A MULTIVARIATE MIXED HIDDEN MARKOV MODEL FOR BLUE WHALE BEHAVIOUR AND RESPONSES TO SOUND EXPOSURE[1]

BY STACY L. DERUITER[*,†], ROLAND LANGROCK[‡,†], TOMAS SKIRBUTAS[†],
JEREMY A. GOLDBOGEN[§], JOHN CALAMBOKIDIS[¶],
ARI S. FRIEDLAENDER[‖,**] AND BRANDON L. SOUTHALL[**]

*Calvin College[*], University of St Andrews[†], Bielefeld University[‡],
Stanford University[§], Cascadia Research Collective[¶],
Oregon State University[‖] and SEA, Inc.[**]*

Characterization of multivariate time series of behaviour data from animal-borne sensors is challenging. Biologists require methods to objectively quantify baseline behaviour, and then assess behaviour changes in response to environmental stimuli. Here, we apply hidden Markov models (HMMs) to characterize blue whale movement and diving behaviour, identifying latent states corresponding to three main underlying behaviour states: shallow feeding, travelling, and deep feeding. The model formulation accounts for inter-whale differences via a computationally efficient discrete random effect, and measures potential effects of experimental acoustic disturbance on between-state transition probabilities. We identify clear differences in blue whale disturbance response depending on the behavioural context during exposure, with whales less likely to initiate deep foraging behaviour during exposure. Findings are consistent with earlier studies using smaller samples, but the HMM approach provides a more nuanced characterization of behaviour changes.

1. Introduction. Measuring and describing behaviour changes by marine mammals in response to underwater noise is a key step in understanding the potential harmful effects of this type of human disturbance on marine ecosystems [Ellison et al. (2012), Shannon et al. (2015)]. The effects of mid-frequency military sonars, airguns used for geophysical exploration, and shipping noise are of particular concern [National Research Council (2005), Southall et al. (2007)]. We focus on potential behaviour changes in response to military mid-frequency active sonar sounds (MFAS, broadly defined) and pseudo-random noise (PRN) in the same frequency range. Here, mid-frequency is broadly 1–10 kHz, but the narrower frequency band containing the primary energy of the MFAS and PRN stimuli is between 3–4.5 kHz. Previous work has suggested that blue whales respond simi-

larly to these MFAS and PRN sounds [Friedlaender et al. (2016), Goldbogen et al. (2013a)].

Controlled exposure experiments (CEEs) are experiments in which animals are tracked—either visually or using animal-borne tags—before, during, and after controlled exposure to the sound stimuli of interest. CEEs are an empirical way to obtain direct measurements of behaviour and potential behavioural responses [Tyack, Gordon and Thompson (2003)]. Analysis of CEE data then requires qualitative [Miller et al. (2012), Sivle et al. (2015)] or quantitative [Antunes et al. (2014), DeRuiter et al. (2013), Goldbogen et al. (2013a)] analysis of multivariate time-series data. These data often include many synoptic variables collected at very high temporal resolution. In some cases, particularly for species of great interest in terms of environmental management, datasets may include only a few individual animals [e.g., DeRuiter et al. (2013), Miller et al. (2015), Stimpert et al. (2014)]. With rapid technological advances in animal-borne sensors, there have been increases in the amount and complexity of these data. However, the development of quantitative analysis methods for these data has often lagged behind [Holyoak et al. (2008)].

In these types of analyses, it is particularly challenging to find methods that combine the relevant data streams in an objective, repeatable way, accounting for individual differences between animals, respecting dependence over time and between streams, and providing output that has a natural interpretation in relation to animal motivation (reasons why an animal performs certain behaviours) and behavioural state. The biological research goal is not simply to provide qualitative descriptions of behaviour, but to conduct hypothesis-driven research into the relationship between acoustic disturbance and behaviour. Given this focus, there is a particular need for models that summarise short-term observations in terms of behavioural states and animal motivation. Differences in behaviour due to a host of contextual variables such as environmental conditions, social conditions, or whale-to-whale variability can be of a magnitude equal to or larger than changes in response to acoustic disturbance [Ellison et al. (2012), Friedlaender et al. (2015)]. Consequently, an adequate model for basic behaviour (and individual variation in that behaviour) is often a prerequisite for assessing the effects of sound on behaviour. However, to date, few tools have been presented that allow such characterisation of multivariate animal behaviour time-series data in terms of behavioural states (even neglecting covariates such as acoustic disturbance). Such models could facilitate biological interpretation, and also provide inputs for models of Population Consequences of Disturbance (PCoD) [National Research Council (2005), Schick et al. (2013), New et al. (2013, 2014), King et al. (2015)].

Hidden Markov models (HMMs) are well suited to characterising multivariate time-series data in terms of a set of latent states [Zucchini, MacDonald and Langrock (2016), Zucchini, Raubenheimer and MacDonald (2008)], which in this case serve as proxies for the underlying behavioural states of the animals. HMMs have previously been applied to multivariate animal behaviour data in order to identify

latent behavioural states [Bagniewska et al. (2013), McKellar et al. (2014)], and somewhat analogous Bayesian hierarchical state-switching models have been used to achieve similar goals [Isojunno and Miller (2015), McClintock et al. (2013)], but no previous work has identified such states in animal behaviour while accounting for individual differences between tagged animals and incorporating effects of an environmental covariate modulating between-state transition rates. Here, we apply HMMs to describe a relatively large dataset on blue whale (*Balaenoptera musculus*) movement and diving behaviour, including a computationally efficient discrete random effect to account for differences between tag records. The model also incorporates covariates affecting the state transition probability matrix, enabling a quantitative analysis of the potential effects of experimental sounds (MFAS and PRN) on the between-state transition probabilities.

2. Blue whale CEE response measurements.

2.1. *Background*: *Research program and tags*.   Our dataset includes observations of 37 blue whales collected offshore of southern California, USA, as part of the Southern California Behavioral Response Study (SOCAL-BRS). SOCAL-BRS is an interdisciplinary research collaboration designed to study marine mammal behaviour and reactions to sound. This study is the first to include CEEs using MFAS transmissions from operational Navy vessels using full-scale sonar systems involved in previous marine mammal strandings, in addition to simulated sonar transmissions from transducer arrays deployed from research vessels. Its overall objective is to provide a better scientific basis for estimating risk and minimizing effects of active sonar for the U.S. Navy and regulatory agencies. The project's experimental methods have been described in detail elsewhere [Southall et al. (2012)], and additional methodological detail is also provided in Supplement A [DeRuiter et al. (2017a)].

Whales were tagged with animal-borne data loggers [DTAGs, Johnson and Tyack (2003), or, in 2 cases, B-probes, Greeneridge Sciences, Inc., Santa Barbara, CA] that recorded acoustic data and high-resolution animal-movement data. In total, the dataset includes 37 individual whales and 1054 dives, 168 of which overlapped with sound exposure periods (see Table 1 for details of tag deployments). In addition to the tag data, surface observations of animal positions were collected visually by observers following the whales in a small boat. Boat-based visual observations to determine spatial locations of animal surface positions were consistent throughout all phases of data analysed (before, during, and after CEEs).

2.2. *Controlled exposure experiment protocols*.   Data were collected before, during, and after 21 CEEs, in which 31 of the 37 whales were exposed to either MFAS or PRN sounds (Table 1). Tag data and visual observations were collected for a minimum of one hour before and after exposures, unless precluded by

TABLE 1

Summary of tag deployments on blue whales analysed in this study. Abbreviations used in this table: *ID, identification*; *Dur., duration*; *h, hours*; *CEE, controlled exposure experiment*; *Simul., simulated*; *MFAS, mid-frequency active sonar*; *PRN, pseudo-random noise*

Number	Whale ID	Date	Dur. (h)	Dives	CEE Dives	Exposure
1	bw10_235a	2010-08-23	1.4	26	7	Simul. MFAS
2	bw10_235b	2010-08-23	1.9	31	9	Simul. MFAS
3	bw10_235_Bprobe_019	2010-08-23	1.5	24	9	Simul. MFAS
4	bw10_238a	2010-08-26	1.5	26	9	Simul. MFAS
5	bw10_239b	2010-08-27	6.0	53	5	Simul. MFAS
6	bw10_240a	2010-08-28	1.6	54	16	Simul. MFAS
7	bw10_240b	2010-08-28	2.7	16	3	Simul. MFAS
8	bw10_241_Bprobe_034	2010-08-29	2.3	27	8	Silent
9	bw10_241a	2010-08-29	2.9	16	3	Silent
10	bw10_243a	2010-08-31	4.8	22	3	PRN
11	bw10_243b	2010-08-31	4.4	22	3	PRN
12	bw10_244b	2010-09-01	1.6	11	2	PRN
13	bw10_244c	2010-09-01	2.4	17	6	PRN
14	bw10_245a	2010-09-02	5.2	36	3	PRN
15	bw10_246a	2010-09-03	4.0	27	2	Simul. MFAS
16	bw10_246b	2010-09-03	4.4	15	2	Simul. MFAS
17	bw10_251a	2010-09-08	3.0	26	7	PRN
18	bw10_265a	2010-09-22	5.2	44	3	Simul. MFAS
19	bw10_266a	2010-09-23	2.5	17	5	PRN
20	bw11_210a	2011-07-29	3.3	17	3	Simul. MFAS
21	bw11_210b	2011-07-29	4.9	33	5	Simul. MFAS
22	bw11_211a	2011-07-30	3.7	24	2	PRN
23	bw11_213b	2011-08-01	4.9	51	6	Simul. MFAS
24	bw11_214b	2011-08-02	4.9	93	10	PRN
25	bw11_218a	2011-08-06	2.6	17	4	PRN
26	bw11_218b	2011-08-06	3.6	23	2	PRN
27	bw11_219b	2011-08-07	2.1	10	0	None
28	bw11_220a	2011-08-08	1.1	6	3	Simul. MFAS
29	bw11_220b	2011-08-08	2.1	15	6	Simul. MFAS
30	bw11_221a	2011-08-09	4.3	33	4	PRN
31	bw11_221b	2011-08-09	2.8	17	4	PRN
32	bw12_292a	2012-10-18	4.4	14	3	PRN
33	bw13_191a	2013-07-10	4.8	41	6	Real MFAS
34	bw13_207a	2013-07-26	3.4	37	8	Silent
35	bw13_214b	2013-08-02	3.5	21	0	None
36	bw13_217a	2013-08-05	2.3	13	0	None
37	bw13_259a	2013-09-16	5.4	67	7	Simul. MFAS

unplanned early tag detachment, darkness, or when the animal was lost. During CEEs, individual sound stimuli (about 1.5 seconds in duration) were transmitted by a custom-built underwater transducer array [Southall et al. (2012)] approxi-

mately once every 25 seconds for 30 minutes total (or 58 minutes of transmissions from a US Navy vessel for the single real MFAS CEE). For further details, consult Supplement A [DeRuiter et al. (2017a)].

2.3. *Whale behaviour data.* Using the high-resolution, multivariate tag data, a number of variables were chosen to summarise the whales' behaviour: dive duration, post-dive surface duration, maximum depth, number of feeding lunges, and variability of whale heading. Two additional variables were based on the visual observations of spatial locations: step length and turning angle in the horizontal dimension. Supplement A [DeRuiter et al. (2017a)] provides details of data processing procedures. All variables were computed on a dive-by-dive basis; in other words, the input data for modelling were time series for which the sampling unit was one dive. (Here, a "dive" was defined as any excursion from the surface to 10 m depth or greater.) Figures 1–2 show example time-series plots of the data for two of the 37 whales. Similar plots for all whales are included in Supplement B [DeRuiter et al. (2017b)].

