MARINE MAMMAL SCIENCE, 21(2):243–259 (April 2005)

MULTIVARIATE REGRESSION OF SATELLITE-LINKED DIVE RECORDER DATA: SIMULTANEOUS ANALYSIS OF ALL BINS

M. A. SIMPKINS[1]
K. L. LAIDRE[2]
National Marine Mammal Laboratory, NOAA/Alaska Fisheries Science Center,
7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.

P. J. HEAGERTY
Department of Biostatistics, University of Washington, Box 357232
Seattle, Washington 98195, U.S.A.

ABSTRACT

Statistical analysis of diving behavior data collected from satellite-linked dive recorders (SDRs) can be challenging because: (1) the data are binned into several depth and time categories, (2) the data from individual animals are often temporally autocorrelated, (3) random variation between individuals is common, and (4) the number of dives can be correlated among depth bins. Previous analyses often have ignored one or more of these statistical issues. In addition, previous SDR studies have focused on univariate analyses of index variables, rather than multivariate analyses of data from all depth bins. We describe multivariate analysis of SDR data using generalized estimating equations (GEE) and demonstrate the method using SDR data from harbor seals (*Phoca vitulina*) monitored in Prince William Sound, Alaska between 1992 and 1997. Multivariate regression provides greater opportunities for scientific inference than univariate methods, particularly in terms of depth resolution. In addition, empirical variance estimation makes GEE models somewhat easier to implement than other techniques that explicitly model all of the relevant components of variance. However, valid use of empirical variance estimation requires an adequate sample size of individual animals.

Key words: harbor seal, *Phoca vitulina*, satellite telemetry, diving behavior, multivariate regression, generalized estimating equations.

Behavioral studies of diving marine animals are complicated because it is difficult to observe animals' behavior under water, particularly in the open ocean. To overcome this difficulty, researchers have turned to telemetry to collect information on the behavior of marine species. One of the most common telemetric instruments

[1] Current address: U. S. Marine Mammal Commission, 4340 East-West Highway, Suite 905, Bethesda, Maryland 20814, U.S.A.; e-mail: msimpkins@mmc.gov.
[2] Current address: Greenland Institute of Natural Resources, Box 570, 3900 Nuuk, Greenland.

deployed on diving marine animals is the satellite-linked dive recorder (SDR; Mate 1989, Stewart *et al.* 1989). These instruments collect large amounts of behavioral data from tagged animals freely swimming and diving at sea and have directly resulted in a breadth of information collected on marine animal movements (*e.g.*, Lowry *et al.* 1998, Mate *et al.* 1999), haul-out behavior (*e.g.*, Gjertz *et al.* 2001, Bengtson and Cameron 2004), diving and surfacing behavior (*e.g.*, Folkow and Blix 1999, Heide-Jørgensen *et al.* 2001), foraging ecology (*e.g.*, Laidre *et al.* 2003, Lea and Dubroca 2003), and habitat selection (*e.g.*, Lowry *et al.* 2000, Laidre *et al.* 2004). SDRs summarize diving behavior into categories or "bins" on a microprocessor to accommodate satellite bandwidth constraints and transmit binned data to polar-orbiting satellites when animals surface (Harris *et al.* 1990).

Statistical analysis of data collected from SDRs is inherently difficult. SDRs bin data from individual dives into depth and duration categories and tally the number of dives in each category for a series of sampling periods (usually 6 h each). The resulting categorical data are best understood as multivariate count or multinomial data, in which each dive of an animal can be assigned to a specific category or bin. SDR studies are longitudinal in nature, with instruments deployed on several individuals and data collected through time from each animal. Such longitudinal studies must account for random variation among individuals to make valid inference regarding the patterns of behavior for the entire population. Additionally, SDR data often are temporally autocorrelated, such that previous data values are predictive of future values (*i.e.*, current behavior depends on past behavior). This temporal autocorrelation impacts the estimated standard error of regression parameters and must be acknowledged either by explicit representation in the model, or through appropriate adjustment of standard errors. Although these statistical complexities are readily addressed individually, addressing all of these statistical problems in one multivariate analysis is complex.

Early SDR studies provided basic descriptions of diving behavior based on qualitative analyses (*e.g.*, Mate *et al.* 1995, Stewart *et al.* 1996). More recent studies have used quantitative techniques to analyze univariate data derived from SDR data. Several studies have used simple summary statistics to analyze mean dive depths, which were estimated from dive tallies in each depth bin (*e.g.*, Krutzikowsky and Mate 2000, Loughlin *et al.* 2003); other studies have conducted more thorough univariate analyses of index variables (Frost *et al.* 2001; Hastings *et al.* 2004). A few studies have looked at the data in a multivariate fashion by conducting analyses on data from each depth bin (*e.g.*, Burns *et al.* 1999, Laidre *et al.* 2003), or by using multivariate ANOVA techniques to test for differences across depth bins (*e.g.*, Nordøy *et al.* 1995, Teilmann *et al.* 1999). However, many SDR studies have ignored one or more of the statistical problems inherent in SDR data.

Modern regression techniques are readily available in common statistical packages and can be applied to analyze SDR data in a statistically valid manner. One common technique, mixed-effects regression, can be used to analyze longitudinal univariate data by explicitly modeling both the random variation between individuals and the correlation structure for repeated measurements within each individual's record in a linear (LME) or non-linear model (Breslow and Clayton 1993). Many questions about diving behavior can be addressed using simple univariate regressions within the mixed-effects framework, either modeling univariate data obtained directly from SDRs (maximum depth, duration or surface time), or by deriving an index variable from the binned SDR data to provide a univariate proxy. For example, many researchers estimate mean, median, or modal dive depths for each sample period and

Table 1. Three simple scenarios of dive distributions for male and female seals to illustrate the nature of multivariate questions. Multivariate questions ask whether the effect of a predictor variable changes across depth bins. In statistical terms, these questions ask whether there is a significant interaction between a predictor variable and depth bins. Scenario 1 does not suggest an interaction between sex and depth bin, but scenarios 2 and 3 do.

