Journal of Mammalogy Advance Access published December 27, 2015

Journal of Mammalogy, xx(x):1–13, 2015
DOI:10.1093/jmammal/gyv193

Estimating the age of Risso's dolphins (*Grampus griseus*) based on skin appearance

Karin L. Hartman,* Anja Wittich, Juan Juan Cai, Frank H. van der Meulen, and José M. N. Azevedo

Department of Biology, University of the Azores, Rua Mãe de Deus 13, 9501-801 Ponta Delgada, Azores, Portugal (KLH, JMNA)
Nova Atlantis Foundation, Risso's Dolphin Research Center, Rua Dr. Freitas Pimentel 11, 9930-309 Santa Cruz das Ribeiras, Lajes do Pico, Azores, Portugal (KLH, AW)
Delft Institute of Applied Mathematics (DIAM), Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands (JJC, FHM)

* *Correspondent: oceanwatch@gmail.com*

One of Risso's dolphin's distinctive characteristics is the tendency to "lighten" with age due to the accumulation of unpigmented scars. These accumulated scars may provide an indication of age. Photographic skin recaptures gathered from 61 free-ranging animals over a period of 15 years were analyzed to develop a skin classification model in 6 skin stages. Classification of photographic skin captures following this model was tested by 15 experts and 13 nonexpert rankers, with a general probability of agreement of 79%. The duration of each skin stage was estimated using a statistical model based on the recorded dates in which individual animals were known to have entered and/or exited a given stage. A Bayesian approach was used to combine available photographic skin recapture data using expert knowledge as prior to predict the duration for each skin stage and thus the mean age at each stage. Results suggest that animals may live more than 45 years, which is in agreement with published information based on dental layers. The proposed skin stages can be correlated with reproduction, with the transition to stage 3 linked to the onset of maturity. Adult females are less scarified than males and were not observed in the whiter skin classes. The proposed skin stage model is noninvasive and easy to apply and could be a valuable tool in further studies of population structure and dynamics of Risso's dolphins.

Key words: age estimation, Bayesian estimation, coloration stages, noninvasive methods, pigmentation, Risso's dolphin, scarification

Knowing the age of animals is a critically important tool in the study of mammalian life history. Data on growth rates, age at sexual maturity, and longevity are necessary to evaluate the health and productivity of populations (Berta et al. 2006). Furthermore, linking age with long-term behavioral patterns, social organization, reproductive strategies, and habitat use can unravel fundamental aspects of a species' biology.

Current methodologies for age estimations in cetaceans are mainly limited to postmortem techniques using teeth, ear plugs, and/or eye lenses (e.g., Klevezal and Klejnenberg 1967; Lockyer 1972; George et al. 1999). A common technique is counting dental growth layer groups (GLGs), where each layer group represents 1 year. This has been applied to Risso's dolphins (e.g., Chen et al. 2011; Bloch et al. 2012) and other species, such as common bottlenose dolphins (*Tursiops truncatus*—Sergeant 1959; Hohn 1980), pilot whales (*Globicephala macrorhynchus*—Kasuya and Marsh 1984 and *G. melas*—Sergeant 1962), and spinner dolphins (*Stenella longirostris*—Myrick et al. 1986). However, these methods are impossible to apply on wild-ranging animals in a marine environment.

Noninvasive methods to estimate the age of marine mammals include evaluating size from photographs (Fearnbach et al. 2011) and analyzing variations of skin color over time. For some species, color variations can be used as a proxy for age. This was done for narwhals (*Monodon monoceros*), which lighten with age (Hay and Mansfield 1989), allowing Auger-Méthé et al. (2010) to use the proportion of white on the skin to relate age with nicks and notches of the dorsal ridge for photoidentification purposes. Other authors have used the ontogenetic development of color patterns to differentiate broad age categories. In a long-term study of Indian Ocean bottlenose dolphins (*Tursiops aduncus*), Smolker et al. (1992) used the pattern of ventral speckling to define 3 broad age classes: calves (no ventral speckles), subadults (no or only very few ventral speckles), and adults (moderate to heavy ventral speckling). Similarly, the 4 phases of spotting described by Perrin (1970) for the Pantropical spotted dolphins (*Stenella attenuata*) have been subdivided in early and late stages whereas these have been correlated with the relative size, affiliation, and behavior for the Atlantic spotted dolphin (*Stenella frontalis*) by Herzing (1997). This resulted in 5 age classes: neonate, calf, juvenile, young adult, and old adult. More interestingly, by closely monitoring individuals over long periods, the development of the color patterns can be linked with age in a more precise manner. For the Shark Bay bottlenose dolphins, Krzyszczyk and Mann (2012) have recently been able to describe how the pattern of speckle intensity and location on the body can be used to estimate the age of an individual animal: the 1st speckles appear on the genital area around 10 years of age, and heavy speckles on the chest and throat signal individuals over 29 years old.

In general, the skin of cetaceans is more sensitive to cuts and scratches than the skin of other mammals, as they lack the natural protection of fur. Various factors, such as accidents, parasites, predators, and intraspecific tooth rakes, leave their marks on the skin (McCann 1974; Lockyer and Morris 1990; MacLeod 1998). The amount of unpigmented skin left by scarring varies widely among cetacean species but is often observed in odontocetes. Scarring accumulation is observed in species such as the narwhal, the sperm whale (*Physeter macrocephalus*), Risso's dolphins (*Grampus griseus*), and several beaked whale species (MacLeod 1998).

On Risso's dolphins, a deep diving species feeding mainly on mesopelagic squids (Clarke 1986), scarring is extremely visible and accumulates primarily on the animals' dorsal and lateral surfaces (Würsig and Jefferson 1990; MacLeod 1998; Kruse et al. 1999; Hartman et al. 2008). The skin of these animals changes during its life: calves are born silvery grey and will turn from pale-grey as juveniles (Flower 1872; Kruse et al. 1999; Bloch et al. 2012; Hartman et al. 2014) to dark brown or black as young adults (Hartman et al. 2008), subsequently becoming almost white as older adults (Lien and Katona 1990; Kruse et al. 1999; Hartman et al. 2008; Bearzi et al. 2010). This discoloration process is mainly caused by the teeth of other Risso's dolphins during social interactions, leaving linear marks on the skin of the body and the dorsal fin. These scars turn white, because of reduced skin pigmentation (MacLeod 1998) and remain visible for long periods, enabling its use for photoidentification purposes (Hartman et al. 2008). Although their teeth are reduced to only 3–7 pairs at the front of the lower jaw (Clarke 1986; Lien and Katona 1990), erupted teeth remain present during all subsequent stages of life and in both sexes (Würsig and Jefferson 1990; MacLeod 1998). However, the specialized cephalopod diet of teuthophagous cetaceans does not require teeth (Clarke 1986). This supports the argument of MacLeod (1998) concerning the social function of teeth in *Grampus*: the unpigmented scars may function as an indicator of "male quality" or male dominance and is therefore used to avoid risks of escalating aggressive encounters between unevenly matched individuals. Results from a social structure study in the Azores indicated that the adult white animals that formed the observed stable cluster pods (pods being defined as long-term stable groups) were probably all males (Hartman et al. 2008).

Existing age estimations for Risso's dolphins have been made from hunted and stranded animals using the GLGs technique (Kruse et al. 1999; Amano and Miyazaki 2004; Taylor et al. 2007; Chen et al. 2011; Bloch et al. 2012), but no correlation with skin appearance has been made, making this information difficult to apply in a noninvasive field study. During photographic identification, individuals were observed to accumulate scars, but no individuals have been followed yet from birth to old age, so an estimate of the length of skin stage, and thereby of the age range during a particular skin stage, is not available. However, the data accumulated over 15 years of known individuals may be used to estimate the length of different skin stages using a Bayesian approach.

In the present study, we combine dated photographic skin recaptures that document changes of skin coloration and socially mediated scarring patterns from long-term resighted and identified animals in order to: 1) propose a classification of skin stages and 2) using a Bayesian approach, to estimate the duration of each skin stage that can be used to establish relative age in this species.

Materials and Methods

Study area.—The data for this study were collected in the coastal waters (approximately 0–10 km offshore) around Pico Island, in the Azores Archipelago, Portugal, covering approximately 540 km^2. Surveys were conducted from 2000 to 2014. Risso's dolphins were located with guidance from 12 fixed look-out posts situated around the island, with the main look-out located in Ribeiras. A land team guided the research vessel towards groups of dolphins. Observations were carried out with a 4.2-m RHIB with 25-hp engine or a 6.7 fiberglass motorboat with a 140-hp engine. The research for this study was carried out in accordance with guidelines of the American Society of Mammalogists (Sikes et al. 2011).

Photo ID.—Identification photographs were taken using film cameras (Minolta X700, 70–200 mm 36/400 ISO slide films) and SLR digital cameras (Nikon D70-D200-D300s, 70–300 mm zoom lens). Over 1,250 Risso's dolphins were individually identified (Hartman et al. 2015), using distinctive characteristics like notches, nicks, amputations, and the distinctive scarification patterns on the dorsal fin (see Hartman et al. 2008 for a detailed overview of the photoidentification methods used).