3. Modelling the state-switching behaviour using HMMs.

3.1. *A baseline model.*

3.1.1. *Motivation.* The data suggest that at least two (and probably three) distinct behavioural states drive the magnitude of the observations made [Goldbogen et al. (2013b)]. For example, the deepest, longest dives usually contain multiple lunges (feeding events), and represent deep feeding behaviour; there are also shorter, shallow feeding dives that usually contain a single lunge; finally, some relatively short, shallow dives without lunges have very low heading variance, and may represent near-surface travelling. Furthermore, there is autocorrelation in the observed time series, due to persistence of behavioural states, with animals tending to repeat the same type of dive several times before switching to a different behavioural state.

HMMs naturally accommodate both these features. A basic N-state HMM in discrete time involves two components: (1) an observed state-dependent process and (2) an unobserved N-state Markov chain, with the observations of the state-dependent process assumed to be generated by one of N component distributions corresponding to the N states of the Markov chain. In HMM applications to animal behaviour data, the states of the Markov chain can often naturally be interpreted as proxies for the behavioural states of animals, although there is not necessarily a one-to-one correspondence between nominal HMM states and biologically meaningful behavioural states [Langrock et al. (2012, 2015), Zucchini, Raubenheimer and MacDonald (2008)].
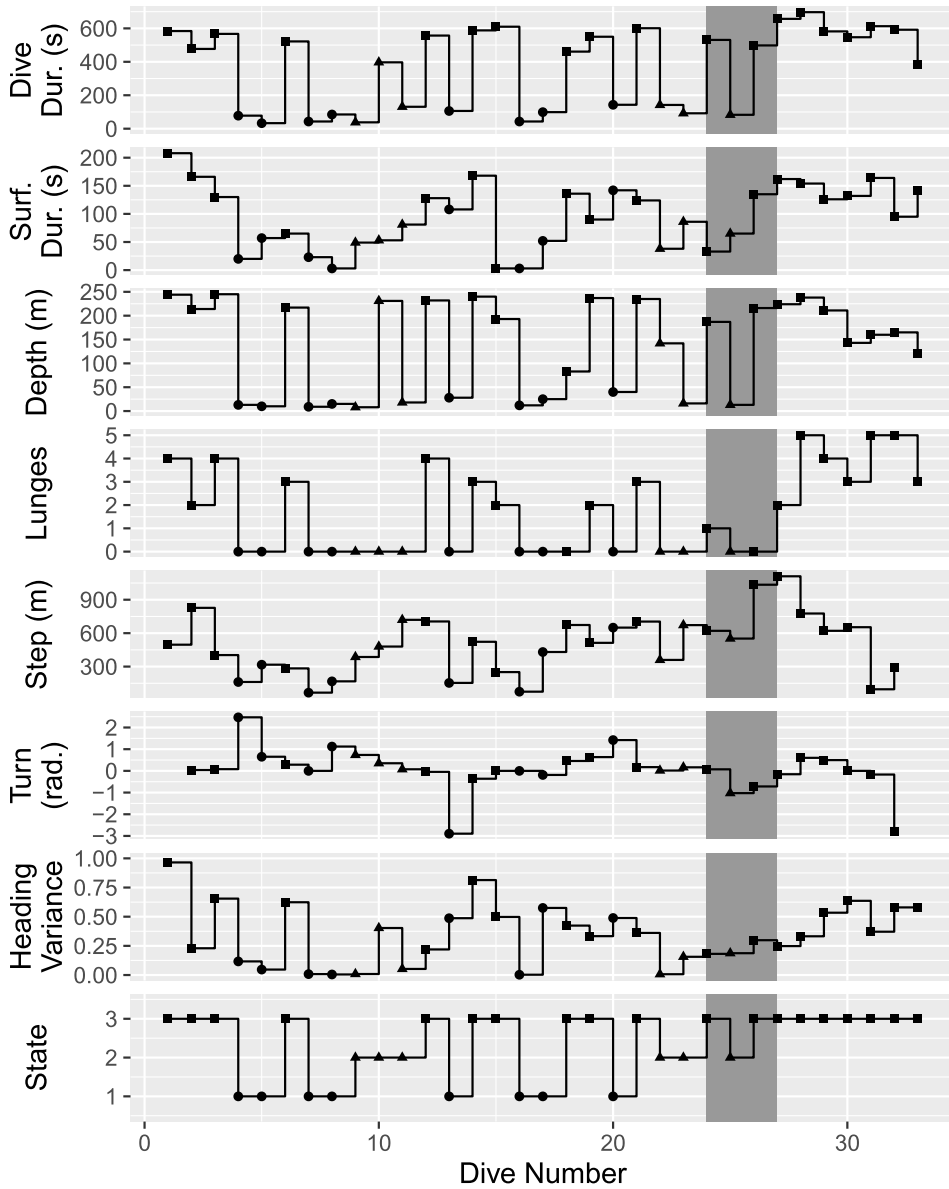
FIG. 1.   *Time-series plot of the input data for whale number* 30. *The CEE exposure period is shaded in darker grey*; *this whale was exposed to a pseudo-random noise* (*PRN*) *CEE. The bottom panel shows the most probable state for each dive according to the best model* (*with* 3 *states*, 4 *contexts*, *and a common effect of acoustic disturbance*). *These states are also indicated in the other panels by symbols*: *circles for state* 1, *triangles for state* 2, *and squares for state* 3.
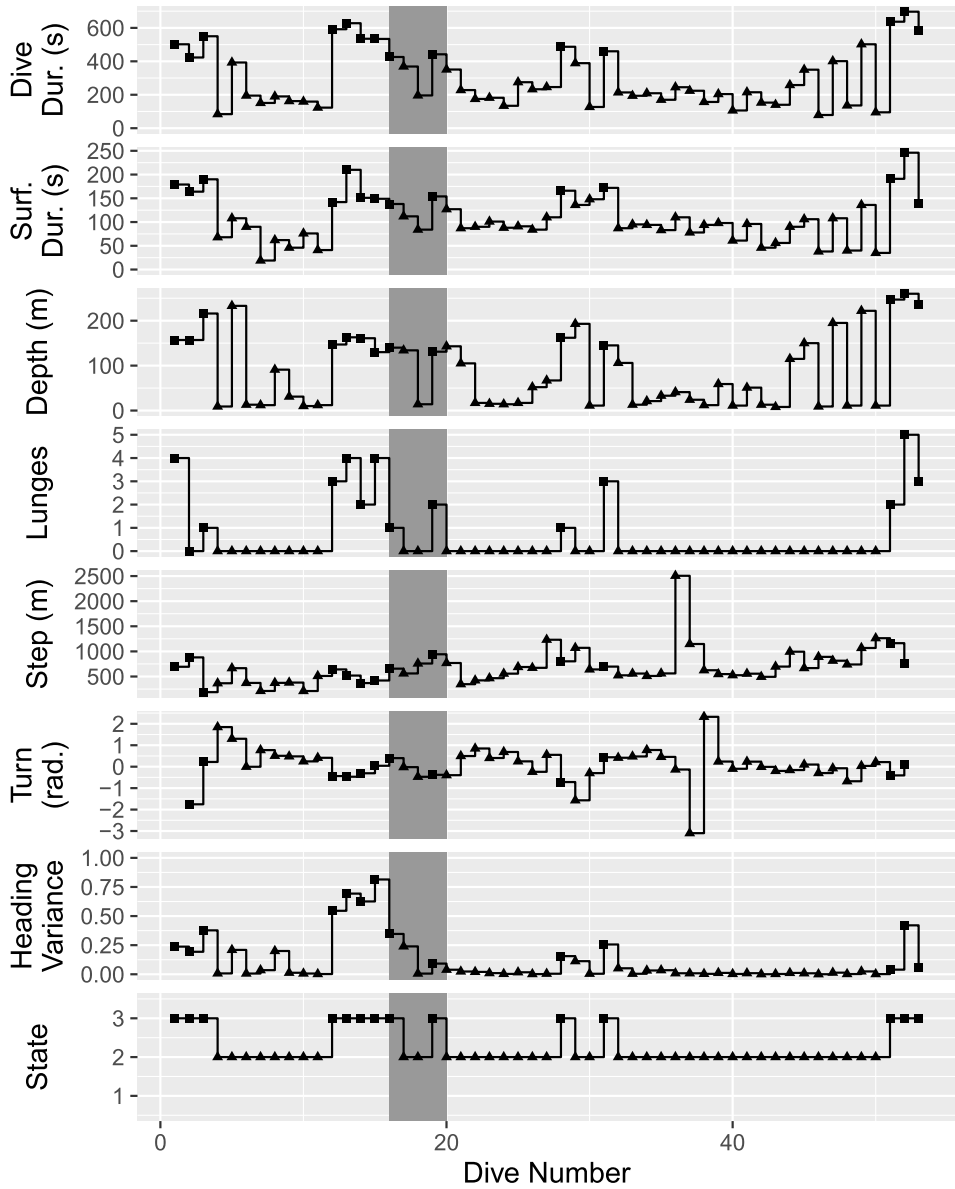
Fig. 2. *Time-series plot of the input data for whale number* 5. *The CEE exposure period is shaded in darker grey; this whale was exposed to a simulated MFA sonar CEE. The bottom panel shows the most probable state for each dive according to the best model (with 3 states, 4 contexts, and a common effect of acoustic disturbance). These states are also indicated in the other panels by symbols: circles for state* 1, *triangles for state* 2, *and squares for state* 3.

3.1.2. *Formulation of the initial model.* For each whale, the observed state-dependent process is multivariate and will be denoted by $\{\mathbf{X}_{wd\cdot}\}_{d=1,2,\ldots,D_w}$, where $\mathbf{X}_{wd\cdot} = (X_{wd1}, \ldots, X_{wdP})$ is the vector of variables observed for dive d performed by whale w, and D_w is the total number of dives performed by that whale. In our application, we have $P = 7$ data streams. The N-state Markov chain giving the (behavioural) states that underlie the dives performed by whale w will be denoted by $\{S_{wd}\}_{d=1,2,\ldots,D_w}$, where S_{wd} gives the state associated with dive d. Thus, with our model, we classify dives into a finite number of categories [Bagniewska et al. (2013), Hart et al. (2010)], rather than modelling diving activity at a finer, within-dive scale [Langrock et al. (2014)]. We will focus on models with $N = 3$ states, with a justification of this choice to be provided below.

We assume a basic first-order dependence structure for the state process, and, for whale w, summarize the state transition probabilities $\gamma_{ij}^{(w)} = \Pr(S_{wd} = j \mid S_{w,d-1} = i)$ in the transition probability matrix (t.p.m.)

$$\mathbf{\Gamma}^{(w)} = (\gamma_{ij}^{(w)}) = \begin{pmatrix} \gamma_{11}^{(w)} & \cdots & \gamma_{1N}^{(w)} \\ \vdots & \ddots & \vdots \\ \gamma_{N1}^{(w)} & \cdots & \gamma_{NN}^{(w)} \end{pmatrix}.$$

We initially assume that $\mathbf{\Gamma}^{(w)} = \mathbf{\Gamma} = (\gamma_{ij})$ for all w (i.e., complete pooling of the individual whales' time series: all whales have the same state-switching dynamics), but this assumption will be relaxed later on. The initial state distribution, denoted by (the row vector) $\boldsymbol{\delta} = (\Pr(S_{w1} = 1), \ldots, \Pr(S_{w1} = N))$, is also initially assumed to be common to all individuals, and, like $\mathbf{\Gamma}$, is estimated from the data.

