Linking northern fur seal dive behavior to environmental variables in the eastern Bering Sea

Ruth Joy,[1],† Michael G. Dowd,[2] Brian C. Battaile,[3,5] Pamela M. Lestenkof,[3,6] Jeremy T. Sterling,[4] Andrew W. Trites,[3] and Richard D. Routledge[1]

[1]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6 Canada
[2]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia B3H 4R2 Canada
[3]Marine Mammal Research Unit, University of British Columbia, Vancouver, British Columbia V6T 1Z4 Canada
[4]National Marine Mammal Laboratory, Alaska Fisheries Science Center, Seattle, Washington 98115–6349 USA

Abstract.   Northern fur seals (*Callorhinus ursinus*) breeding on the Pribilof Islands, Alaska have declined dramatically over the past 40 years. Effective conservation of northern fur seals depends on understanding the foraging behavior of adult females whose foraging success is linked to pup survival. We determined the foraging behavior for 11 tagged lactating female northern fur seals from the Pribilof Islands using a state-space modeling approach with an autoregressive movement model. To interpret at-sea behavior in the context of oceanic habitat, we spatially and temporally matched high-resolution reconstructed tracks to a set of environmental covariates that included: commercial groundfish catch, sea surface temperature, primary productivity, wind speed, depth and time of day. We used a Bayesian hierarchical framework to implement a multinomial regression model to link behavior to environmental covariates and account for the mismatch of scale between fur seal behavior and the environmental variables by incorporating an error-in-covariates approach into the hierarchical model. The Bayesian framework allowed us to build a single model to synthesize the information from all the northern fur seal foraging tracks and the available information about the underlying environmental conditions. Application of the approach indicated that the behavioral states for the northern fur seal were significantly related to the Alaska commercial groundfish catch, particularly walleye pollock (*Gadus chalogramma*).

[5] Present address: Alaska Science Center, U.S. Geological Survey, Anchorage, Alaska 99508-4626 USA.
[6] Present address: Aleut Community of St Paul Island Tribal Government, Ecosystem Conservation Office, St. Paul Island, Alaska 99660 USA.
† **E-mail:** rutherfordjoy@gmail.com

INTRODUCTION

The population of northern fur seals (NFS) in the Pribilof Islands of Alaska has declined dramatically during the past 40 years, and

continues to decline without any obvious reason yet identified (Towell et al. 2006, Lee et al. 2014, Towell et al. 2014). Conservation efforts require a foundation of scientific understanding about NFS ecology, a key element of which is foraging

strategy. The foraging strategy of adult female NFS is of particular interest as success in finding enough prey to eat can be linked not only to adult survival but also to pup survival (Antonelis et al. 1997). For a lactating female, foraging success is a time-limited trial in which enough energy must be obtained from prey caught in a changing ocean environment to maintain a dependent and growing pup left behind at her natal rookery. Failure implies trouble meeting the energy needs of her pup and/or herself, and could affect their survival. The relationship between lactating NFS movement and habitat is shaped by foraging success and the physiological constraints of feeding a stationary, land-based pup. Understanding the relationship between movement, behavior, prey species, and the oceanic environment is a critical part of understanding population processes (Bowler and Benton 2005, Patterson et al. 2009). Indeed, the National Marine Fisheries Service's NFS conservation plan has identified this as a highest-level priority (NMFS 2007).

We examined a set of lactating northern fur seals that were tagged on the Pribilof Islands in the Bering Sea during the breeding season of 2005 and 2006 in order to link foraging behavior to environmental conditions and prey fields. Marine animal movement is a complex 3-D process that does not always simplify into lower dimensions, and there is mounting evidence to demonstrate the potential perils of inferring animal behavior based on horizontal trajectories alone (e.g., McClintock et al. 2013 and references therein). For instance, horizontal straightness indices (i.e., area restricted searches) poorly correlate with feeding behaviors (Austin et al. 2006, Weimerskirch et al. 2007). First passage time is a movement metric that measures animal passage through a horizontal region of fixed radius and has been linked to environmental variables, but it can be confounded by the slow speeds of resting behaviors, and the fast speeds of foraging behavior along tortuous paths; both behaviors lead to similar times to traverse a fixed radius region. Depth information has been widely used to determine dive type, but is open to subjectivity in determining what dive types or dive metrics to link to foraging behavior (e.g., Gentry et al. 1986, Goebel 2002). More recently, classification methods that use state-space mod-

els have been used to infer behavioral states (and switching) using horizontal location information (e.g., Jonsen et al. 2007). Extensions of standard state-space methods have been developed to link behavior to environmental covariate information (Dragon et al. 2012, Breed et al. 2012), but it remains a difficult problem to capture the vertical dimension of animal movement and its link to the environment. In this study, we used both horizontal and vertical movement information: high-resolution tracks were constructed from tag orientation and speed data, and depth information was used to classify fur seal behavior using the approach proposed by Dowd and Joy (2011).

The dominant prey in fur seal diets is juvenile walleye pollock (*Gadus chalogramma*; Perez and Bigg 1986, Sinclair et al. 1994, Zeppelin and Ream 2006, Zeppelin and Orr 2010), a species of groundfish that is commercially caught as adults. However, direct information on prey fields is difficult to collect and is not always available. This is also true for other environmental variables that affect foraging such as bathymetry (Antonelis et al. 1997, Call et al. 2008) and the shelf break (Goebel et al. 1991, Robson et al. 2004, Sterling and Ream 2004), lunar cycle (Ream et al. 2005), thermocline depth, and surface fronts (Nordstrom et al. 2013a, Sterling et al. 2014). The extent to which these variables influence prey fields is poorly understood (Ream et al. 2005, Kuhn et al. 2010).

Quantitatively linking behavior to changes in the environment at the landscape level is important for understanding foraging strategies and fur seal ecology. Bayesian hierarchical models have been suggested as an appropriate statistical framework for doing this (Schick et al. 2008). In this study, we developed and applied such a Bayesian hierarchical model. The primary advantage of the framework was that the uncertainty in both seal behavior and the environmental covariates could be fully accounted for by building a hierarchy of conditional models to describe the complexity of our data and the processes that generated them (Cressie et al. 2009). Our framework emphasizes the uncertainty in modeling behavior through the incorporation of errors-in-covariates, and population inference using individual information. While our application of this method was on the
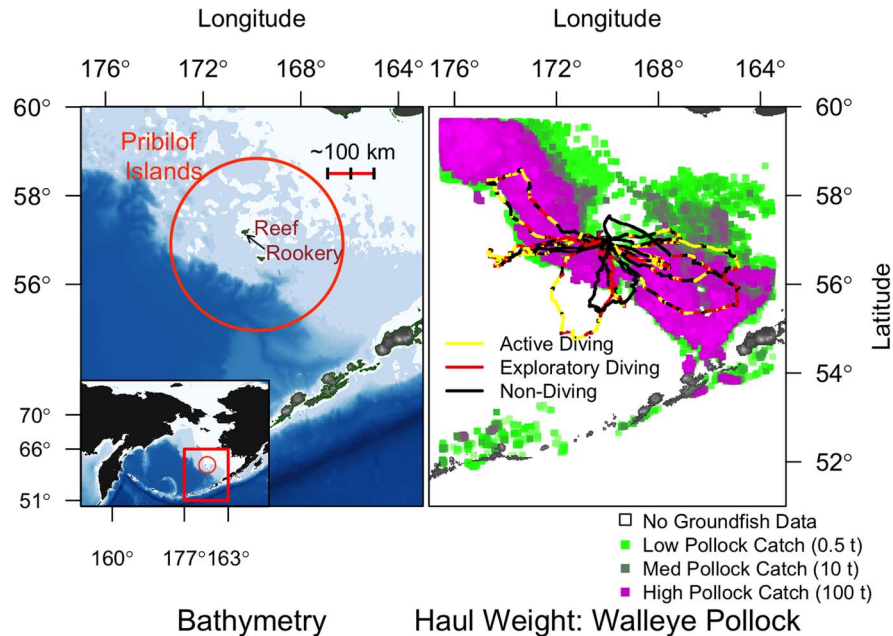
Fig. 1. The left panel shows Reef Rookery, St Paul Island, Pribilof Islands, Alaska. Reef Rookery is located on a peninsula that extends off the southern coast of St Paul Island. The intensity of blue represents bathymetry of the surrounding ocean. Adjacent waters are considered "on-shelf" and are typically <200 m. The right panel shows haul weight of walleye pollock calculated from the US Department of Commerce domestic observer data of the Alaska groundfish fishery for 2004–2007 date-restricted to be during northern fur seal summer pupping dates: July 9–November 11, 2004–2007. There is significant missing catch data off-shelf as there is no groundfish industry here, thus our analysis is biased towards on-shelf behaviors. Identified behaviors of northern fur seal along eleven at-sea foraging tracks of lactating female northern fur seals from St Paul Island, Alaska are also shown.

behavior of northern fur seals at a breeding site (rookery) on the Pribilof Islands, it is important to emphasize that our approach for analyzing animal movement in relation to the environment is adaptable for other populations and species.

MATERIALS AND METHODS

Our central goal was to understand how the ocean environment influences the behavior of a population of northern fur seals in the Bering Sea. We used biotelemetry data as well as external sources of information on the oceanic environment. The approach, as detailed below, included determining high-resolution positional information from the tag data, identifying environmental variables along the tracks, performing behavioral inference using a state-space model, and synthesizing all this information using a hierarchical Bayesian statistical model to quanti-

fy the relationship between environment and behavior.

Tag data

During the 2005 and 2006 breeding seasons, 18 lactating northern fur seals were captured at Reef Rookery on St Paul Island (57.18° N, 170.27° W; left panel of Fig. 1; 5 in 2005, 13 in 2006). Tags were attached mid-dorsally to each fur seal using methods described in Boyd and Croxall (1992). Three types of tags were used: (1) An archival dead-reckoner tag (Driesen & Kern GmbH, Bad Bramstedt, Germany), (2) An ARGOS (Advanced Research and Global Observation Satellite Platform Transmitter Terminal) satellite transmitter tag (Spot5, Wildlife Computers, Redmond, Washington, USA), and (3) A VHF radio transmitter tag (A2920 Glue On, Advanced Telemetry Systems, Isanti, Minnesota, USA). The dead-reckoner tags were 10-channel loggers with a

32-MB archive that recorded time, depth, speed (using a swim paddle system), temperature, light, pitch, roll, compass heading (in three-dimensions, 3-D), and body orientation (belly up or belly down). These were programmed to collect data at 2-second or 5-second intervals for the duration of the foraging trip (specifically, two of the 2005 tags were set to sample at 5-second intervals, and the remaining 2005 and all 2006 tags were programmed to sample at 2-second intervals). The satellite tag information (latitude and longitude) was used to calibrate dead-reckoner route calculations. The radio tag allowed the animals to be relocated when they returned to the rookery. All fur seals were recaptured in order to remove the tracking devices, and retrieve the time series data logged during their at-sea foraging trips.

