1 Appendices

2 Alternative logistic parameterization

3 We reparameterized the standard logistic function so that the two parameters are precisely
4 what we seek. Define $F(t) \in (0,1)$ as the fraction molted as a function of day t. The
5 standard logistic function is

$$F(t) = \frac{e^{a+bt}}{1 + e^{a+bt}} = \frac{e^{t'}}{1 + e^{t'}}, \tag{1}$$

6 where a and b are the parameters and $t' = a + bt$ is written for convenience. First define one
7 new parameter, c, as the day of mid-molt, ie the day when molt reaches 50%, $F(c) = 0.5$.
8 That happens at $e^{t'} = 1$, when $t' = 0$ so

$$c = -\frac{b}{a}. \tag{2}$$

Next, define a parameter d as the time between 5% and 95% molt, ie the time it takes the
animal to molt. Find t' at $F = 0.05$ and 0.95,

$$F(t') = \frac{e^{t'}}{1 + e^{t'}} = \begin{cases} 0.95 & \text{when } t' = \ln 0.95 - \ln 0.05 = \ln 19 \\ 0.05 & \text{when } t' = \ln 0.05 - \ln 0.95 = -\ln 19, \end{cases} \tag{3}$$

where ln is the natural logarithm. That means $t = -a/b \pm (\ln 19)/b$. The desired parameter
d is the difference between those two values of t,

$$d = \frac{2 \ln 19}{b}. \tag{4}$$

9 The logistic function never reaches 0 or 1, and we chose 5% and 95% based on the limits of
10 our observations. Had we chosen instead 1% to 99% as the definition of molt time,
11 parameter d would be 1.56× higher ($2 \ln 99$).

12 The model was fitted using the new parameters c and d, and all results are presented with
13 those parameters. Within the model, an algorithm in a subroutine converted c and d back
14 to a and b (Eqs. **??**, **??**) to make use of the standard logistic formula (Eq. **??**). It was
15 therefore not necessary to rewrite Equation **??** using c and d.

16 Modeling molt date

17 Our goal here was to estimate the mean molt date and its variance. We used data from
18 2016-2022 because we had detailed observations of many individuals in each of those years,

19 and preliminary calculations indicated there was no year-to-year variation in molt date over
20 that period. To assure valid estimates, however, we used a model including both a year
21 term and a term for individual animals. That avoided concerns about repeated measures
22 from individuals whose molt was observed in more than one year. The terms for year and
23 individual required fitting the logistic curve to molt progress in an hierarchical framework,
24 with two levels: individual and year.

25 For each individual i in year j, the model included a pair of molt parameters (c_{ij}, d_{ij}),
26 which we write θ_{ij} for brevity. There were hyper-means $\widehat{\theta}_i$ for each individual and $\widehat{\theta}_j$ for
27 each year, along with a grand hyper-mean $\boldsymbol{\theta}$, the mean molt across all individuals and
28 years. Within each group (individual or year), we assumed parameters followed a Gaussian
29 distribution, so there were corresponding hyper-standard-deviations σ_a, σ_y. There was a
30 single σ_a for all animals, assuming all had the same year-to-year variation (a standard
31 assumption in multi-level models), and one σ_y across years.

32 Fitting individual parameters required an error term for molt observations, ϵ, which was
33 assumed to be constant for all individuals in every year; ϵ was estimated along with the
34 molt parameters. The probability of one observation of a molt fraction, M_{ijt}, on day t for
35 individual i in year j, given the parameters, is

$$P_{ijt} = \mathcal{N}(M_{ijt}, Mean = F(t, \theta_{ij}), SD = \epsilon), \tag{5}$$

36 where $F(t, \theta_{ij})$ is the logistic prediction on day j given parameters θ_{ij} (Eqs. **??, ??, ??**). $\mathcal{N}$
37 means the normal probability of observing $M_{i,j,t}$ given the mean and standard deviation ϵ.
38 A Gaussian error for the observed fraction molted was not our first choice, since M is
39 constrained to $[0, 1]$. We tested alternative errors (beta, logit, half-normal), and the
40 Gaussian worked best, leading to strong parameter estimates quickly. The fact that M is
41 never < 0 though the Gaussian error would accept it does not cause errors. Overall, we
42 found an estimate of $\epsilon = 0.10$, showing that observations of percent molt were $\pm 10\%$.