To allow use of an unconstrained numerical optimizer to find the maximum likelihood estimates, we reparametrise all model components in terms of unconstrained parameters. In particular, we let

$$\gamma_{ij} = \frac{\exp(\alpha_{ij})}{1 + \sum_{l \neq i} \exp(\alpha_{il})},$$

and perform the maximization of the likelihood with respect to the real-valued parameters α_{ij}, $i, j = 1, \ldots, N$, $i \neq j$. This reparametrisation based on the multinomial logit function will be convenient also when extending the model to incorporate random effects and covariates (see subsequent sections).

We assume that the different time series (whales) observed are independent of each other. For the state-dependent processes, we assume that, given the state underlying dive d, the observation vector $\mathbf{X}_{wd\cdot}$ is conditionally independent of past and future observations and states. Furthermore, for the observation vector $\mathbf{X}_{wd\cdot}$, we assume *contemporaneous conditional independence* given the states, that is,

$$f(\mathbf{x}_{wd\cdot} \mid S_{wd} = s_{wd}) = \prod_{p=1}^{P} f(x_{wdp} \mid S_{wd} = s_{wd}).$$

Note that we use f as a general symbol for a density or probability function. Unconditionally, the $P = 7$ component variables will still be dependent on each other because the Markov chain induces dependence between them. For the blue whale data considered here, we found that the correlation implied by the HMMs fitted under the contemporaneous conditional independence assumption was slightly lower than the empirical correlation between the component variables [see Supplement B, DeRuiter et al. (2017b)]. This indicates that the assumption is adequate, since our interest does not centre precisely on the correlation between the data streams, but on the state-switching dynamics. In addition, there is generally no practical alternative to contemporaneous conditional independence: unless a multivariate normal distribution can be assumed (which rarely is the case), there is usually no simple multivariate distribution available to specify the correlation structure between variables within states. For more details on the assumption of contemporaneous conditional independence, see Zucchini, MacDonald and Langrock (2016).

As a consequence of the contemporaneous conditional independence assumption, to formulate a model for the blue whale dataset, we simply need to specify one univariate distribution for each of the seven variables observed. (The parameters of each distribution then depend on the underlying state.) We model the dive depth, dive duration, post-dive surface duration and step length with gamma distributions, since those variables all take on positive real values and may have right-skewed distributions. For the gamma distributions, we employ a parametrisation in terms of the mean μ and standard deviation σ; if required, these parameters can be easily converted to the more standard shape (k) and scale (θ) parametrisation according to $k = \mu^2/\sigma^2$ and $\theta = \sigma^2/\mu$. For the number of lunges, we use a Poisson distribution, a standard choice for count data. We select a von Mises distribution—which is a circular analogue of the normal distribution—for the turning angle data. Finally, we model the heading variance with a beta distribution, which, like the data, has support on the interval [0, 1].

3.1.3. *Model fitting.* A key property of HMMs is that an efficient algorithm, the forward algorithm, can be used to evaluate the likelihood. The forward algorithm exploits the dependence structure of the HMM to calculate the likelihood recursively, from the start to the end of the observed time series. At each time step the likelihood, as well as the probability of each state at the current time step, is updated based on the following: the t.p.m., the state-dependent distributions, and the probabilities of each state at the preceding time-step (using the initial state distribution for the first observed time-step). This renders it computationally feasible and convenient to estimate the model parameters by numerically maximising the likelihood [MacDonald (2014)]. For the model specified above—with both t.p.m., $\boldsymbol{\Gamma}$, and initial state distribution, $\boldsymbol{\delta}$, being identical across individuals—the use of

the forward algorithm leads to the likelihood expression

$$(1) \qquad \mathcal{L} = \prod_{w=1}^{37} (\boldsymbol{\delta} \mathbf{Q}(\mathbf{x}_{w1\cdot}) \boldsymbol{\Gamma} \mathbf{Q}(\mathbf{x}_{w2\cdot}) \cdots \boldsymbol{\Gamma} \mathbf{Q}(\mathbf{x}_{wD_w\cdot}) \mathbf{1}),$$

where $\mathbf{Q}(\mathbf{x}_{wd\cdot})$ is an $N \times N$ diagonal matrix with the kth entry on the diagonal given by $\prod_{p=1}^{P} f(x_{wdp} \mid S_{wd} = k)$ (i.e., the joint density of the observations $x_{wd1}, \ldots, x_{wdp}$, given state k), and $\mathbf{1} \in \mathbb{R}^N$ is a column vector of ones.

We note that an advantage of the HMM formulation described above is its ability to deal with missing and partially missing data. In ecological datasets in particular, on some occasions data may be available from some variables in the observed multivariate state-dependent process, but missing for other variables. For example, in the current dataset, observations for some whales span nightfall. After dark, surface visual observations ceased, and so all data derived from visually observed positions are missing. However, the tag-derived variables are recorded as usual regardless of darkness. As another example, some whales in the study were tagged with B-probes rather than DTAGs. The B-probe devices do not have tri-axial accelerometers and magnetometers, and so they do not collect any heading data. An HMM—with an assumption of contemporaneous conditional independence, as detailed earlier in Section 3.1.2— can easily accommodate such partially observed data in the state-dependent process. If any x_{wdp} (the observation of variable p for dive d by whale w) is missing, we simply have $f(x_{wdp} \mid S_{wd} = k) = 1$ in equation (1) [Zucchini, MacDonald and Langrock (2016)]; in consequence, the missing data point makes no contribution to $\mathbf{Q}(\mathbf{x}_{wd\cdot})$, and thus does not affect the overall likelihood, but observations of other variables for that same dive do still contribute.

Confidence intervals for the parameter estimates were obtained based on the inverse of the observed Fisher information, that is, the inverse of the Hessian of the negative log-likelihood at its minimum. Since the model was parametrised in terms of unconstrained parameters, the Fisher information provides approximate standard errors for those unconstrained parameters. Assuming approximate normality of the estimators, confidence intervals were then constructed for the unconstrained parameters, before transforming the resulting interval boundaries to the scale of the constrained parameters, that is, those of interest.

For the blue whale dataset, in the case of three states ($N = 3$), it took five minutes to numerically maximize this model's likelihood using `nlm()` in R [R Core Team (2015)] (on an octa-core i7 CPU, at 2.7 GHz and with 4 GB RAM).

3.1.4. *Results and interpretation of the states.* For the baseline model with $N = 3$ states, the resulting parameter estimates with associated 95% confidence intervals are presented in Table 2 and Figure 3. The estimated t.p.m. and the corresponding stationary distribution (as implied by the t.p.m.) are

$$\hat{\boldsymbol{\Gamma}} = \begin{pmatrix} 0.931(0.894, 0.952) & 0.014(0.005, 0.039) & 0.055(0.034, 0.087) \\ 0.018(0.005, 0.065) & 0.785(0.709, 0.839) & 0.197(0.141, 0.266) \\ 0.071(0.047, 0.108) & 0.100(0.070, 0.139) & 0.829(0.783, 0.864) \end{pmatrix}$$

TABLE 2

Parameter estimates and 95% confidence intervals (in parentheses) for the state-dependent distributions for the 3-state model with complete pooling. For each variable, the distribution fitted to the data is indicated in parentheses. An == symbol indicates that the relevant parameter was constrained to take on the specified value, rather than estimated

Variable	State 1	State 2	State 3
Duration (gamma)	$\mu = 140$ (132, 148)	$\mu = 334$ (305, 367)	$\mu = 516$ (503, 530)
	$\sigma = 80$ (73, 88)	$\sigma = 212$ (187, 241)	$\sigma = 130$ (120, 140)
Surface Duration (gamma)	$\mu = 70$ (64, 77)	$\mu = 86$ (78, 95)	$\mu = 151$ (144, 158)
	$\sigma = 68$ (61, 76)	$\sigma = 55$ (47, 65)	$\sigma = 69$ (63, 75)
Maximum Depth (gamma)	$\mu = 32$ (30, 35)	$\mu = 68$ (59, 79)	$\mu = 170$ (164, 176)
	$\sigma = 24$ (21, 26)	$\sigma = 65$ (55, 77)	$\sigma = 60$ (56, 65)
Step Length (gamma)	$\mu = 189$ (175, 204)	$\mu = 675$ (629, 725)	$\mu = 406$ (376, 439)
	$\sigma = 134$ (121, 149)	$\sigma = 305$ (268, 348)	$\sigma = 287$ (258, 318)
Turning Angle (von Mises)	$\mu == 0$	$\mu == 0$	$\mu == 0$
	$\kappa = 1.0$ (0.9, 1.2)	$\kappa = 3.1$ (2.5, 3.7)	$\kappa = 0.8$ (0.6, 1.0)
Heading Variance (beta)	$a = 1.0$ (0.8, 1.1)	$a = 0.5$ (0.4, 0.6)	$a = 1.7$ (1.5, 1.9)
	$b = 2.1$ (1.8, 2.4)	$b = 5.4$ (4.2, 7.1)	$b = 1.6$ (1.4, 1.8)
Number of Lunges (Poisson)	$\lambda = 0.7$ (0.6, 0.8)	$\lambda = 0.0$ (0.0, 0.1)	$\lambda = 3.4$ (3.2, 3.6)

and (0.433(0.311, 0.557), 0.199(0.134, 0.282), 0.368(0.283, 0.455)), respectively. The latter implies that, according to the fitted model, in the long run 43%, 20%, and 37% of the dives are made in states 1, 2, and 3, respectively. The initial state distribution was estimated as $\hat{\delta} = (0.169\ (0.075, 0.337), 0.171\ (0.076, 0.326), 0.660\ (0.475, 0.792))$. The discrepancy between estimated initial state distribution and stationary distribution implied by the t.p.m. indicates that the time of tagging is not independent of the behavioural state process of the individuals, with many more initial dives being allocated to state 3 than would be the case under stationarity. We note that the confidence intervals obtained all seem plausible. However, given the rough surface of the likelihood, with many local maxima, some caution is warranted when considering and interpreting such curvature-based uncertainty quantification.

This simple model succeeds in capturing several of the important features of the data. In particular, it identifies three biologically relevant dive types. State 1 includes shorter, shallower dives with 0–1 lunges per dive, short step length, and variable turning angle and heading variance; these are likely shallow foraging dives. State 2 comprises dives of moderate length and depth, without lunges, and with long step lengths and small turning angles and heading variance; these probably represent directed travel by the whales (without feeding). Finally, state 3 likely includes deep foraging dives, with long duration, deep depth, many lunges, moderate step length, and variable turning angle and heading variance. These three
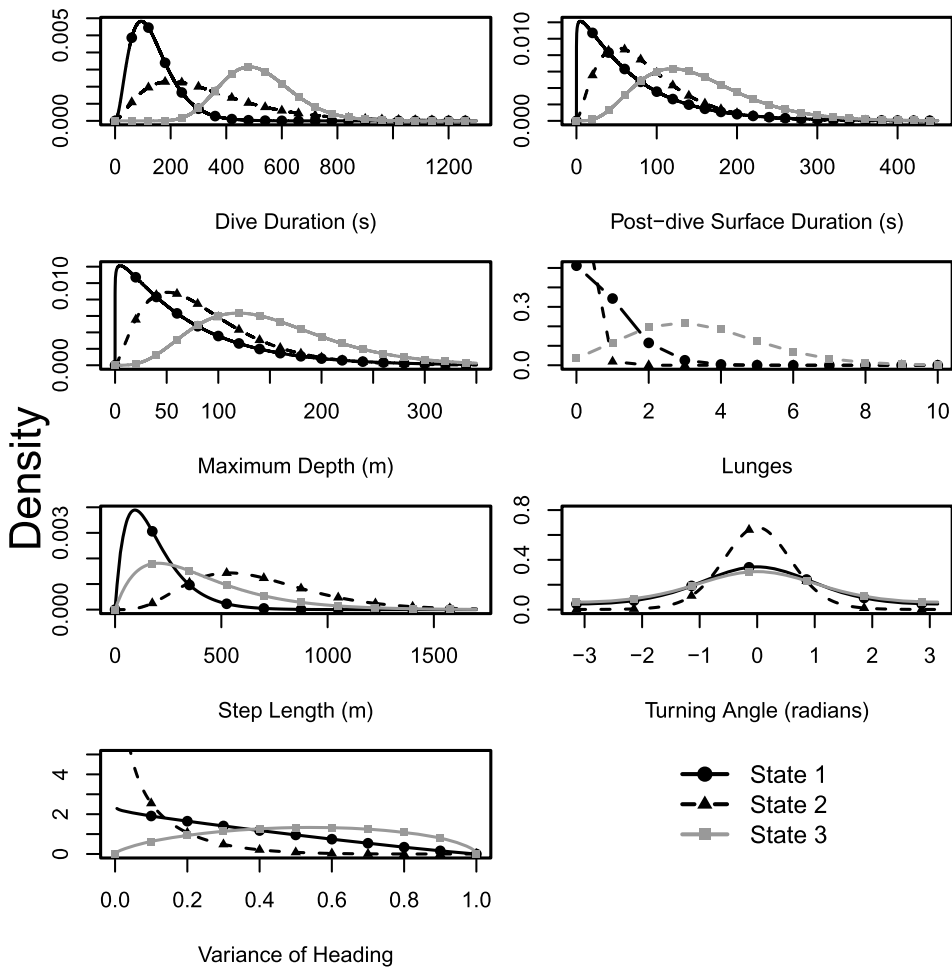
FIG. 3.   *Fitted state-dependent distributions for the 3-state HMM with complete pooling.*

dive types provide an adequate, if simplistic, classification of the main dive types performed by all the whales in the dataset.