Track reconstruction

Information from the dead-reckoner tags and the ARGOS transmitters were combined to produce high-resolution foraging tracks. The dead-reckoner channels recorded compass bearing, inclination, body orientation, depth, and speed-through-the-water for each female. We used the dead-reckoning channel vectors of compass bearing, speed, and depth to reconstruct the 2- and 3- dimensional swim paths of the fur seals (e.g., Wilson et al. 1993, Ropert-Coudert et al. 2002, Shiomi et al. 2008). We processed the tag's speed channel to correct for cumulative errors due to bias in the speed paddle's position. Zero-speed periods were identified as regions with near-zero variance, and linear interpolation was used to re-calibrate the speed record. Each foraging track was then corrected for any speed and direction offsets and ocean drift by constraining the dead-reckoner track to lie between ARGOS satellite locations which have limitations related to accuracy of location and infrequency of satellite locations per day (Loughlin et al. 1999, Patterson et al. 2008). We used only high-quality locations (ARGOS Location Classes 1, 2, 3 i.e., location error < 1.0 km; Mate et al. 1998) to compute speeds to limit the possibility of large location errors. We did not formally accommodate this source of errors in our track reconstructions as 1 km errors were deemed small relative to the finest scale covariate data (1 minute of a degree or 1.9-km resolution). We then translated

ARGOS locations into Universal Transverse Mercator (UTM) coordinates, and translated the dead-reckoner track into polar coordinates by rotating the angle of movement, and rescaling the radial coordinate to match the direction and great circle distance between consecutive ARGOS locations. When applied to all fur seals, this yielded a 2- or 5-second resolution track reconstruction for each that in-filled between the ARGOS satellite locations. In this way, the track was linked in space and time to the environmental conditions encountered while foraging at-sea. All location and movement analyses were conducted in UTM units and back-transformed to geographic coordinate units for presentation purposes.

Fur seal behavior

Fur seal behavior along a foraging track was determined using the state-space modeling approach proposed by Dowd and Joy (2011). The analysis proceeds by differencing the tag's depth channel to create a measure of vertical velocity. A vertical movement model based on a second order auto-regressive model

$$w_t = a_1 w_{t-1} + a_2 w_{t-2} + e_t$$

was used where w_t represents the vertical velocity, and a_1 and a_2 are the autoregressive parameters. The error process was assumed to be a zero-mean white noise Gaussian process with variance parameter σ_w^2. A unit time increment here corresponded to the 2- or 5-second resolution of the tag data. The data along each fur seal track is sectioned into 26-minute windows (with 13-minute overlap between adjacent windows) allowing us to identify approximately stationary ensembles of dives from which to determine behavior. For each time window, the parameters a_1, a_2 and σ_w^2 are estimated, taking into account observation error. This was accomplished using an augmented state-space model and multiple iterated filtering (see Dowd and Joy 2011 for full details of the methodology). The along-track parameter estimates were then further smoothed using a locally optimized kernel smoother (Herrmann 2013, R library "lokern") to identify longer period behavior, and compensate for short-term random parameter fluctuations.

The estimated along-track movement parameters derived from the tag data were then
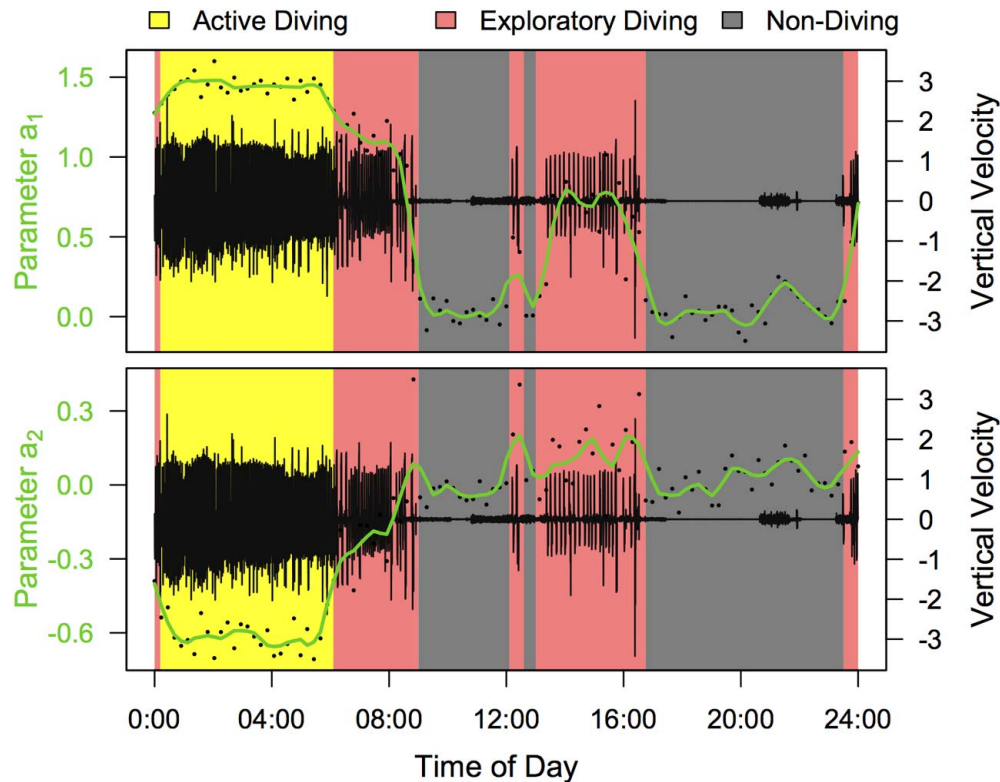
Fig. 2. Time series for a single day, August 18, 2006 for a single track. Shown are the estimated AR(2) model parameters a_1 and a_2; kernel smoothed (blue line) and original estimates (black dots). These are overlaid with the time series of vertical velocity (black lines). The yellow, pink and grey blocks correspond to active diving, exploratory diving, and non-diving, respectively (as diagnosed from values of a_1 and a_2).

classified into behavior types. We considered three discrete behavior classes: (1) non-diving; (2) active diving; and (3) exploratory diving. The non-diving behavior corresponded to near-zero values for the estimated process error variance σ_w^2, and includes sleeping and resting states, as well as surface transiting; these behaviors are characterized by a lack of engagement in the immediate environment. The behavioral states corresponding to active and exploratory diving are diagnosed from the estimated values for the auto-regressive parameters a_1 and a_2. Time series theory allows one to define the dynamic system behavior of an AR(2) model based on the values of a_1 and a_2 (Priestley 2004, Shumway and Stoffer 2006). We defined the a_1, a_2 values that corresponded to periodic behavior to represent active diving (i.e., regular and repeated dives). The other a_1 and a_2 values that corresponded to non-periodic behavior were identified as exploratory diving (i.e., less regular and intermittent dives).

See Dowd and Joy (2011) for further details.

Fig. 2 provides a concrete illustration of this behavioral classification using the movement parameters. It shows the vertical velocity data, the estimated values of a_1 and a_2, and the corresponding behavioral classifications for a single day (August 18, 2006) for one fur seal track showing a clear correspondence between the estimated parameter values and the behavioral type. Thus we have taken the output of a state space model that related movement characteristics in vertical speed to coherent bouts or segments of behavior. In the analysis that follows, we concatenate adjacent 26-minute windows of similar behavior into a single observation of the classified behavior that then acts as the response variable for that segment of track. For each fur seal, the series of behaviors observed along the length of track is then associated with a set of space and time positions in the Bering Sea that reflect the habitat where

the fur seal engaged in each behavior. This will be explicitly linked to the spatio-temporal environmental covariate information in the following section (*At-sea environmental conditions*), using the Bayesian hierarchical model (*Hierarchical Bayesian model* below).

At-sea environmental conditions

We considered a variety of available environmental data from the Bering Sea that could plausibly be linked to northern fur seal foraging behavior. We included information on fisheries (as a proxy for prey fields), as well as physical and biological oceanographic characteristics that are detailed below.

Commercial groundfish catch and walleye pollock.—A major goal of the study was to ascertain the extent (if any) to which fur seal behavior was associated with fish abundance and, in particular, walleye pollock density. As a proxy for fish abundance, we used the US Department of Commerce domestic observer data of the Alaska groundfish fishery for 2004–2007 (NMFS 2012). We limited the fish catch data to be between July 9, the beginning date of the breeding season (Trites 1992), and November 11, the median dispersal dates for pups on St Paul Island in 2005 (Lea et al. 2009). We spatially linked the fur seal tracks to individual groundfish haul at the nearest minute of longitude and latitude (≤1.9 km). We selected two variables of interest: (1) the haul weight of walleye pollock (85% of the total catch weight was walleye pollock), and (2) the total catch weight (including both retained and discarded species). Where multiple hauls were linked to a single track segment, we took the median pollock haul weight and total catch weight to represent that segment of track (right panel of Fig. 1). If no catch was linked to a segment of track (i.e., off-shelf regions), this was considered missing covariate data. Missing catch data was not in-filled with zeros so that the catch data could be used as a proxy for abundance of prey, rather than directly representing fisheries catch (and confounding issues of 0 catch with 0 effort).

Sea surface temperature.—Sea surface temperature (SST) is a primary oceanographic variable that is easily measured (Nordstrom et al. 2013b), and may influence the behavior of northern fur seals (Nordstrom et al. 2013a). Here, temperature was recorded directly by the dead-reckoner tag

at 2- or 5-second intervals. To isolate SST, we removed all the data that were observed outside of the top two meters of the ocean surface. The median sea surface temperature along each behavioral segment for each fur seal was extracted and assigned as an environmental covariate.

Primary productivity.—Ocean primary productivity was obtained from the NOAA CoastWatch net primary productivity for the Gulf of Alaska and Bering Sea. This is based on satellite-collected chlorophyll-a concentration and photo-synthetically available radiation (PAR) measurements, corrected for the amount of organic carbon used by planktonic organisms in respiration (Behrenfeld and Falkowski 1997). We used gridded spatial maps at 1/6th degree resolution (~18.5 km) processed as 8-day time-averages over the domain and period of interest. Primary productivity was assigned to the foraging track by linearly interpolating these data so as to match the time and location of the fur seal track.

Wind speed.—To account for the potential effect of sea-state on foraging behavior, wind speed was determined along the fur seal tracks. We used the National Climatic Data Center wind product, which blends satellite wind speeds from multiple platforms such as scatterometers, and passive microwave radiometers (Zhang et al. 2006). For each 24-hour period, we extracted daily wind speeds using the highest resolution 1/4 degree (~28 km) gridded wind fields. These were linearly interpolated to match the location of each behavioral segment at the relevant time.

Ocean depth.—The bathymetry of this region of the Bering Sea where the foraging tracks are located is characterized by shallow on-shelf waters <200 m deep, and off-shelf waters with ocean depths of 3000 m or more. Note the maximum dive depth of a northern fur seal is ~200 m (Gentry et al. 1986). Hence, we simplified bathymetry into an on-shelf/off-shelf categorical variable (We also tried the definitions of Call et al. (2008) that defined a three-level categorical depth variable: inner-middle shelf (0–100 m), outer shelf (100–200 m), and off-shelf (>200 m), but the additional resolution was not informative in our study). Under this definition, "off-shelf" waters include the deeper regions at the shelf break associated with foraging behavior in northern fur seals (Goebel et al. 1991, Robson

et al. 2004, Sterling and Ream 2004). We matched longitude and latitude of the track to the nearest depth measurement from the ETOPO1 bathymetry model (Amante and Eakins 2009), which gives bathymetry with resolution 1/60th of a degree, (~1.9 km). We then calculated the median depth for each of the behavior segments along the at-sea track. For example, if the seal was in an active diving mode between midnight and 4am, we took the median depth of the ocean during that period of the track, and classified it as on-shelf or off-shelf.

Time of day.—Time for each behavioral state was obtained from the dead-reckoner tag at the start of each behavioral segment. As northern fur seals have circadian patterns in behavior (Ream et al. 2005), we transformed the local time of day (*t*) into a circular variable of the form

$$\text{Time} = A\cos(\omega t) + B\sin(\omega t)$$

where $\omega = 2\pi/24$ hr. This ensured that the interpretation of time at 24:00 would be equivalent to that at 00:00. This provides a sinusoidal covariate function representing time of day, wherein the coefficients A and B are the focus of inference.

