1-1-2006

A database for the study of marine mammal behavior: Gap analysis, data standardization, and future directions

Scott A. Shaffer
University of California - Santa Cruz, scott.shaffer@sjsu.edu

D P. Costa
University of California - Santa Cruz

A Database for the Study of Marine Mammal Behavior: Gap Analysis, Data Standardization, and Future Directions

Scott A. Shaffer and Daniel P. Costa

***Abstract*—A relational database was compiled that contained published information on the diving behavior and/or movement patterns of marine mammals to facilitate a modeling effort of the ESME (Effects of Sound of Marine Environment) program. A total of 448 references from reports, books, and peer-reviewed journal articles were obtained. The metadata describing each animal studied, location of the study, and equipment used were entered into the database as well as empirical data describing the diving behavior and movement patterns of each animal. In total, the database contained 1,815 entries from 51 different marine mammal species or subspecies. The majority of animals were seals and sea lions with 1,560 entries from 29 individual species. More than half the number of animals studied were from high latitude regions (e.g., Arctic and Antarctic). Other problem areas identified were: 1) data reduction in summaries, 2) inability to easily summarize qualitative and quantitative data, and 3) lack of standardization in data reporting. A solution is to create a common access data archive where researchers contribute raw published or unpublished geospatially referenced data sets. This would improve access to original datasets with large volumes of data, which overall, enhances the power to develop robust behavioral or ecological models that could help define critical habitats of marine mammals.**

***Index Terms*—Database Development, Diving, Live Access Server, Marine Mammals, Tracking**

I. Introduction

Predictive models are an important part of science and implicit with any modeling effort is the acquisition of data, whether empirically or theoretically derived, that can be used as dependent or independent variables within a model. Concomitantly, the objectives and predictive power desired ultimately depend on the quality, quantity, variability, and breadth of the data used in the model. Therefore collections of data (i.e., a database) serve an important function of any modeling effort. This paper describes the development of a database that supplied published data for a model that emulated the behavior of marine mammals exposed to different sound fields (Houser, and Shyu and Hillson, this issue).

Naval exercises that generate sounds of a certain frequency range and/or amplitude have attracted much attention due to the potential impact on marine mammals [1-4]. The effect of anthropogenic sounds on marine mammals is particularly controversial because marine mammals are protected by US federal laws (Marine Mammal Protection Act of 1972 and Endangered Species Act). Furthermore, marine mammals are high profile animals that receive considerable media attention [2, 3, 5]. Consequently, the Office of Naval Research (ONR) created a program called Effects of Sound on the Marine Environment (ESME) to evaluate and model the influence of sound propagation on marine mammal species (Shyu and Hillson, this issue).

One initiative of the ESME program was to collate all available published information on behavioral aspects related to diving and/or tracking of free-ranging marine mammals. This information was integrated into ESME models (Shyu and Hillson, this issue) to emulate the behavior of marine mammals under specified sound fields, environmental conditions, and geographic locations, and for various marine mammal species (Houser, this issue). In addition, this comprehensive survey provided the Navy with a review of which marine mammal species had been studied, where they were studied, and which methodologies had been used. This type of information was deemed essential for any research program that examined the effects of noise on marine mammals [6]. Thus, when used alone, or in combination with data from other disciplines (e.g., oceanography, fisheries science, etc.), the database could assist the development of predictive models for defining the behavior and distribution of marine mammals.

II. Approach

The approach for the creation of this database was to compile all available published papers and reports that presented data

Manuscript received August 16, 2004. This work was supported in part by the U.S. Office of Naval Research under Grants N00014-00-l-0880 & N00014-03-l-0651, and funds from the Sloan and Moore Foundations. The paper was originally communicated by S. A. Shaffer at the Environmental Consequences of Underwater Sound conference, May 2003, San Antonio, Texas.

S. A. Shaffer is with the Dept. of Ecology and Evolutionary Biology, University of California, 100 Shaffer Road, Santa Cruz, CA 95060-5730 USA (phone: 831-459-2691; fax: 831-459-3383; e-mail: shaffer@biology.ucsc.edu).