Each encounter with an individual Risso's dolphin, where the skin appearance could be observed and photographed adequately, was defined as a photographic skin recapture. For the collection of photographic skin recapture data, high-quality photographs were essential. These were defined as 100% sharp and taken at a distance of approximately 10–20 m from the dolphin. A clear view of the dorsal fin and saddle area, the head, and the front part of the back was required, the so-called scoring area. Furthermore, the scoring area had to be visible without interferences of water or sunlight glimmerings (Fig. 1). Additional data included the identification of an individual and the date and time of the photographic skin recapture.

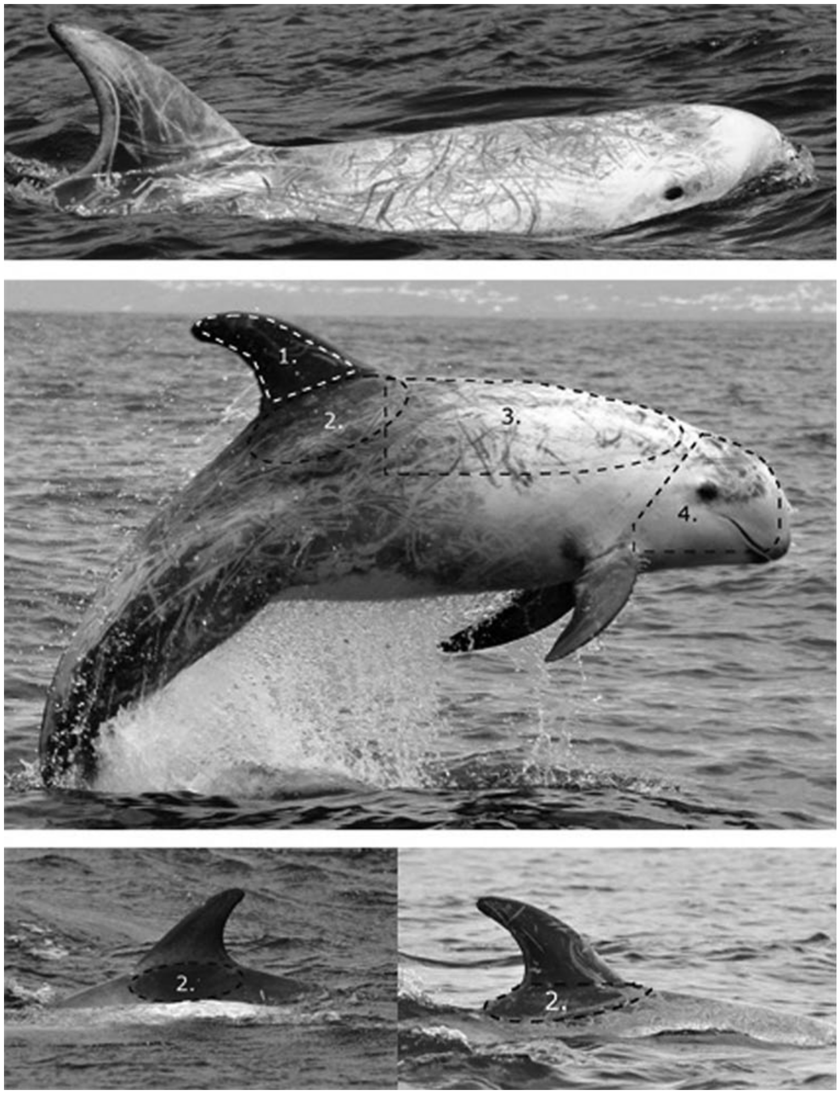

Fig. 1.—Scoring areas used for skin recaptures: dorsal fin of Risso's dolphins (1), saddle area (2), frontal back (3), and head (4).

Photographic data sets and skin stages.—Approximately 4,800 h of ocean-based fieldwork were spent with the Risso's dolphins, over a period of 15 years. Resident individuals of both sexes (Hartman et al. 2015) were frequently photographically recaptured during the season and over consecutive years. Therefore, the changes of discoloration and scarification over years of residents were used for the determination of a skin stage model. Based on analysis of a time-series of photos of known individuals, 6 skin stages have been determined for the purpose of ranking, each consisting of an early and a late phase: 1) Unscratched, 2) Limited Scarification, 3) Moderate Scarification, 4) Severe Scarification, 5) Marbled, and 6) White (Fig. 2).

When determining the skin stages of the animals based on photographic skin recaptures, a small number of skin stages would be easier to apply but would be less precise. On the other hand, a larger number would be more precise but would make the ranking model more difficult to apply. Given that the whitening of a Risso's dolphin skin is a continuous process, any subdivision in stages will have a subjective component. We therefore decided to test whether the 6 skin stages could be generally recognized without prior (expert) knowledge.

Two groups of raters were invited to test whether the skin classification model was usable. The 1st group consisted of 15 experts (biologists working with cetaceans) and the 2nd group consisted of 13 nonexperts. This distinction was made in order to evaluate if prior knowledge, expertise or training were required to apply the skin classification model. Raters were given a manual with the description of each stage coupled with illustrative photographs. For each of the skin stages, a set of 10 photographs of different individuals with unknown gender was assembled, in a total of 60 skin recaptures, presented in random order. Raters were asked to classify each picture in one of the 6 stages. To assess rater consistency, the probability of agreement for 2 arbitrarily chosen raters was estimated using an approach introduced in Fleiss (1971) (see also De Mast and Van Wieringen 2007). Following that approach, a kappa-type index was calculated that compares the measurement system with a chance measurement system in which raters classify objects completely at random, independent of the object's true

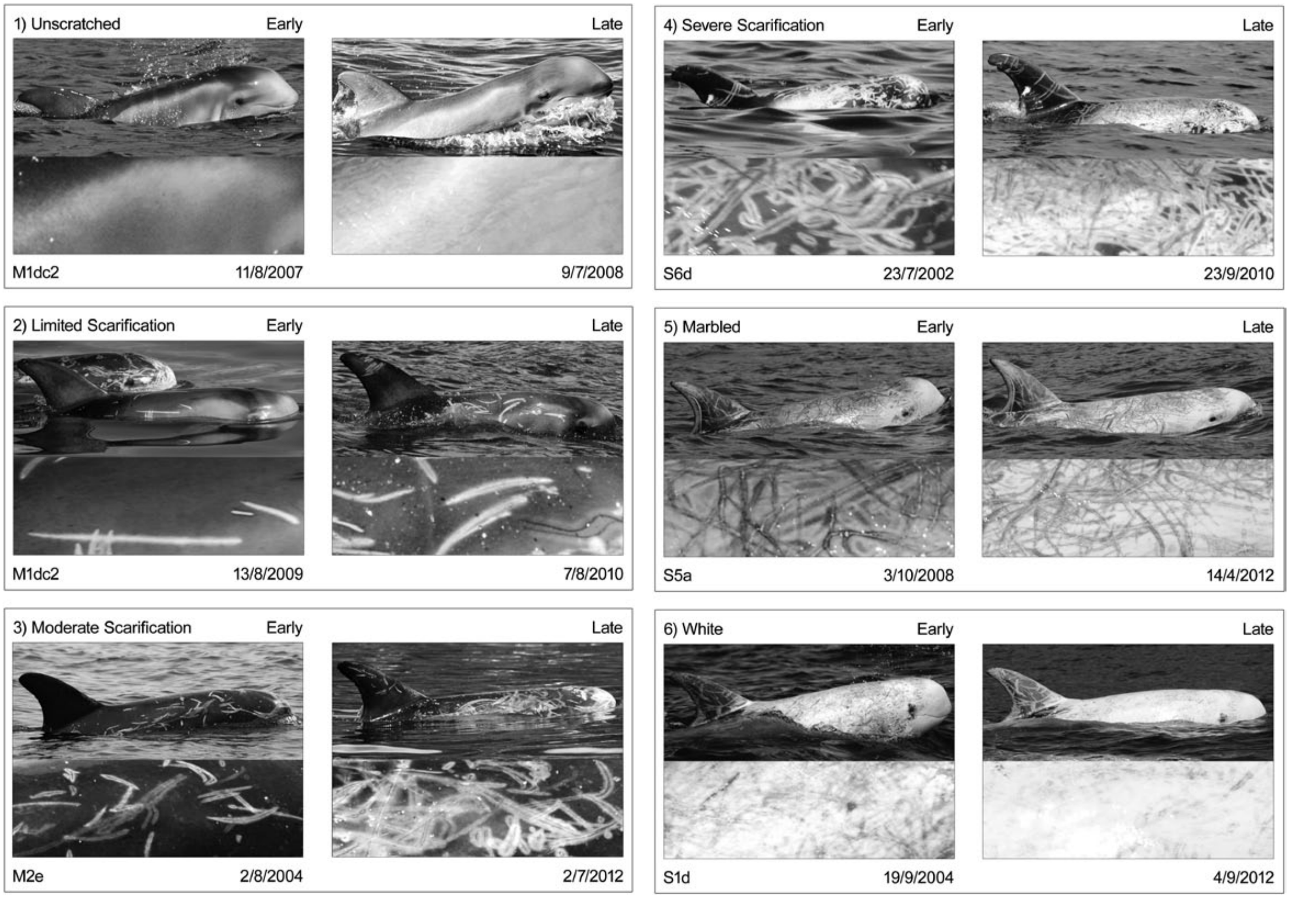

Fig. 2.—Six skin stages of Risso's dolphins, illustrating early (entry) and late (exit) phases to each stage by photo recaptures of the same individual on different dates. The code name of each animal is given on the lower left of each box. The bottom panel of each set of photos represents a zoom of the picture in the top panel.