Scenario	Depth bin	Males (dives)	Females (dives)
1	1	6	3
	2	4	2
2	1	6	2
	2	4	8
3	1	6	1
	2	4	4

use those values as proxies for the multivariate binned data for each period (*e.g.*, Frost *et al.* 2001, Krutzikowsky and Mate 2000, Loughlin *et al.* 2003).

Some questions about diving behavior are not amenable to univariate regressions; these questions are explicitly multivariate in nature. These "multivariate" questions can be defined broadly as questions concerning changes in the distribution of dives among depth bins as a function of predictor variables. More specifically, multivariate questions ask whether the effect of a predictor variable changes across depth bins. In statistical terms, these questions ask whether there is a significant interaction between a predictor variable and depth bins. For example, consider the simple multivariate question, "Does the distribution of dives differ between males and females?" Table 1 shows example data for three possible scenarios based on the simplest case of only two depth bins. In scenario 1, males perform twice as many dives as females, but the distribution of dives is similar across depth bins for both sexes (3:2; bin 1:bin 2). An accurate statistical model for scenario 1 would include both sex and depth bin as predictors of the total number of dives but no interaction between sex and depth bin. In scenario 2, males and females perform the same total number of dives (10), but the distribution of dives differs across depth bins (3:2 *vs.* 1:4; bin 1:bin 2, males *vs.* females). A statistical model for scenario 2 would require an interaction between sex and depth bin, but there would be no effect of sex on total dives. In scenario 3, males and females differ both in the total number of dives (10:5; male:female) and in the distribution of dives across depth bins (3:2 *vs.* 1:4; bin 1:bin 2, males *vs.* females). A statistical model for scenario 3 would require an interaction between sex and depth bin, and there would be an effect of sex on total dives.

Here, we consider the more complex question of how diving depths change over a diel cycle. If the study animals prey upon organisms that undergo diel vertical migrations, more dives would be expected to occur at shallow depth bins at night and at deeper bins during the day (*e.g.*, Croxall *et al.* 1985, Croll *et al.* 1992, Le Boeuf *et al.* 1993). Although one could analyze the number of shallow and deep dives separately using several univariate regressions (*e.g.*, Laidre *et al.* 2003), such an approach ignores the correlation or interdependence between the numbers of dives to different depth bins. This interdependence occurs because each sample represents a fixed time span (usually 6 h) and each dive exhausts some of that time span. Therefore, a finite number of dives can occur within the sampling period, and the number of dives to one depth bin is influenced by the number of dives to other bins during the same period. In addition, deeper dives generally are longer in duration, thus one deep dive exhausts more of the sampling period than a shallower dive would (leaving less time available for other dives during the sampling period). Ignoring the interdependence between depth bins by fitting a univariate regression to data from each depth bin is not a palatable approach to multivariate questions. Multivariate, longitudinal, mixed-effects regression methods are not widely available (forcing researchers to use multiple univariate regressions and ignore interdependence; Laidre *et al.* 2003). Generalized estimating equations (GEE), however, can be used to analyze multivariate longitudinal data (Liang and Zeger 1993, Hanley *et al.* 2003). The GEE approach estimates the standard error of regression estimates empirically using the observed individual-to-individual variation within each sample group (in this case across animals), rather than explicitly modeling the various components of the variance (random variation between individuals, repeated measures correlation, and correlation between bins).

In this study, we introduce the use of GEE methods for multivariate analysis of SDR data. As a case study, we use data from a study of harbor seal (*Phoca vitulina*) diving behavior conducted in Prince William Sound, Alaska, between 1992 and 1997. These data were initially analyzed using univariate mixed-effects regressions of index variables (Frost *et al.* 2001; Hastings *et al.* 2004). We describe the functional limitations of both GEE and mixed-effects regression approaches and discuss strategies for focusing SDR statistical analyses on questions that can be answered using the two techniques.

Methods

In many cases, questions regarding diving behavior can be refined so that they can be answered using univariate statistical techniques (*e.g.*, Frost *et al.* 2001, Krutzikowsky and Mate 2000, Loughlin *et al.* 2003). Here, we focus on questions that require multivariate techniques.

Problems with Zero and Missing Values

SDR data often include bins with very few dives, resulting in bins with a large number of zero values. A preponderance of zero values can cause over-dispersion in regression models, invalidating inherent assumptions about the mean-variance relationship that is specified by generalized linear models (McCullagh and Nelder 1989). For example, the common Poisson regression model assumes that the variance is equal to the mean. However, heterogeneity among individuals will lead to a variance that is larger than the mean. This over-dispersion can be corrected in a variety of ways, including the use of negative-binomial or zero-inflated Poisson models, and *post-hoc* adjustments to standard error estimates (Huber 1967, White 1982). A more difficult problem to address is any potential pattern in missing values that may be associated with, or caused by, changes in the behavior of the monitored animals. For example, SDR data may be recorded less frequently when the monitored animals are diving more extensively and spending less time at the surface, where SDR transmissions can successfully reach orbiting satellites to be recorded. In this scenario, fewer data points would be received and recorded when animals were diving more, and analyses of these data would be biased, reflecting only behavior when animals were diving less frequently.

For our test case using harbor seal data, preliminary data exploration suggested that fewer data points were recorded during the winter when animals seemed to be diving more extensively (*i.e.*, more missing values during winter). To account for this, we included season as an explanatory variable in our models, so that seasonal changes in behavior would be evident in the analytical results when sufficient data were available in each season. If data were rare during one season, making statistical estimates unreliable, we would have considered eliminating that season from the analysis and focusing only on behaviors in seasons with better data coverage. With respect to zero values, data exploration indicated an obvious trend of proportionally more zeros recorded in deeper bins. In the last two bins (150–200 m and >200 m), zeros constituted >90% of the 6-h records. We pooled data from the deepest three bins into a single large bin, which included all dives >100 m and approximately 70% zero values.