In all models from this point forward, we consider only formulations with three states. Although this is a somewhat arbitrary choice, the selected model captures the most important structure in the data, is biologically informative and interpretable, serves the desired scientific purpose, and is generally consistent with previous classifications of blue whale behavioural states [Goldbogen et al. (2013a)]. For the given data, as is often found when applying HMMs to complex real data, formal model selection criteria favour models with much higher numbers of states (results not reported), resulting in a model that would be essentially uninterpretable and infeasible to work with [cf. Dean et al. (2012), Van de Kerk et al. (2015), Langrock et al. (2015)]. In fact, standard model selection criteria (such as AIC,

BIC, or cross-validation) tend to fail at choosing the correct number of states of an HMM applied to a complex real data set. HMMs are used to analyse complex time-series data, and, like all statistical models, they are simplistic representations of the real data-generating process. However, these models have a relatively complicated hierarchical structure, and so the simplifications occur in various places in the model formulation. To name a few possibilities, the families of state-dependent distributions may be too inflexible; the Markov property might be too strong an assumption; variation over time (cyclic or not) may not be adequately modelled; covariate information might be missing; there might be additional dependence structure between successive observations; and there might be outliers. When not accounted for in the model, any of these features can lead to model selection criteria favouring models with overly complex state architectures, that is, more states than adequate relative to the actual biological process, simply to compensate for the simplistic model formulation. In other words, the additional states essentially relate to structure in the data that should, were it been feasible, ideally have been incorporated directly in the model formulation. While these more complex models may perform better in terms of forecasts, they can also obscure patterns in the data and make insight into the underlying biological process more difficult, and so they are clearly undesirable when characterisation rather than prediction is the goal.

Given these challenges, we chose to focus on three-state models for the practical reasons mentioned. Since it is not the aim of the current paper to reveal how many states are required in order to explain the observed variance, but rather to investigate how key behavioural patterns may change in response to sonar, we think that it is both legitimate and sensible to choose the number of states in this pragmatic way, which focuses on interpretability and practicality of the resulting model. However, to see what we might be missing by focusing on the simpler three-state model, we investigated the differences between the fitted three- and four-state models. In the given example, when moving from the three-state to the four-state model, the 'deep foraging' state is split into two states, which are somewhat distinct in terms of the maximum depths, dive times, and surface times, but, crucially, imply almost exactly the same number of foraging lunges (though not necessarily identical foraging rates) (see Supplement B [DeRuiter et al. (2017b)] for plots comparing the state-dependent distributions for three- and four-state models). Thus, from a biological perspective, it seems more appropriate to merge these two model states into a single deep foraging state, as done by the model with three states. The resulting minor lack of fit between observed and modelled state-dependent distributions for the deep foraging state is acceptable, since we focus on drawing inference regarding the state-switching dynamics rather than on predictive performance of the time-series model.

Although the initial three-state model captures several important features of the data, in the raw data, it is immediately evident that the proportion of dives of each type, and the transition rates between them, vary strongly from whale to whale [see Figures 1–2 and Supplement B, DeRuiter et al. (2017b)]. Accordingly, as the

next step in model development, we chose to relax the assumption of homogeneity between whales (tag records).

3.2. *Accounting for heterogeneity.*

3.2.1. *Motivation.* Comparing the time-series dive data between whales, several clear patterns indicate that the assumption of homogeneity among individuals is inadequate. Some whales have one predominant dive type. For example, as shown in Supplement B [DeRuiter et al. (2017b)], whales 10, 11, 12, and 36 perform almost solely deep, multi-lunge dives, while whales 1, 5, and 17 perform numerous consecutive shallow dives without lunges and with larger turning angles. Other whales alternate more regularly between dive types. For example, whales 2, 6, 19, 24, and 30 alternate dives with and without lunges, without very long series of either.

This variation may be evidence of genuine individual differences between whales, although that conclusion is difficult to confirm without re-tagging the same individual multiple times in different situations [Goldbogen et al. (2013a)]. Even in the absence of individual differences as such, it is of biological interest to account for differences between tag recordings. Blue whale behaviour, particularly feeding behaviour, is strongly driven by environmental covariates such as the density and depth at which prey are present [Goldbogen et al. (2013b), Hazen, Friedlaender and Goldbogen (2015)]. These covariates are very difficult to measure, and are missing from many observations of whale behaviour (including those of most individuals in our study). Variation in prey abundance and depth, among other environmental covariates, may drive whales to occupy a corresponding behavioural 'context' [Friedlaender et al. (2016)]. In other words, blue whales forage differently depending on prey (krill) abundance: when prey are shallow and more dispersed, whales feed with low lunge counts and maximize oxygen conservation; when deep, dense krill patches are available, whales likely go into oxygen debt in order to lunge as many times as possible and maximize energy gain [Goldbogen et al. (2015), Hazen, Friedlaender and Goldbogen (2015)]. The increased maneuvering necessary to harvest low-prey-density patches relative to dense ones is also significant [Goldbogen et al. (2015), Hazen, Friedlaender and Goldbogen (2015)]. Within this framework, it seems reasonable to assume that there is a finite, relatively small number of behavioural contexts adopted by blue whales in our study region.

3.2.2. *Formulation of a model with discrete-valued random effects.* Maruotti and Rydén (2009), McKellar et al. (2014), and Towner et al. (2016) suggest the use of discrete-valued random effects in HMMs to capture this type of heterogeneity across component time series. This approach has two main advantages over the more common use of continuous-valued random effects: (1) the computational effort is substantially reduced, and (2) restrictive distributional assumptions on the random effects are avoided. Furthermore, interpretation is often more intuitive than for continuous-valued random effects. For example, each outcome of a

discrete random effect incorporated in the t.p.m. corresponds to one type of state-switching pattern, which in turn corresponds to one of the behavioural contexts described earlier.

To incorporate discrete random effects into our blue whale dive behaviour model, we assume that

$$\mathbf{\Gamma}^{(w)} = \mathbf{\Gamma}_{\xi^{(w)}} = (\gamma_{ij\xi^{(w)}})$$

with an i.i.d. discrete random effect $\xi^{(w)}$ such that $\Pr(\xi^{(w)} = k) = \pi_k$ for $k = 1, \ldots, K$, and

$$\gamma_{ijk} = \Pr(S_{wd} = j \mid S_{w,d-1} = i, \xi^{(w)} = k) = \frac{\exp(\alpha_{ijk})}{1 + \sum_{l \neq i} \exp(\alpha_{ilk})},$$

for $k = 1, \ldots, K$. With this mixture model for the state-switching dynamics, there are K possible $\mathbf{\Gamma}$s, $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K$, and each individual whale's time series is assumed to be driven by exactly one of them. The mixture weight π_k, $k = 1, \ldots, K$, gives the probability that the t.p.m. $\mathbf{\Gamma}_k$ underlies the state-switching dynamics for a given whale. The number of mixture components, K, controls the flexibility of the model to account for heterogeneity and can be chosen using model selection criteria. The initial state distributions are also assumed to be context-specific, that is, $\boldsymbol{\delta}^{(w)} = \boldsymbol{\delta}_{\xi^{(w)}}$, where the K different initial state distributions, $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K$, are estimated alongside the other model parameters.

3.2.3. *Model fitting.*   The likelihood of the above HMM with discrete-valued random effects is

$$(2) \quad \mathcal{L} = \prod_{w=1}^{37} \sum_{k=1}^{K} \mathcal{L}_k^{(w)} \pi_k = \prod_{w=1}^{37} \sum_{k=1}^{K} (\boldsymbol{\delta}_k \mathbf{Q}(\mathbf{x}_{w1.}) \mathbf{\Gamma}_k \mathbf{Q}(\mathbf{x}_{w2.}) \cdots \mathbf{\Gamma}_k \mathbf{Q}(\mathbf{x}_{wD_w.}) \mathbf{1} \pi_k).$$

Here $\mathcal{L}_k^{(w)}$ denotes the conditional likelihood of the observations made for whale w given the kth behavioural context (i.e., given $\xi^{(w)} = k$ and hence $\mathbf{\Gamma}^{(w)} = \mathbf{\Gamma}_k$); otherwise the notation is the same as in (1).

Fitting this more complicated model via naive application of a numerical optimizer presented several difficulties. Due to the fairly large number of parameters, the optimisation was slow and subject to numerical instability in terms of identification of local rather than global maxima of the likelihood. To allow for an efficient fitting of the model, we used several complementary strategies. First, we wrote the HMM forward algorithm using the Rcpp package [Eddelbuettel (2013)], which provides a simple way to use C++ code in conjunction with R software. This substitution of C++ code for R code greatly reduced the computation time required for model fitting. Second, we carefully selected appropriate starting values following a two-stage procedure:

I. In the first stage, we determined reasonable starting values for the parameters of the state-dependent distributions. Exploratory analysis indicated that these parameter values are stable for any number of possible outcomes of the discrete random effect K. Thus, we ran the optimizer for $K = 1$ and used the resulting estimates of state-dependent distribution parameters as starting values for the rest of the optimization stages.

II. In the second stage of model fitting, the goal was to find reasonable starting values for the parameters determining the Markov state process, that is, the entries of the transition probability matrices Γ_k and the corresponding initial distributions δ_k. Exploratory analysis indicated that the estimators of these parameters are highly unstable due to local maxima in the likelihood surface, and so selecting good starting values is crucial. We fixed the state-dependent distribution parameters at their starting values, and allowed only the parameters of the state process to vary during this optimisation stage. (Fixing the parameters of the state-dependent distributions decreased the required computing time by about a factor of 20.) To select starting values for the parameters of the state process, we ran the optimization 15,000 times, using random starting values in each run. The resulting sets of parameter estimates related to the hidden Markov chain were ranked in terms of their corresponding data likelihood, and the best 100 sets were used as starting values for the next stage of model fitting (comprehensive optimisation).