Hierarchical Bayesian model

We followed a hierarchical Bayesian modeling approach to examine possible relationships between behavior and the environmental covariates that describe fur seal at-sea habitat. This statistical methodology addresses the linking of behavior to habitat by decomposing the overall problem into a set of simpler and more tractable problems. Specifically, the complex joint probability distributions is expressed as a series of conditional models that aim to explain the relationship between data and a causal process as follows (see Berliner 1996, Cressie et al. 2009):

$$P(\text{Parameters}|\text{Data}) \propto P(\text{Data}|\text{Process, Parameters})$$
$$\times P(\text{Process}|\text{Parameters}) \times P(\text{Parameters})$$

where $P(\bullet)$ represents a probability distribution. The target distribution $P(\text{Parameters}|\text{Data})$ that embodies our solution allows us to determine the parameters of our model linking behavior to the environment, making use of all available data. The first level of the hierarchy, $P(\text{Data}|\text{Process, Parameters})$, is the likelihood function or observation model that describes the observation process. The second level, $P(\text{Process}|\text{Parameters})$, describes the state process model that links an individual's behavior to parameters; one can also model explicitly the uncertainty in the environmental covariates at this level. The third, or highest, level is a parameter model, $P(\text{Parameters})$ that expresses our uncertainty and prior information on the parameters. Successful application of the hierarchical framework requires that careful attention be given to the specification of the uncertainty introduced at each of these levels as outlined below.

Observation model

Our observation model linked the set of observed outcomes, $\boldsymbol{y}_{ij}$, to their outcome probabilities, $\boldsymbol{p}_{ij}$, using a multinomial probability model. Specifically, the observation vector, $\boldsymbol{y}_{ij} = (y_{ij}^{(1)}, y_{ij}^{(2)}, \ldots, y_{ij}^{(K)})$, characterized the behavioral state of the *i*th fur seal over the *j*th segment of track. Each $y_{ij}^{(k)}$ for $k = 1$ to K was equal to 1 if the animal was observed in the *k*th behavioral state, and 0 otherwise. Correspondingly, the probability of observing the *i*th fur seal at the *j*th location engaged in the *k*th behavioral mode was written in vector form as $\boldsymbol{p}_{ij} = (p_{ij}^{(1)}, p_{ij}^{(2)}, \ldots, p_{ij}^{(K)})$.

Local process model

The process model linked the probability of being in a particular behavioral state, or the latent probability vector $\boldsymbol{p}_{ij}$, to the environmental covariates. This was done using a multinomial regression model (McCullagh and Nelder 1989) according to the following procedure. Let $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijp})^T$ be the vector containing the true values for each of the p covariates that characterize the environment of the *i*th fur seal over the *j*th segment of track (^{T} designates the vector transpose). We modeled the logarithm of the ratio of the probability of each category relative to that of a baseline category; this is done since the total probability is one, and hence one of the behavioral categories is redundant. We selected the most commonly observed behavior (here, *non-diving*) as our baseline category ($k = 1$), following Agresti (1990). The multinomial logit linear regression model is then

$$\log\left(\frac{P(y_{ij}^{(k)}) = 1}{P(y_{ij}^{(1)}) = 0}\right) = \boldsymbol{X}_{ij}\boldsymbol{\beta}_i^{(k)}$$

where $\boldsymbol{\beta}_i^{(k)} = \beta_{i1}^{(k)}, \beta_{i2}^{(k)} \ldots, \beta_{ip}^{(k)}$ is the ith fur seal's parameter vector of regression coefficients corresponding to behavior class k.

The process and observation models were combined by defining the joint likelihood of the three behavioral outcomes using the observation model and the multinomial regression, for all K track segments for the ith fur seal, can then be expressed as:

$$P(\boldsymbol{p}_i|\boldsymbol{X}_i, \boldsymbol{\beta}_i) = \exp\left\{ \sum_{k=2}^{K}\left(\boldsymbol{\beta}_i^{(k)}\left[\sum_{j=1}^{J_i} y_{ij}^{(k)} \boldsymbol{X}_{ij} \right] \right) - \sum_{j=1}^{J_i} \log\left(1 + \sum_{k=2}^{K} e^{\boldsymbol{X}_{ij}\boldsymbol{\beta}_i^{(k)}} \right) \right\}$$

where J_i is the total number of behavioral segments for the ith fur seal.

One key aspect of this study was the need to account for errors in the covariate data (Stephens and Dellaportas 1992, Carroll et al. 2006). The usual assumption in regression models is that all covariates have been measured without error, but violations of this assumption generate biased and inaccurate inferences (Gustafson 2003). In our case, these arose due to intrinsic measurement errors in the environmental variables, as well as the use of behavioral segments as the analysis units. These lead to errors of representativeness introduced through spatial and temporal averaging (i.e., aggregated data; see Gotway and Young 2002).

Here, we explicitly accounted for errors in the environmental covariates. To do so, we designate the observed value of the covariates along the behavioral segment as $\boldsymbol{W}_{ij}$, where $\boldsymbol{W}_{ij}$ is related to its true value, $\boldsymbol{X}_{ij}$, according to

$$\boldsymbol{W}_{ij} = \boldsymbol{X}_{ij} + \boldsymbol{U}_{ij}$$

where $\boldsymbol{U}_{ij}$ describes the observation error associated with the ith fur seal on the jth track segment and we assumed that $\boldsymbol{U}_{ij} \sim \mathcal{N}(0, \sigma_{iU}^2)$ so that the variance σ_{iU}^2 is specific to an individual fur seal. Note that this is a random effect model wherein the true values are determined through an error-prone but unbiased measure.

Errors in environmental covariates can be incorporated into the hierarchical model as a prior distribution for $\boldsymbol{X}_i$. We selected the multivariate normal distribution ($\mathcal{MVN}$) as the form of that prior; i.e.,

$$P(\boldsymbol{X}_i) \sim \mathcal{MVN}(\boldsymbol{W}_i, \sigma_{iU}^2 \boldsymbol{I}_{J_i})$$

where $\boldsymbol{I}_{J_i}$ denotes an identity matrix of dimension J_i. The full Bayesian specification also requires us to specify a prior on its variance σ_{iU}^2. Specifically, we select an independent conjugate inverse gamma distribution or $\sigma_{iU}^2 \sim I\mathcal{G}(a_U, b_U)$. The parameters we chose for this distribution were $a_U = 3$, and $b_U = 1$, and limited the sampling range for the prior to be within a factor of 5 standard deviations of the prior distribution mean. In practice, limiting the sampling range for the prior only affected a small number of cases for which there was limited information in the covariate data to estimate σ_{iU}^2. A detailed description of all Bayesian priors used in our hierarchical model appear in the Appendix.

The following environmental covariates were modeled with error: total catch weight of groundfish; haul weight of walleye pollock; primary productivity; sea surface temperature; and wind. Total catch weight and haul weight of walleye pollock were positively correlated, $r = 0.85$. Therefore, to limit the detrimental effects of multicolinearity, these two groundfish measures were never both incorporated into the same model. The following covariates were considered to be measured without error: bathymetry (on-shelf/off-shelf); and time of day.

Parameter model

The parameter model linked the lower-level models for each individual fur seal, and so may be viewed as a population model. The main idea is that by linking the regression models for the individual fur seals via a shared prior, we incorporate population variability (Cressie et al. 2009), and borrow model strength between animals (Ntzoufras 2009). This was done by selecting a joint prior for the lower-level regression parameters, $\boldsymbol{\beta}_i$. That is, the regression parameters specific to a single animal (e.g., $\boldsymbol{\beta}_i$) are thought of as a random sample of coefficients from a distribution of possible values, and the posterior of each $\boldsymbol{\beta}_i$ is a weighted mean of the animal's regression parameters and the overall population effect.

We assumed that each of the components of $\boldsymbol{\beta}_i$ was normally distributed, and the joint prior distribution of $\boldsymbol{\beta}_i$ was then a multivariate normal likelihood, i.e.,

$$P(\boldsymbol{\beta}_i) \sim \mathcal{MVN}(\boldsymbol{B}, \boldsymbol{\Sigma}_\beta). \qquad (1.1)$$

Again, the fully Bayesian specification requires priors to be specified for the higher-level parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_\beta$ of the $\mathcal{MVN}$ likelihood of Eq. 1.1. The following inverse Wishart (IW) and $\mathcal{MVN}$ distributions were selected as priors for these parameters,

$$P(\boldsymbol{\Sigma}_\beta) \quad \sim \quad IW(v_0, \boldsymbol{V}_0)$$

$$P(\boldsymbol{B}|\boldsymbol{\Sigma}_\beta) \quad \sim \quad \mathcal{MVN}\left(\boldsymbol{B}_0, \frac{1}{a_0}\boldsymbol{\Sigma}_\beta\right)$$

with hyperparameters $v_0 = m + 3$ (to ensure that $v_0 > m$), covariance matrix $\boldsymbol{V}_0 = v_0 \times \boldsymbol{I}_m$ and $a_0 = 0.01$ (to ensure the covariance matrix is uninformative). Here, $\boldsymbol{I}_m$ denotes an identity matrix of dimension m (i.e., $\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{I}_m$ are $m \times m$ matrices). Note that the priors for both $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_\beta$ were conjugate to the higher-level $\mathcal{MVN}$ likelihood in Eq. 1.1. Full derivations of conditional distributions appear in the Appendix.

Bayesian computation and statistical assessment

We have presented a hierarchical multinomial Bayesian error-in-covariate model to describe the relationship between a fur seal's behavior and her environment. To solve for the target posterior distribution, we implemented a Markov Chain Monte Carlo (MCMC) Metropolis-within-Gibbs sampling algorithm that was coded using R statistical software (R Core Development Team 2012). The MCMC was run for 1,000,000 iterations (discarding 10,000 iterations for burn-in). The partial correlation coefficient was used to determine the decimation rate for the MCMC chain (i.e., we fit autoregressive models of successively higher-orders until the lag suggested non-significant partial correlations between chain components). As a consequence, chains were thinned to every 50th iteration. To check robustness, we initialized the chains from three different starting points using different random seeds. Convergence was assessed by inspection of trace plots, and using the Gelman-Rubin statistic ($\hat{R}$) as modified by Brooks and Gelman (1998; using R library "coda") as a quantitative measure of convergence.

Model comparison and model adequacy

Model comparison and selection was undertaken using information-theoretic approaches. In our study, we were interested in understanding fur seal behavior in general and not simply the behavior of the sampled fur seals. Since inference around population parameters was our focus, we compared the likelihood of each model's population parameters. Towards this end, we used the following metrics: (1) Akaike Information Criterion to compare model fit (AIC, Akaike 1973); (2) the posterior mean deviance as a Bayesian measure of fit ($\overline{D(\boldsymbol{\beta})}$, Dempster 1974, Spiegelhalter et al. 2002); and (3) m_{DIC} as a diagnostic of leverage (Spiegelhalter et al. 2002). Note that these quantities provide useful comparisons between candidate models, but do not provide insight into model adequacy.

We examined model adequacy through distributional summaries. We compared posterior predictive distributions of replicated data (produced under the model assumptions) with the observed data (Rubin 1984, Gelman et al. 2004). Assessment of the overall goodness of fit used posterior predictive P values (p_{pp}, Meng 1994, Carlin and Louis 2000). Due to uncertain large sample properties of posterior predictive P values (see discussion in Bayarri and Berger 2000), we assumed that a P value around 0.5 indicated that the distributions of the replicated and actual data were close, while a value close to zero or one indicated strong differences between them (Gelman et al. 2004).