Multivariate Analysis

Multivariate analysis can be conducted using a simplified GEE process. GEE methods provide valid inference for linear, Poisson, and logistic regression models when parameters are estimated using longitudinal data (Liang and Zeger 1993). There are two features that characterize GEE: a "working correlation" model that is used to weight the longitudinal vector of observations from each individual, and an empirical variance estimator (sometimes called a "robust" variance estimator) that provides a valid large-sample estimate of the variance-covariance matrix for the regression estimates even when an incorrect correlation model is specified. Based on this last feature, a generalized linear model (GLM) with Poisson error variance can be used to estimate parameter values and a working correlation model, which assumes complete independence among data points (clearly, incorrect for multivariate longitudinal data). The GEE empirical variance estimator can then be used to calculate the correct variance-covariance matrix (Appendix).

A backwards stepwise process can be used to select a parsimonious multivariate model, starting with the most complex model and removing predictor variables and interactions that are not significant in a stepwise fashion. GEE methods do not provide a maximum log-likelihood estimate, so model selection must be based on statistical tests of estimated regression coefficients rather than likelihood-based procedures such as likelihood-ratio tests or information summaries such as AIC (Akaike information criterion). For this analysis, we used a slight generalization of the standard one degree of freedom t-test to assess significance for groups of parameters:

$$t^2 = \beta' \mathrm{cov}(\beta)\beta \tag{1}$$

where: β represents the vector of q parameter estimates to be tested, cov(β) represents the variance-covariance matrix for β, and β' is the transpose of β. We considered the parameters, β, to be significant when the t^2-value was greater than 95% of a chi-squared distribution with q degrees of freedom (*i.e.*, $\alpha = 0.05$).

Valid statistical tests of parameter significance require the sample size of individuals to be large relative to the number of levels fitted for any given parameter or interaction between parameters (*e.g.*, a quadratic parameter would have two fitted levels). To ensure the validity of statistical tests, the hypotheses should be refined to minimize the number and complexity of parameters to be estimated before beginning a multivariate analysis.

For SDR data, the number of depth bins included in multivariate analysis can dramatically affect the complexity of fitted parameters, especially interactions between depth bin and other variables. To determine the appropriate number of depth bins to include in a multivariate analysis, model selection should initially incorporate analysis of the total dive count, or equivalently analysis based on pooling data across all bins. This model with all depth bins pooled can be used as a base model against which all models that include depth effects can be compared. A variety of options for pooling depth bins should be considered, ranging from one zone that includes all depth bins (the base model) to a maximal number of zones constrained to ensure that the number of estimated parameters for any given variable (including any interactions between depth zones and that variable) does not exceed the sample size of individuals.

Harbor Seal Case Study

The harbor seal data used in this study consisted of SDR records from seals collected in Prince William Sound, Alaska, between January 1992 and December 1997. The SDRs were programmed with consistent depth bins for all years, with six bins: 4–20 m, 20–50 m, 50–100 m, 100–150 m, 150–200 m, and deeper than 200 m. SDRs recorded the number of dives to each of these depth bins during four 6-h periods each day: 2100–0300, 0300–0900, 0900–1500, and 1500–2100 local solar time.

For the case study, we focused on the hypothesis that harbor seals exhibit diel changes in diving depths consistent with foraging on diel migrating prey. Thus, our primary interest was in the interaction between depth bin and time of day. We expected harbor seals to perform more shallow dives at night and more deep dives during the day. Based on previous analyses (Frost *et al.* 2001, Hastings *et al.* 2004), it was apparent that sex, age, and season could also affect harbor seal diving behavior, so we included those variables in our regression. As possible, we minimized the number of parameters estimated for each predictor variable. We modeled seasonal effects using four seasonal categories, rather than months or weeks. Seasons were defined as: prebreeding (1 April–14 May), breeding (15 May–31 July), postmolt (1 September–30 November), and winter (1 December–31 March), following Hastings *et al.* (2004). Reducing the number of depth bins to alleviate the problem of excessive zeros had the added effect of reducing the number of parameters estimated for bins and bin interactions with other variables. We also included sex, two age classes, and four time periods for each day.

Harbor seals dove primarily to the shallowest depth bin and dove less frequently to each sequentially deeper depth bin. We expected more variation in parameter estimates for shallower bins, which included more dives, than for deeper bins. Therefore, we chose zonal options to determine whether each shallow depth bin should be analyzed separately, or pooled with the deeper bins. We started the analysis by pooling all depth bins into one zone and analyzing total dives for each 6-h sample. We then re-analyzed the data with two depth zones: the shallowest depth bin (0–20 m) and all the deeper bins combined (>20 m). We continued this process, reanalyzing the data with three depth zones (4–20 m, 20–50 m, and >50 m) and with all four depth bins (recall that the three deepest bins were previously pooled to alleviate problems with excessive zeros in those bins).

For each zonal option, we first conducted a GLM regression including all four explanatory variables (age, sex, season, time of day) and depth zone interactions with each variable. We modeled the dive counts in each bin using a Poisson error model and accounted for any over-dispersion in the residuals by use of the empirical variance estimation with GEE (Appendix). Following each regression, we corrected the GLM standard errors and variance-covariance matrix using GEE empirical variance estimation (Appendix). We then conducted several statistical tests to determine if any variables should be removed from the model, using the generalized t-test described above (equation 1). For all zonal options, we decided which variables and/or depth zone interactions to retain in the model based on the significance of each variable in the model. During model selection, we removed individual variables, interactions, and depth zones that were not significant in the model until all remaining terms were significant in the final model.

For each zonal option, we conducted two tests for each independent variable, answering the following questions: (1) "Did the variable have a significant impact on the model on its own or through an interaction with depth zones?" (*e.g.*, test H_0: sex = sex*zone = 0, where * indicates an interaction and zone represents all depth zones in the model), and (2) "Did the variable's effect vary by depth zone?" (*i.e.*, was there a significant interaction with depth zone; *e.g.*, test H_0: sex*zone = 0). For each zonal option with more than one depth zone, we also tested whether the model was significantly different from a simpler model with one fewer depth zones. Specifically, we tested for differences between parameter estimates for the two depth zones that had been pooled in the simpler model. For example, in the third zonal option, the >50 m depth zone was a "new" zone that had been pooled with the 20–50 m zone in the second zonal option (>20 m). If parameter estimates for the new, >50 m, zone were not different from those for the 20–50-m zone, then we inferred that the two zones should remain pooled (*i.e.*, splitting the two zones was not justified and the second zonal option would be selected because it was more parsimonious than the third zonal option). Conversely, if parameter estimates for the new zone were different, then we inferred that the two zones should be split (*i.e.*, the third zonal option would be preferred over the second zonal option).