value. This kappa index has been advocated by (among others) Brennan and Prediger (1981) and De Mast and Van Wieringen (2007). The kappa index was calculated by first subtracting 1/6 from the estimated probability of agreement, followed by division of the outcome by 5/6. The agreement analysis was performed for each skin stage separately and for all skin stages combined, and for each rater group and all raters combined as well. Next, the estimated probability of agreement with an expert (KLH) was calculated, to identify if classification was performed correctly. All computations were done using the statistical software package R.

Statistical estimation of duration of skin stages.—Since skin appearance changes with time, the ideal dataset for estimating the duration of each skin stage would come from the longitudinal study of a large group of individuals followed from birth, on which changes in coloration would have been recorded. In cetacean studies, however, longitudinal studies such as these are practically impossible. Even re-encountering a given individual is a matter of chance, given the high mobility of these animals and the logistic difficulties of searching for them in the ocean. In the present study, we were nevertheless able to utilize an unusually long dataset gathered on a relatively small area with a resident population (Hartman et al. 2008, 2014, 2015). From the ca. 1,250 identified individuals, we have selected 61, that were resighted and from which photographic skin recaptures were collected for over 10 years in a 15-year time period, and from which a total of 387 high-quality skin recaptures were obtained. This set includes 10 animals of known age since they were followed from the newborn stage up to a minimum of 2 and a maximum of 9 years.

Since the photographic skin recaptures were recorded at irregular time intervals, the moments of transition between skin stages were unknown. The data could nevertheless provide a lower bound on the duration of a skin stage for a particular individual if there were at least 2 recaptures available within this skin stage. An upper bound on the duration was available if both an entrance and an exit of a certain skin stage was observed. The duration of each skin stage was thus predicted based on the observed lower and upper bounds. Fig. 3 illustrates how these bounds were obtained. For example, 8 photographic skin recaptures of individual S2c were available for analysis. Up to the 5th recapture, it was identified as in skin stage 5; hence the duration of stage 5 is at least as long as the period between the 1st and the 5th recaptures. Similarly, it has been in stage 6 for at least the time of the 6th recapture up until the 8th recapture.

Because the scarification patterns are believed to have a social function, the differences that could occur between males and females in the pace of discoloration were taken into account

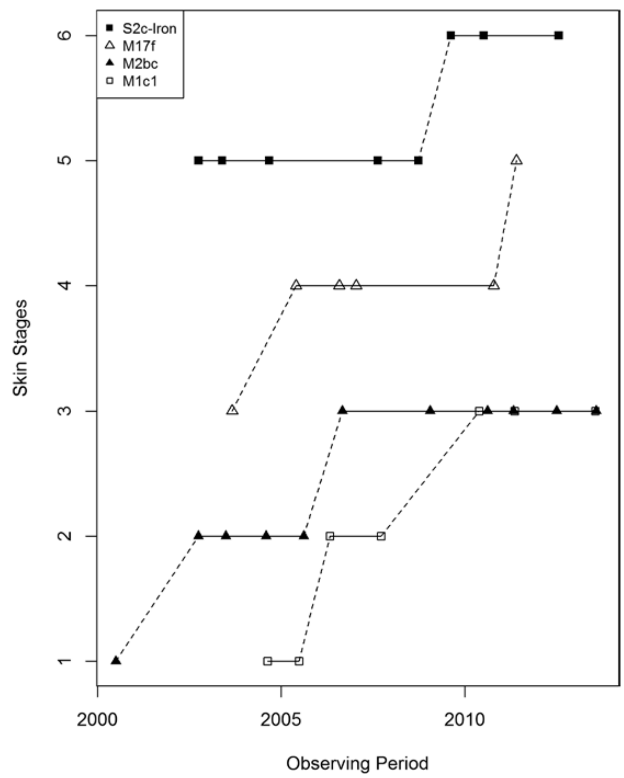

Fig. 3.—Examples of 4 individuals with photographic skin recaptures over multiple years, showing how the lower and upper bounds of the time of transition between skin stages were obtained, with subsequent recaptures of an individual joined by a line. Each dot represents a photographic skin recapture.

by estimating the duration and age of skin stages for males and females separately. The gender for all adults used in the dataset was confirmed by molecular techniques (K. Hartman, pers. obs.). In skin stages 1 and 2, the number of sexed animals was not sufficient to provide separate age estimations for each gender, as done for skin stages 3 and 4. In the higher skin stages, very few stage 5 and no stage 6 females were available, so only male data could be analyzed. Hence, we consider the following 8 skin stages: stage 1 (1 all), stage 2 (2 all), stage 3 female (3 F), stage 4 female (4 F), stage 3 male (3 M), stage 4 male (4 M), stage 5 male (5 M), and stage 6 male (6 M).

Bayesian statistical estimation.—Since only bounds on durations are observed, direct estimation of the mean duration by averaging is not feasible. For this reason, a statistical model for the observed data was introduced. Statistical inference was done under the Bayesian approach (see Box and Tiao 1992 and Gill 2002). This approach enabled to include available prior knowledge on the durations. This knowledge was used to specify a prior distribution on all parameters of the statistical model. This is nontrivial and we refer to Appendix I for mathematical details. Computational details are gathered in Appendix 2. As there are tuning parameters in the prior specification, a sensitivity analysis was carried out to judge their effect on predicted durations. The tuning parameter that reflects the weight assigned to prior knowledge (denoted by γ in Appendix I) has a strong effect on the results for stages with few observations (in particular if these observations contain only lower bounds on the duration). For stages with somewhat larger numbers of observations, the effect of the prior vanishes, as is generally the case.

Results

Classification of photographic skin recaptures.—The classification results of photographic skin recaptures by expert and nonexpert raters are summarized in Fig. 4. Visually, it is clear that agreement was better in skin stages 1 up to 4 and that disagreements were limited to neighboring stages. A few pictures presented in the test were consistently ranked incorrectly and seemed overall problematic to judge by the rankers. Nevertheless, the kappa results (Table 1) showed an overall substantial interrater agreement (probability of agreement 79%). Both biologists and nonexperts found less agreement in assessing age classes 5 and 6. Over all photographs, agreement among biologists was found to be slightly higher than among nonexperts, though the difference is small. Besides consistency, we also calculated the fraction of times that the classification coincided with the expert knowledge. For biologists, this was 0.80 (kappa = 0.77) and for nonexperts, it was 0.77 (kappa = 0.73).

Bayesian estimation of duration and age.—Fig. 5 shows the predictive density of the durations of the skin stages based on Bayesian analysis. We refer to Table 2 for an indication of the variability of results with different prior tuning. The mean predicted duration for skin stage 1 (all sexes) was 2.6 years, followed by 3.4 years for skin stage 2. According to our prediction, female dolphins stayed in stage 3 for 15.2 years; in contrast, males remained in this skin class for 7.8 years. A similar pattern was observed for females in skin stage 4 (12.7 years) versus the males (8.5 years). For skin class 5 and 6, only observations for male dolphins were obtained (10.5 and 11.8 years, respectively). The simulated ages were calculated by summing up the simulated durations. This leads to the density estimates of predicted ages presented in Fig. 6. Individuals in skin stage 1 (all sexes), were 2.6 years old on average, whereas the mean age for those in skin stage 2 was 6 years. For females in skin stage 3, a mean age of 21.2 years was estimated, with a last prediction of 33.8 age in skin class 4. Males were on average 13.8 years old in skin class 3, followed by 22.2 years in skin class 4. The average age in skin class 5 for males was 32.7 and finally 44.5 years in skin class 6.