Finally, we carried out a full numerical maximisation of the likelihood. Initially, we simply ran the numerical maximiser, selecting the initial values for the state-dependent distribution parameters and the Markov chain parameters as described earlier. To further decrease the chance of finding only local maxima rather than the global maximum, we then additionally jittered the parameter estimates and re-ran the optimiser. Specifically, for each of the 100 best models identified in the previous stage, we added a small amount of noise to each parameter estimate, and used these jittered values as the initial values to re-run the optimiser. This was done five times for each of the 100 best models identified. This jittering procedure is similar to a bootstrap restart as suggested by Wood (2001); we also implemented the latter, but for our data it did not yield any improvements.

A possible alternative to this brute-force, but effective, method for finding the global maximum could be to run a few dozen iterations of the expectation-maximisation (EM) algorithm (as it is known to be slightly less sensitive to the choice of initial values) before switching to (much faster) direct maximisation of the likelihood, as was done, for example, in Bulla et al. (2012). However, in light of the complexity of the current model and the resulting requirement for substantial technical derivations for implementation of the EM algorithm, we have not pursued this option in the given application.

3.2.4. *Model selection on the number of behavioural contexts.* All of the sets of parameter estimates obtained were ranked according to their corresponding likelihoods, and we selected the maximum likelihood estimates to define the best

*Log-likelihood and AIC values for the different models considered. The last column also gives the
difference in AIC for each model relative to the base model* (3 *states, one behavioural context,
and no effect of acoustic exposure*)

Type of model	K	Number of parameters	log$\mathcal{L}$	AIC	Δ AIC
No effect of acoustic exposure	1	44	−25736.6	51561.1	0
	2	53	−25682.9	51471.8	−89.3
	3	62	−25665.4	51454.8	−106.3
	4	71	−25651.0	51444.0	−117.1
	5	80	−25643.3	51446.5	−114.6
Common acoustic exposure effect	4	77	−25642.5	**51438.9**	−122.2
Context-specific acoustic exposure effect	4	95	−25636.8	51463.7	−97.4

model for each value of K. We considered K (the number of behavioural contexts) to take on candidate values from 1 to 5, selecting the optimal K based on AIC scores.

Table 3 provides model selection results, including log-likelihoods and AIC values for the various models considered. According to the AIC scores, the optimal number of behavioural contexts, K, is four for these blue whale data. Rather than reporting model parameter estimates at this point, we defer those results to the next section where we consider one final refinement of the model: inclusion of an effect of acoustic exposure on the transition probability matrices.

3.3. *Investigating the effect of sound exposure.*

3.3.1. *Motivation.* As detailed in Section 2 and Supplement A [DeRuiter et al. (2017a)], most of the whales included in our dataset were subject to CEEs in which they were exposed to PRN or MFAS sounds (real or simulated). Previous analysis of a subset of the current dataset has suggested that blue whale behavioural responses to CEEs did not uniformly occur, but that when they did, they were variable, complex, and included responses such as the cessation of deep feeding behaviour and initiation of directed travel [Goldbogen et al. (2013a)]. To allow for such responses, we modified our model formulation to include the effect of an acoustic exposure covariate on the transition probability matrix (or matrices). We note that models such as this one, which incorporate both random effects and covariates, fall into the category of mixed HMMs as discussed in Altman (2007).

Although the number of CEEs included in our dataset is quite large compared to all other such studies, it is still small from a statistical perspective, and so we did not attempt to differentiate between exposures of different types (14 simulated

MFA, 1 real MFA, and 14 PRN CEEs). Rather, we sought to assess the utility of HMMs to investigate potential changes in blue whale state-switching behaviour in response to these somewhat similar mid-frequency sounds. Previous studies [Goldbogen et al. (2013a)] and ongoing analyses are considering individual responses by specific sound type to compare potential differential responses, including identifying changes as a function of received sound level. Here, we assumed that the effects of all CEE sounds were identical regardless of sound type and received sound exposure level, and that silent control CEEs had no effect on behaviour. While these assumptions undoubtedly represent simplifications of reality, the results of previous work suggest that they should probably be approximately appropriate for this species [Goldbogen et al. (2013a)].

3.3.2. *Model formulation of a model that incorporates sound exposure.* In models with an acoustic exposure covariate, altered t.p.m.'s are in place during times when whales were subject to sonar exposure. For each behavioural context, the between-state transition probabilities $\gamma_{ijkd}^{(w)} = \Pr(S_{wd} = j \mid S_{w,d-1} = i, \xi^{(w)} = k)$ are now defined as

$$(3) \qquad \gamma_{ijkd}^{(w)} = \frac{\exp(\alpha_{ijk} + \beta_{ijk} z_{wd})}{1 + \sum_{l \neq i} \exp(\alpha_{ilk} + \beta_{ilk} z_{wd})},$$

where

$$z_{wd} = \begin{cases} 1, & \text{if whale } w \text{ was exposed to sonar during dive } d; \\ 0, & \text{otherwise.} \end{cases}$$

In other words, for each of the K contexts, there are two t.p.m.'s: one associated with baseline (undisturbed) periods, and the other with CEE (potentially disturbed) periods.

We consider several possible constraints on the β parameters, which quantify the strength and direction of the relationships between exposure and each transition probability. In the most general case, the effect of acoustic exposure is context-specific so that each β_{ijk} has a unique value for each k, $k = 1, \ldots, K$; the effect of exposure is different for each t.p.m. entry within every behavioural context. We also consider the possibility of a common acoustic exposure effect, where the β_{ijk} have constant values across all contexts, that is, $\beta_{ij1} = \cdots = \beta_{ijK}$ for all $i, j = 1, 2, 3$. Finally, we constructed confidence intervals around the parameter estimates via the observed Fisher information (or via profile likelihood, if the observed Fisher information indicated a standard error of zero for a given estimate). Thus we implicitly consider the possibility that acoustic exposure may have effects only on specific contexts or specific interstate transitions.

3.3.3. *Model fitting*.   The likelihood of the model including both discrete-valued random effects and the sonar exposure covariate has exactly the same form as stated in equation (2), except that now each t.p.m. additionally varies over dives as it depends on the associated covariate value z_{wd}.

As the mixed HMM requires estimation of even more parameters than the model with the discrete random effects, fitting it entails technical and numerical challenges equal to or greater than those encountered with the discrete random effect model. To obtain reliable model parameter estimates, we followed the same procedure outlined in Section 3.2.3.

3.3.4. *Results*.   Table 3 provides model selection results, including the models with common and context-specific acoustic exposures. The best model as selected by AIC has $K = 4$ behavioural contexts and an effect of acoustic exposure that is common across all behavioural contexts.

Figure 4 shows the state-dependent distributions for the final (best) model. State 1 includes the shortest, shallowest dives, with short step lengths and few lunges; it may correspond to shallow foraging behaviour. State 2 features intermediate depth and duration, long step length, small turning angle, low heading variance, and no lunges; perhaps it can be labelled travelling behaviour. State 3 has the deepest, longest dives, with intermediate step length, large variability of heading, and multiple lunges; it is consistent with deep foraging behaviour. These distributions are very similar to those of the initial model (the simplest one, with $N = 3$, $K = 1$, and no discrete random effect or exposure effect; Figure 3), confirming that the main differences between the initial and final models are in the transition probabilities, not the properties of the states themselves.

For the final, best model, the baseline t.p.m. is different for each of the four behavioural contexts, and each of the four matrices is affected in the same way by acoustic disturbance (that is, the β_{ijk} parameters are the same for each context; see Section 3.3.2 for details). The resulting eight baseline and exposure t.p.m.'s are presented graphically in Figure 5. Table 4 presents the estimated initial state distributions for each context. Context 1 has very high persistence in states 2 and 3, and immediate transition from state 1 to 3 (if 1 is ever observed); context 2 has high persistence in all three states, but with occupancy of state 3 quite rare (cf. Table 4); context 3 has persistence in state 2 and some alternation between states 1–3; and context 4 has the highest probability of switching between states, with frequent transitions to state 3 from other states.

Table 5 presents the maximum likelihood estimates of the parameters related to the state-dependent process, of the weights associated with the discrete random effects, and of the parameters describing the effect of sonar exposure. The table also includes 95% confidence intervals for the parameter estimates. As for the baseline model, these confidence intervals were obtained from the observed Fisher information, except for the interval provided for the parameter β_{13}, which was obtained
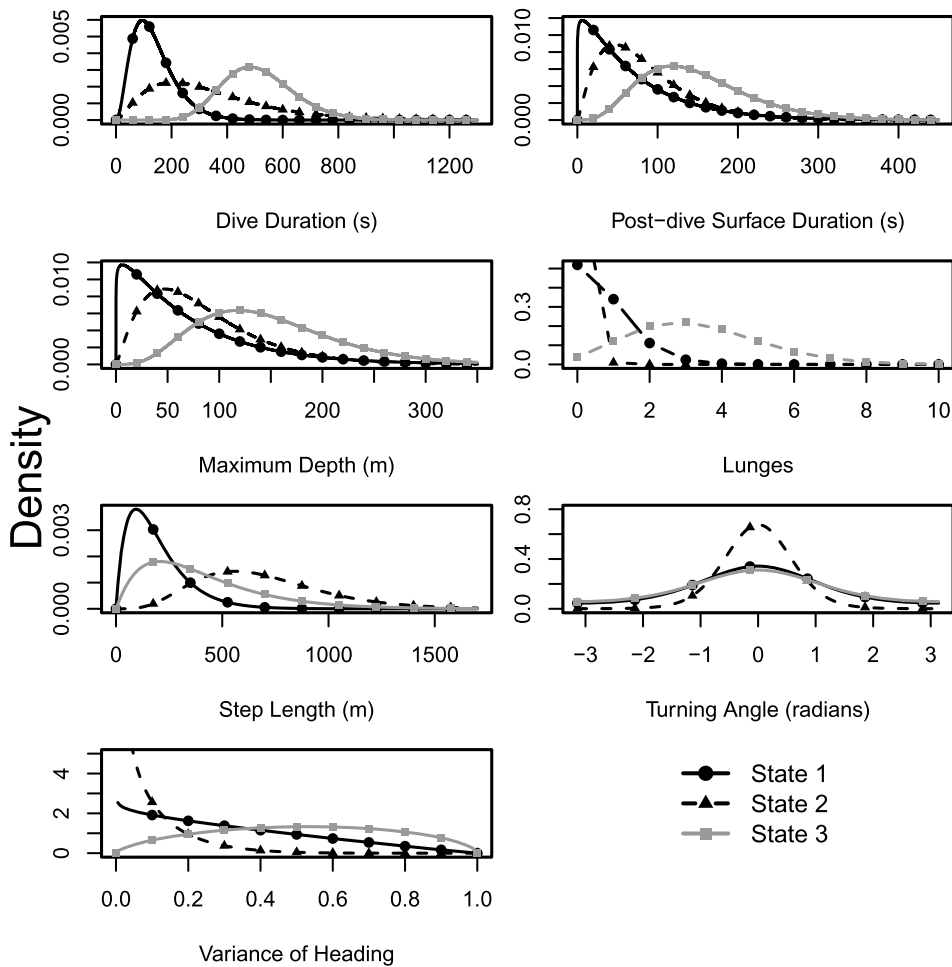
FIG. 4.   *Fitted model estimates of the state-dependent distributions for the 3-state HMM with 4 contexts and a common effect of acoustic disturbance.*