For example, for the third zonal option we tested whether parameter estimates (including all interactions between depth zones and other variables) were significantly different between 20–50 m and >50 m depth zones. To simplify analysis, we forced the model to fix parameter estimates for the 20–50-m zone at 0 (one level of any factor variable, such as depth zone, must be fixed during parameter estimation). We could then test whether the parameter estimates for the >50 m depth zone (zone3), including interactions, were different than 0 (test H_0: zone3 = zone3*β = 0, where β represents all independent variables in the model). We also tested whether interactions with the >50 m depth zone were different than 0 (test H_0: zone3*β = 0).

Results

SDR records were collected from 49 seals in Prince William Sound, Alaska, between January 1992 and December 1996 (Table 2). A total of 13,068 data records (*i.e.*, one data record from each 6-h period for each seal) were used in this analysis (Table 2). Each data record included counts of the number of dives to each depth bin, or depth zone, thus the sample size of counts used in analysis was a multiple of the number of data records (the multiple varied depending on the number of depth bins or zones in the analysis).

Table 2. Harbor seal data used in case study (Frost *et al.* 2001; Hastings *et al.* 2004). The sex, age, and year of capture are shown for each seal, along with the sample size (n) of 6-h histogram records. Each 6-h record included counts of the number of dives to each depth bin, or depth zone, thus the sample size of counts, c, used in analysis was a multiple of n (number of depth bins or zones in the analysis multiplied by n).

Seal ID	Year	Sex	Age	n
9201	1992	M	SUB	83
9202	1992	M	SUB	134
9203	1992	F	AD	80
9204	1992	M	SUB	111
9301	1993	M	AD	136
9302	1993	M	AD	263
9303	1993	M	AD	209
9304	1993	F	AD	120
9305	1993	M	AD	167
9306	1993	M	AD	159
9307	1993	F	AD	269
9308	1993	M	AD	769
9309	1993	M	SUB	408
9310	1993	M	AD	446
9311	1993	M	AD	416
9312	1993	M	AD	322
9401	1994	F	SUB	307
9402	1994	M	AD	650
9403	1994	M	AD	496
9404	1994	M	SUB	296
9405	1994	F	AD	271
9406	1994	F	AD	138
9407	1994	F	SUB	229
9408	1994	F	AD	250
9501	1995	M	SUB	158
9502	1995	F	SUB	204
9503	1995	F	AD	187
9504	1995	M	AD	257
9505	1995	F	SUB	152
9506	1995	F	AD	174
9507	1995	F	SUB	322
9509	1995	F	AD	452
9510	1995	F	AD	562
9511	1995	M	SUB	223
9512	1995	F	AD	425
9513	1995	F	SUB	27
9601	1996	F	SUB	118
9602	1996	F	SUB	158
9603	1996	F	SUB	123
9604	1996	M	SUB	57
9605	1996	F	AD	207
9606	1996	F	AD	204
9607	1996	F	SUB	127
9609	1996	F	AD	811
9610	1996	M	SUB	227
9611	1996	M	SUB	209
9612	1996	M	SUB	150
9613	1996	F	AD	588
9614	1996	F	SUB	217

Table 3. Parameter estimates are shown for the final multivariate model. Empirical standard errors (SE) are shown for each estimate along with standard errors estimated by a generalized linear model (GLM), which assumed complete independence between data points. Note that one level of each factor variable was fixed during parameter estimation; age = adult, depth bin = bin 3, sex = female, and time of day = midnight were all fixed at 0 (and are not shown in the table). Depth bins were coded as: bin 1 = 4–20 m, bin 2 = 20–50 m, bin 3 = 50–100 m, and bin 4 > 100 m. Time of day was coded as: midnight = 2100–0300, morning = 0300–0900, midday = 0900–1500, and evening = 1500–2100 local solar time. An asterisk between two variables indicates an interaction term.

Parameter	Estimate	GLM SE	Empirical SE
(Intercept)	1.493	0.008	0.376
Subadult	0.348	0.003	0.073
Depth bin 1	2.141	0.008	0.453
Depth bin 2	0.583	0.010	0.386
Depth bin 4	−0.874	0.013	0.398
Male	0.559	0.007	0.325
Morning	0.014	0.009	0.161
Midday	−0.029	0.010	0.252
Evening	−0.009	0.009	0.105
Depth bin 1*Male	−0.959	0.008	0.421
Depth bin 2*Male	−0.160	0.009	0.356
Depth bin 4*Male	0.391	0.011	0.352
Depth bin 1*Morning	−0.438	0.010	0.193
Depth bin 2*Morning	−0.338	0.012	0.163
Depth bin 4*Morning	0.219	0.015	0.196
Depth bin 1*Midday	−0.500	0.011	0.301
Depth bin 2*Midday	−0.670	0.013	0.258
Depth bin 4*Midday	0.814	0.015	0.293
Depth bin 1*Evening	−0.267	0.010	0.123
Depth bin 2*Evening	−0.248	0.012	0.120
Depth bin 4*Evening	0.180	0.015	0.122

Three of the four parameters of interest (age, sex, and time of day) were significant in the final model and the best model included all four depth bins (Table 3). Depth interactions occurred with sex and time of day, but were not present with age. Harbor seals consistently performed more dives to the shallowest bin than any other depth bin. Sub-adult seals consistently dove more frequently to all depth bins than adults (Fig. 1). Female seals performed more dives to the shallowest depth bin than males (37.9 dives for females *vs.* 25.4 dives for males for adults in period 0, Fig. 2). However, males made more dives to the other depth bins than females (differences on the order of 2.4–3.8 dives for adults in period 0). Standard errors estimated empirically using GEE were one to two orders of magnitude larger (8 to 55 times larger) than the naive standard errors estimated by GLM (Table 3).