D. P. Costa is with the Dept. of Ecology and Evolutionary Biology, University of California, 100 Shaffer Road, Santa Cruz, CA 95060-5730 USA (e-mail: costa@biology.ucsc.edu).

on the diving and/or movement patterns (i.e., tracking) of marine mammals worldwide. Online searches were performed using databases that catalog peer-reviewed publications such as BIOSIS, Current Contents, Medline, and Zoological Records. The following keywords were used when conducting searches: dive, diving, track, tracking, telemetry, satellite, radio, seal, sea lion, dolphin, whale, or any combination thereof. Known research labs working in this field were also contacted for preprints and reports not cataloged in the online databases. A bibliography of all publications was created in EndNote® [7] and all papers were collected, with exception of obscure foreign journals (<20 publications). The data from each publication were extracted and entered into a relational database (e.g., Microsoft Access 2000). At a minimum, available metadata for each individual animal studied was entered, which included details about species, age, age class, sex, reproductive season, geographic information (e.g., hemisphere, major ocean basins, oceanic zones) of capture or release sites, and the type of equipment used to monitor diving and movement patterns. For empirically derived data (i.e., diving behaviors and tracking data), measures of central tendency (i.e., mean, variance, median, mode, minimum and maximum) were entered. This included parameters on diving behavior such as diving depth, duration, surface time, and diving frequency. A complete list of data fields and the percentage of each represented in the database is given in Table 1.

TABLE 1
Summary of data types and the proportion of each extracted from publications included in the database. Each proportion is out of 1,815 entries.

Data Type	Proportion Represented (%)
Metadata	
Record Number	100
Species	100
Sex	90.9
Age	10.2
Age Class (adult, juvenile, pup, etc.)	73.9
Year Of Study	97.0
Subject ID	97.4
Local Season	84.4
Ocean (major ocean basin)	99.9
Hemisphere	100
Oceanic Sector (offshore, inshore, etc.)	8.7
Geographic Area (bay, gulf, fjord, etc.)	97.5
Initial Body Mass	59.3
Final Body Mass	18.5
Author Of Publication	100
Year Of Publication	100
Publication Title	100
Diving Data	
Trips To Sea	15.0
Days At Sea	45.8
Number Of Dives	52.6
Rate Of Diving (dives per hour)	34.2
% Time Diving	20.6
Descent Rate (swim speed on dive)	14.4
Ascent Rate (swim speed on dive)	14.4
Dive Depth	
Mean	47.1
Mode	9.1
Median	15.2
Maximum	39.7
Dive Time	
Mean	46.2
Mode	9.0
Median	14.8
Maximum	36.4
Bottom Time	
Mean	16.3
Mode	7.9
Median	11.2
Maximum	11.2
Surface Time	
Mean	25.8
Mode	7.9
Median	8.8
Maximum	12.3
Min Depth For A Dive	42.8
Time Depth Recorder (TDR) Type	79.9
Resolution Of TDR	23.6
Diurnal Pattern?	14.5
Surface Swim Speed	12.9
Raw Data Available	8.8
Tracking Data	
Rate Of Travel	12.3
Total Distance	12.5
Maximum Range (distance from origin)	17.0
Type Of Tracking Tag	53.8

III. RESULTS AND DISCUSSION

The search for published papers, reports, book chapters, and books totaled to 448 references (413 references on diving behavior and 35 on movement patterns). The bibliography contained references from 1965 [8] through November 2002 [9], the date of the last search (Fig. 1). Only 155 publications yielded data that could be extracted without any required interpretation or modification of the published data set (see below for more details). These data were then entered into the database, which totaled to 1,815 entries (i.e., single animals) comprised of 29 pinniped (seals and sea lions) and 16 cetacean (whales and dolphins) species, plus the dugong, *Dugong dugong*, and European otter, *Lutra lutra*. The majority of species (58.3%) were studied at latitudes above 50° North or South (Table 2). The group with the greatest representation of species in the database was pinnipeds (1,560 entries) of which, Antarctic fur seals, *Arctocephalus gazella* (288 entries), Weddell seals, *Leptonychotes weddelli* (258 entries), and harbor seals, *Phoca vitulina* (247 entries) comprised the majority of entries. There were only 241 entries (13.3% of total) for cetaceans of which, the majority of data were from white whales, *Delphinapterus leucas* (49 entries), harbor porpoises, *Phocoena phocoena* (42 entries), and narwhals, *Monodon monoceros* (30 entries).