Sample size effects

We ran simulations to examine the effect of sample size (i.e., the number of tagged fur seals) on model inference. This was done by drawing sets of simulated fur seals with various sample sizes from the model's (population) posterior. The analysis was then repeated for synthetic data sets ranging from 5 to 100 fur seals. We then compared these posteriors to the original posterior via the Kullback-Leibler (K-L) divergence statistic (Kullback and Leibler 1951). This allowed us to quantify the discrepancy in posteriors that was a consequence of sample size. The K-L divergences were calculated using the R library "flexmix" (Leisch 2004).

RESULTS

Foraging tracks

Of the 18 lactating female northern fur seals captured at Reef Rookery on St Paul Island

Table 1. Diagnostics and information criteria for the various Bayesian hierarchical models fit. The final, selected model is noted in boldface.

Fitted model	AIC†	$D(\bar{\boldsymbol{\beta}})$‡	m§	m_{DIC}¶	p_{pp}#
Time only	346.5	351.3	6	31.9	0.80
Time and main effect models					
Time + log(Total catch)	290.1	349.9	8	38.9	0.55
Time + log(Pollock)	290.0	344.0	8	37.8	0.72
Time + On/Off shelf	298.7	344.5	8	40.8	0.84
Time + Primary productivity	382.0	344.0	8	40.2	0.04
Time interaction models					
Time × log(Total catch)	285.6	322.8	12	48.5	0.32
Time × log(Pollock)	**284.3**	**323.2**	**12**	**44.3**	**0.72**
Time × On/Off shelf	303.6	330.8	12	51.8	0.62
Time × Primary productivity	292.0	334.7	12	60.0	0.00
Time × log(Pollock) + On/Off shelf	332.3	314.1	14	66.2	0.51

† AIC measures model fit (smaller implies better fit).
‡ $D(\bar{\boldsymbol{\beta}})$ measures model adequacy (smaller implies better).
§ m represents the number of population parameters in the upper level model.
¶ m_{DIC} is an estimate of leverage.
# p_{pp} is a measure of model predictive power; a value greater than 0.05 implies there is no evidence model is predicting poorly.

during the pupping seasons of 2005 and 2006, 10 individuals yielded complete data sets from which the foraging paths and behavior information could be retrieved. One of these animals provided data from two complete foraging trips. The incomplete data records of the other animals were due to tag failures including: water entering the housing, the speed paddle breaking off, fish vertebrae lodging between the speed paddle and housing, and battery failure. Hence, our data set for the analysis comprised of 11 complete northern fur seal foraging tracks. Foraging trips ranged in length from 5.5 days to 11.2 days, with a median of 7 days spent at sea. An average of 40 ARGOS positional fixes and 404,400 archival data records were obtained per trip.

The right panel of Fig. 1 maps the track reconstructions of 11 fur seal at-sea trips (with colors representing behavioral classifications). The fur seals traveled an average maximum linear distance from St Paul Island of 279 km, with a maximum distance away of 391 km. In general, these fur seals dispersed from the rookery in all directions, except directly to the north. They covered a wide area of both on-shelf and off-shelf waters in the Bering Sea and showed no preference for foraging in any single area. Five animals stayed on the continental shelf, one animal moved along the shelf break foraging in the canyons, and five animals went across the shelf break into the deep water of the central Bering Sea. The one animal for which we had two successive trips swam to a similar off-shelf location in both her at-sea trips.

Behavioral inference

The track reconstructions of Fig. 1 show the results from the behavioral classification along the foraging tracks. The track segments corresponding to the three behavioral classes (active diving, exploratory diving, and non-diving) identified according to the state-space algorithm of Dowd and Joy (2011). Far from the rookery, a mix of behaviors with a preponderance of active diving can be seen for all tracks. There was relatively more exploratory diving nearer the rookery on the outgoing legs, while the near-rookery incoming legs are characterized by non-diving transiting behavior. Overall, the fur seals spent 35% of the time at sea engaged in either active (14%) or exploratory dives (21%). Fur seals spent 65% of time at the surface engaged in non-diving behaviors. The 11 foraging trips had a median of 41 segments of different behavior modes, with those animals taking longer foraging trips exhibiting a greater number of behavior changes.

Relating behavior to environment

We used the hierarchical Bayesian model (see *Materials and methods: Hierarchical Bayesian model*) to investigate the link between fur seal behavior and the environment. A forward selection based model building strategy was undertaken using

Table 2. Posterior summaries for higher level model coefficients, B, in the final, selected model (Time $\times$ log(Pollock); Table 1). Parameters with significant Bayesian P values (* $P < 0.05$) are noted in boldface.

Regression parameter	B†	Median $q_{0.50}$	Credible interval ($q_{0.025}$, $q_{0.975}$)	$\hat{R}$‡
Intercept	$B_{1:2}$	0.928	(−0.564, 2.473)	1.01
	$B_{1:3}$	0.649	(−0.983, 2.277)	1.00
cos(Time)	$B_{1:2}$	**−1.336***	**(−2.594, −0.179)***	**1.02**
	$B_{1:3}$	0.413	(−0.874,1.797)	1.00
sin(Time)	$B_{1:2}$	−1.146	(−3.216, 0.913	1.00
	$B_{1:3}$	−2.098	(−4.434, 0.159)	1.04
log(Pollock)	$B_{1:2}$	0.61	(−0.597, 1.897)	1.03
	$B_{1:3}$	0.478	(−1.011, 1.974)	1.03
cos(Time) $\times$ log(Pollock)	$B_{1:2}$	0.068	(−1.045, 1.131)	1.01
	$B_{1:3}$	0.257	(−1.005, 1.549)	1.01
sin(Time) $\times$ log(Pollock)	$B_{1:2}$	−1.239	(−2.866, 0.252)	1.00
	$B_{1:3}$	**−2.479***	**(−4.993, −0.339)***	**1.05**

† $B_{1:2}$ denotes the regression parameters corresponding to the logit response $\log\left(p_{ij}^{(2)}/p_{ij}^{(1)}\right)$, or the log odds of explanatory diving vs. baseline non-diving. $B_{1:3}$ denotes the regression parameters corresponding to the logit response $\log\left(p_{ij}^{(3)}/p_{ij}^{(1)}\right)$, or log odds of active diving vs. baseline non-diving).

‡ $\hat{R}$ is the Gelman-Rubin Bayesian measure of convergence; values near 1 implies good convergence of the MCMC chain.

the candidate information metrics for model adequacy and selection (see *Materials and methods: Bayesian computation and statistical assessment*). Detailed results are given in Table 1.

We first fit main effect models to each of our at-sea environmental covariates. We then fit a series of models that all included time of day (Time) and incorporated the other covariates one at a time (in an additive fashion). The time of day models fit with Pollock, Total Catch and Bathymetry (On/Off Shelf) were found to be the best subset according to the model selection criteria. There was no relation to Primary productivity. We failed to get satisfactory convergence ($\hat{R} \gg 1$) in the models with Sea surface temperature and Wind, and these results are omitted from Table 1.

We next fit a set of models that include Time and one other main effect, and a Time $\times$ main effect interaction. The two models with Time interacting with a measure of groundfish catch (both Pollock and Total Catch) had the best fit diagnostics. The model that included Time of day and Primary productivity failed the posterior predictive test, suggesting that the model did not fit the data well, despite having adequate higher-order parameter convergence in their respective MCMC's (i.e., $\hat{R} \approx 1$).

One additional step of model complexity was also considered. We fit a three-covariate model based on the best time-interaction model, and bathymetry (Time $\times$ Pollock + On/Off Shelf; Table 1). This model was penalized aggressively by the AIC statistic, suggesting the additional complexity was not warranted despite a fair goodness of fit to the data. Table 1 reports a subset of the results from the model fitting.

Our final selected model from this procedure included the main effects of time of day and (log) haul weight of walleye pollock and included an interaction term of these covariates: Time $\times$ log(Pollock) (bold line in Table 1). It is notable that the model that substituted Total Catch for Pollock was a similarly good model with only marginally poorer fit (Time $\times$ log(Total catch) in Table 1). Coefficient magnitudes, and signs were similar for both models. Detailed summary statistics for posterior distributions of the regression coefficients can be found in Table 2). Fig. 3 graphically depicts magnitude and 95% credible intervals for each of the population parameters, and shows significance for the main effect of time of day and the interaction of time of day and (log) haul weight of walleye pollock.

To illustrate the effect of incorporating errors in covariates, we also fit the final selected model but assumed the covariates had no error (Fig. 3). By including errors in covariates, the regression parameters for those covariates with error were shrunk towards zero, and tended to show a loss of precision (i.e., there is increase in the uncertainty estimates). Also, this figure shows the parameter estimates of exactly measured covariates (i.e., those not modeled with error such as time of day) were significantly biased when the covariate errors in the remaining covariates were
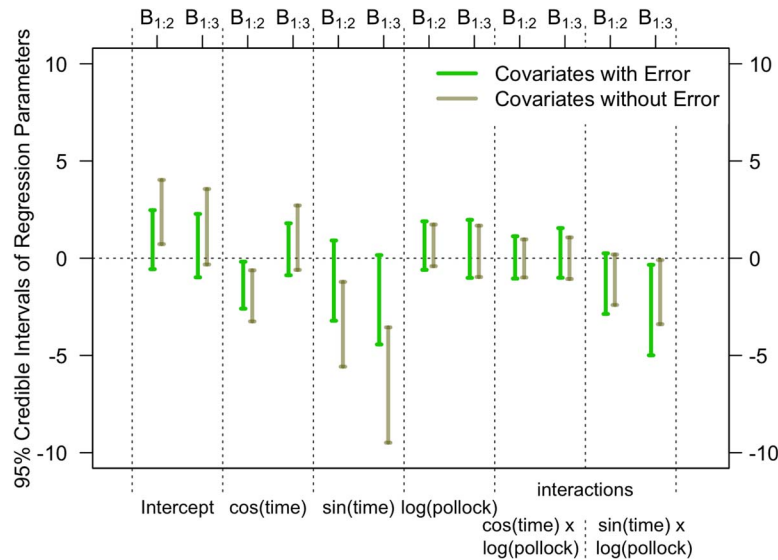
Fig. 3. The 95% credible intervals for 12 regression parameters $\boldsymbol{B}$ from the population level model that includes time of day and (logged) haul size of walleye pollock (Time × Pollock) modeled with and without error. Each pair of $\boldsymbol{B}$ coefficients show the credible intervals of the population-level parameters linking northern fur seal behavior to at-sea habitat. $B_{1:2}$ denotes the regression parameters corresponding to the logit response $\log\left(p_{ij}^{(2)}/p_{ij}^{(1)}\right)$, or the log odds of exploratory diving vs. baseline non-diving. $B_{1:3}$ denotes the regression parameters corresponding to the logit response $\log\left(p_{ij}^{(3)}/p_{ij}^{(1)}\right)$, or log odds of active diving vs. (baseline) non-diving.

ignored. The implication is that more of the parameters would be significant if error in the covariates had been omitted, and those parameters would be biased. However, the model fit diagnostics for the case of no errors in covariates were found to be significantly worse (without error: AIC = 297.0, $\overline{D(\boldsymbol{\beta})}$ = 329.4, p_{pp} = 0.46 vs. with error: AIC = 284.3, $\overline{D(\boldsymbol{\beta})}$ = 323.2, p_{pp} = 0.72).
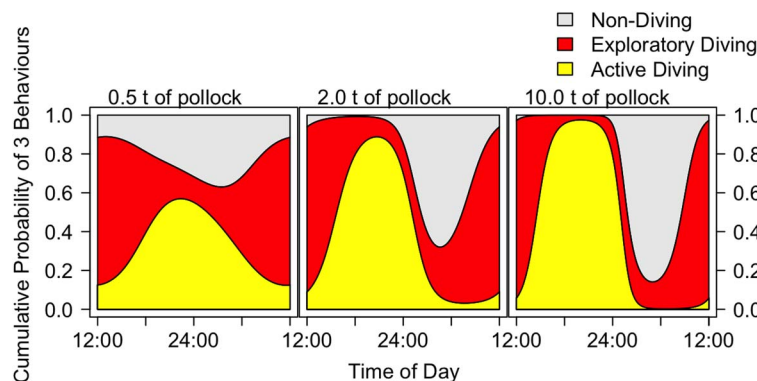
Fig. 4. Probabilities of behavior modes active diving, exploratory diving, and non-diving in response to time of day (shown for a 24-hour period starting at local noon), and increasing commercial catch size of walleye pollock. The left, middle and right panels of the graph show the predicted relative probabilities of behaviors in areas of small-sized walleye pollock hauls (0.5 tonnes), medium-sized hauls (2 tonnes), and large-sized hauls of pollock (10 tonnes). Note that the x-axis depicts time of day where the axis starts and finishes at noon to highlight the maximum amplitude of active foraging (in yellow) at night.