There was a significant interaction between time of day and depth bin consistent with the hypothesis that harbor seals exhibit diel differences in diving behavior (Fig. 3, Table 3). Seals performed more dives to the shallowest bin around midnight (37.9 dives *vs.* 22.3 dives at midday for adult females) and more dives to the deepest bin around midday (4.1 dives *vs.* 1.9 dives at midnight for adult females). Although the parameter estimates were statistically different, the effect was small, particularly in the changes in the number of dives to the deepest bin.

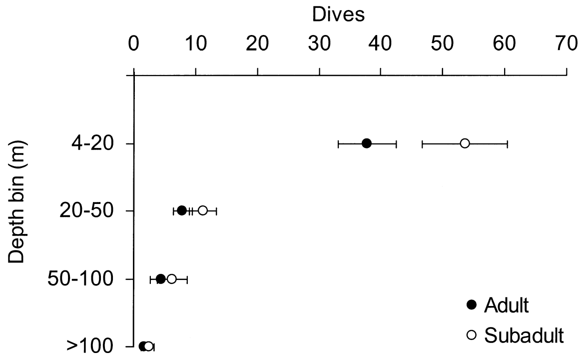

Figure 1. Number of dives predicted for adult and subadult harbor seals based on multivariate GEE regression (standardized to female seals at midnight). Error bars represent ±1 SE based on GEE empirical variance estimation.

Discussion

Multivariate analysis of SDR data provides more thorough, detailed information about the diving behavior of tagged animals than can be obtained from univariate analyses. In particular, multivariate analysis of bin counts investigates whether the distribution of dives varies across individuals or environmental conditions characterized by the predictor variables. Multivariate information can be used to describe aspects of behavior that otherwise would require several univariate analyses. For example, Frost *et al.* (2001) used three univariate analyses of index variables to describe harbor seal diving behavior: effort (time spent wet per 6-h period), diving focus (a measure of the dominance of one depth bin in the dive-tally histograms for each 6-h period), and preferred diving depth (the modal depth bin in the dive-tally histograms for 6-h periods with "focused" diving). The multivariate analysis presented here provides information about changes in the number of dives to all depth bins, rather than simply changes in the modal depth bin. Further, the entire distribution of dives across depth bins can be examined with multivariate analyses, rather than simply measuring changes in the dominance of the modal

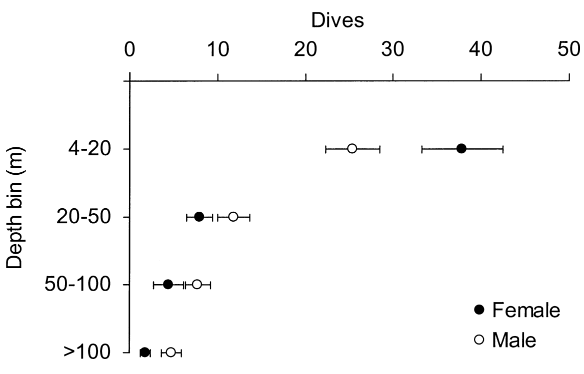

Figure 2. Number of dives predicted for female and male harbor seals based on multivariate GEE regression (standardized to adult seals at midnight). Error bars represent ±1 SE based on GEE empirical variance estimation.

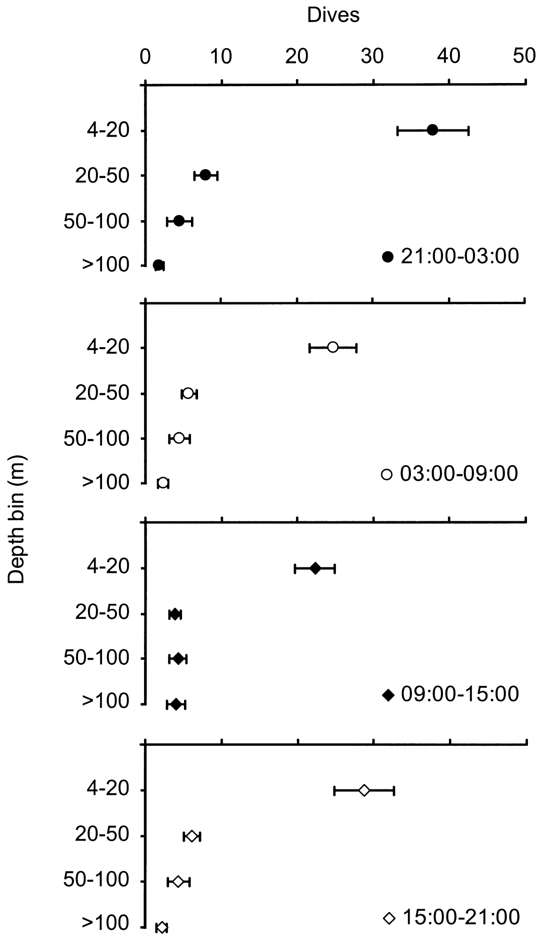

Figure 3. Number of dives predicted for harbor seals at four different times of day (6-h periods shown on panels) based on multivariate GEE regression (standardized to adult female seals). Error bars represent ±1 SE based on GEE empirical variance estimation.

depth bin (diving focus). Finally, diving effort to all depth bins can be approximated by the number of dives to each bin, though repeating the analysis with time-at-depth data would provide a more accurate description of diving effort.