Discussion

A noninvasive method was developed to estimate Risso's dolphins age based on cumulative skin scarification as documented by photographic skin recaptures. Visual rating of 6 skin stages was well possible and was validated based on ratings by both experts and nonexperts. Based on these skin stages, Bayesian analysis provided average estimated duration and age of an animal in a skin stage. Skin stages for infants and juveniles were not considered sex dependent. Skin stages of females lasted longer and no females were observed in the final stage,

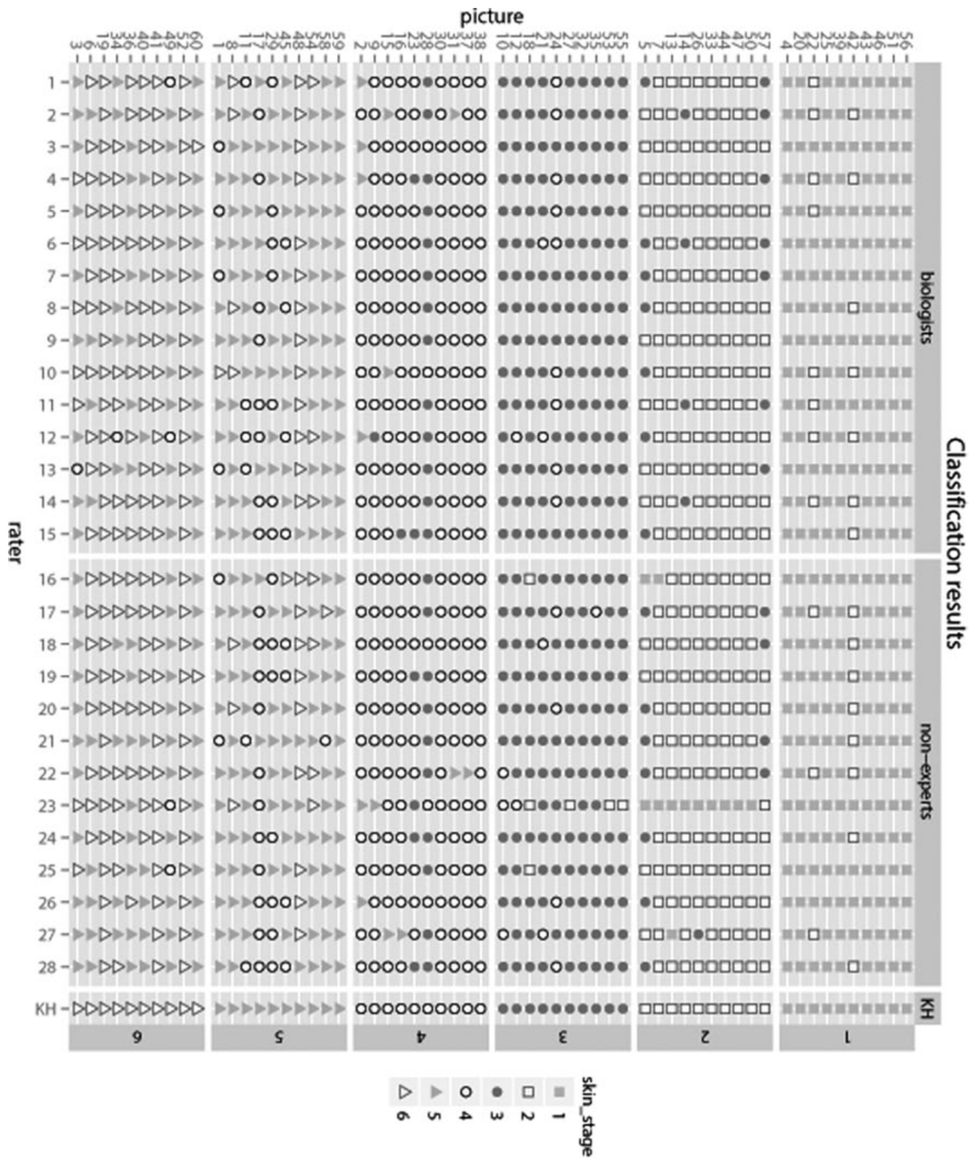

Fig. 4.—Classification by independent raters of the skin stage for a test set of 60 pictures. Raters were divided in 2 groups (biologists and nonexperts). Expert knowledge classification (KLH) is also given.

Table 1.—Skin stage classification: estimated probability of agreement between raters. The corresponding kappa index is given in brackets; values have been multiplied by 100.

Skin stage	Experts	Nonexperts	All raters
1	89 (87)	91 (89)	90 (88)
2	85 (82)	73 (67)	79 (75)
3	91 (89)	77 (73)	84 (81)
4	84 (81)	81 (77)	83 (80)
5	66 (59)	64 (57)	65 (58)
6	72 (66)	73 (67)	73 (67)
all	81 (78)	77 (72)	79 (75)

while skin stages of males were shorter and they did reach the final stage.

Data.—It might be argued that the lower and upper bounds of the skin stages were initially determined on the basis of expert knowledge and were thus subjective. However, the dataset used for this study contains data on 10 animals that were followed from birth until transition to the next skin stages (i.e., stages 2 and 3). Also, the results of the Bayesian tests are consistent with the known age of these animals.

The main methodological issue in the present method of age estimation is that not enough transitions were observed in the more advanced skin stages. This is related to the increasing duration of each stage, which exceeds the duration of the study itself and reduces the probability that a transition is recorded. Our data set is nevertheless longer and more intense than any other that we know of, so these data are the best available at this moment to provide skin stage and age estimates.

As this study was limited to a known population of Risso's dolphins in the Azores, there is the possibility that all conclusions drawn are specific to this geographical area and possibly not reflective of all Risso's dolphins populations worldwide. Other populations of Risso's dolphins around the world may

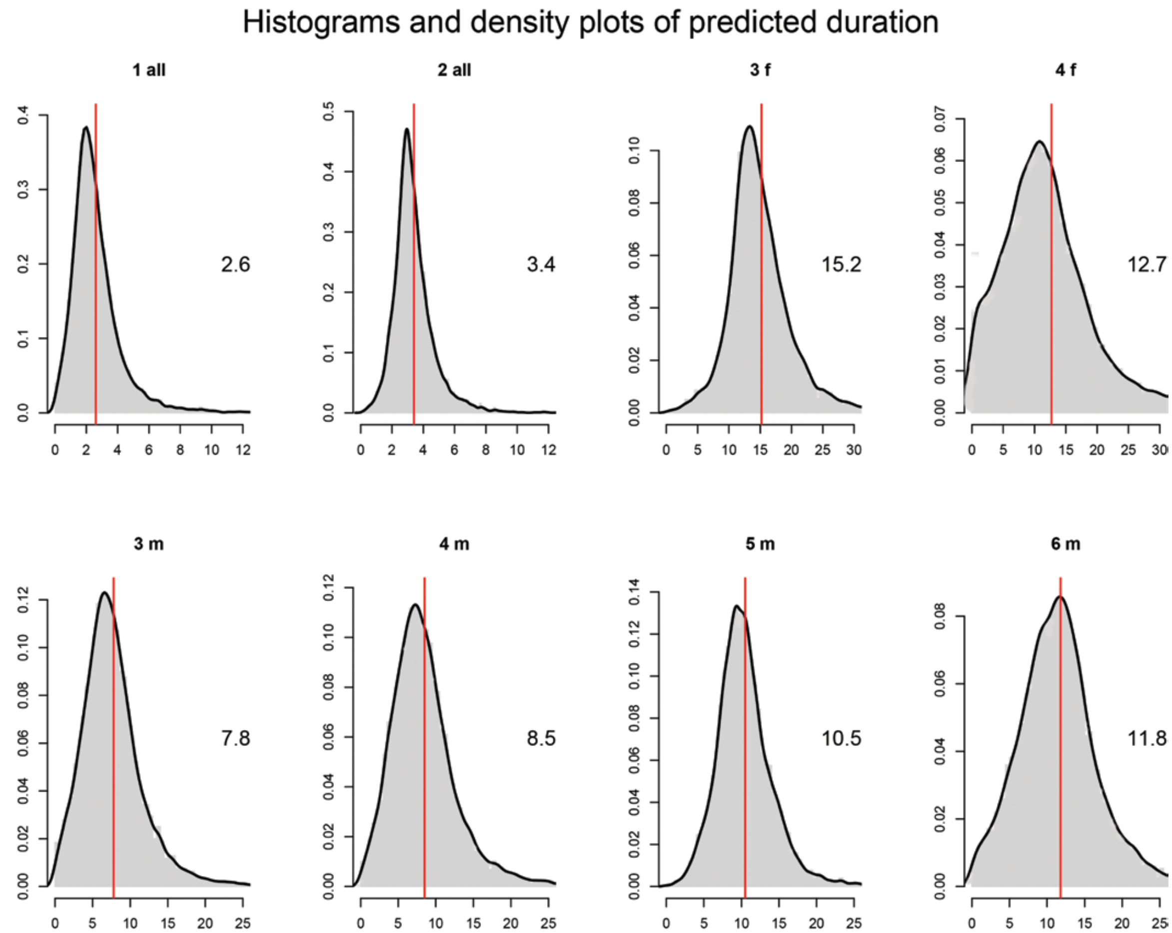

Fig. 5.—Kernel density plots for simulated future (predicted) durations. The mean value is printed in the figure and indicated by a vertical line. Stage 1 (1 all), stage 2 (2 all), stage 3 female (3 F), stage 4 female (4 F), stage 3 male (3 M), stage 4 male (4 M), stage 5 male (5 M), and stage 6 male (6 M).