Fig. 5. Effect of increasing sample size on Kullback-Leibler (K-L) divergence for one selected population-level parameter (cos(T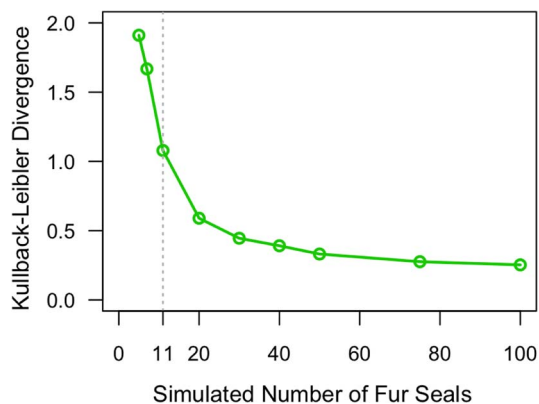ime) × log(Pollock): B_{13}). K-L divergence is 0 if and only if the posterior distribution for a simulated dataset is identical to the model from which it was generated. The dashed vertical line at 11 simulated fur seals corresponds to the sample size for this study.

This comparison leads to the conclusion that including covariate error in the model framework provides a significantly better model fit, as well as less biased parameter estimates.

The model results for the daily cycle of behavior in response to changes in time of day and commercial (log) haul size of walleye pollock are shown in Fig. 4. This figure highlights the increase in the relative probability of engaging in active diving behaviors in response to increasing size of pollock catch. This was accompanied by a near-zero probability of non-diving behaviors during those times of active diving in high catch areas. One caution in the interpretation of this figure is that our model did not predict the length of time that the northern fur seals engaged in particular behaviors, but rather predicted the probability of a behavior beginning at various times of the day. This figure suggests that the peak in the start of exploratory dive behaviors is around noon, and that fur seals are most likely to begin active diving behaviors at just past 8:00 in the evening, and most likely to end diving and begin non-diving behaviors around 5:00 in the morning. Comparing results for a haul size of 10 tonnes vs. 2 tonnes suggests there is a higher probability of starting active diving behaviors sooner in the evening, and hence a lower (in fact, near-zero) probability of non-dive behaviors at

that time.

Central to the Bayesian paradigm is the notion that as the data quantity (and quality) increase; the posterior becomes less sensitive to prior assumptions (Cressie et al. 2009). Fig. 5 shows a result from the sample size assessment using the Kullback-Liebler divergence metric, K-L. This indicates that the gains in model performance stabilize as the number of tagged fur seals exceeds 20, after which the K-L divergence becomes a relatively flat function. The figure therefore suggests that significant improvement in the inferences could have been gained if about twice as many complete records were obtained for analysis.

DISCUSSION

This study examined the at-sea foraging behavior of a population of lactating northern fur seals on St Paul Island, Alaska. Information on fur seal behavior was obtained from electronic tags for 11 complete foraging trips. Detailed 3-D tracks were reconstructed using high-resolution depth, speed and compass directional information from attached tags, calibrated to positional ARGOS fixes. Fur seal behavior was inferred with a state-space modeling approach using vertical velocity data and movement model; this was the first comprehensive application of the approach proposed by Dowd and Joy (2011) and allowed for successful identification of coherent behavioral segments (non-diving, active and exploratory diving) along the foraging tracks. A comprehensive hierarchical Bayesian framework was then developed to relate dive behavior to the marine environment in which they forage.

Our data were collected during the summer pupping seasons when female northern fur seals are tied to the rookery. Such restricted foraging constrains their ability to explore the foraging habitat in terms of both time spent away, and distance travelled. That is, they must select foraging habitat to obtain energy to feed themselves and their pup, while still returning to the rookery in time to feed a fasting pup. Female northern fur seals in our study embarked on relatively long pelagic journeys averaging 279 km in linear distance from the rookery, and lasting from 5.5 to 11.2 days. This distance is consistent with, but slightly longer than, the

linear distances reported in Nordstrom et al. (2013*a*) (228 km) and Goebel et al. (1991) (200 km). The length of time of the foraging trip is consistent with reported lengths from other studies (e.g., Gentry and Holt 1986, Nordstrom et al. 2013*a*; from 1 to 14 days). The tagged fur seals did not appear collectively to prefer any one region around St Paul Island, other than avoiding the continental shelf immediately to the north. Individuals may, however, have preferences: the one female for which we had two consecutive tracks went to a similarly located feeding area off-shelf in both foraging trips. Call et al. (2008) showed that 27 of 36 female fur seals with repeated trips followed the same general direction they used in their previous trips. Other studies have suggested that northern fur seals from Reef Rookery use all hydrographic domains around St Paul Island (e.g., Robson et al. 2004), including both on-shelf and off-shelf habitats (Loughlin et al. 1987, Goebel et al. 1991, Sterling and Ream 2004, Call et al. 2008). Scat samples from the same rookery contained both on-shelf (e.g., walleye pollock, Pacific herring) and off-shelf species (e.g., squid, myctophids; Zeppelin and Orr 2010). Our behavior maps showed constant movement and consistently changing diving behaviors throughout the at-sea foraging trips. Active and exploratory dive behaviors were found over much of the foraging paths, and non-diving transiting was most prevalent near the rookery, particularly on the return leg.

Our study suggests that the foraging behavior of lactating northern fur seals may be influenced by the abundance of walleye pollock. Walleye pollock of year-class 0 and 1 are the most common prey in scat studies of the northern fur seal diet (Perez and Bigg 1986, Sinclair et al. 1994, Sinclair et al. 1996, Zeppelin and Ream 2006), and walleye pollock comprised 89.3% of fur seal diet as measured in scats collected within a month of our 2006 tagged females foraging trips from Reef Rookery (Zeppelin and Orr 2010). We found that, in locations with more abundant walleye pollock, northern fur seals were more likely to be actively diving, especially at night. Our proxy for walleye pollock abundance was the US Department of Commerce domestic observer data of the Alaska groundfish industry aggregated over both time and space for the region of interest. It is unlikely that this reflects

the real-time abundance of prey that the fur seals actually encountered along their foraging route, and is confounded with fishing effort. However, it may still reflect persistent prey distributions that are attractive to foraging northern fur seals. Nordstrom et al. (2013*a*) also linked foraging of northern fur seals to oceanic surface fronts (eddies and filaments) and Kuhn (2011), to thermocline depth. These features may also be proxies for the spatial distribution of prey species.

While we used commercial groundfish haul data, we had no direct measure of bycatch of juvenile pollock, nor did we have any small scale evidence that commercial boats were fishing at the same time and location as our tagged fur seals. Hence, we cannot draw conclusions about the potential for competition occurring between the commercial fishery and northern fur seals. However, we can say that being in the same area where fishing occurred would not necessarily imply competitive interactions unless it was known that the fur seals were feeding on the same fish targeted by the fishery, and whether the abundance of that target prey was limited. Evidence from scat analysis suggests that juvenile pollock of year-class 0 and 1 (2–20 cm) are preferred by foraging fur seals (Zeppelin and Ream 2006), which are smaller than the >40 cm adults taken by the commercial fishery (Ianelli et al. 2007). As adult walleye pollock are known to prey on age-0 juveniles of their own species (Bailey 1989), it is reasonable to suggest that the link between the distribution of commercially harvested adult pollock and northern fur seals may be confounded due to their sharing of a common prey resource. Alternatively, Gudmundson et al. (2006) investigated diet overlap at a breeding rookery with commercially targeted year-classes of walleye pollock using prey remains in spews, and found significant overlaps with the fishery. Although the percentage of the diet that is regurgitated versus the percentage that is passed through the digestive tract is unknown, the spews indicate that northern fur seals do eat adult size-classes of walleye pollock.

Our results highlight the strong diel pattern in the foraging behavior of northern fur seals, and its link to prey abundance. We found that female fur seals showed strong preferences for active diving at night, while preferring non-dive be-

haviors such as resting or transiting in the mornings. Afternoons were typically associated with exploratory dive behaviors. Kuhn et al. (2010) proposed that shallow foraging bouts at night may be related to nighttime migrations of juvenile pollock as they follow vertically migrating zooplankton to the surface. Deeper foraging behavior has been associated with fur seals targeting concentrated groups of juvenile pollock at the thermocline (Kuhn 2011, Nordstrom et al. 2013a). We speculate that if northern fur seals were aware of suitable feeding areas (from prior experience or environmental cues), they could reduce search time and energetic costs by employing strategies (e.g., active or exploratory diving) at particular times along foraging paths that were successful on a previous feeding trip.

We used an auto-regressive movement model to link dive information to foraging behavior. This provided a mechanistic way of categorizing dive types, using interpretable movement signatures as diagnosed from the parameter values of our AR(2) vertical movement model. This contrasts with other classification schemes that distinguished dive types based on depth (e.g., Gentry et al. 1986, Goebel 2002), or used the torturous paths associated with prey patchiness (Benoit-Bird et al. 2013a, Benoit-Bird et al. 2013b). Parameters corresponding to periodic solutions of the AR(2) process were termed *active diving*. These included shallow, repeated dives with short surface-time intervals, as well as deeper, repeated *U*-shaped dives with comparatively longer surface recovery times. Such dive types have been attributed to foraging behaviors in an array of top marine predators such as gray seals (*Halichoerus grypus*; Austin et al. 2006), southern elephant seals (Gallon et al. 2013), Australian fur seals (*Arctocephalus pussillus*; Arnould and Hindell 2001), harbor seals (*Phoca vitulina*; Baechler et al. 2002), and others. Parameters corresponding to aperiodic solutions of the AR(2) process, or a correlated vertical random walk, were termed exploratory diving. That is, the typical surface time between dives is longer and more irregular relative to the dive time, and the vertical profile of the dive may be more V-shaped. It is less clear what the underlying motivation for intermittent V-shaped dives is, but others have attributed them to foraging on larger pelagic fish or squid in northern gannets (*Morus bassanus*; Garthe et al.

2000), or non-foraging activities, including predator avoidance, and explorations in crabeater seals (*Lobodon carcinophagus*; Bengston and Stewart 1992). The non-diving and non-foraging behaviors such as resting, sleeping, grooming and surface transiting were described, in the AR(2) solution space, as corresponding to periods for which vertical movement followed a white-noise process.