A significant outcome of the survey was the lack of data for most marine mammal species, but especially cetaceans. This is attributed to the fact that cetaceans are entirely aquatic, which makes them more difficult to capture or monitor at sea. In contrast, pinnipeds are amphibious so they spend a certain proportion of their lives on land (primarily during breeding and molting), making them more accessible to researchers [10]. Tag attachment is also facilitated by the application of adhesives to the pelage of pinnipeds, whereas cetaceans have smooth hairless skin so this method is not applicable. Cetacean researchers must therefore use temporary attachments (e.g. suction cups) or more invasive mounts that

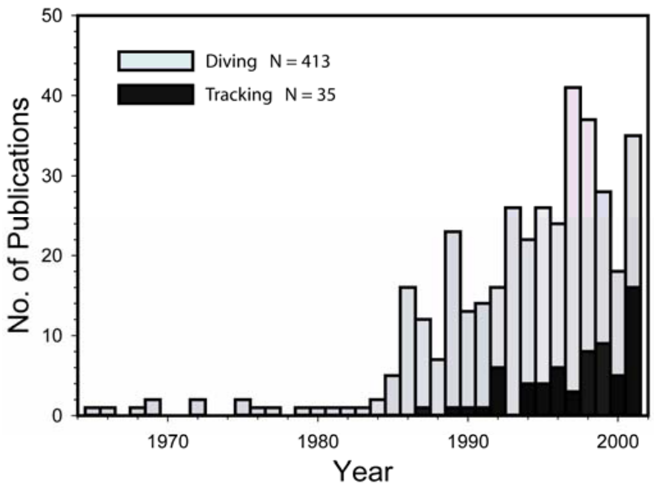

Fig. 1. Publication rate of diving and tracking studies of marine mammals. The number of studies increases with the rapid development in microprocessor technology and the availability of personal computers. The overlapping bars are not cumulative totals.

pierce the skin [11, 12], which may or may not be satisfactory. Interestingly, cetaceans are the largest order of marine mammals with over 70 species, yet they are the least studied for the given number of species that comprise the order. At least one group of cetaceans, the beaked whales, are suspected of being impacted by anthropogenically derived sounds in the oceans [3, 13, 14], but more research on cetaceans, which simultaneously examines diving behavior and sound exposure is required to address this issue.

The data survey also revealed that there was a substantial bias in the geographic areas where diving studies involving marine mammals occurred. For example, over 50% of the species studied ranged within Antarctic or Arctic regions (Table 2) and less than 30% of all species studied ranged along the continental US. Thus future research efforts should focus on species that occur within the US EEZ because this is where marine mammals are potentially impacted by anthropogenic noise from large transport ships, fishing vessels, petrochemical exploration, and naval exercises.

In addition to the gap analysis, there were a number of issues concerning the presentation, extraction, and utility of data presented in publications. For example, there was a significant loss of information in some publications that presented a single summary for all individuals studied compared to others where data were summarized for each individual animal that was studied (e.g., a mean ± SD for 10 individuals *vs*. 10 means ± SDs; one of each animal). A single summary for multiple animals resulted in a substantial loss of individual variability and overall, a lack of utility for any subsequent analyses. Thus, group-summarized data were not included in the database, which is a major reason why the database only contained empirical data from less than half the number of papers in the bibliography. An interesting corollary in marine mammal research has been the increase in the number of individuals studied as microprocessor-based data loggers become smaller and less expensive. If this trend continues, it is likely that future publications will present data in a summarized format in lieu of publishing large tables or

TABLE 2

The distribution by region (major ocean or sea only) of marine mammals studied using electronic devices to quantify diving behavior. These data are from a total of 155 studies (Shaffer and Costa 2002) published between 1965 and 2002 (November).

Ocean or Sea	Hemisphere		Total
	Northern	Southern	
Pacific	269	120	389
Atlantic	278	31	309
Arctic	270	--	270
Southern	--	683	683
Indian	--	23	23
Equatorial Pacific	26	--	26
North Sea	72	--	72
Bering Sea	33	--	33
Totals	948	857	1,805

appendices of data for each individual studied. Ultimately, this will make it more difficult to create databases and/or analyses based on data extracted from publications.

Another problem encountered was the lack of standardization and/or presentation of empirical and metadata. Several studies were not clear about the definition of terms used to describe the data in a publication and there was often insufficient detail about data analysis procedures, software used, data filtering, and in some cases, the instruments used to study at-sea behavior. The type of statistical parameters presented in a study also varied. As an example, some studies presented a mean and variance only, while others presented a suite of parameters (e.g., mean, median, mode, range, etc.) that described the central tendency of a particular behavior. Finally there were limitations on the ability to extract information from figures (e.g., 2-dimensional dive profiles, or geospatial images of animals tracked at sea) when the corresponding quantitative data were not given. It was common for studies to present figures showing the partial record of a diving animal, or the track of a single animal at sea (e.g., Fig. 2), without also providing the corresponding quantitative data. Unfortunately this rendered the qualitative information (i.e., dive profiles or tracks) to be of limited value. This was most problematic for data that described the geospatial movements of animals at sea because it is difficult to provide meaningful quantitative summaries of the data.