Table 2.—Mean predicted duration for different values of γ, λ. n_s = number of observations in a stage; n_c = number of observations for which both upper and lower bound are available; γ = tuning parameter representing the chance of misspecified prior knowledge (a value of 1 indicating complete misspecification); λ = tuning parameter in the prior specification. More precisely, it is the prior mean on the coefficient of variation for the duration of all stages.

Stage	n_s	n_c	(γ, λ)				
			(0.5, 0.5)	(0.5, 1)	(0.5, 2)	(0.1, 1)	(0.9, 1)
1 all	10	10	2.6	2.6	2.6	1.8	17.5
2 all	14	4	3.4	3.3	3.4	3.4	3.5
3 F	13	0	15.2	15.1	15.2	12	51.7
4 F	7	0	12.7	12.5	12.4	10.4	43.5
3 M	9	3	7.8	7.7	7.7	7.3	9.2
4 M	13	3	8.5	8.5	8.6	7.5	11.6
5 M	17	1	10.5	10.4	10.4	8.2	19.3
6 M	12	0	11.8	11.7	11.7	9.5	26.4

not necessarily follow this age classification. Factors such as the opportunity for multiple social interactions, water temperature, and water quality may impact whitening and scar healing, as might local diet and other environmental factors.

Age.—The Bayesian statistics provide an age estimate of both females and males. In our study, using a noninvasive method, the mean estimated age of females at stage 4 was 33.8 years and of males at stage 6 (white) was around 44.8 years. This is consistent with the earlier studies since they estimate that Risso's dolphins can at least be 30 years old. In the literature based on dead individuals, the oldest individual known to date was determined to be 38 years old using the GLGs technique (Taylor et al. 2007), and several other studies have reported individuals reaching over 30 years of age (Kruse et al. 1999; Amano and Miyazaki 2004; Taylor et al. 2007; Chen et al. 2011; Bloch et al. 2012). However, our study expands the estimates from the

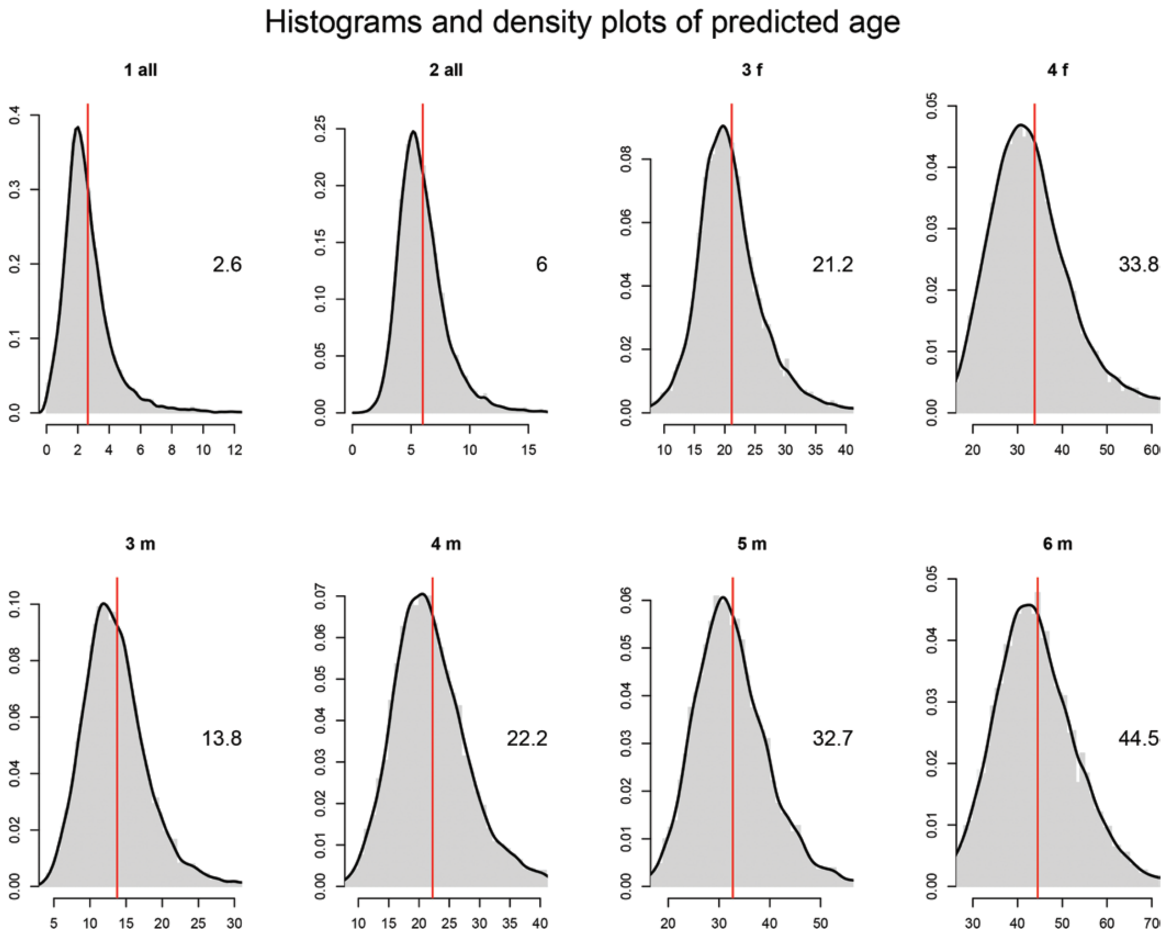

Fig. 6.—Kernel density plots for simulated future (predicted) ages. The mean value is printed in the figure and indicated by a vertical line. Stage 1 (1 all), stage 2 (2 all), stage 3 female (3 F), stage 4 female (4 F), stage 3 male (3 M), stage 4 male (4 M), stage 5 male (5 M), and stage 6 male (6 M).

literature by indicating that the average estimated mean age of males in the highest skin stage is about 44.5 years. In addition, the skin stages of females last much longer and therefore are more likely to result from an underestimation of female age.

Maturation.—Previous authors (Amano and Miyazaki 2004; Chen et al. 2011; Bloch et al. 2012) have used stranded or killed animals to extract biological information, namely on age and maturity stages. Their data suggest that calves are weaned before 2 years of age, which fits well with the transition from Uniform to Limited Scarification at about that age. We therefore propose that stage 1 corresponds to nursing calves and skin stage 2 to juveniles. Based on this same information, it was possible to conclude that the age at maturation in Risso's dolphins is between 8 and 10 years in females and 8 and 11 years in males. Animals in this skin stage can therefore be considered subadults or young adults (skin stage 3), whereas all those in skin stages 4–6 are adults. This information has been included in Table 3.

Differences between females and males.—Basing age estimation on skin stages, using repeated observations of known individuals to estimate the age at which the transition between stages is made, is a method previously used in other species (e.g., Krzyszczyk and Mann 2012). However, in those species, the transitions are ontogenetically determined whereas in Risso's dolphins, the striking pattern of scratches that leads to whitening is environmentally and socially determined, i.e., it results in some part from encounters with prey but mostly from social interactions with conspecifics. Our behavioral data indicate that males fight each other much more frequently than females, as previously pointed out by MacLeod (1998). This is consistent with the skin stages found and their duration. Through this study, we found that males progress more rapidly through the skin stages and that some males reach the final "white" stage 6, while females reach the whiter stages more rarely and much later than males, if at all. This is supported by our larger dataset of all sexed individuals that did not provide sufficient photographic skin recaptures to be part of this study: we have no record of a white female (stage 6) and very few long-term data on females in skin stage 5 (n = 2; excluded from the dataset used for the Bayesian analysis). This suggests that a white skin (in particular stage 6) may be a male feature.

Why 6 skin stages?—The social structure of Risso's dolphins around Pico Island can be best described as a unique "stratified community" where adult animals form highly stable social units, grouped by age and sex classes. However, younger adult

Table 3.—Characteristics of *Grampus griseus* skin stages.

Skin stage	Maturity stage	Base skin appearance	Scarification			
			Dorsal fin	Saddle area	Frontal back	Head
1: Unscratched	Nursing calf	Typical yellow snout and 6–10 vertical fetal folds covering the central body. Gray skin.	No scars visible	No scars visible	No scars visible	No scars visible
2: Limited Scarification	Juvenile	Pale-grayish to dark brown	Few isolated scars	No scars visible	Isolated linear scars; also cephalopod marks and other spots.	Few isolated scars
3: Moderate Scarification	Subadult	Overall dark brown to black	Few isolated scars	No scars visible	Some areas with overlapping scarification, but many isolated linear scars visible; also cephalopod marks and other spots.	Many isolated scars
4: Severe Scarification	Adult	Overall dark brown to black	Many isolated scars	With isolated scars	Very few isolated scars, most scarification is overlapping but original skin still visible.	Overlapping scarification areas, dark skin visible
5: Marbled	Adult	Overall dark brown to black	Overlapping scarification areas	Overlapping scars, dark skin visible	No isolated scars and almost no dark skin visible.	White, covered with overlapping scarification
6: White	Adult	Overall dark brown to black	Overlapping scarification areas	Overlapping scars, dark skin visible	Covered in multiple scarification layers, no dark skin visible.	White, covered with overlapping scarification

animals maintain weaker associated bonds when compared with the associations of older adults. Adult females cluster in stable pods when nursing and calving (Hartman et al. 2008, 2014).