Previous univariate analyses of harbor seal diving behavior modeled variation in modal, or preferred, depth bin (Frost *et al.* 2001, Hastings *et al.* 2004). Those analyses, however, could not determine the cause for changes in modal depth bins. For example, Hastings *et al.* (2004) found evidence that modal depth was shallower at midnight during pre-breeding and breeding seasons. Hastings *et al.* (in press), however, could not determine if this change was caused by seals performing more shallow dives, fewer deep dives, or both at midnight. Our multivariate analysis indicated that seals performed more shallow dives at midnight than during other periods, without a substantial decrease in the number of deeper dives. The multivariate analysis also indicated that subadults dove more frequently to all depth bins than adults, which was reflected in lower diving focus for subadults than adults in previous univariate analyses (Hastings *et al.* 2004). The multivariate analysis confirmed an interaction between sex and depth bin, with female seals performing more dives to the shallowest depth bin and fewer dives to all other bins than males. This relationship was reflected in higher diving focus in females (Hastings *et al.* 2004; and for only adult seals in Frost *et al.* 2001) because the distribution of female dives was skewed toward the shallowest bin compared to male dives. Previous univariate analyses did not find a significant relationship between modal depth bin and sex (Frost *et al.* 2001, Hastings *et al.* 2004). Multivariate analysis did not find a seasonal effect on diving behavior, which was found in both previous univariate analyses. Inconsistencies between multivariate and univariate analyses may relate to systematic differences between analyzing changes in which depth bin was the modal bin and analyzing changes in the number of dives to each bin.

As mentioned above, the number of dives to various depth bins changed with time of day differently for shallow and deep bins as expected for predation on diel-migrating prey. Dives to the shallowest depth bin increased at midnight, and dives to the deepest bin increased during midday. The relative change in number of dives between the shallowest and deepest bins, however, was not consistent with seals feeding on diel-migrating prey. Seals did perform substantially more dives to the shallowest bin during midnight (37.9 dives *vs.* 22.3 dives at midday for adult females), but they only performed a few more dives to the deepest bin around midday (4.1 dives *vs.* 1.9 dives at midnight for adult females). This pattern of increased shallow dives at night is more consistent with seals targeting shallow prey at all times with more effort at midnight. This additional effort at night is difficult to interpret. It could reflect lower success rates at night, which cause seals to dive more to compensate. Alternatively, it could reflect higher success rates at night, which cause seals to focus their foraging effort (*i.e.*, increase the number of dives) at night. Of course, foraging effort would be better quantified by analyses of dive durations or time-at-depth data, which we did not analyze here.

Clearly, multivariate analytical methods provide more information, particularly in terms of depth resolution, than univariate methods. In addition, empirical variance estimation makes GEE models somewhat easier to implement than more complex linear or generalized linear mixed-effects models, which need to explicitly model all the relevant components of variance. Reliance on empirical variance estimation, however, makes GEE models sensitive to the number of sample groups or individual animals. In general, GEE models are robust with sample sizes greater than 40–50, although further corrections have shown reasonable performance with as few as 10–20 clusters or individuals (Mancl and DeRouen 2001). Studies of diving behavior have notoriously low sample sizes of individual animals, and in some cases GEE methods may not be feasible. We suggest, however, that low sample sizes of individual animals not only limit the feasibility of GEE techniques, but also limit the ability to interpret the results from other statistical techniques (especially in cases where behavioral variation is high between individuals; *e.g.*, Boveng *et al.* 1996). When confronted with low sample sizes of individuals, analysts should consider limiting themselves to fairly simple hypotheses that can be adequately answered in the face of large variability between individuals.

When possible, we recommend focusing SDR analyses on hypotheses that can be answered using univariate techniques. Univariate longitudinal analyses can be conducted using either GEE or linear mixed-effects (LME) approaches, and LME regressions may allow for inference with lower sample sizes of animals. Large cluster sizes of data within individuals, however, can be problematic for LME regressions, which must algorithmically invert the covariance matrix to obtain maximum likelihood estimates. SDR studies often have very large cluster sizes (100s or 1,000s of observations). Because of numerical limitations in some cases, GEE techniques may be preferable to LME for univariate analyses. LME analyses also are sensitive to over-dispersion caused by a preponderance of zeros in the data, while GEE analyses can account for over-dispersion within the empirical variance-estimation routine. Similar to GEE, LME methods can also be generalized for analysis of binary outcomes or count outcomes (Breslow and Clayton 1993). The simple "working independence" GEE methods used here, however, are sensitive to varying cluster sizes between individuals because all data points are weighted equally in the initial regression, which determines the parameter estimates. If individuals have widely different cluster sizes, GEE parameter estimates will be biased toward the values suggested by those individuals with the largest cluster sizes. LME methods and GEE methods that adopt a non-diagonal correlation structure are not as easily biased. In addition, both LME and GEE methods can include weighting factors to account for different cluster sizes. GEE methods also are more sensitive to missing values than a correctly specified LME. Both GEE and LME methods, however, are sensitive to behaviorally induced patterns in missing values. For example, if SDR data are less frequently recorded when animals are diving more extensively, analyses will be biased and reflect the behavior when animals were diving less frequently. See Diggle *et al.* (2002) for an overview of issues and approaches to missing longitudinal data.

When the hypotheses require a truly multivariate analysis, like the case study here, GEE methods are necessary (although more complex likelihood-based or Bayesian techniques may also be applied). Multivariate analyses should focus on specific, testable hypotheses and avoid over-parameterization by reducing the pool of potential predictor variables, the number of parameters estimated within each variable, and the number of depth bins included in the analysis. When undergoing this reduction/simplification process, the analyst should explore the data first to determine reasonable compromises. In some cases, the prior knowledge or the hypothesis itself may determine which variables or depth bins to include in the analysis. In our case, the hypothesis determined the primary relationship of interest, which was the interaction between depth bins and time of day. Prior knowledge suggested that we include age, sex, and season within the pool of potential predictor variables. We were able to include all four depth bins within our suite of depth-zonal options because our sample size of individuals was large enough to support an analysis of that complexity. We were careful to reduce the number of parameters estimated for each variable. For example, we modeled time-of-year as a four-category seasonal variable, rather than a 12-category month variable. If the effect of time-of-year was "smooth," we could have modeled that variable as a polynomial or spline function of date.