using the profile likelihood [Venzon and Moolgavkar (1988)] since here the observed Fisher information indicated a standard error of zero, due to a ridge of the likelihood in this dimension. [The reason for this ridge is that, for any sufficiently large negative estimate of the regression coefficient $\beta_{13\cdot}$, the transformation via the inverse of the multinomial logit link in (3) effectively results in a zero transition probability from state 1 to state 3, which corresponds to the maximum likelihood estimate.] Among the acoustic exposure parameters (βs), only $\beta_{13\cdot}$ and $\beta_{21\cdot}$ have 95% confidence intervals that do not contain zero. The corresponding parameter estimates indicate that during acoustic exposure the rate of already rare transitions from state 1 to state 3 falls to zero ($\beta_{13\cdot} = -27.1$), balanced by an increased
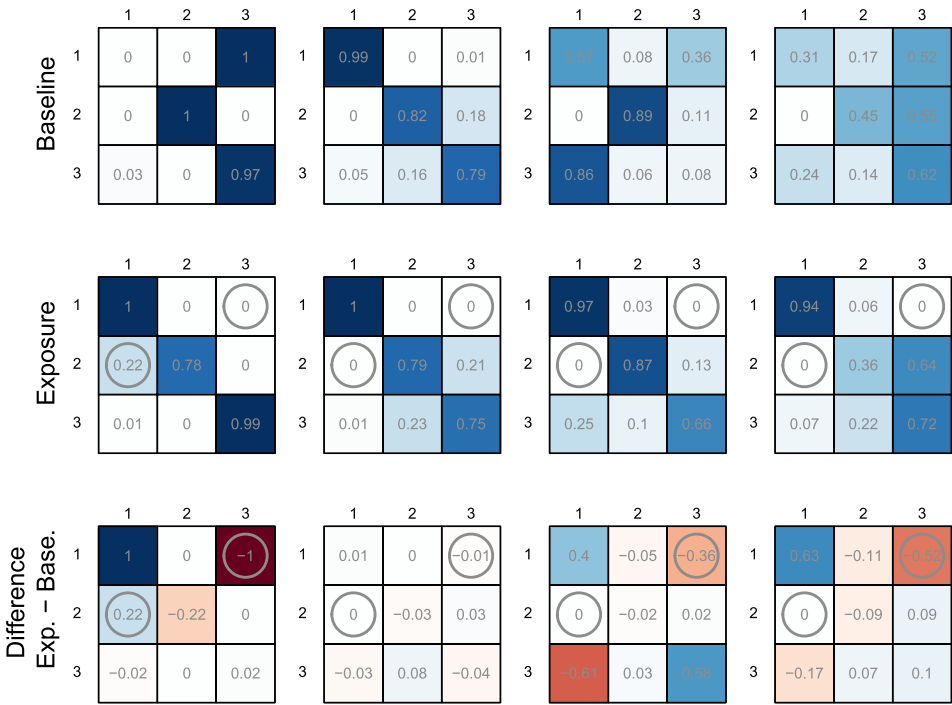
FIG. 5.   *Fitted model estimates of the transition probability matrices for the 3-state HMM with 4 contexts (one column per context). For each context, the first row shows the t.p.m.s (in the absence of acoustic disturbance), the second row shows the t.p.m. with acoustic disturbance, and the third row shows the difference between the two. The magnitudes of the transition probabilities are indicated by the fill colour, as well as the printed numeric values in each cell. For the t.p.m.'s with acoustic disturbance, circles around the probabilities indicate entries corresponding to acoustic-disturbance parameters whose 95% confidence intervals do not contain zero.*

probability of remaining in state 1. There is also an increase in the rate of transitions from state 2 to state 1 ($\beta_{21.}$), at the expense of a decreased persistence in state 2.

TABLE 4
Estimated initial state distributions $\delta^{(w)}$ for the best model, with N = 3, K = 4, and a common effect of acoustic disturbance

Context	State 1	State 2	State 3
$k = 1$	0.84	0.00	0.16
$k = 2$	0.63	0.37	0.00
$k = 3$	0.50	0.50	0.00
$k = 4$	0.66	0.00	0.34

TABLE 5
Parameter estimates, along with 95% CIs, for the best model as selected by AIC ($N = 3$ states, $K = 4$ behavioural contexts, and an effect of acoustic exposure that is common across all behavioural contexts)

Parameter	State	Lower bound	Estimate	Upper bound
Dive Duration (μ)	1	131	139	147
Dive Duration (σ)	1	71.4	77.8	84.7
Dive Duration (μ)	2	312	342	375
Dive Duration (σ)	2	190	217	247
Dive Duration (μ)	3	502	515	528
Dive Duration (σ)	3	120	129	139
Surface Duration (μ)	1	65.2	71.3	77.91
Surface Duration (σ)	1	61.2	68.1	75.7
Surface Duration (μ)	2	76.4	84.8	94.1
Surface Duration (σ)	2	48.2	56.5	66.2
Surface Duration (μ)	3	143	150	157
Surface Duration (σ)	3	63.6	69.2	75.3
Maximum Depth (μ)	1	30.0	32.2	34.5
Maximum Depth (σ)	1	21.2	23.3	25.7
Maximum Depth (μ)	2	58.6	67.7	78.1
Maximum Depth (σ)	2	54.3	64.4	76.4
Maximum Depth (μ)	3	164	169	176
Maximum Depth (σ)	3	56.3	60.9	65.8
Number of Lunges (λ)	1	0.492	0.656	0.651
Number of Lunges (λ)	2	0.001	0.010	0.095
Number of Lunges (λ)	3	3.06	3.31	3.43
Step Length (μ)	1	179	193	208
Step Length (σ)	1	125	138	152
Step Length (μ)	2	649	693	741
Step Length (σ)	2	269	304	343
Step Length (μ)	3	379	409	441
Step Length (σ)	3	259	287	318
Turning Angle (κ)	1	0.850	1.01	1.19
Turning Angle (κ)	2	2.61	3.18	3.86
Heading Variance (a)	1	0.826	0.940	1.07
Heading Variance (b)	1	1.77	2.05	2.36
Heading Variance (a)	2	0.435	0.518	0.617
Heading Variance (b)	2	4.89	6.55	8.78
Heading Variance (a)	3	1.42	1.66	1.86
Heading Variance (b)	3	1.34	1.58	1.83
Mass, rand. eff. comp. 1 (π_1)		0.189	0.333	0.494
Mass, rand. eff. comp. 2 (π_2)		0.280	0.449	0.607
Mass, rand. eff. comp. 3 (π_3)		0.013	0.054	0.187
Mass, rand. eff. comp. 4 (π_4)		0.070	0.163	0.320
Acoustic Exposure ($\beta_{12.}$)		-2.56	-0.997	0.561
Acoustic Exposure ($\beta_{13.}$)		$-\infty$	-27.1	-6.86
Acoustic Exposure ($\beta_{21.}$)		15.0	17.0	19.0
Acoustic Exposure ($\beta_{23.}$)		-0.829	0.147	1.123
Acoustic Exposure ($\beta_{31.}$)		-3.22	-1.24	0.733
Acoustic Exposure ($\beta_{32.}$)		-0.507	0.390	1.29

For each whale, we decode the states locally (i.e., for each data point, we find the most likely state under the fitted model and given the observations) as follows:

$$\Pr(S_{wd} = j \mid \mathbf{X}_{w\cdot\cdot}, \mathbf{z}_{w\cdot})$$

$$= \sum_{k=1}^{K} \Pr\big(S_{wd} = j, \xi^{(w)} = k \mid \mathbf{X}_{w\cdot\cdot}, \mathbf{z}_{w\cdot}\big)$$

$$= \sum_{k=1}^{K} \Pr\big(S_{wd} = j \mid \xi^{(w)} = k, \mathbf{X}_{w\cdot\cdot}, \mathbf{z}_{w\cdot}\big) \Pr\big(\xi^{(w)} = k \mid \mathbf{X}_{w\cdot\cdot}, \mathbf{z}_{w\cdot}\big)$$

$$= \sum_{k=1}^{K} \Pr\big(S_{wd} = j \mid \xi^{(w)} = k, \mathbf{X}_{w\cdot\cdot}, \mathbf{z}_{w\cdot}\big) \frac{\mathcal{L}_k \pi_k}{\mathcal{L}}.$$

Bayes' theorem is applied in the last step. The calculation of the first factor simply corresponds to the local decoding of a basic HMM with no random effects and is performed as described in Zucchini, MacDonald and Langrock (2016) (Section 5.3.1). All remaining quantities are evaluated as previously described. These decoded states are included in Figures 1–2, and corresponding figures for all whales analysed are included in the web-based Supplementary Materials.

4. Discussion. Our analysis of the blue whale data demonstrates the successful application of HMMs to multivariate animal behaviour data. While multivariate HMMs and other hidden state models are already in use for ecological data, most applications involve fitting models to animal movement data, such as animal tracks in two or three dimensions [for example, see Langrock et al. (2012), McKellar et al. (2014), Langrock et al. (2014), McClintock et al. (2012), Patterson et al. (2008), Johnson et al. (2008), Morales et al. (2004)]. Examples of using HMMs to classify and characterise broader animal behaviour states are much less prevalent [but see Bagniewska et al. (2013), Isojunno and Miller (2015)]. Here, we demonstrate the power and flexibility of this approach. By applying HMMs to free-ranging marine mammals in a range of behavioural states and sound exposure regimes, we illustrate how they can incorporate random effects to account for individual or contextual differences, and also quantify the effects of environmental covariates (here, sound exposure) in modulating behavioural transitions. An additional strength of this modelling approach, particularly for biological data sets, is its ability to simply and effectively deal with partially observed multivariate data without omitting any observed data points.

Our best model for the blue whale data incorporates a discrete random effect to account for differences between observed whales, whether these are due to genuine individual variation or the effects of unobserved contextual variables. Large inter-individual variability is a prominent, obvious feature of our dataset and many

other ecological datasets. In fact, the magnitude of these differences is often equal to or greater than the changes induced by experimental manipulation, and so accounting for them can dramatically improve model fit, as evidenced by the very large improvement in AIC score between models without and with the random effect. More importantly, failing to account for individual differences (or contexts, as we call them here) may make it difficult to detect relatively small-magnitude behaviour modulations caused by other factors (such as acoustic disturbance). In particular, not accounting for individual heterogeneity can result in biases and invalid standard errors of estimators that describe an effect of interest. While it is easy to argue that random effects to account for individual variation are a key component of animal behaviour models, fitting traditional continuous random effects can be computationally challenging, especially for models like HMMs that are already complex and parameter-rich. The discrete random effect formulation used in this study makes inclusion of between-subject variability, which is key for improving model fit, practically feasible and reasonably quick, within a simple maximum-likelihood estimation framework.

In addition to accounting for inter-subject differences, we illustrate here a simple formulation allowing the inclusion of a covariate affecting the t.p.m., which could easily be generalised to multiple covariates if the data allow. In the blue whale case, previous research suggested that the most likely effect of acoustic disturbance would be on rates of transition between behavioural states [Goldbogen et al. (2013a)]. However, one could also consider including covariates affecting one or more of the state-dependent distributions, if appropriate for the data. In the case of acoustic disturbance in particular, it would be of considerable biological interest to consider whether response intensity scales with stimulus type (simulated or real MFAS, or PRN) or with the received level of the sound stimulus at the whale. In fact, while many regulations assume that responsiveness depends on exposure level, considerable and evolving data support the conclusion that myriad contextual factors related to the exposure configuration and internal behavioural features of the subject can strongly affect response probability [Ellison et al. (2012)]. A wealth of studies indicate differential responses to different sound types within the same species [for example, Miller et al. (2012), Goldbogen et al. (2013a), Nowacek, Johnson and Tyack (2004), and reviews Southall et al. (2007), DeRuiter (2010)], although in some cases sound frequency and type have not been confirmed as important predictors of response [Antunes et al. (2014)]. Potential effects of signal type and received level could be easily accommodated in our model formulation, essentially resulting in a multiple multinomial logistic regression of the transition probabilities on the stimulus type and intensity. In addition to the stimulus type and sound level considerations mentioned above, it would be of biological interest to assess whether responses extended beyond the end of the exposure period or were cumulative across multiple exposures. Unfortunately, sample size but also computational limitations prevent us from implementing these models and fitting

them to the blue whale dataset. This circumstance highlights one limitation of hidden state models like our mixed multivariate HMM: they are parameter-rich and can be challenging to fit to data, such that computational considerations and sample size often limit the complexity of candidate models. It is therefore interesting, and perhaps surprising, that we see a relatively consistent behavioural response in our data across several stimulus types (as further detailed below). One reason may be the relatively limited range of exposure levels; none of the CEEs resulted in received levels exceeding 157 dB re. 1 μPa rms [computed as detailed in DeRuiter et al. (2013)], and so relatively little information is lost in considering the acoustic disturbance as 'on' or 'off' and in disregarding the measured intensity. Alternatively, animals may be responding more to the presence of a relatively nearby sound source producing several sound types rather than distinguishing among the types themselves.