During the observation period, it became apparent that the majority of stable cluster pods consist of individuals with the same scarification appearance and, thus, skin stage. This is an indication that individuals form clusters of more or less the same age during subadulthood, (skin stage 3) and the adult 1 phase (skin stage 4). Remaining together and being involved in the same social events (intraspecies-scarring), they share the same scar history and look equal at the level of coloration composition.

Looking at the results of our ranking test, there was an overall substantial agreement among raters on ranking the photographs provided in the skin stages that were defined. The greatest disagreements were seen in distinguishing between the 2 highest scarified stages. This difficulty would disappear if these 2 stages were combined. Although some females may enter skin stage 5, no female was ever recorded in skin stage 6. Moreover, we argue that they should be maintained because they encompass a large proportion of an individual male's life, and they may correspond to biologically different realities. For example, individuals ranked in skin class 4 and 5 were regularly observed in reproductive activities, with one or more females being present, whereas skin class 6 animals were mostly observed resting during the day, with rarely females present. For that reason, bundling skin class 5 and 6 in 1 "white class" would probably lead to a loss of social and new details concerning the behavioral ecology of this population for further analysis at the level of complex mating behavior as has been suggested by Hartman et al. (2015).

Altogether, the present skin stage classification and the corresponding age and maturity stage distributions provide an important noninvasive tool for the study of the population dynamics of Risso's dolphins in the Azores and elsewhere. Coupled with long-term studies like the present one, they can help estimating age-specific population parameters and track demographic changes in a population, in the context of natural or anthropogenic impacts. It would be of interest to apply this method on other populations in order to compare results on a global scale.

Acknowledgments

This research was supported by Burgers Zoo, The Netherlands, INNO-WWF, The Netherlands, FIN, The Netherlands, WING, The Netherlands, and Lucy Burgers Stichting, The Netherlands and received crucial private donations from A. Tiepel, T. van der Harst, L. Feikema, and F. van Hooff. We thank Nikon Europe and Foto Natura (D. Ellinger) for their generous and important material support. This research was conducted under scientific license issued by the Azores Government (Licenses N4-CN/2004, N9-CN/2005, N98-2009/DRA, N40-2011/DRA, N71-2012/DRA, 36/2013/DRA, 49/2014/DRA). We thank all students and volunteers who helped during their internships with data collection in the field, for which we are very grateful. M. Ruyter designed Fig. 2. Several pictures for this test were kindly provided by L. Steiner and J. Quaresma. Many thanks to all the volunteers for taking the time to score our test sets. Biologist rankers: M. Acosta, B. Adam, V. Au, J. Azevedo, D. Burrier, I. Chen, A. Cosentino, S. Eisfeld, J. Faria, M. Fernandez, A. Gannier, S. Geelhoed, J. Gimenez, N. Hodgins, M. Mariani, A. Martinez Ruiz, P. Palavecino, L. Polo, E. Remonato, L. Steiner, and C. Yzoard. Nonexpert rankers: L. Feikema, M. N. Hartman, M. Mrusczok, L. Neijmeijer, E. Nuyten, M. Odendahl, A. Plistilova, S. Rybicki, D. Suijker,

P. Telleman, T. van der Stap, J. Verhoeven, D. Weilerman, and U. Yu. Useful comments on drafts and other reviews were given by A. Cecchetti, M. Fernandez, P. van der Harst, L. Niemeijer, L. Steiner, L. Sterck, and P. Telleman.

Literature Cited

Amano, M., and N. Miyazaki. 2004. Composition of a school of Risso's dolphin, *Grampus griseus*. Marine Mammal Science 20:152–160.

Auger-Méthé, M., M. Marcoux, and H. Whitehead. 2010. Nicks and notches of the dorsal ridge: promising mark types for the photo-identification of narwhals. Marine Mammal Science 26:663–678.

Bearzi, G., R. R. Reeves, E. Remonato, N. Pierantonio, and S. Airoldi. 2010. Risso's dolphin *Grampus griseus* in the Mediterranean Sea. Mammalian Biology 76:385–400.

Berta, A., J. L. Sumich, and K. M. Kovacs. 2006. Chapter 14: population structure and dynamics. Pp. 416–445 in Marine mammals—evolutionary biology (A. Berta, J. L. Sumich, and K. M. Kovacs). Academic Press, Burlington, Massachusetts.

Bloch, D., G. Desportes, P. Harvey, C. Lockyer, and B. Mikkelsen. 2012. Life history of Risso's dolphin (*Grampus griseus*) (G. Cuvier, 1812) in the Faroe Islands. Aquatic Mammals 38:250–266.

Box, G. E. P., and G. C. Tiao. 1992. Bayesian inference in statistical analysis. John Wiley and Sons, New York.

Brennan, R. L., and D. J. Prediger. 1981. Coefficient kappa: some uses, misuses, and alternatives. Educational and Psychological Measurement 41:687–699.

Chen, I., A. Watson, and L. Chou. 2011. Insights from life history traits of Risso's dolphins (*Grampus griseus*) in Taiwanese waters: shorter body length characterizes northwest Pacific population. Marine Mammal Science 27:E43–E64.

Clarke, M. R. 1986. Cephalopods in the diet of odontocetes. Pp. 281–321 in Research on dolphins (M. M. Bryden and R. Harrison, eds.). Clarendon Press, Oxford, United Kingdom.

De Mast, J., and W. N. Van Wieringen. 2007. Measurement system analysis for categorical measurements: agreement and kappa-type indices. Journal of Quality Technology 39:191–202.

Jaynes, E. T. 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge, United Kingdom.

Fearnbach, H., J. W. Durban, D. K. Ellifrit, and K. C. Balcomb III. 2011. Size and long-term growth trends of endangered fish-eating killer whales. Endangered Species Research 13:173–180.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin 756:378–382.

Flower, W. H. 1872. On Risso's dolphin, *Grampus griseus* (Cuv.). The Transactions of the Zoological Society of London 8:1–21.

George, J. C., et al. 1999. Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. 1999. Canadian Journal of Zoology 77:571–580.

Gill, J. 2002. Bayesian methods: a social and behavioral approach. Chapman & Hall/CRC, Boca Raton, Florida.

Hartman, K. L., M. Fernandez, and Azevedo, J. M. N. 2014. Spatial segregation of calving and nursing Risso's dolphins (*Grampus griseus*) in the Azores, and its conservation implications. Marine Biology 161:1419–1428.

Hartman, K. L., M. Fernandez, A. Wittich, and J. M. N. Azevedo. 2015. Sex differences in residency patterns of Risso's dolphins (*Grampus griseus*) in the Azores: causes and management implications. Marine Mammal Science 31:1153–1167.

Hartman, K. L., F. Visser, and A. J. E. Hendriks. 2008. Social structure of Risso's dolphins (*Grampus griseus*) at the Azores: a stratified community based on highly associated social units. Canadian Journal of Zoology 86:294–306.

Hay, K. A., and A. W. Mansfield. 1989. Narwhal, *Monodon monoceros* Linnaeus, 1758. Pp. 145–176 in Handbook of marine mammals. Vol. 4 (S. H. Ridgway and R. Harrison, eds.). Academic Press. London, United Kingdom.

Herzing, D. L. 1997. The life history of free-ranging Atlantic spotted dolphins (*Stenella frontalis*): age classes, color phases, and female reproduction. Marine Mammal Science 13:576–595.

Hohn, A. A. 1980. Age determination and age related factors in the teeth of Western North Atlantic bottlenose dolphins. Scientific Reports of the Whales Research Institute 32:39–66.

Kasuya, T., and H. Marsh. 1984. Life history and reproductive biology of the short-finned pilot whale, *Globicephala macrorhynchus*, off the Pacific coast of Japan. Report of the International Whaling Commission, Special Issue 6:259–309.

Klevezal, G. A., and S. E. Klejnenberg. 1967. Opredelenie vostrata mlekopitayushchikh po sloistym strukturam zubov i kosti. Izdatel'stvo Nauka, Moscow [Transl.] 1969. Age determination of mammals from annual layers in teeth and bones. Israeli Program for Science Translations, Jerusalem, Israel.

Kruse, S., D. K. Caldwell, and M. C. Caldwell. 1999. Risso's dolphin, *Grampus griseus* (G. Cuvier, 1812). Pp. 183–212 in Handbook of marine mammals. Vol. 6 (S. H. Ridgway and R. Harrison, eds.). Academic Press, London, United Kingdom.