Given the limitations of extracting data from publications outlined above, what are some potential solutions to enhance data access, utility, and preservation? Perhaps one of the best solutions to improve data access and utility is by developing large data archives of published or unpublished raw data. Such archives can improve access to large volumes of data, enhance data mining, and allow for greater development of analytical models [15, 16]. Moreover large data archives reduce information entropy or data loss [16] and provide greater flexibility in the interchange between qualitative and quantitative information. Thus, funding agencies should provide more incentives for researchers to contribute their data to large common access data repositories [15]. Likewise, researchers should become more receptive to the idea of archiving published or unpublished datasets.

Another solution is to standardize the way diving and tracking data are presented in future publications. This ultimately increases the ability to compare data without underlying assumptions related to analysis and/or treatment [15]. Furthermore, it sets a precedent for future studies to provide a minimum amount of empirical and metadata in published papers and reports.

Although our literature-based database has limited utility, there are a number of existing data archives for the natural and physical sciences (e.g., LTER, OBIS, JGOFS, NCDC, NGDC, Cornell Library of Natural Sounds, see review in [15]) that illustrate what could be developed for marine mammal research. Currently, there are a number of new initiatives developing to archive large volumes of geo-spatially referenced data for marine animals. The OBIS-SeaMAP program (http://seamap.env.duke.edu/) has become a database repository of historical tracking and survey data for birds, reptiles, and marine mammal studies. The OBIS-SeaMAP program also provides tools to map animal distributions in relation to oceanographic features. Another research program called Tagging of Pacific Pelagics or TOPP (http://toppcensus.org/), is using bio-logging science to study the distribution and abundance of marine mammals, birds, sea turtles, and numerous fish and shark species in the North Pacific Ocean. Part of this effort involves the creation of a database/server that combines remotely sensed environmental data served by the Pacific Fisheries Environmental Laboratory (PFEL; http://las.pfeg.noaa.gov/las/), with real-time or archived tracking data collected by the tagged animals. The objective of the TOPP database/server is to provide users with environmental data that is spatio-temporally matched with animal locations and behavior. Ultimately, these data will be used in analytical models to help researchers define the physical processes that influence animal distributions in the oceans.

An exciting avenue of research that can be achieved with large ecological databases like TOPP or OBIS-SeaMAP is the ability to combine behavioral data collected by animals tracked at sea with information on environmental conditions obtained remotely by orbiting satellites. For example, an elephant seal tracked at sea using satellite telemetry, can be examined in relation to different environmental features that reveal interesting biological patterns (Fig. 2). In Fig. 2A, the seal traveled northward over fairly deep water toward the Aleutian Islands but no real pattern is evident with bathymetry. However, if the same track is overlaid onto averaged sea surface temperature (SST) or chlorophyll concentration (Fig. 2 B & C), it is apparent that the seal traveled to the Aleutian Islands because biological productivity, assumed from cooler SSTs and higher chlorophyll concentration, is higher in that region of the North Pacific Ocean. In this example, the seal's geo-position was obtained daily by satellite uplinks that were subsequently archived on the PFEL Live Access Server (LAS; http://las.pfeg.noaa.gov/las/main.pl). In addition to serving the tracking data, the LAS serves complementary environmental data in OPeNDAP format that is platform independent. Thus, the LAS can produce images for qualitative analyses and provides all associated quantitative data which can be used for more sophisticated analytical models. These analyses were only possible because the TOPP database contains large volumes of raw data for each individual studied. Hence, this level of analyses would not have been possible using the literature-based dataset compiled for the ESME program.