43 There were additional likelihood functions for the hyper-parameters, for example,

$$H_i = \mathcal{N}(\theta_{ij}, Mean = \widehat{\theta}_i, SD = \sigma_a) \tag{6}$$

44 is the probability of the individual's estimated mean $\widehat{\theta}_i$ given its annual estimates θ_{ij} and
45 the within-animal standard deviation σ_a. There were also likelihood functions for the year
46 means $\widehat{\theta}_j$ and the grand mean $\boldsymbol{\theta}$. The only prior probabilities for any parameters were the
47 trivial requirements that standard deviations be > 0.

48 The full likelihood for observations of one individual included both the probability of
49 observations given the logistic model (Eq. **??**) and the probability of the hyper-mean (Eq.
50 **??**). The likelihood of hyper-parameters, however, did not depend on individual
51 observations, only the estimates of all individual means. The same applied to observations
52 within and across years. This is the key to the multi-level approach (**????**), because some
53 individuals had too few observations to produce a useful fit of the logistic model on their
54 own, while others with many observations throughout the molt produced strong estimates
55 of logistic parameters. By including the hyper-parameters, those with many observations

56 supported those with few (**?**). This worked well with the molt data, because parameter d,
57 governing how fast they molted, varied little among females, so in animals with poor data,
58 the population-wide $\hat{d}$ helped shape the curve. But c, the date of molt, varied widely
59 between animals, meaning that molt day for animals with poor data was poorly constrained.

Parameter fitting

61 We used a Bayesian, Monte-Carlo parameter-fitting method, sampling the posterior
62 distributions by repeated Metropolis updates. The model for adult females included 383
63 individuals over 7 years, with a total of 837 female-year combinations, meaning 837 logistic
64 parameters, 390 hyper-means, plus a grand mean and 3 standard deviations. At each step
65 of the Monte Carlo chain, one of those parameters was updated by a random draw, and a
66 new likelihood was calculated (Eq. **??**, **??**). The Metropolis algorithm explores parameter
67 space, accepting new parameter combinations close to the maximum likelihood but not at
68 the maximum. Models for adult females, juvenile males (164 animal-years), and juvenile
69 females (178 animal years) were run separately.

70 The completed Markov chains from each model provided the posterior distributions for all
71 parameters and any statistics derived from parameters. Chains were run 20000 steps, and
72 examined visually for mixing; they converged quickly. The initial 16000 steps were
73 discarded as burn-in. We report the mean of post-burn-in chains as best estimates, and
74 quartiles 0.025 and 0.975 for 95% credible intervals. Hypotheses were tested by checking
75 overlap of credible intervals.

Modeling arrival and departure

77 A model for the tenure of every individual throughout the spring haul-out was separate
78 from the molt model because it is an independent topic. We decided, however to take
79 advantage of estimates from the molt model in several ways that both strengthened and
80 simplified the tenure estimates. First, most of the adults carried dye marks on their
81 unmolted fur from the winter breeding season, then lost the mark upon molting. Animals
82 were thus harder to detect after molting. Arrival and departure models were thus
83 separated, the former using all observations prior to an individual's estimated molt day
84 (parameter c in the molt model), the latter using observations after the molt day. We also
85 used a subset of all animals having molt date estimated with high precision, those with 95%
86 credible intervals of $c < 6$ days wide. This is a well-observed subset and thus leads to
87 stronger estimates of arrival and departure dates.

88 A second way we took advantage of the molt results arose because we found no year-to-year
89 variation in molt date. For this reason, we omitted a year term from the arrival and
90 departure models. Moreover, we simplified the model by omitting a hierarchy for
91 individuals across years. The single multi-level feature was the animal-year combination.
92 Thus, the estimate for each individual was separate in every year. In well-observed animals,

93 arrival and departure dates were strongly constrained by observations, and in our current
94 analysis, we are not interested in variation among individuals.

95 Both arrival and departure estimates arose directly from observations in a straightforward
96 way. If an animal was first seen on 1 May, the arrival date must be $\leq$ 1 May. If that animal
97 was then observed most days until molting, arrival was probably close to 1 May, but if the
98 animal was only seen every fifth day, there would be a good chance it arrived several days
99 earlier. The delay between true arrival and observed arrival thus depends on detection
100 probability, which is a simple calculation. After the first observation, we counted all days
101 an individual was observed divided by all days observers were in the field. It was important
102 to exclude the first day from this calculation, otherwise the detection estimate would be
103 biased upward. The departure estimate is exactly the same in reverse.