Krzyszczyk, E., and J. Mann. 2012. Why become speckled? Ontogeny and function of speckling in Shark Bay bottlenose dolphins (*Tursiops* sp.) Marine Mammal Science, 28:295–307.

Lien, J., and K. S. Katona. 1990. A guide to the photographic identification of individual whales based on their natural and acquired markings. American Cetacean Society, San Pedro, California.

Lockyer, C. 1972. The age at sexual maturity of the Southern Fin Whale (*Balaenoptera physalus*) using annual layer counts in the ear plug. Journal du Conseil International pour l'Exploration de la Mer 34:276–294.

Lockyer, C., and R. Morris. 1990. Some observations on wound healing and persistence of scars in *Tursiops truncatus*. Pp. 113–118 in Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters (P. H. Hammond, S. A. Mizroch, and G. P. Donovan, eds.). Report of the International Whaling Commission. Special Issue 12. Cambridge, United Kingdom.

MacLeod, C. D. 1998. Intraspecific scarring in odontocete cetaceans: an indicator of male 'quality' in aggressive social interactions? Journal of Zoology (London) 244:71–77.

McCann, C. 1974. Body scaring on Cetacea - odontocetes. Scientific Reports of the Whales Research Institute, Tokyo 26:145–155.

Myrick, A. C., Jr., A. A. Hohn, J. Barlow, and P. Sloan. 1986. Reproductive biology of female spotted dolphins, *Stenella attenuata*, from the eastern tropical Pacific. Fishery Bulletin 84:247–259.

Perrin, W. F. 1970. Color pattern of the eastern Pacific spotted porpoises, *Stenella graffmani*, of the Eastern Pacific and Hawaii, with comments on delphinid pigmentation. Zoologica 54:135–149.

Sergeant, D. E. 1959. Age determination of odontocete whales from dentinal growth layers. Norsk Hvalfangsttid 48:273–288.

Sergeant, D. E. 1962. The biology of the pilot or pothead whale (*Globicephala melaena*) (Traill) in Newfoundland waters. Bulletin Fisheries Research Board of Canada 132:1–84.

Sikes, R. S., W. L. Gannon, and the Animal Care and Use Committee of the American Society of Mammalogists. 2011. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. Journal of Mammalogy 92:235–253.

Smolker, R. A., A. F. Richards, R. C. Connor, and J. W. Pepper. 1992. Sex differences in patterns of association among Indian Ocean bottlenose dolphins. Behaviour 123:38–69.

Tanner, M. A., and W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82:528–540.

Taylor, B. L., S. J. Chivers, J. P. Larese, and W. F. Perrin. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. Southwest Fisheries Science Center Administrative Report LJ-07-01:21.

Würsig, B., and T. A. Jefferson. 1990. Methods of photo-identification for small cetaceans. Pp. 43–52 in Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters (P. H. Hammond, S. A. Mizroch, and G. P. Donovan, eds). Report of the International Whaling Commission. Special Issue 12. Cambridge, United Kingdom.

Yu, Y., and X. L. Meng. 2011. To center or not to center: that is not the question – an Ancillarity-Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. Journal of Computational and Graphical Statistics 20:531–570.

Submitted 4 August 2015. Accepted 20 November 2015.

Associate Editor was Jeanette A. Thomas.

Appendix I

Bayesian statistical approach.—Denote the number of dolphins in stage s by n_s. Index the dolphins by $i = 1, \ldots n_s$. Let Y_i be the duration of stage s for dolphin i and assume that $Y_1, \ldots Y_{n_s}$ are independent and identically distributed. This means that (i) information on duration of one particular dolphin does not provide information on the duration of another dolphin and (ii) all dolphins have the same stage characteristics. We model Y_i with a gamma distribution. There are 3 advantages of this choice. First, it is commonly used for modeling lifetimes and is mathematically tractable. Second, the family of gamma distributions includes both decreasing and unimodal probability density. Hence, it is a rather general and broad family. Third, it includes increasing hazard rates, which seems intuitively plausible for our application. The density of the gamma distribution, denoted by $Ga(\alpha,\beta)$, is given by

$$f(x;\alpha,\beta) = \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right) x^{\alpha-1} e^{-\beta x}, x \geq 0.$$

We assume

$$Y_i \sim Ga\left(\alpha_s, \frac{\alpha_s}{\mu_s}\right),$$

this parametrization implies that the expectation and the variance of Y_i are given by μ_s and $\frac{\mu_s^2}{\alpha_s}$, respectively.

Denote the data for stage s with

$$D_s = \{YL_i, YU_i), i = 1, \ldots, n_s\}.$$

Here, YL_i and YU_i denote the observed lower and upper bounds for Y_i, respectively. In case no upper bound is available, we set $YU_i = +\infty$. The goal is to predict duration based on D_s. We take a Bayesian approach to overcome the difficulty due to the fact that only bounds are observed and to exploit expert knowledge on the duration.

The Bayesian approach requires specification of a prior distribution for all unknown parameters. In this case, the unknown parameters consists of the tuple (α_s, μ_s) with $s \in \{1, \ldots, 8)$. For μ_s, we choose the $Ga(A_s, B_s)$ as prior distribution. To calibrate the parameters A_s and B_s, we made use of the expert knowledge of KLH to obtain the lower and upper bounds of the mean age duration for each stage. Denote the provided bounds by bl_s and bu_s, respectively. Then A_s and B_s are determined such that $P(bl_s \leq \mu_s \leq bu_s)$ where $\gamma \in (0,1)$ is a turning parameter that indicates the chance of misspecified prior knowledge (a value close to 1 indicating complete misspecification).

It is nontrivial to find an a priori distribution for α_s. We use a maximum entropy prior for α_s. Maximum entropy prior is widely used in Bayesian analysis (see, e.g., Jaynes 2003). Note that the coefficient of variation of Y_s, $cv = \frac{E(Y_s)}{\sqrt{\mathrm{Var}(Y_s)}} = \frac{1}{\sqrt{\alpha_s}}$.

If a priori, we expect the value of cv to be λ, then the maximum entropy prior distribution for cv is an exponential distribution with mean λ. This implies that the prior distribution for α_s is given by,

$$P(\alpha_s \leq x) = P\left(cv \geq \frac{1}{\sqrt{x}}\right) = \exp\left(-\frac{1}{\lambda\sqrt{x}}\right), x \geq 0.$$

We assume that the coefficients of variation of Y_s are all the same for all stages and have an exponential distribution with mean λ. Then the prior density for α_s, for $s = 1, \ldots, 8$, is given by

$$\pi(\alpha_s;\lambda) = \frac{1}{2\lambda\alpha_s\sqrt{\alpha_s}} \exp\left(-\frac{1}{\lambda\sqrt{\alpha_s}}\right), \alpha_s \geq 0$$

(this follows upon differentiating the previous display). Note that here λ is also a turning parameter.

The prior specification is completed by assuming that the priors on α_s and μ_s are independent. The joint prior density of (α_s, μ_s) is then $p(\mu_s, \alpha_s; A_s, B_s, \lambda) = f(\mu_s; A_s, B_s)\pi(\alpha_s; \lambda)$. As such, the prior depends on 2 tuning parameters: γ and λ.

The Bayesian approach stipulates that all inference should be based upon the posterior density of (α_s, μ_s). Unfortunately, as often in Bayesian analysis, the posterior density cannot be evaluated in a closed form. To deal with this problem, we use stochastic simulation. More precisely, we apply a data augmentation algorithm for drawing samples from the posterior of (α_s, μ_s). Data augmentation is a Markov Chain Monte Carlo method (see, e.g., Tanner and Wong 1987). Details of the algorithm are included in Appendix 2.

