddell seal. Bulletin of the Scripps Institution of Oceanography. Volume 28. University of California Press, Berkeley, CA.

DeLong, R. L., and B. S. Stewart. 1991. Diving patterns of northern elephant seal bulls. Marine Mammal Science 7:369–384.

Goebel, M. E., J. L. Bengtson, R. L. DeLong, R. L. Gentry and T. R. Loughlin. 1991. Diving patterns and foraging locations of female northern seals. Fishery Bulletin 89:171–179.

Hindell, M. A., D. J. Slip and H. R. Burton. 1991. The diving behavior of adult male and female southern elephant seals, *Mirounga leonina* (Pinnipedia: Phocidae). Australian Journal of Zoology 39:595–619.

Kooyman, G. L. 1968. An analysis of some behavioral and physiological characteristics related to diving in the Weddell seal. Pages 227–261 *in* G. A. Llano and W. L. Schmitt, eds. Biology of the antarctic seas. Antarctic Research Series, Volume 3. American Geophysical Union, Washington, D.C.

Kooyman, G. L. 1975. A comparison between day and night diving in the Weddell seal. Journal of Mammalogy 56:563–574.

Kooyman, G. L. 1981. Weddell seal: consummate diver. Cambridge University Press, Cambridge.

Kooyman, G. L., M. A. Castellini, R. W. Davis and R. A. Maue. 1983. Aerobic diving limits of immature Weddell seals. Journal of Comparative Physiology B 151: 171–174.

LE BOEUF, B. J., D. P. COSTA, A. C. HUNTLEY AND S. D. FELDKAMP. 1988. Continuous, deep diving in female northern elephant seals, *Mirounga angustirostris*. Canadian Journal of Zoology 66:446–458.

MILLIGAN, G. W., AND M. C. COOPER. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50:159–179.

SAS INSTITUTE INC. 1990. SAS/STAT Users Guide, Version 6, Fourth Edition, Volume 1 and 2. SAS Institute Inc., Cary, NC.

TESTA, J. W. In press. Overwinter movements and diving behavior of female Weddell seals (*Leptonychotes weddellii*) in the SW Ross Sea, Antarctica. Canadian Journal of Zoology.

TESTA, J. W., S. E. B. HILL AND D. B. SINIFF. 1989. Diving behavior and maternal investment in Weddell seals (*Leptonychotes weddellii*). Marine Mammal Science 5:399–405.

JASON F. SCHREER[1] AND J. WARD TESTA,[2] Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska 99775-1080. [1] Present address: Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. [2] Present Address: Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, Alaska 99518. Received 25 January 1994. Accepted 23 July 1994.