The GEE approach described here (also see Appendix) allows for multivariate analysis of SDR data, including the number of dives to each depth bin within the analysis in a robust and statistically appropriate manner. Previous analyses were restricted to univariate analyses, which provided limited information about the diving behavior of animals. GEE methods are readily available, computationally simple, and flexible, but they do require moderate to large sample sizes of individuals. LME methods can be used for robust univariate analyses, and in many cases, univariate analyses can adequately test focused hypotheses. Regardless of analytical methods, analysts should be careful to address several common statistical problems that can occur in SDR studies: (1) temporal autocorrelation within data from one individual, (2) random variation between individuals, (3) interdependence in values between depth bins, (4) over-dispersion caused by a preponderance of zeros in one or more depth bins, and (5) behaviorally induced patterns in missing values. Ignoring these problems can result in incorrect, biased results at worst and artificially deflated standard errors, with associated over-parameterization of regression models, at best. In our analysis, a completely naive GLM analysis would result in standard errors that were one to two orders of magnitude too low (Table 3), and model selection using a naive GLM analysis would suggest that all potential predictor variables and interactions were significant.

Acknowledgments

We thank the Alaska Department of Fish and Game for providing the case study data, particularly Kathy Frost. Original data collection was conducted as part of the *Exxon Valdez* Oil Spill Restoration Program and was funded by the *Exxon Valdez* Oil Spill Trustee Council. Capture and handling of harbor seals was authorized by National Marine Fisheries Service Permits 770 and 1000 issued to the Alaska Department of Fish and Game. Many people assisted in the capture and handling of seals, including: Vladimir Burkanov, Jennifer Burns, Maggie and Mike Castellini, Rob DeLong, Brian Fadely, Kathy Frost, Tracey Gotthardt, Shane Kanatous, Lloyd Lowry, Dan Reed, Bob Small, Ken Taylor, Pam Tuomi, Jay Ver Hoef, and Randy Zarnke. This manuscript was improved by comments from Michael Cameron, Rod Hobbs, and two anonymous reviewers on earlier drafts.

Literature Cited

Bengtson, J. L., and M. F. Cameron. 2004. Seasonal haulout patterns of crabeater seals (*Lobodon carcinophaga*). Polar Biology 27:344–349.

Boveng, P. L., B. G. Walker and J. L. Bengston. 1996. Variability in Antarctic fur seal dive data: Implications for TDR studies. Marine Mammal Science 12:543–554.

Breslow, N. E., and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 8:9–25.

Burns, J. M., M. A. Castellini and J. W. Testa. 1999. Movements and diving behavior of weaned Weddell seal (*Leptonychotes weddellii*) pups. Polar Biology 21:23–36.

Croll, D. A., A. J. Gaston, A. E. Burger and D. Konnoff. 1992. Foraging behavior and physiological adaptation for diving in thick-billed murres. Ecology 73:344–356.

Croxall, J. P., I. Everson, G. L. Kooyman, C. Ricketts and R. W. Davis. 1985. Fur seal diving behavior in relation to vertical distribution of krill. Journal of Animal Ecology 54:1–8.

Diggle, P. J., P. Heagerty, K. Y. Liang and S. L. Zeger. 2002. Analysis of longitudinal data. Oxford University Press, New York, NY.

Folkow, L. P., and A. S. Blix. 1999. Diving behaviour of hooded seals (*Cystophora cristata*) in the Greenland and Norwegian Seas. Polar Biology 22:61–74.

Frost, K. J., M. A. Simpkins and L. F. Lowry. 2001. Diving behavior of subadult and adult harbor seals in Prince William Sound, Alaska. Marine Mammal Science 17:813–834.

Gjertz, I., D. Griffiths, B. A. Krafft, C. Lydersen and Ø. Wiig. 2001. Diving and haul-out patterns of walruses *Odobenus rosmarus* on Svalbard. Polar Biology 24:314–319.

Hanley, J. A., A. Negassa, M. D. deB. Edwardes and J. E. Forrester. 2003. Statistical analysis of correlated data using generalized estimating equations: an orientation. American Journal of Epidemiology 157:364–375.

Harris, R. B., S. G. Fancy, D. C. Douglas, G. W. Garner, S. C. Amstrup, T. R. McCabe and L. F. Pank. 1990. Tracking wildlife by satellite: Current systems and performance. United States Department of the Interior, Fish and Wildlife Service, Fish and Wildlife Technical Report 30:1–52.

HASTINGS, K. K., K. J. FROST, M. A. SIMPKINS, G. W. PENDLETON, U. G. SWAIN AND R. J. SMALL. 2004. Regional differences in diving behavior of harbor seals in the Gulf of Alaska. Canadian Journal of Zoology 82:1755–1773.

HEIDE-JØRGENSEN, M. P., N. HAMMEKEN, R. DIETZ, J. ORR AND P. R. RICHARD. 2001. Surfacing times and dive rates for narwhals (*Monodon monoceros*) and belugas (*Delphinapterus leucas*). Arctic 54:284–298.

HUBER, P. J. 1967. The behavior of maximum likelihood estimators under nonstandard conditions. Pages 221–233 *in* L. M. LeCam and J. Neyman, eds. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1. University of California Press, Berkeley, CA.

KRUTZIKOWSKY, G. K., AND B. R. MATE. 2000. Dive and surfacing characteristics of bowhead whales (*Balaena mysticetus*) in the Beaufort and Chukchi seas. Canadian Journal of Zoology 78:1182–1198.

LAIDRE, K. L., M. P. HEIDE-JØRGENSEN, R. DIETZ, R. C. HOBBS AND O. A. JØRGENSEN. 2003. Deep-diving by narwhals *Monodon monoceros*: Differences in foraging behavior between wintering areas. Marine Ecology Progress Series 261:269–281.

LAIDRE, K. L., M. P. HEIDE-JØRGENSEN, M. L. LOGDSON, R. C. HOBBS, P. HEAGERTY, R. DIETZ, O. A. JØRGENSEN AND M. A. TREBLE. 2004. Seasonal narwhal habitat associations in the High Arctic. Marine Biology 145:821–831.

LEA, M.-A., AND L. DUBROCA. 2003. Fine-scale linkages between the diving behaviour of Antarctic fur seals and oceanographic features in the southern Indian Ocean. ICES Journal of Marine Science 60:990–1002.