In this study of blue whale behavioural responses to acoustic disturbance, it is difficult to draw firm conclusions about the biological significance of the changes observed during acoustic exposure: the difference in AIC scores between the model with and without the exposure covariate is only about 5.1. While the model with an effect of acoustic disturbance has considerably more support from the data than the model without, the improvement in model fit resulting from the exposure covariate is modest compared to that of the random effect. This modest improvement could result from the limited number of CEE dives observed. It might also indicate that the disturbance effect has a relatively small magnitude, and/or affects only some of the individuals, and/or affects only certain behaviour states, as evidenced by the confidence intervals presented in Table 5. These results are again consistent with those of Goldbogen et al. (2013a).

The mixed HMM fitted here allows us not only to state that the acoustic disturbance had an effect, but also to characterise that effect quantitatively. The types of changes that occur, according to the common acoustic exposure effect model, are consistent with previous work. We find that whales are less likely to initiate deep foraging behaviour during acoustic disturbance, in accordance with Goldbogen et al. (2013a), who analysed a subset of the data presented here. Beaked whales [DeRuiter et al. (2013), Miller et al. (2015), Stimpert et al. (2014), Tyack et al. (2011)] and sperm whales [Isojunno and Miller (2015)] have also been shown to avoid deep-feeding behaviour during disturbance. In each case, more research is needed to understand whether this response helps whales conserve energy, avoid detection, or something else. We also found that the rate of transition from state 2 (directed travel) to state 1 (shallow dives, with no or few feeding lunges) was elevated during exposure. This effect on the transition rate was largest in behavioural context 1, but in context 1 whales were very unlikely ever to be in state 2; and in all other contexts, the baseline probability of transition from state 2 to 1 was so low that the 'elevated' probability during exposure was still effectively zero. If confirmed, a change from directed travel to shallow dives might relate to increased

vigilance or variability of behaviour during lower-level acoustic disturbance related to consideration of an appropriate course of action, as seen in some beaked whales [DeRuiter et al. (2013), Miller et al. (2015)]. However, since this finding deals with very rare or unlikely events, more research would clearly be needed to confirm the results and their interpretation.

Another important consideration for the interpretation of the model results is the grouping of dives with and without lunges together in state 1, as it relates to our observations of the natural history of these whales. By selecting a three-state model to reasonably group dives into categories, the model groups no-, single-, and several-lunge dives within a single shallow-diving state, indicating that these dives are quite similar (aside from number of lunges), or at least that the differences between them are less than those between shallow, travelling, and deep-feeding dives. Although this model categorization may differ slightly from a priori distinctions between shallow feeding, deep feeding, and nonfeeding [as in Goldbogen et al. (2013a)], we believe this grouping is still generally consistent with the behavioural state of shallow-diving and feeding behaviour in blue whales. Recent data, for instance, indicates that there may be significantly greater energetic benefit (as well as cost) in deep versus shallow foraging [Hazen, Friedlaender and Goldbogen (2015)].

The results of this study are consistent with previous observations that behavioural responsiveness to acoustic disturbance is highly context-specific: animals adjust the intensity of their response depending not only on the intensity of the sound they experience, but also on the environmental and behavioural conditions [Ellison et al. (2012)]. This study illustrates that even if the type of response is consistent across behavioural contexts, the apparent intensity of the response displayed will vary depending on the baseline behaviour (Figure 5). The detection and further characterisation of these behavioural responses is of strong biological and management interest since it can contribute to our understanding and mitigation of any potential negative effects of anthropogenic noise. Since the behaviour changes observed relate to alterations of foraging behaviour, they have particular potential to affect animals' health and fitness, and perhaps lead to population consequences. Although the HMM analysis results are reassuringly consistent with previous work on blue whale responses to acoustic disturbance, the models presented here offer several advantages and features that complement previous regression-based approaches [Friedlaender et al. (2016), Goldbogen et al. (2013a), Hazen, Friedlaender and Goldbogen (2015)]. The regression-based approach relies on principal components analysis to decompose the dependent multivariate behaviour data into independent response variables, and then fits generalised additive mixed-effect models to each one. This approach does account for whale-to-whale variation and quantify differences in principal components scores in response to different sound exposure types. However, biological interpretation of the principal components axes is notoriously challenging [James and McCulloch (1990)], and characterisation of behaviour or discussion of results in terms of

dive types or behavioural states must be post-hoc and relatively subjective, as the models do not explicitly include such structure. The HMM approach naturally integrates data from multiple input data streams, accounting for the time-series nature of the data and for some dependence between data streams (subject to the contemporaneous conditional independence assumption). It provides a concise summary of blue whale dive behaviour with and without effects of acoustic exposure, in the form of a finite set of possible dive types (states) and their accompanying characteristics (state-dependent distributions), all of which here have straightforward biological interpretations. In addition, based on the fitted HMM, one can readily compute simulated whale behaviour time series in the presence and absence of acoustic exposure, which may greatly facilitate incorporation of these results into impact assessments and models of population consequences of acoustic disturbance that use agent-based models [Donovan et al. (2012), Houser (2006)]. Mixed HMMs prove to be a flexible, powerful tool for the characterisation of animal behavioural states, including behavioural responses to acoustic disturbance or other environmental stimuli.

SUPPLEMENTARY MATERIAL

Supplement A: Detailed experimental methods for "A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure" (DOI: 10.1214/16-AOAS1008SUPPA; .pdf). This supplement provides additional detail on the field data collection protocols used to collect the dataset.

Supplement B: Supplemental figures for "A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure" (DOI: 10.1214/16-AOAS1008SUPPB; .pdf). This supplement provides figures of the input data (and decoded states from the final model) for all 37 whales in the dataset.

REFERENCES

ALTMAN, R. M. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *J. Amer. Statist. Assoc.* **102** 201–210. MR2345538

ANTUNES, R., KVADSHEIM, P. H., LAM, F. P. A., TYACK, P. L., THOMAS, L., WENSVEEN, P. J. and MILLER, P. J. O. (2014). High thresholds for avoidance of sonar by free-ranging long-finned pilot whales (*Globicephala melas*). *Mar. Pollut. Bull.* **83** 165–180.

BAGNIEWSKA, J. M., HART, T., HARRINGTON, L. A. and MACDONALD, D. W. (2013). Hidden Markov analysis describes dive patterns in semiaquatic animals. *Behav. Ecol.* **24** 659–667.

BULLA, J., LAGONA, F., MARUOTTI, A. and PICONE, M. (2012). A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series. *J. Agric. Biol. Environ. Stat.* **17** 544–567. MR3041884

DEAN, B., FREEMAN, R., KIRK, H., LEONARD, K., PHILLIPS, R. A., PERRINS, C. M. and GUILFORD, T. (2012). Behavioural mapping of a pelagic seabird: Combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *J. R. Soc. Interface* 20120570. DOI:10.1098/rsif.2012.0570.

DERUITER, S. L. (2010). Marine animal acoustics. In *An Introduction to Underwater Acoustics* (X. Lurton, ed.) 425–474. Praxis Publishing Limited, Chichester, UK.

DERUITER, S. L., SOUTHALL, B. L., CALAMBOKIDIS, J., ZIMMER, W. M. X., SADYKOVA, D., FALCONE, E. A., FRIEDLAENDER, A. S., JOSEPH, J. E., MORETTI, D., SCHORR, G. S., THOMAS, L. and TYACK, P. L. (2013). First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biol. Lett.* **9** 20130223. DOI:10.1098/rsbl.2013.0223.

DERUITER, S. L., LANGROCK, R., SKIRBUTAS, T., GOLDBOGEN, J. A., CALAMBOKIDIS, J., FRIEDLAENDER, A. S. and SOUTHALL, B. L. (2017a). Supplement to "A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure." DOI:10.1214/16-AOAS1008SUPPA.

DERUITER, S. L., LANGROCK, R., SKIRBUTAS, T., GOLDBOGEN, J. A., CALAMBOKIDIS, J., FRIEDLAENDER, A. S. and SOUTHALL, B. L. (2017b). Supplement to "A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure." DOI:10.1214/16-AOAS1008SUPPB.

DONOVAN, C. R., HARRIS, C., HARWOOD, J. and MILAZZO, L. (2012). A simulation-based method for quantifying and mitigating the effects of anthropogenic sound on marine mammals. In *Proceedings of Meetings on Acoustics* **17** 070043. Acoustical Society of America, Melville, NY.

EDDELBUETTEL, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York.

ELLISON, W. T., SOUTHALL, B. L., CLARK, C. W. and FRANKEL, A. S. (2012). A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. *Conserv. Biol.* **26** 21–28.

FRIEDLAENDER, A. S., GOLDBOGEN, J. A., HAZEN, E. L., CALAMBOKIDIS, J. and SOUTHALL, B. L. (2015). Feeding performance by sympatric blue and fin whales exploiting a common prey resource. *Mar. Mamm. Sci.* **31** 345–354.

FRIEDLAENDER, A. S., HAZEN, E. L., GOLDBOGEN, J. A., STIMPERT, A. K., CALAMBOKIDIS, J. and SOUTHALL, B. L. (2016). Prey-mediated behavioral responses of feeding blue whales in controlled sound exposure experiments. *Ecol. Appl.* **26** 1075–1085.

GOLDBOGEN, J. A., SOUTHALL, B. L., DERUITER, S. L., CALAMBOKIDIS, J., FRIEDLAENDER, A. S., HAZEN, E. L., FALCONE, E. A., SCHORR, G. S., DOUGLAS, A., MORETTI, D. J., KYBURG, C., MCKENNA, M. F. and TYACK, P. L. (2013a). Blue whales respond to simulated mid-frequency military sonar. *Proc. Biol. Sci.* **280** 20130657.

GOLDBOGEN, J. A., FRIEDLAENDER, A. S., CALAMBODKIS, J., MCKENNA, M. F., SIMON, M. and NOWACEK, D. P. (2013b). Integrative approaches to the study of baleen whale diving behavior, feeding performance, and foraging ecology. *Bioscience* **63** 90–100.

GOLDBOGEN, J. A., HAZEN, E. L., FRIEDLAENDER, A. S., CALAMBOKIDIS, J., DERUITER, S. L., STIMPERT, A. K. and SOUTHALL, B. L. (2015). Prey density and distribution drive the three-dimensional foraging strategies of the largest filter feeder. *Funct. Ecol.* **29** 951–961.

HART, T., MANN, R., COULSON, T., PETTORELLI, N. and TRATHAN, P. (2010). Behavioural switching in a central place forager: Patterns of diving behaviour in the macaroni penguin (*Eudyptes chrysolophus*). *Mar. Biol.* **157** 1543–1553.

HAZEN, E. L., FRIEDLAENDER, A. S. and GOLDBOGEN, J. A. (2015). Blue whales (*Balaenoptera musculus*) optimize foraging efficiency by balancing oxygen use and energy gain as a function of prey density. *Sci. Adv.* **1** e1500469.