A general hierarchical data analysis and modeling framework is proposed here for interpreting marine animal tag data. It starts with track identification and behavioral inference using a state-space model, and then uses these results within a Bayesian hierarchal model to link fur seal behavior to the ocean environment. The analysis is spatially implicit (not explicit). That is, we did not attempt to answer the question of where in the ocean did the fur seals chose to forage or not forage. With only 11 tracks over an area of >125,000 km², this would have been too ambitious. Instead, we addressed the simpler question: given that the fur seal swam through a particular location, could we predict her most likely behavior based on a set of local environmental variables?

We note also that it is possible to consider the entire approach within the Bayesian hierarchical framework (e.g., Bestley et al. 2013). However, for our case it made sense to separate the components, doing the behavioral inference first and then the Bayesian hierarchical model. The main reason was the scale mismatch between the high-resolution movement information (seconds), and that of the behavioral inference and linkage to covariates (hours+). This implies that incorporating the state-space-based behavioral inference as part of the hierarchy of the Bayesian model, while conceptually appealing, would have been computationally infeasible due to the MCMC approaches employed. However, by taking the output from the state-space model as input to the hierarchical Bayesian model, there is some risk of misclassification of behavior.

Although it might be possible to write a model that has misclassification error in the response as well as measurement errors in the covariates, these models would have non-separable model errors. The effect of misclassification of a binary response in a logit regression model is to introduce a misclassification parameter that must

be additionally estimated (Neuhaus 1999) as well as to reduce significance of the naive estimates of regression parameters (Küchenhoff et al. 2006). Therefore in this model, the potential error in the classification of behavior would have the same effect as modeling the covariate error by shrinking regression parameters of model parameters towards zero.

Our Bayesian hierarchical model has some important features that should be incorporated into any study attempting to link animal behavior to environmental conditions. First, it focuses on population level inference using a collection of individuals. This was done by imposing a level in the hierarchy that accounted for individual variation. Second, errors in the environmental covariates were modeled explicitly. This is extremely important since not only are the variables themselves subject to measurement errors but they often have a complex error structure since they are frequently themselves data products derived from multiple sources. As well, there is also a great deal of uncertainty associated with spatially and temporally projecting them to match the fur seal tracks. In our study, this is the case for the groundfish catch data. Finally, we remark that since all our MCMC algorithms were coded directly, we had full control of the statistical model building process. This allowed us to tailor the Bayesian computational approaches to the specific problem at hand, and so account for the fact that these sampling-based algorithms are sensitive to assumptions made, and hence must be optimized for reliable posterior inference.

In summary, our data and analysis framework allowed us to link northern fur seal behavior to their ocean environment. The analysis could be further expanded. The sample size analysis suggests that having another 10 tracks would have generated enhanced precision. There are other potentially useful channels on the deployed tags that we have not made use of, particularly the high-resolution 3-D body orientation information that could provide further insight to behavior. We anticipate that other oceanographic information on, for example, frontal eddies and filament locations, could provide additional valuable insight into the association between fur seals and the Bering Sea environment. The general issues of scale, and temporal and spatial

dependence, remain outstanding issues in all such problems linking fine-scale movement and behavior to large-scale environmental variables. It can, however, be treated with an errors-in-covariate approach within a Bayesian hierarchical framework like the one used here. Finally, we feel that examining movement and foraging behavior of the maternal fur seal in relation to her pup weight gains, rookery residency times, and other life-history and bioenergetic characteristics would further contribute to the understanding of fur seal ecology and pup survival.

ACKNOWLEDGMENTS

LITERATURE CITED

Agresti, A. 1990. Categorical data analysis. John Wiley and Sons, New York, New York, USA.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Academiai Kiado, Budapest, Hungary.

Amante, C., and B. W. Eakins. 2009. ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis. NGDC-24. National Geophysical Data Center, NOAA, US Department of Commerce, Boulder, Colorado, USA.

Antonelis, G. A., E. H. Sinclair, R. R. Ream, and B. W. Robson. 1997. Inter-island variability in the diet of female northern fur seals (Callorhinus ursinus) in the Bering Sea. Journal of Zoology 242:435–451.

Arnould, J. P. Y., and M. A. Hindell. 2001. Dive behavior, foraging locations, and maternal attendance patterns of Australian fur seals (Arctocepha-

lus pusillus doriferus). Canadian Journal of Zoology 79:35–48.

Austin, D., W. D. Bowen, J. I. McMillan, and S. J. Iverson. 2006. Linking movement, diving, and habitat to foraging success in a large marine predator. Ecology 87:3095–3108.

Baechler, J., C. A. Beck, and W. D. Bowen. 2002. Dive shapes reveal temporal changes in the foraging behavior of different age and sex classes of harbour seals (Phoca vitulina). Canadian Journal of Zoology 80:1569–1577.

Bailey, K. M. 1989. Interaction between the vertical distribution of juvenile walleye Pollock (Theragra chalcogramma) in the eastern Bering Sea, and cannibalism. Marine Ecology Progress Series 53:205–213.

Bayarri, M. J., and J. O. Berger. 2000. P-values for composite null models. Journal of the American Statistical Association 95(452):1127–1142.

Behrenfeld, M. J., and P. G. Falkowski. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnology and Oceanography 42:1–20.

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

Benoit-Bird, K. J., B. C. Battaile, C. A. Nordstrom, and A. W. Trites. 2013a. Foraging behavior of northern fur seals closely matches the hierarchical patch scales of prey. Marine Ecology Progress Series 479:283–302.

Benoit-Bird, K. J., B. C. Battaile, S. A. Heppell, B. Hoover, D. Irons, N. Jones, K. J. Kuletz, C. A. Nordstrom, R. Paredes, R. M. Suryan, C. M. Waluk, and A. W. Trites. 2013b. Prey patch patterns predict habitat use by top marine predators with diverse foraging strategies. PloS ONE 8(1):e53348.

Berliner, L. M. 1996. Hierarchical Bayesian time-series models. Fundamental Theories of Physics 79:15–22.

Bestley, S., I. D. Jonsen, M. A. Hindell, C. Guinet, and J.-B. Charrassin. 2013. Integrative modeling of animal movement: incorporating in situ habitat and behavioral information for a migratory marine predator. Proceedings of the Royal Society B 280:1–19.

Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behavior to spatial dynamics. Biological Reviews 80:205–225.

Boyd, I. L., and J. P. Croxall. 1992. Diving behavior of lactating Antarctic fur seals. Canadian Journal of Zoology 70:919–928.

Breed, G. A. D. P., I. Costa, D. Jonsen, P. W. Robinson, and J. Mills-Flemming. 2012. State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecological Modeling 235:49–58.

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Call, K. A., R. R. Ream, D. Johnson, J. T. Sterling, and R. G. Towell. 2008. Foraging route tactics and site fidelity of adult female northern fur seal (Callorhinus ursinus) around the Pribilof Islands. Deep-Sea Research Part II 55:1883–1896.

Carlin, B., and T. A. Louis. 2000. Bayes and empirical Bayes methods for data analysis. Second edition. CRC Press, Boca Raton, Florida, USA.

Carroll, R. J., D. Rupert, L. A. Stefanski, and C. M. Crainiceanu. 2006. Measurement error in nonlinear models: a modern perspective. Second edition. CRC Press, Boca Raton, Florida, USA.

Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Dempster, A. P. 1974. The direct use of likelihood for significance testing. Pages 335–352 in O. Barndorff-Nielsen, P. Blaesild, and G. Schou, editors. Proceedings of the Conference on Foundational Questions in Statistical Inferences, Aarhus, Denmark, May 7–12, 1973. Department of Theoretical Statistics, University of Aarhus, Aarhus, Denmark.

Dowd, M. G., and R. Joy. 2011. Estimating behavioral parameters in animal movement models using a state-augmented particle filter. Ecology 92:568–575.

Dragon, A.-C., A. Bar-Hen, P. P. Monestiez, and C. Guinet. 2012. Horizontal and vertical movements as predictors of foraging success in a marine predator. Marine Ecology Progress Series 447:243–257.

Gallon, S., F. Bailleul, J.-B. Charrassin, C. Guinet, C.-A. Bost, Y. Handrich, and M. Hindell. 2013. Identifying foraging events in deep diving southern elephant seals, Mirounga leonina, using acceleration data loggers. Deep-Sea Research II 88:14–22.

Garthe, S., S. Benvenuti, and W. A. Montevecchi. 2000. Pursuit plunging by northern gannets (Sula bassana) feeding on capelin (Mallotus villosus). Proceedings of the Royal Society B 267:1717–1722.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis. Second edition. CRC Press, Boca Raton, Florida, USA.

Gentry, R. L., and J. R. Holt. 1986. Attendance behavior of northern fur seals. Pages 41–60 in R. L. Gentry and G. L. Kooyman, editors. Fur seals: maternal strategies on land and at sea. Princeton University Press, Princeton, New Jersey, USA.

Gentry, R. L., G. L. Kooyman, and M. E. Goebel. 1986. Feeding and diving behavior of northern fur seals. Pages 61–78 in R. L. Gentry and G. L. Kooyman, editors. Fur seals: maternal strategies on land and

at sea. Princeton University Press, Princeton, New Jersey, USA.

Goebel, M. E. 2002. Northern fur seal lactation, attendance and reproductive success in two years of contrasting oceanography. Dissertation. University of California, Santa Cruz, California, USA.

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

Gotway, C. A., and L. J. Young. 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97:632–648.

Gudmundson, C. J., T. K. Zeppelin, and R. R. Ream. 2006. Application of two methods for determining diet of northern fur seals (Callorhinus ursinus). Fishery Bulletin 104:445–455.

Gustafson, P. 2003. Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. CRC Press, Boca Raton, Florida, USA.

Herrmann, E. 2013. Lokern: kernel regression smoothing with local or global plug-in bandwidth. R package version 1.1-5. http://CRAN.R-project.org/package=lokern

Ianelli, J. N., S. Barbeaux, T. Honkalehto, S. Kotwicki, K. Aydin, and N. Williamson. 2007. Eastern Bering Sea walleye pollock. Pages 41–138 in Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions. North Pacific Fisheries Management Council, Anchorage, Alaska, USA.

Jonsen, I. D., R. A. Myers, and M. C. James. 2007. Identifying leatherback turtle foraging behavior from satellite-telemetry using a switching state-space model. Marine Ecology Progress Series 337:255–264.

Küchenhoff, H., S. M. Mwalili, and E. Lesaffre. 2006. A general method for dealing with misclassification in regression: the misclassification SIMEX. Biometrics 62:85–96.

Kuhn, C. E. 2011. The influence of subsurface thermal structure on the diving behavior of northern fur seals (Callorhinus ursinus) during the breeding season. Marine Biology 158:649–663.

Kuhn, C. E., Y. Tremblay, R. R. Ream, and T. S. Gelatt. 2010. Coupling GPS tracking with dive behavior to examine the relationship between foraging strategy and fine-scale movements in northern fur seals (Callorhinus ursinus). Endangered Species Research 12:125–139.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of the Institute of Statistical Mathematics 22:79–86.

Lea, M.-A., D. Johnson, R. Ream, J. Sterling, S. Melin, and T. Gelatt. 2009. Extreme weather events influence dispersal of naive northern fur seals. Biology Letters 5:252–257.

Lee, O. A., V. Burkanov, and W. H. Neill. 2014. Population trends of northern fur seals (Callorhinus ursinus) from a metapopulation perspective. Journal of Experimental Marine Biology and Ecology 451:25–34.