104 The parameters needed are A_k, the arrival date, and δ_{k1} detection probability, both
105 referring to individual-year combination k. The 1 indicates detection before the molt. The
106 likelihood of δ was estimated as

$$P(\delta_{k1}) = \mathcal{B}(T, O), \tag{7}$$

107 where T is the total number of days with observations and O the number on which animal
108 k was observed, with $\mathcal{B}$ meaning the binomial probability. Given δ, the probability of a
109 date A was

$$P(A_{k1}) = \mathbf{G}(F - A, \delta_{k1}). \tag{8}$$

110 F is the first date observed and $\mathbf{G}$ is a geometric distribution with probability δ. It is the
111 probability of failing to observe the animal $F - A$ times before the first success. Since there
112 was a single hierarchy for arrival date across all female-year combinations (ie no year term),
113 there was just a single hyper-mean and hyper-standard-deviation (as in Eq. **??**). The
114 likelihood functions for the model of departure D were equivalent, given post-molt
115 detection δ_{k2}, a last day L of observation, and departure hyper-parameters. Arrival and
116 departure parameters were fitted using the same Monte Carlo procedure described for the
117 molt model. A tenure in every case was calculated as $D - A$ from every step of the Monte
118 Carlo parameter chain.

Model verification

120 We graphed the fitted logistic curve for every individual's molt sequence in the years
121 2020-2021. These are included at the end of the Supplement. This allowed a visual check
122 for cases where the curve missed data points. An example is animal 2020-49700, where the
123 animal was recorded as unmolted, then fully molted once, unmolted once again, and finally
124 fully molted; at least one must be an error. The model accommodated by fitting a gradual
125 molt. Indeed, our visual screening revealed that poor fits such as this one were nearly
126 always due to an error mistaking 100% molt for 0%, or vice versa, with the error usually
127 producing an unusually gradual or unusually abrupt molt.

128 We thus screened all cases in adult females where the molt duration exceeded 9 days, and in

129 juveniles where it exceeded 12 days, over all years, 2016-2022. This led us to conclude that
130 any molt duration > 10 days in adults and > 15 days in juveniles were caused by errors.

131 The tenure chart (main text, Fig. 2) highlights other outliers, and we thoroughly screened
132 the most extreme. One obvious outlier appears at the top left of Figure 2, apparently
133 arriving 70 days before molting. She was an interesting exception, because she had an
134 horrendous and recent shark scar when she arrived to breed in January, probably losing her
135 fetus in the attack. After departing in late February, she returned in early April for 6 days,
136 was not seen for 66 days, and finally returned in June to go through a normal molt
137 sequence. She broke the rule that animals always come ashore continuously during the
138 molt, but she was still healing in March and likely could not migrate normally. The second
139 extreme outlier (down Figure 2) also arrived 70 days before molting, but her early
140 observation was easy to discount as a mis-identification. Other outliers in arrival might be
141 valid, but can also be attributed to errors in molt score. We also checked two cases in
142 Figure 2 in which females appeared to start molting before they arrived. Both are most
143 likely due to a single mistaken molt score. There were few such outliers, we did not remove
144 any from the data presented.

Post-hoc tests of molt date versus age and breeding status

146 To test whether molt timing (date and duration) varied with breeding status, we took
147 advantage of posterior distributions for every molt-sequence (animal-year) in the Bayesian
148 model, since these reflect statistical confidence. Statistics were recalculated using 200
149 randomly-selected sets of post-burn-in parameters from the Monte Carlo chains. Credible
150 intervals were then calculated as 95th percentiles of those 200 replicates; statistical
151 significance was inferred from non-overlapping intervals.

Consistency of molt scores

153 Observations were frequent, and two or more observers worked together on most days, so
154 there were numerous cases where two different people recorded a molt score for the same
155 animal on the same day. There were occasions where one observer recorded data for the
156 same animal twice, but we discarded those from this analysis in case they were duplicates.
157 That left 1220 replicate estimates of molt, including 1079 with two observations, 135 with
158 three, and six in which four different observers noted the same animal on a day. For the
159 1079 pairs, the two estimates were highly correlated, and in the vast majority of unmolted
160 or fully molted animals, the two observers agreed exactly (Fig. A1). Among the 141 with
161 > 2 observations, there were 117 where the entire set was identical; the average standard
162 deviation across scores within a set was 1.06% molt.