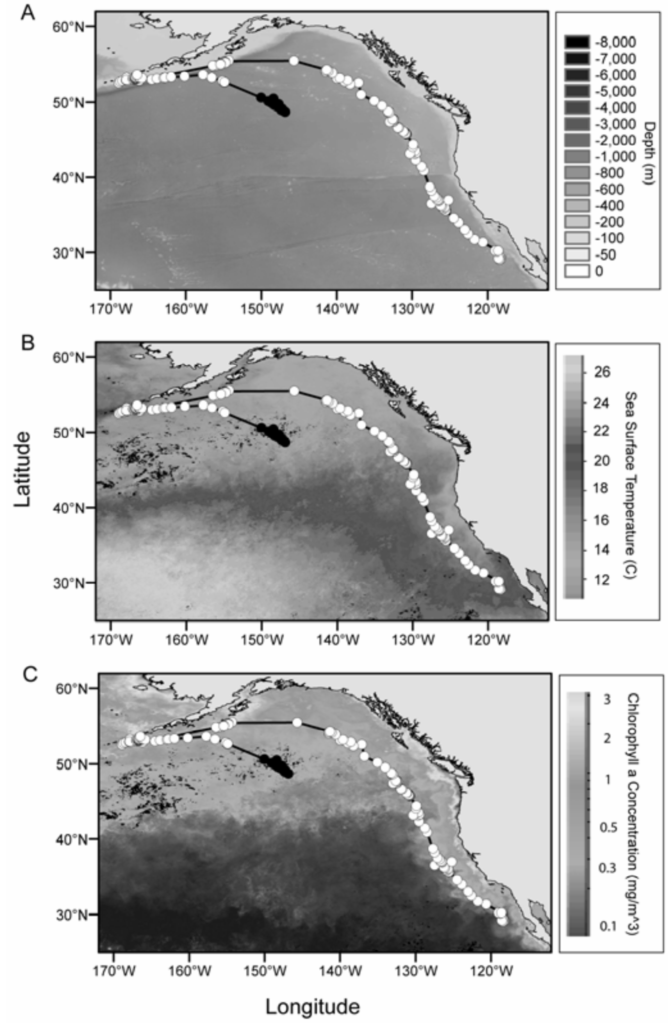

Fig. 2. The at-sea distribution of a single male elephant seal tracked using satellite telemetry for 130 days, from its origin at Guadalupe Island, Mexico (28.9°N, 118.3°W). In each figure, the track of the seal is overlaid with a different oceanographic feature: A) bathymetry, B) sea surface temperature (SST), and C) chlorophyll concentration. The data for SST and chlorophyll concentration were MODIS (Moderate Resolution Imaging Spectroradiometer), 4 km resolution, for the time period in the track (denoted by filled circles) when the seal was returning to Guadalupe Island.

IV. Conclusion

Databases form an important part of data management because they improve the power of modeling for a variety of applications. In this case, behavioral data of marine mammals were compiled and used in applications to model the effects of sound on these organisms. However the organization of these data into a database also identified a number of gaps in the current knowledge of marine mammal behavior like the lack of data on cetaceans and the need for more research on species that range throughout low latitudes (e.g. equatorial waters) and areas within the US EEZ. Further examination of the database also revealed that there is a substantial loss of information in the way data are reported in publications (e.g. summarized data). Consequently, there are significant limitations on the utility of our literature-based dataset. Moreover, these issues clearly illustrate the need for the development of a common repository to archive data on marine mammal diving behavior and tracking data. Such an archive would prevent the loss of data for posterity, standardize the collection and importation of future data, and promote collegiality and collaboration among researchers. More importantly, it would provide greater power to determine critical habitats and further exploration of ecological models.

ACKNOWLEDGMENT

The authors thank S. Simmons and S. Fowler for assistance with data entry, M. Braun of Wildlife Computers for suggestions on database design, and A. Walli for assistance with GIS images. The authors also thank R. Gisiner and M. Hastings of the Office of Naval Research for hosting the ECOUS meeting where this paper was communicated. Lastly, the authors thank the editors of this issue and two anonymous reviewers for comment on earlier drafts. The database created for the ESME program was submitted to OBIS SeaMAP for archiving on 21 March 2005. Go to http://seamap.env.duke.edu/ or contact the authors directly for copies of the database and bibliography.

REFERENCES

[1] D. A. Croll, C. W. Clark, J. Calambokidis, W. T. Ellison, and B. R. Tershy, "Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales," *Anim. Conserv.*, vol. 4, 2001, pp. 13-27.

[2] A. Fernández, M. Arbelo, R. Deaville, I. A. P. Patterson, P. Castro, J. R. Baker, E. Degollada, H. M. Ross, P. Herráez, A. M. Pocknell, F. Rodríguez, F. E. Howie, A. Espinosa, R. J. Reid, J. R. Jaber, D. W. Martin, A. A. Cunningham, and P. D. Jepson, "Whales, sonar and decompression sickness - Reply," *Nature (London)*, vol. 15 April 2004, pp. 1-2.