Leisch, F. 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. Journal of Statistical Software 11:1–18.

Loughlin, T. R., J. L. Bengtson, and R. L. Merrick. 1987. Characteristics of feeding trips of female northern fur seals. Canadian Journal of Zoology 65:2079–2084.

Loughlin, T. R., W. J. Ingraham, Jr., N. Baba, and B. W. Robson. 1999. Use of a surface-current model and satellite telemetry to assess marine mammal movements in the Bering Sea. Pages 615–630 in T. R. Loughlin and K. Ohtani, editors. Dynamics of the Bering Sea. University of Alaska Sea Grant, Fairbanks, Alaska, USA.

Mate, B. R., R. Gisiner, and J. Mobley. 1998. Local and migratory movements of Hawaiian humpback whales tracked by satellite telemetry. Canadian Journal of Zoology 76:863–868.

McClintock, B. T., D. J. F. Russell, J. Matthiopoulos, and R. King. 2013. Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. Ecology 94:838–849.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall/CRC, London, UK.

Meng, X. L. 1994. Posterior predictive P-values. Annals of Statistics 22:1142–1160.

Neuhaus, J. 1999. Bias and efficiency loss due to misclassified responses in binary regression. Biometrika 86:843–855.

NMFS. 2007. Conservation plan for the Eastern Pacific stock of northern fur seal (Callorhinus ursinus). National Marine Fisheries Service, Juneau, Alaska, USA.

NMFS. 2012. National Observer Program Annual Report—FY 2011. NOAA Technical Memorandum NMFS F/SPO–123. National Marine Fisheries Service, Juneau, Alaska, USA.

Nordstrom, C. A., B. C. Battaile, C. Cotté, and A. W. Trites. 2013a. Foraging habitats of lactating northern fur seals are structured by thermocline depths and submesoscale fronts in the eastern Bering Sea. Deep-Sea Research II 88:78–96.

Nordstrom, C. A., K. J. Benoit-Bird, B. C. Battaile, and A. W. Trites. 2013b. Northern fur seals augment ship-derived ocean temperatures with higher temporal and spatial resolution data in the eastern Bering Sea. Deep Sea Research Part II: Topical Studies in Oceanography 94:257–273.

Ntzoufras, I. 2009. Bayesian modeling using WinBUGS. John Wiley & Sons, Hoboken, New Jersey, USA.

Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn. 2009. Classifying movement behavior in relation to environmental conditions using hidden Markov models. Journal of Animal Ecology 78:1113–1123.

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. State-space models of individual animal movement. Trends in Ecology & Evolution 23:87–94.

Perez, M. A., and M. A. Bigg. 1986. Diet of northern fur seals, *Callorhinus ursinus*, off western North America. Fishery Bulletin 84:957–971.

Priestley, M. B. 2004. Spectral analysis and time series. Academic Press, London, UK.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ream, R. R., J. T. Sterling, and T. R. Loughlin. 2005. Oceanographic features related to northern fur seal migratory movements. Deep Sea Research Part II: Topical Studies in Oceanography 52:823–843.

Robson, B. W., M. E. Goebel, J. D. Baker, R. R. Ream, T. R. Loughlin, R. C. Francis, G. A. Antonelis, and D. P. Costa. 2004. Separation of foraging habitat among breeding sites of a colonial marine predation, the northern fur seal (*Callorhinus ursinus*). Canadian Journal of Zoology 82:20–29.

Ropert-Coudert, Y., A. Kato, Y. Naito, J. Baudat, A. Bost, and Y. Le Yaho. 2002. Swim speed of free ranging Adélie penguins *Pygoscelis adeliae* and its relation to the maximum depth of dives. Journal of Avian Biology 33:94–99.

Rubin, D. B. 1984. Bayesian justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics 12:1152–1172.

Schick, R. S., S. R. Loarie, F. Colchero, B. D. Best, A. Boustany, D. A. Conde, P. N. Halpin, L. N. Joppa, C. M. McClellan, and J. S. Clark. 2008. Understanding movement data and movement processes: current and emerging directions. Ecology Letters 11:1338–1350.

Shiomi, K., K. Sato, H. Mitamura, N. Arai, Y. Naito, and P. J. Ponganis. 2008. Effect of ocean current on the dead-reckoning estimation of 3-D dive paths of emperor penguins. Aquatic Biology 3:265–270.

Shumway, R. H., and D. S. Stoffer. 2006. Time Series analysis and its applications: with R examples. Third edition. Springer, New York, New York, USA.

Sinclair, E. H., G. A. Antonelis, B. W. Robson, R. R. Ream, and T. R. Loughlin. 1996. Northern fur seal, *Callorhinus ursinus*, predation on juvenile pollock, *Theragra chalcogramma*. NMFS-126:167–178 U.S. Department of Commerce, NOAA, Seattle, Washington, USA.

Sinclair, E. H., T. R. Loughlin, and W. Pearcy. 1994. Prey selection by northern fur seals (*Callorhinus ursinus*) in the eastern Bering Sea. Fishery Bulletin 92:132–156.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2002. WinBUGS user manual. Version 1.4. Medical Research Council Biostatistics Unit, Cambridge, UK.

Stephens, D. A., and P. Dellaportas. 1992. Bayesian analysis of generalized linear models with covariate measurement error. Pages 813–820 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian statistics 4. Oxford University Press, Oxford, UK.

Sterling, J. T., and R. R. Ream. 2004. At-sea behavior of juvenile male northern fur seals (*Callorhinus ursinus*). Canadian Journal of Zoology 82:1621–1637.

Sterling, J. T., A. M. Springer, S. J. Iverson, S. P. Johnson, N. A. Pelland, D. S. Johnson, M.-A. Lea, and N. A. Bond. 2014. The sun, moon, wind, and biological imperative–shaping contrasting wintertime migration and foraging strategies of adult male and female northern fur seals (*Callorhinus ursinus*). PloS ONE 9:e93068.

Towell, R. G., R. R. Ream, J. Bengston, and J. Sterling. 2014. 2014 northern fur seal pup production and adult male counts on the Pribilof Islands, Alaska. F/AKC3:RT. Alaska Fisheries Science Centre, National Marine Mammal Laboratory, Seattle, Washington, USA.

Towell, R. G., R. R. Ream, and A. E. York. 2006. Decline in northern fur seal (*Callorhinus ursinus*) pup production on the Pribilof Islands. Marine Mammal Science 22:486–491.

Trites, A. W. 1992. Reproductive synchrony and the estimation of mean date of birth from daily counts of northern fur seal pups. Marine Mammal Science 8:44–56.

Weimerskirch, H., D. Pinaud, F. Pawlowski, and C.-A. Bost. 2007. Does prey capture induce area-restricted search? A fine-scale study using GPS in a marine predator, the wandering albatross. American Naturalist 170:734–743.

Wilson, R. P., K. Puetz, C. A. Bost, B. M. Culik, R. Bannasch, T. Reins, and D. Adelung. 1993. Diel dive depth in penguins in relation to diel vertical migration of prey: Whose dinner by candlelight? Marine Ecology Progress Series 94:101–104.

Zeppelin, T. K., and A. J. Orr. 2010. Stable isotope and scat analyses indicate diet and habitat partitioning in northern fur seals *Callorhinus ursinus* across the eastern Pacific. Marine Ecology Progress Series 409:241–253.

Zeppelin, T. K., and R. R. Ream. 2006. Foraging habitats based on the diet of female northern fur seals (*Callorhinus ursinus*) on the Pribilof Islands, Alaska. Journal of Zoology 270:565–576.

Zhang, H. M., J. J. Bates, and R. W. Reynolds. 2006. Assessment of composite global sampling: Sea surface wind speed. Geophysical Research Letters 33. doi: 10.1029/2006GL027086

SUPPLEMENTAL MATERIAL

APPENDIX

Derivations of Target Distributions needed for the Hierarchical Bayesian Models

Likelihood

$$P(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{Z}_i) \propto \exp\left\{ \sum_j \boldsymbol{y}_{ij}(\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}) - \sum_j \log(1 + e^{\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}}) \right\}$$

Prior Distributions

$$\boldsymbol{\beta}_i \sim \mathcal{MVN}(\mathbf{1B}, \boldsymbol{\Sigma}_\beta)$$

$$\boldsymbol{X}_i \sim \mathcal{MVN}(\boldsymbol{W}_i, \sigma^2_{iU}\boldsymbol{I}_{J_i})$$

$$\sigma^2_{iU} \sim IG(\mathrm{a}_U, \mathrm{b}_U)$$

$$\boldsymbol{\Sigma}_\beta \sim I\mathcal{W}(\nu_0, \mathbf{V}_0)$$

$$\mathbf{B} \sim \mathcal{MVN}(\mathbf{B}_0, \frac{1}{\mathrm{a}_0}\boldsymbol{\Sigma}_\beta)$$

Posterior Distributions
Joint Posterior Distribution

$$P(\boldsymbol{\beta}, \boldsymbol{X}, \sigma^2_U, \mathbf{B}, \boldsymbol{\Sigma}_\beta|\boldsymbol{y}_i, \boldsymbol{Z}_i, \cdot)$$

$$\propto \prod_{i=1}^{n}(P(\mathbf{y}_i|\boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{Z}_i)P(\boldsymbol{X}_i|\boldsymbol{W}_i, \sigma^2_{iU}) \times$$

$$P(\boldsymbol{\beta}_i|\mathbf{B}, \boldsymbol{\Sigma}_\beta)P(\sigma^2_{iU}|\mathrm{a}_U, \mathrm{b}_U)) \times$$

$$P\left(\mathrm{B}, \boldsymbol{\Sigma}_\beta|\nu_0, \mathbf{V}_0, \mathbf{B}_0, \frac{1}{\mathrm{a}_0}\right)$$

$$\propto \prod_{i=1}^{n}\exp\left(\sum_{j=1}^{m_i}\boldsymbol{y}_i(\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}) - \sum_{j=1}^{m_i}\log(1 + e^{\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}})\right) \times$$

$$\frac{1}{\sqrt{(2\pi\sigma^2_{iU})}}\exp\left((2\sigma^2_{iU})^{-1}\sum_{j=1}^{m_i}(\boldsymbol{X}_{ij} - \boldsymbol{W}_{ij})^2\right) \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-n/2}\exp\left(-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{1B})\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}_i - \mathbf{1B})\right) \times$$

$$(\sigma^2_{iU})^{\mathrm{a}_U+1}\exp(-\mathrm{b}_U/\sigma^2_{iU}) \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-(\nu_0+p+1)/2}\exp\left(-\frac{1}{2}(\mathbf{V}_0\boldsymbol{\Sigma}_\beta^{-1})\right) \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{B} - \mathbf{B}_0)\left\{\frac{1}{\mathrm{a}_0}\boldsymbol{\Sigma}_\beta\right\}^{-1}(\boldsymbol{\beta}_i - \mathbf{1B})\right).$$

Lower-level posterior distributions.
Marginal Posterior Distribution of $\boldsymbol{\beta}$

$$P(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \mathbf{B}, \boldsymbol{\Sigma}_\beta)$$

$$\propto \exp\left\{\sum_{j=1}^{m_i}\boldsymbol{y}_{ij}(\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}) - \sum_{j=1}^{m_i}\log(1 + e^{\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}})\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-n/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{1B})\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}_i - \mathbf{1B})\right\}$$