HOLYOAK, M., CASAGRANDI, R., NATHAN, R., REVILLA, E. and SPIEGEL, O. (2008). Trends and missing parts in the study of movement ecology. *Proc. Natl. Acad. Sci. USA* **105** 19060–19065.

HOUSER, D. S. (2006). A method for modeling marine mammal movement and behavior for environmental impact assessment. *IEEE J. Oceanic Eng.* **31** 76–81.

ISOJUNNO, S. and MILLER, P. J. O. (2015). Sperm whale response to tag boat presence: Biologically informed hidden state models quantify lost feeding opportunities. *Ecosphere* **6** 1–46.

JAMES, F. C. and McCULLOCH, C. E. (1990). Multivariate analysis in ecology and systematics: Panacea or Pandora's box?. *Ann. Rev. Ecolog. Syst.* **21** 129–166.

JOHNSON, M. P. and TYACK, P. L. (2003). A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Oceanic Eng.* **28** 3–12.

JOHNSON, M. P., HICKMOTT, L. S., SOTO, N. A. and MADSEN, P. T. (2008). Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (*Mesoplodon densirostris*). *Proceedings of the Royal Society of London B*: *Biological Sciences* **275** 133–139.

KING, S. L., SCHICK, R. S., DONOVAN, C., BOOTH, C. G., BURGMAN, M., THOMAS, L. and HARWOOD, J. (2015). An interim framework for assessing the population consequences of disturbance. *Methods Ecol. Evol.* **6** 1150–1158.

LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. and MORALES, J. M. (2012). Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions. *Ecology* **93** 2336–2342.

LANGROCK, R., MARQUES, T. A., BAIRD, R. W. and THOMAS, L. (2014). Modeling the diving behavior of whales: A latent-variable approach with feedback and semi-Markovian components. *J. Agric. Biol. Environ. Stat.* **19** 82–100. MR3257903

LANGROCK, R., KNEIB, T., SOHN, A. and DERUITER, S. L. (2015). Nonparametric inference in hidden Markov models using P-splines. *Biometrics* **71** 520–528.

MACDONALD, I. L. (2014). Numerical maximisation of likelihood: A neglected alternative to EM? *Int. Stat. Rev.* **82** 296–308. MR3248241

MARUOTTI, A. and RYDÉN, T. (2009). A semiparametric approach to hidden Markov models under longitudinal observations. *Stat. Comput.* **19** 381–393.

McCLINTOCK, B. T., KING, R., THOMAS, L., MATTHIOPOULOS, J., McCONNELL, B. J. and MORALES, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecol. Monogr.* **82** 335–349.

McCLINTOCK, B. T., RUSSELL, D. J. F., MATTHIOPOULOS, J. and KING, R. (2013). Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. *Ecology* **94** 838–849.

McKELLAR, A. E., LANGROCK, R., WALTERS, J. R. and KESLER, D. C. (2014). Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird. *Behav. Ecol.* **26** 148–157.

MILLER, P. J. O., KVADSHEIM, P. H., LAM, F.-P. A., WENSVEEN, P. J., ANTUNES, R., ALVES, A. C., VISSER, F., KLEIVANE, L., TYACK, P. L. and SIVLE, L. D. (2012). The severity

of behavioral changes observed during experimental exposures of killer (*Orcinus orca*), long-finned pilot (*Globicephala melas*), and sperm (*Physeter macrocephalus*) whales to naval sonar. *Aquat. Mamm.* **38** 362–401.

MILLER, P. J. O., KVADSHEIM, P. H., LAM, F. P. A., TYACK, P. L., CURE, C., DERUITER, S. L., KLEIVANE, L., SIVLE, L. D., VAN IJSSELMUIDE, S. P., VISSER, F., WENSVEEN, P. J., VON BENDA-BECKMANN, A. M., MARTIN LOPEZ, L. M., NARAZAKI, T. and HOOKER, S. K. (2015). First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. *R. Soc. Open Sci.* **2** 140484.

MORALES, J. M., HAYDON, D. T., FRAIR, J., HOLSINGER, K. E. and FRYXELL, J. M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology* **85** 2436–2445.

NATIONAL RESEARCH COUNCIL (2005). *Marine Mammal Populations and Ocean Noise*: *DeterminIng When Noise Causes Biologically Significant Effects*. National Academies Press, Washington, DC.

NEW, L. F., MORETTI, D. J., HOOKER, S. K., COSTA, D. P. and SIMMONS, S. E. (2013). Using energetic models to investigate the survival and reproduction of beaked whales (family Ziphiidae). *PLoS ONE* **8** e68725.

NEW, L. F., CLARK, J. S., COSTA, D. P., FLEISHMAN, E., HINDELL, M. A., KLANJŠČEK, T., LUSSEAU, D., KRAUS, S., MCMAHON, C. R., ROBINSON, P. W., SCHICK, R. S., SCHWARZ, L. K., SIMMONS, S. E., THOMAS, L., TYACK, P. and HARWOOD, J. (2014). Using short-term measures of behaviour to estimate long-term fitness of southern elephant seals. *Mar. Ecol. Prog. Ser.* **496** 99–108.

NOWACEK, D. P., JOHNSON, M. P. and TYACK, P. L. (2004). Proceedings of the Royal Society of London B: Biological Sciences. *Proceedings of the Royal Society B—Biological Sciences* **271** 227–231.

PATTERSON, T. A., THOMAS, L., WILCOX, C., OVASKAINEN, O. and MATTHIOPOULOS, J. (2008). State-space models of individual animal movement. *Trends Ecol. Evol.* **23** 87–94.

R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/.

SCHICK, R. S., NEW, L. F., THOMAS, L., COSTA, D. P., HINDELL, M. A., MCMAHON, C. R., ROBINSON, P. W., SIMMONS, S. E., THUMS, M., HARWOOD, J. and CLARK, J. S. (2013). Estimating resource acquisition and at-sea body condition of a marine predator. *J. Anim. Ecol.* **82** 1300–1315.

SHANNON, G., MCKENNA, M. F., ANGELONI, L. M., CROOKS, K. R., FRISTRUP, K. M., BROWN, E., WARNER, K. A., NELSON, M. D., WHITE, C., BRIGGS, J., MCFARLAND, S. and WITTEMYER, G. (2015). A synthesis of two decades of research documenting the effects of noise on wildlife. *Biol. Rev. Camb. Philos. Soc.* **91** 982–1005.

SIVLE, L. D., KVADSHEIM, P. H., CURÉ, C., ISOJUNNO, S., WENSVEEN, P. J., LAM, F.-P. A., VISSER, F., KLEIVANE, L., TYACK, P. L., HARRIS, C. M. and MILLER, P. J. O. (2015). Severity of expert-identified behavioural responses of humpback whale, minke whale and northern bottlenose whale to naval sonar. *Aquat. Mamm.* **41** 469–502.

SOUTHALL, B. L., BOWLES, A. E., ELLISON, W. T., FINNERAN, J. J., GENTRY, R. L., GREENE, C. R., KASTAK, D., KETTEN, D. R., MILLER, J. H., NACHTIGALL, P. E., RICHARDSON, W. J., THOMAS, J. A. and TYACK, P. L. (2007). Marine mammal noise exposure criteria: Initial scientific recommendations. *Aquat. Mamm.* **33** 411–521.

SOUTHALL, B. L., MORETTI, D., ABRAHAM, B., CALAMBOKIDIS, J., DERUITER, S. L. and TYACK, P. L. (2012). Marine mammal behavioral response studies in southern California: Advances in technology and experimental methods. *Mar. Technol. Soc. J.* **46** 48–59.

STIMPERT, A. K., DERUITER, S. L., SOUTHALL, B. L., MORETTI, D. J., FALCONE, E. A., GOLDBOGEN, J. A., FRIEDLAENDER, A., SCHORR, G. S. and CALAMBOKIDIS, J. (2014).

Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii*, exposed to simulated sonar. *Sci. Rep.* **4** 7031.

TOWNER, A., LEOS-BARAJAS, V., LANGROCK, R., SCHICK, R. S., SMALE, M. J., JEWELL, O., KASCHKE, T. and PAPASTAMATIOU, Y. P. (2016). Sex-specific and individual preferences for hunting strategies in white sharks. *Funct. Ecol.* **30** 1397–1407.

TYACK, P., GORDON, J. and THOMPSON, D. (2003). Controlled exposure experiments to determine the effects of noise on marine mammals. *Mar. Technol. Soc. J.* **37** 41–53.

TYACK, P. L., ZIMMER, W. M. X., MORETTI, D., SOUTHALL, B. L., CLARIDGE, D. E., DURBAN, J. W., CLARK, C. W., D'AMICO, A., DIMARZIO, N., JARVIS, S., MCCARTHY, E., MORRISSEY, R., WARD, J. and BOYD, I. L. (2011). Beaked whales respond to simulated and actual navy sonar. *PLoS ONE* **6** e17009.

VAN DE KERK, M., ONORATO, D. P., CRIFFIELD, M. A., BOLKER, B. M., AUGUSTINE, B. C., MCKINLEY, S. A. and OLI, M. K. (2015). Hidden semi-Markov models reveal multiphasic movement of the endangered Florida panther. *J. Anim. Ecol.* **84** 576–585.

VENZON, D. J. and MOOLGAVKAR, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Appl. Stat.* **37** 87.

WOOD, S. N. (2001). Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting. *Biometrics* **57** 240–244.

ZUCCHINI, W., MACDONALD, I. L. and LANGROCK, R. (2016). *Hidden Markov Models for Time Series*: *An Introduction Using R* (*Second Edition*). Chapman & Hall, London.

ZUCCHINI, W., RAUBENHEIMER, D. and MACDONALD, I. L. (2008). Modeling time series of animal behavior by means of a latent-state model with feedback. *Biometrics* **64** 807–815.

S. L. DERUITER
MATHEMATICS AND STATISTICS DEPARTMENT
CALVIN COLLEGE
3201 BURTON STREET SE
GRAND RAPIDS, MICHIGAN 49506
USA
E-MAIL: stacy@calvin.edu

T. SKIRBUTAS
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF ST ANDREWS
THE OBSERVATORY, BUCHANAN GARDENS
ST ANDREWS KY16 9LZ
UNITED KINGDOM
E-MAIL: skirbutas.tomas@gmail.com

J. CALAMBOKIDIS
CASCADIA RESEARCH COLLECTIVE
218 1/2 4TH AVENUE W.
OLYMPIA, WASHINGTON 98501
USA
E-MAIL: calambokidis@cascadiaresearch.org

R. LANGROCK
DEPARTMENT OF BUSINESS ADMINISTRATION
    AND ECONOMICS
BIELEFELD UNIVERSITY
POSTFACH 10 01 31
BIELEFELD
GERMANY
E-MAIL: roland.langrock@uni-bielefeld.de

J. GOLDBOGEN
HOPKINS MARINE STATION
STANFORD UNIVERSITY
120 OCEANVIEW BOULEVARD
PACIFIC GROVE, CALIFORNIA 93950
USA
E-MAIL: jergold@stanford.edu

A. FRIEDLAENDER
OREGON STATE UNIVERSITY
FISHERIES AND WILDLIFE
MARINE MAMMAL INSTITUTE
HATFIELD MARINE SCIENCE CENTER
2030 MARINE SCIENCE DRIVE
NEWPORT, OREGON 97365
USA
E-MAIL: Ari.Friedlaender@oregonstate.edu

B. SOUTHALL
SEA, INC.
9099 SOQUEL DRIVE, SUITE 8
APTOS, CALIFORNIA 95003
USA
E-MAIL: brandon.southall@sea-inc.net