LE BOEUF, B. J., D. E. CROCKER, S. B. BLACKWELL, P. A. MORRIS AND P. H. THORSON. 1993. Sex differences in diving and foraging behaviour of northern elephant seals. Symposia of the Zoological Society of London 66:149–178.

LIANG, K. Y., AND S. L. ZEGER. 1993. Regression analysis for correlated data. Annual Review of Public Health 14:43–68.

LOUGHLIN, T. R., J. T. STERLING, R. L. MERRICK, J. L. SEASE AND A. E. YORK. 2003. Diving behavior of immature Steller sea lions (*Eumetopias jubatus*). Fishery Bulletin, U.S. 101:566–582.

LOWRY, L. F., K. J. FROST, R. DAVIS, D. P. DEMASTER AND R. S. SUYDAM. 1998. Movements and behavior of satellite-tagged spotted seals (*Phoca largha*) in the Bering and Chukchi Seas. Polar Biology 19:221–230.

LOWRY, L. F., V. N. BURKANOV, K. J. FROST, M. A. SIMPKINS, R. DAVIS, D. P. DEMASTER, R. SUYDAM AND A. SPRINGER. 2000. Habitat use and habitat selection by spotted seals (*Phoca largha*) in the Bering Sea. Canadian Journal of Zoology 78:1959–1971.

MANCL, L. A., AND T. A. DEROUEN. 2001. A covariance estimator for GEE with improved small-sample properties. Biometrics 57:126–134.

MATE, B. R. 1989. Watching habits and habitats from earth satellites. Oceanus 32:14–18.

MATE, B. R., K. A. ROSSBACH, S. L. NIEUKIRK, R. S. WELLS, A. B. IRVINE, M. D. SCOTT AND A. J. READ. 1995. Satellite-monitored movements and dive behavior of a bottlenose dolphin (*Tursiops truncatus*) in Tampa Bay, Florida. Marine Mammal Science 11:452–463.

MATE, B. R., B. A. LAGERQUIST AND J. CALAMBOKIDIS. 1999. Movements of North Pacific blue whales during the feeding season off southern California and their southern fall migration. Marine Mammal Science 15:1246–1257.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized Linear Models. Chapman and Hall, New York, NY.

NORDØY, E. S., L. FOLKOW AND A. S. BLIX. 1995. Distribution and diving behaviour of crabeater seals (*Lobodon carcinophagus*) off Queen Maud Land. Polar Biology 15:261–268.

STEWART, B. S., S. LEATHERWOOD, P. K. YOCHEM AND M.-P. HEIDE-JOERGENSEN. 1989. Harbor seal tracking and telemetry by satellite. Marine Mammal Science 5:361–375.

STEWART, B. S., E. A. PETROV, E. A. BARANOV, A. TIMONIN AND M. IVANOV. 1996. Seasonal movements and dive patterns of juvenile Baikal seals, *Phoca sibirica*. Marine Mammal Science 12:528–542.

TEILMANN, J., E. W. BORN AND M. ACQUARONE. 1999. Behaviour of ringed seals tagged with satellite transmitters in the North Water polynya during fast-ice formation. Canadian Journal of Zoology 77:1934–1946.

WHITE, H. 1982. Maximum likelihood estimation of misspecified models. Econometrica 50:1–25.

Received: 12 August 2004
Accepted: 15 November 2004

APPENDIX

Generalized estimating equations (GEE) provide valid inference for linear, Poisson, and logistic regression models when parameters are estimated using longitudinal data (Liang and Zeger 1993). There are two features that characterize GEE: a "working correlation" model that is used to weight the longitudinal vector of observations from each individual, and an empirical variance estimator (sometimes called a "robust" variance estimator) that provides a valid large-sample estimate of the variance-covariance matrix for the regression estimates even when an incorrect correlation model is specified. Based on this last feature, we fitted a generalized linear model with Poisson error variance (GLM) to estimate parameter values and a working correlation model, which assumed complete independence among data points (clearly, incorrect for multivariate longitudinal data). We then used the GEE empirical variance estimator to calculate the correct variance-covariance matrix.

The empirical variance estimator used the observed variance among individuals to estimate non-parametrically the variance-covariance matrix of the estimated regression coefficients. The observed variance among individuals was estimated using the estimating functions:

$$U(j) = D(j)W(j)(Y(j) - mu(j)) \tag{2}$$

where: $Y(j)$ was the vector of c dive counts (one dive count for each depth bin/zone per 6-h recording period) for individual j, $Y(j) = [Y(j,1), Y(j,2), \ldots, Y(j, c)]$, and $mu(j)$ was the vector of means given by the regression model, $mu(j) = [mu(j,1), mu(j,2), \ldots, mu(j, c)]$. $W(j)$ was a weight matrix (dimension $c \times c$) designed to adjust for the working correlation structure. In our case, we chose a working correlation model, which assumed complete independence among data points. As a result, equal weight was given to each observation, and $W(j)$ was chosen as the identity matrix. For more complex correlation structures, $W(j)$ is formed as the inverse of the working correlation matrix; inverting large correlation matrices can be problematic. $D(j)$ was a $p \times c$ matrix of derivates where the $[k,s]$ element represented the derivative of $mu(j, s)$ with respect to the k-th parameter (*i.e.*, the derivative of mu with respect to each of p parameters).

The variance-covariance matrix V was estimated using:

$$V = A^{-1}\left(\sum_j U(j)xU(j)'\right)A^{-1} \tag{3}$$

where: $U(j)$ was given by (3), $U(j)'$ was the transpose of $U(j)$, and A^{-1} was calculated as the inverse of:

$$a = \sum_j D(j)W(j)D(j)' \tag{4}$$

Interested readers can find Splus 6.1 code for two integrated functions that use the output from "GLM" and estimate an empirical variance-covariance matrix using these methods at: http://faculty.washington.edu/heagerty/MMS. Similar analyses can also be conducted using GEE options of the SAS procedure GENMOD, or using the "cluster(id)" option with the Poisson regression methods in STATA.