[3] P. D. Jepson, M. Arbelo, R. Deaville, I. A. P. Patterson, P. Castro, J. R. Baker, E. Degollada, H. M. Ross, P. Herráez, A. M. Pocknell, F. Rodríguez, F. E. Howie, A. Espinosa, R. J. Reid, J. R. Jaber, V. Martin, A. A. Cunningham, and A. Fernández, "Gas-bubble lesions in stranded cetaceans," *Nature (London)*, vol. 425, 2003, pp. 575-576.

[4] L. E. Rendell and J. C. D. Gordon, "Vocal response of long-finned pilot whales (Globicephala melas) to military sonar in the Ligurian Sea," *Mar. Mam. Sci.*, vol. 15, 1999, pp. 198-204.

[5] C. A. Pinatadosi and E. D. Thalmann, "Whales, sonar and decompression sickness - reply," *Nature (London)*, vol. 15 April 2004, pp. 1.

[6] N. R. C. (NRC), *Ocean Noise and Marine Mammals*. Washington, D.C.: The National Academies Press, 2003.

[7] S. A. Shaffer and D. P. Costa, "Bibliography of Marine Mammal Diving and Tracking Studies," University of California, Santa Cruz November 2002.

[8] G. L. Kooyman, "Techniques used on measuring diving capacities of Weddell seals," *Polar Rec.*, vol. 12, 1965, pp. 391-394.

[9] M. A. Hindell, R. Harcourt, J. R. Waas, and D. Thompson, "Fine-scale three-dimensional spatial use by diving, lactating female Weddell seals *Leptonychotes weddellii*," *Mar. Ecol. Progr. Ser.*, vol. 231, 2002, pp. 275-284.

[10] D. P. Costa, "The secret life of marine mammals: Novel tools for studying their behavior and biology at sea," *Oceanography*, vol. 6, 1993, pp. 120-128.

[11] R. S. Wells, H. L. Rhinehart, P. Cunningham, J. Whaley, M. Baran, C. Koberna, and D. P. Costa, "Long distance offshore movements of bottlenose dolphins," *Mar. Mam. Sci.*, vol. 15, 1999, pp. 1098-1114.

[12] A. R. Martin and T. G. Smith, "Deep diving in wild, free-ranging Beluga whales, *Delphinapterus leucas*," *Can. J. Fish. Aquat. Sci.*, vol. 49, 1992, pp. 462-466.

[13] A. Frantzis, "Does acoustic testing strand whales?," *Nature (London)*, vol. 392, 1998, pp. 29.

[14] D. I. Evans and G. R. England, "Joint interim report: Bahamas marine mammal stranding event of 15-16 March 2000," National Oceanic and Atmospheric Administration 2001.

[15] W. K. Michener and J. W. Brunt, "Ecological data: design, management and processing," in *Methods in Ecology*, J. H. Lawton and G. W. Likens, Eds. Oxford: Blackwell Scientific, 2000, pp. 180.

[16] W. K. Michener, J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford, "Nongeospatial metadata for the ecological sciences," *Ecol. Appl.*, vol. 7, 1997, pp. 330-342.

Scott A. Shaffer was born in San Diego, California in November 1966. He earned a B.Sc. in biology from San Diego State University, San Diego, California in 1993; a M.Sc. in marine science from the University of California, Santa Cruz, California in 1996; and a Ph.D. in biology from the University of California, Santa Cruz, California in 2000. He is currently a Research Biologist and lecturer at the University of California, Santa Cruz, California in the Department of Ecology and Evolutionary Biology. His main research focuses on the physiological and behavioral ecology of marine birds and mammals and he has authored or coauthored 16 peer-reviewed publications, and 7 reports or popular articles. Dr. Shaffer is also a researcher in the Tagging of Pacific Pelagics research program.

Daniel P. Costa was born in Van Nuys, California in January 1952. He earned a B.A. in biology from the University of California, Los Angeles, California in 1974; and a Ph.D. in biology from the University of California, Santa Cruz, California in 1978. He is currently a Professor of Biology at the University of California, Santa Cruz, California in the Department of Ecology and Evolutionary Biology. His main research interests focus on the physiological and behavioral ecology of marine mammals and birds. He has authored or coauthored over 150 peer-reviewed publications, book chapters, or popular articles. Dr. Costa is also one of the lead principal investigators in the Tagging of Pacific Pelagics research program.