Choosing $\gamma = 0.5$ and $\lambda = 0.5$, we ran the data augmentation algorithm for 10^5 iterations and used the first 10^4 as burn-in samples. As an example, the trace plots for stages $s = 2$ and 3 are given in Figs. 7 and 8, respectively. Here, every 10th

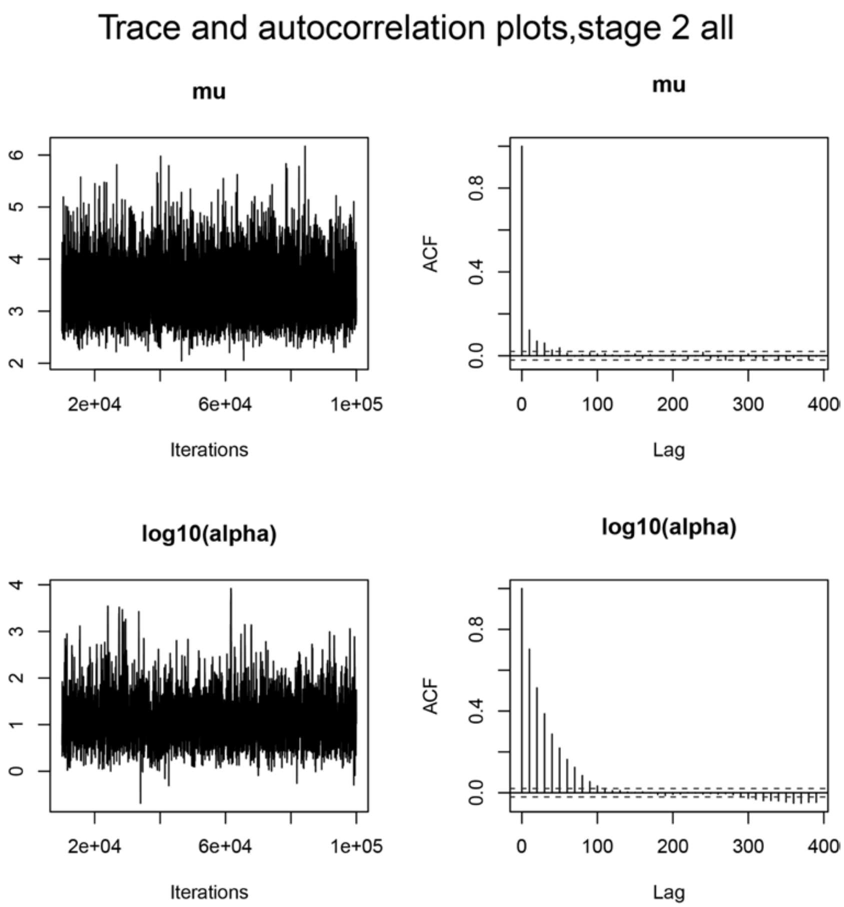

Fig. 7.—Trace and autocorrelation plots for $\log_{10}(\alpha)$ and μ. Data from stage 2.

iteration of the chain is displayed. For updating (α_s, μ_s), we used random-walk Metropolis–Hastings steps with acceptance probabilities tuned to be approximately 0.25.

Let (α_{si}, μ_{si}) be the 10th observation of the posteriors of (α_s, μ_s). The predictive density of duration and age for stage s can be approximated by a kernel density estimator based on samples from $Ga\left(\alpha_{si}, \frac{\alpha_{si}}{\mu_{si}}\right)$. The results are summarized in Figs. 5 and 6.

Appendix II

Data augmentation algorithm to draw from the posterior.—Each skin stage was analyzed independently but in the same manner. In the sequel, we take age stage $s \in \{1, \ldots, 8\}$ to explain our model.

In this section, we present an algorithm for drawing from the posterior of (α_s, μ_s), for $s = 1, \ldots, 8$. Note that when applying the algorithm, we choose $\gamma = \lambda = 0.5$. Thus, the values of (A_s, B_s, λ) are determined.

1. Initialize (α_s, μ_s)
2. For all $i \in I_s = \{1, \ldots, n_s\}$, draw $Y_i \sim Ga\left(\alpha_s, \frac{\alpha_s}{\mu_s}\right)$ conditional on $Y_i \in (YL_i, YU_i)$.
3. Update α_s by proposing $\log \alpha_s^0 = \log\alpha_s + \sigma_1 Z$, where $Z \sim N(0,1)$. Accept α_s^0 with probability equal to $\min(1, P_1)$, where

$$P_1 = \frac{\pi(\alpha_s^0;\lambda)}{\pi(\alpha;\lambda)} \frac{\Pi_{i\in I_s} f\left(Y_i;\alpha_s^0,\frac{\alpha_s^0}{\mu_s}\right)\alpha_s}{\Pi_{i\in I_s} f\left(Y_i;\alpha_s,\frac{\alpha_s}{\mu_s}\right)\alpha_s^0}$$

4. Update μ_s by proposing $\log\mu_s^0 = \log\mu_s + \sigma_2 Z$, where $Z \sim N(0,1)$. Accept μ_s^0 with probability equal to $\min(1, P_2)$, where

$$P_2 = \frac{f(\mu_s^0;A_sB_s)}{f(\mu_s;A_sB_s)} \frac{\Pi_{i\in I_s} f\left(Y_i;\alpha_s,\frac{\alpha_s}{\mu^0_s}\right)\mu_s}{\Pi_{i\in I_s} f\left(Y_i;\alpha_s,\frac{\alpha_s}{\mu_s}\right)\mu_s^0}$$

5. Iterate steps 3 and 4 a large number of times.

Mixing of the chain can be improved by using a strategy advocated in Yu and Meng (2011). The example in section 4.2 of that paper is close to our problem here. In the terminology of Yu and Meng (2011), we add an auxiliary step for μ_s. For $i \in I_s$, define $\eta_i = \frac{\alpha_s}{\mu_s} Y_i$. It is not hard to see that $\eta_i \sim Ga(\alpha_i, 1)$, meaning that the distribution of η_i does not depend on μ_j. We propose to update μ_s conditional on $\{\eta_i, i \in I_s\}$ instead of $\{Y_i, i \in I_s\}$. After some calculations, this results in the following step.

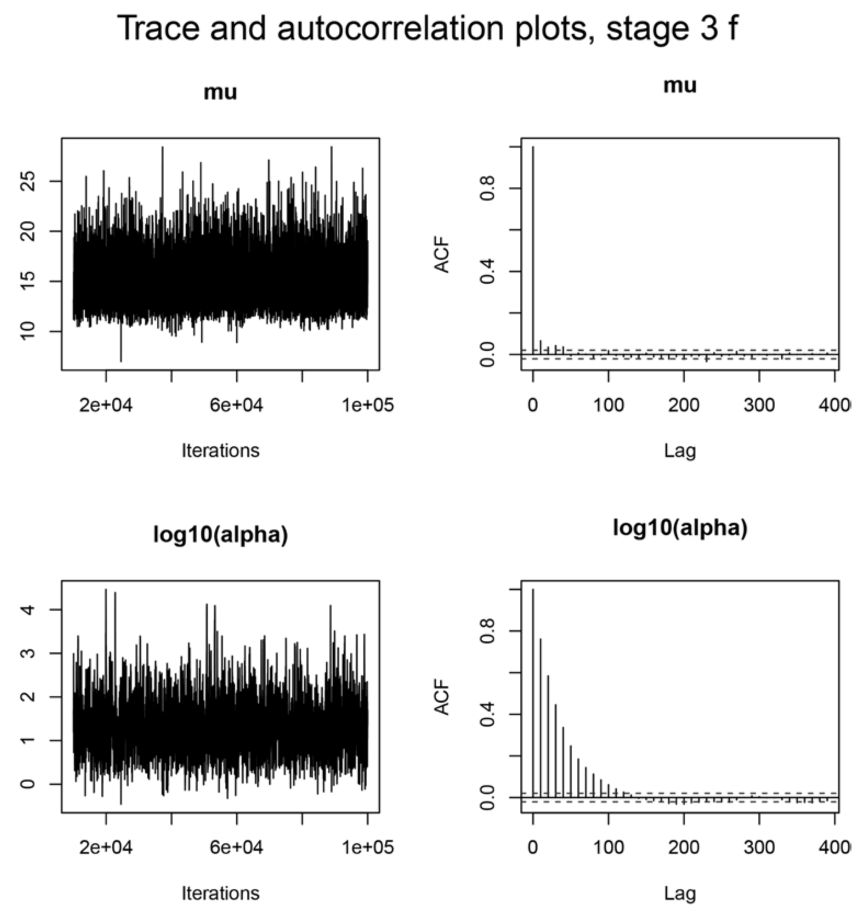

Fig. 8.—Trace and autocorrelation plots for $\log_{10}(\alpha)$ and μ. Data from stage 3.

4′. Draw $\mu_s^0 \sim Ga(A_s, B_s)$ conditional on $\mu_j^0 \in [L, U]$, where

$$L = \alpha_s \max_{i \in I_s} \frac{YL_i}{\eta_i} \text{ and } U = \alpha_s \min_{i \in I_s} \frac{YU_i}{\eta_i}$$

Next, set

$$Y_i = \mu_s^0 \eta_i \big/ \alpha_s .$$

The algorithm then consists of iterating steps (3), (4), and (4′) until the chain reaches its stationary regime. This is the algorithm that has been used for obtaining our results.

We also included a sensitivity study to assess the influence of the 2 tuning parameters γ and λ that appear in the prior specification (Table 2).

Table 2 shows the mean predicted durations for different values of (γ, λ). In Table 2, n_s denotes the number of observations in each stage and n_c denotes the number of observations for which both lower and upper bounds are available.

From columns 4–6, we infer that the choice of λ has a very small effect on the mean predicted durations. However, a small value of γ implies stronger prior beliefs and this effect is clearly reflected in the rightmost 3 columns. The last column shows that incorporation of prior knowledge is necessary to get reasonable estimates. The strong effect of the prior is mainly due to small sample sizes and the fact the most observations only consist of lower bounds.