Figure 1: Consistency of molt score using 1220 cases in which two or more different observers scored an animal on the same day over 2016-2022. The x-axis is one of the two scores, the y-axis the other. At the top right, there were 645 cases where both observations were 100% molt, and the bottom left includes 278 where both were 0% (points are jittered slightly, otherwise those would all appear as a single point). There were 32 cases, the red crosses, where observers were opposite, including 26 where one observation was 0% and the other 100%. The line is the regression ($r^2 = 0.986$). Removing all points at exactly 0% or 100%, scores were still consistent ($r^2 = 0.815$).
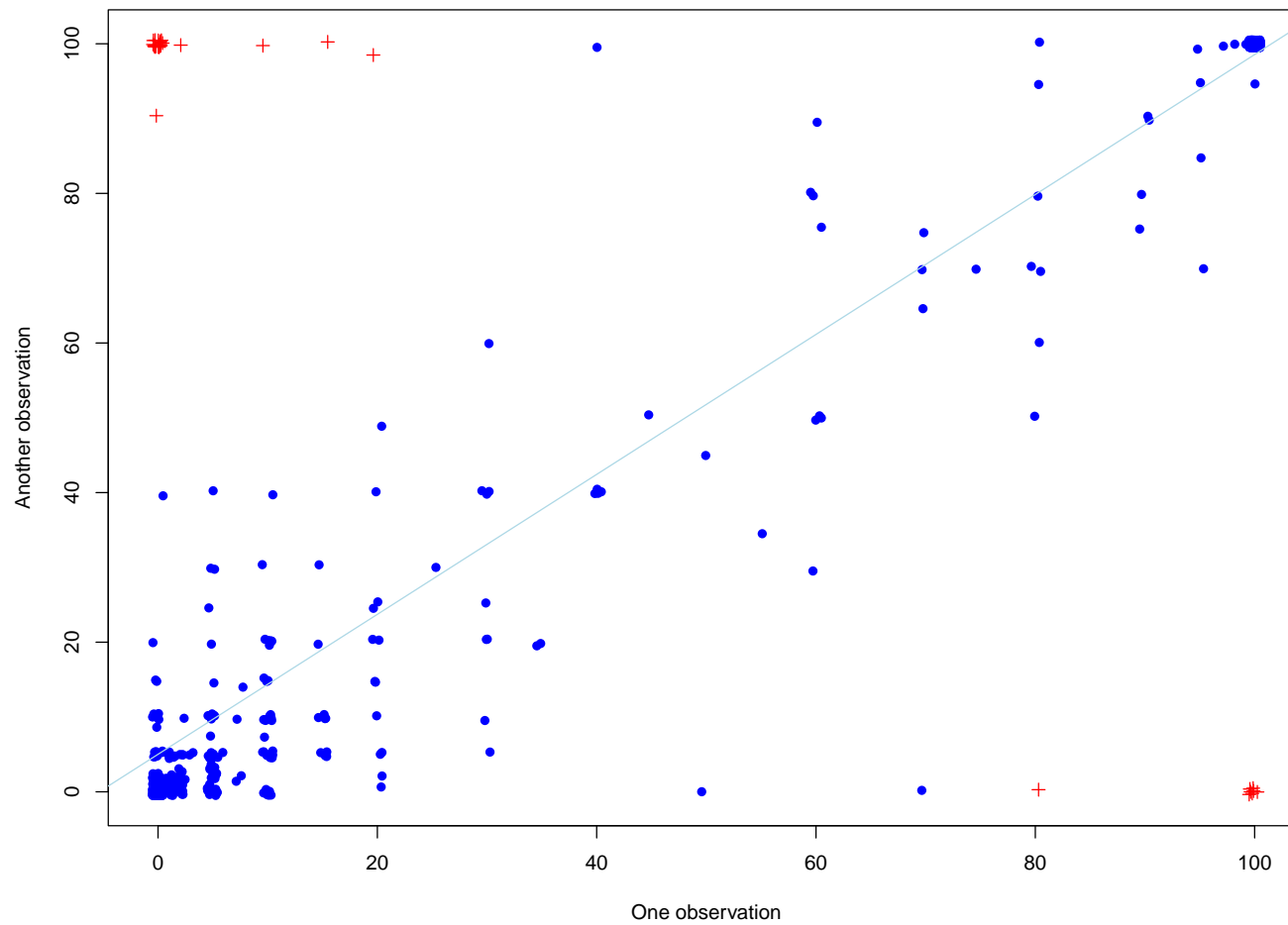
163 **Supplemental References**

Complete model fits

Following are the logistic fits to every adult female's molt sequence in 2020 and 2021.