Marginal Posterior Distribution of Covariate $\boldsymbol{X}_i$

$$P(\boldsymbol{X}_i|\boldsymbol{y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i, \boldsymbol{\beta}_i, \sigma^2_{iU})$$

$$\propto \exp\left\{\sum_{j=1}^{m_i}\boldsymbol{y}_{ij}(\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}) - \sum_{j=1}^{m_i}\log(1 + e^{\boldsymbol{\beta}_{iZ}\boldsymbol{Z}_{ij} + \boldsymbol{\beta}_{iX}\boldsymbol{X}_{ij}})\right\} \times$$

$$\frac{1}{\sqrt{(2\pi\sigma^2_{iU})^{m_i}}}\exp\left\{(2\sigma^2_{iU})^{-1}\sum_{j=1}^{m_i}(\boldsymbol{X}_{ij} - \boldsymbol{W}_{ij})^2\right\}$$

Marginal Posterior Distribution of σ^2_{iU}, Applying $IG(\mathrm{a}_U, \mathrm{b}_U)$ Prior

$$P(\sigma^2_{iU}|\boldsymbol{X}_i, \boldsymbol{W}_i, \mathrm{a}_U, \mathrm{b}_U)$$

$$\propto \frac{1}{\sqrt{(2\pi\sigma^2_{iU})^{m_i}}}\exp\left\{(2\sigma^2_{iU})^{-1}\sum_j(\boldsymbol{X}_{ij} - \boldsymbol{W}_{ij})^2\right\} \times$$

$$(\sigma^2_{iU})^{\mathrm{a}_U+1}\exp\{-\mathrm{b}_U/\sigma^2_{iU}\}$$

$$\propto \exp\left\{\frac{1}{\sigma^2_{iU}}\left[\mathrm{b}_U + \frac{1}{2}\sum_j(\boldsymbol{X}_{ij} - \boldsymbol{W}_{ij})^2\right]\right\}(\sigma^2_{iU})^{\mathrm{a}_U + \frac{m_i}{2} + 1}$$

such that

$$P(\sigma_{iU}^2 \mid \boldsymbol{X}_i, \boldsymbol{W}_i, a_U, b_U)$$

$$\sim IG\left(a_U + \frac{m_i}{2},\ b_U + \frac{1}{2}\sum_{j=1}^{m_i}(\boldsymbol{X}_{ij} - \boldsymbol{W}_{ij})^2\right).$$

Higher-level Conditional Posterior Distributions.
The higher-level priors $\mathbf{B}$, $\boldsymbol{\Sigma}_\beta$ are no longer a function of the data $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}$, but instead a function of the lower-level parameters $\boldsymbol{\beta}_i$, which are sufficient for the original data. Thus the prior for $\boldsymbol{\beta}_i$ becomes the likelihood of the higher level parameters, and since the dimension of $\mathbf{B}$ is $q \times p$, and in our application $q{=}1$ as the higher-level covariate matrix of independent variables contains only the intercept, we can write this likelihood as follows

$$P(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{11} \mid \mathbf{B}, \boldsymbol{\Sigma}_\beta, 1)$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-n/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{1}\mathbf{B})'(\boldsymbol{\beta}_i - \mathbf{1}\mathbf{B})\,\boldsymbol{\Sigma}_\beta^{-1}\right\}.$$

$$\text{(A.1)}$$

To suggest the form of the natural conjugate priors for $\boldsymbol{\Sigma}_\beta$ and $\mathbf{B}$, we factor the likelihood (A.1) into two parts

$$P(\boldsymbol{\beta} \mid \mathbf{B}, \boldsymbol{\Sigma}_\beta, \mathbf{1})$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-n/2}\exp\left\{-\frac{1}{2}(\bar{\boldsymbol{\beta}}_i - \mathbf{1}\bar{\boldsymbol{\beta}})'(\boldsymbol{\beta}_i - \mathbf{1}\bar{\boldsymbol{\beta}}) + \right.$$

$$\left. (\mathbf{B} - \bar{\boldsymbol{\beta}})'\mathbf{1}'\mathbf{1}(\mathbf{B} - \bar{\boldsymbol{\beta}})\boldsymbol{\Sigma}_\beta^{-1}\right\}$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-(n-1)/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{1}\bar{\boldsymbol{\beta}})'(\boldsymbol{\beta}_i - \mathbf{1}\bar{\boldsymbol{\beta}})\boldsymbol{\Sigma}_\beta^{-1}\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{B} - \bar{\boldsymbol{\beta}})'\mathbf{1}'\mathbf{1}(\mathbf{B} - \bar{\boldsymbol{\beta}})\boldsymbol{\Sigma}_\beta^{-1}\right\}$$

where $\bar{\boldsymbol{\beta}} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\beta}$, or the p-column averages of the $\boldsymbol{\beta}$ matrix, and substituting $\mathbf{S} = (\boldsymbol{\beta}_i - \mathbf{1}\bar{\boldsymbol{\beta}})'(\boldsymbol{\beta}_i - \mathbf{1}\bar{\boldsymbol{\beta}})$ gives the factored likelihood:

$$P(\boldsymbol{\beta} \mid \mathbf{B}, \boldsymbol{\Sigma}_\beta, \mathbf{1})$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-(n-1)/2}\exp\left\{-\frac{1}{2}\mathbf{S}\boldsymbol{\Sigma}_B^{-1}\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{B} - \bar{\boldsymbol{\beta}})'\mathbf{1}'\mathbf{1}(\mathbf{B} - \bar{\boldsymbol{\beta}})\boldsymbol{\Sigma}_\beta^{-1}\right\} \quad \text{(A.2)}$$

Marginal Posterior Distribution of $\boldsymbol{\Sigma}_\beta$ and $\mathbf{B}$

The first line of the factored likelihood in (A.2) suggests an Inverse Wishart (IW) kernel for the covariance matrix $\boldsymbol{\Sigma}_\beta$, and the second line suggests a multivariate normal kernel (MVN) for $\mathbf{B}|\boldsymbol{\Sigma}_\beta$. The prior on $\mathbf{B}$ is dependent on the scale parameter as $\boldsymbol{\Sigma}_\beta$ cannot be factored out of (A.2), thus the prior for $\mathbf{B}$ is a conditionally conjugate prior $P(\mathbf{B}|\boldsymbol{\Sigma}_\beta)$. The conjugate IW prior with hyperparameters v_0, $\mathbf{V}_0$ was selected for $\boldsymbol{\Sigma}_\beta$, and the conditional conjugate MVN with hyperparameters $\mathbf{B}_0, \frac{1}{a}\boldsymbol{\Sigma}_\beta$ were selected for $\mathbf{B}|\boldsymbol{\Sigma}_\beta$ as follows

$$\boldsymbol{\Sigma}_\beta \sim IW(v_0, \mathbf{V}_0)$$

$$\mathbf{B}|\boldsymbol{\Sigma}_\beta \sim MVN\left(\mathbf{B}_0, \frac{1}{a_0}\boldsymbol{\Sigma}_\beta\right).$$

The posterior distributions of $\mathbf{B}$, $\boldsymbol{\Sigma}_\beta$ are then derived

$$P(\mathbf{B}, \boldsymbol{\Sigma}_\beta | \cdot)$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-n/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_i - \mathbf{1}\mathbf{B})'(\boldsymbol{\beta}_i - \mathbf{1}\mathbf{B})\boldsymbol{\Sigma}_\beta^{-1}\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-(v_0+p+1)/2}\exp\left\{-\frac{1}{2}\mathbf{V}_0\boldsymbol{\Sigma}_\beta^{-1}\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-p/2}\exp\left\{-\frac{1}{2}(\mathbf{B} - \mathbf{B}_0)'(\mathbf{B} - \mathbf{B}_0)\left(\frac{1}{a_0}\boldsymbol{\Sigma}_\beta\right)^{-1}\right\}$$

$$\text{(A.3)}$$

We combine the terms involving $\mathbf{B}$, letting $\mathcal{W} = (1, \ldots, 1, \sqrt{a_0})'$, $\mathcal{Z} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n, \sqrt{a_0}\mathbf{B}_0)'$, and let

$$\begin{aligned}\tilde{\mathbf{B}} &= (\mathbf{1}'\mathbf{1} + a_0)^{-1}(\mathbf{1}'\mathbf{1}\bar{\boldsymbol{\beta}} + a_0\mathbf{B}_0) \\ &= (n + a_0)^{-1}(n\bar{\boldsymbol{\beta}} + a_0\mathbf{B}_0).\end{aligned}$$

Then (A.3) becomes

$$P(\mathbf{B}, \boldsymbol{\Sigma}_\beta | \cdot)$$

$$\propto |\boldsymbol{\Sigma}_\beta|^{-(v_0+p+1)/2}\exp\left\{-\frac{1}{2}\mathbf{V}_0\,\boldsymbol{\Sigma}_\beta^{-1}\right\} \times$$

$$|\boldsymbol{\Sigma}_\beta|^{-(n+p)/2}\exp\left\{-\frac{1}{2}(\mathcal{Z} - \mathcal{W}\tilde{\mathbf{B}})'(\mathcal{Z} - \mathcal{W}\tilde{\mathbf{B}})\boldsymbol{\Sigma}_\beta^{-1} + \right.$$

$$\left. (\mathbf{B} - \tilde{\mathbf{B}})'(\mathbf{B} - \tilde{\mathbf{B}})\boldsymbol{\Sigma}_\beta^{-1}\right\}$$

$$\propto |\mathbf{\Sigma}_\beta|^{-(v_0+n+p+1)/2} \exp\left\{-\frac{1}{2}\mathbf{V}_0\mathbf{\Sigma}_\beta^{-1}\right.$$

$$\left. +\frac{1}{2}(\mathcal{Z}-\mathcal{W}\tilde{\mathbf{B}})'(\mathcal{Z}-\mathcal{W}\tilde{\mathbf{B}})\mathbf{\Sigma}_\beta^{-1}\right\}\times$$

$$|\mathbf{\Sigma}_\beta|^{1/2}\exp\left\{\frac{1}{2}(\mathbf{B}-\tilde{\mathbf{B}})'(n+a_0)(\mathbf{B}-\tilde{\mathbf{B}})\}\mathbf{\Sigma}_\beta^{-1}\right\}.$$

Through the combining of terms, the posterior distribution for $\mathbf{B}|\mathbf{\Sigma}_\beta$ and $\mathbf{\Sigma}_\beta$ is a product of a $\mathcal{MVN}$ and IW, i.e.,

$$\mathbf{\Sigma}_\beta \sim IW\left(v_0+n, \mathbf{V}_0+(\mathcal{Z}-\mathcal{W}\tilde{\mathbf{B}})'(\mathcal{Z}-\mathcal{W}\tilde{\mathbf{B}})\right)$$

$$\sim IW\left(v_0+n, \mathbf{V}_0+\{(\mathbf{\beta}_i-\mathbf{1}\tilde{\mathbf{B}})'(\mathbf{\beta}_i-\mathbf{1}\tilde{\mathbf{B}})\right.$$

$$\left. +(\tilde{\mathbf{B}}-\mathbf{B}_0)'a_0(\tilde{\mathbf{B}}-\mathbf{B}_0)\}\right)$$

and

$$\mathbf{B}|\,\mathbf{\Sigma}_\beta \sim \mathcal{MVN}\left(\tilde{\mathbf{B}}, \frac{\mathbf{\Sigma}_\beta}{(n+a_0)}\right),$$

where $\tilde{\mathbf{B}}=\frac{n\,\bar{\mathbf{\beta}}+a_0\,\mathbf{B}_0}{n+a_0}$, and $\bar{\mathbf{\beta}}=(\mathbf{1}'\mathbf{1})^{-1}\,\mathbf{1}'\mathbf{\beta}$ are consistent with the definitions given above.