



## Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data

Kylie L. Scales, Elliott L. Hazen, Michael G. Jacox, Christopher A. Edwards, Andre M. Boustany, Matthew J. Oliver and Steven J. Bograd

K. L. Scales (<http://orcid.org/0000-0003-0843-0956>) ([kylie.scales@noaa.gov](mailto:kylie.scales@noaa.gov)), M. G. Jacox and C. A. Edwards, *Inst. of Marine Sciences, Univ. of California Santa Cruz, CA, USA.* – E. L. Hazen, S. J. Bograd, KLS and MGJ, *National Oceanic and Atmospheric Administration (NOAA) Southwest Fisheries Science Center, Environmental Research Division, Monterey, CA, USA.* – A. M. Boustany, *Nicholas School of the Environment, Duke Univ., Durham, NC, USA.* – M. J. Oliver, *School of Marine Science and Policy, Univ. of Delaware, Lewes, DE, USA.*

Understanding and predicting the responses of wide-ranging marine predators such as cetaceans, seabirds, sharks, turtles, pinnipeds and large migratory fish to dynamic oceanographic conditions requires habitat-based models that can sufficiently capture their environmental preferences. Marine ecosystems are inherently dynamic, and animal–environment interactions are known to occur over multiple, nested spatial and temporal scales. The spatial resolution and temporal averaging of environmental data layers are therefore key considerations in modelling the environmental determinants of habitat selection. The utility of environmental data contemporaneous to animal presence or movement (e.g. daily, weekly), versus synoptic products (monthly, seasonal, climatological) is currently debated, as are the trade-offs between near real-time, high resolution and composite (i.e. synoptic, cloud-free) data fields.

Using movement simulations with built-in environmental preferences in combination with both modelled and remotely-sensed (ROMS, MODIS-Aqua) sea surface temperature (SST) fields, we explore the effects of spatial and temporal resolution (3–111 km, daily–climatological) in predictive habitat models. Results indicate that models fitted using seasonal or climatological data fields can introduce bias in presence-availability designs based upon animal movement datasets, particularly in highly dynamic oceanographic domains. These effects were pronounced where models were constructed using seasonal or climatological fields of coarse ( $> 0.25$  degree) spatial resolution. However, cloud obstruction can lead to significant information loss in remotely-sensed data fields. We found that model accuracy decreased substantially above 70% data loss. In cloudy regions, weekly or monthly environmental data fields may therefore be preferable. These findings have important implications for marine resource management, particularly in identifying key habitats for populations of conservation concern, and in forecasting climate-mediated ecosystem changes.

Habitat modelling techniques are increasingly used to understand species' distributions and environmental preferences, and to inform conservation planning. Accurate, representative models with good predictive capabilities are essential for the identification of key habitats for populations of conservation concern (Guisan et al. 2013); effective management of anthropogenic threat (e.g. mitigation of ship strike on whales, Redfern et al. 2013), and forecasting changes in habitat suitability under future climate change (Durner et al. 2009, Hazen et al. 2013a, b, Hannah et al. 2014, Robinson et al. 2015, Willis-Norton et al. 2015). In the oceans, accurate models are especially important where management of marine resources seeks to integrate the dynamic nature of both pelagic ecosystems and anthropogenic activities in near real-time (i.e. Dynamic Ocean Management, DOM; Hobday et al. 2011, Howell et al. 2015, Maxwell et al. 2015).

Habitat-based models for wide-ranging marine predators such as cetaceans, seabirds, sharks, turtles, pinnipeds and

large migratory fish must reflect the multiple, nested spatial and temporal scales over which animal–environment interactions occur to be accurate and therefore useful in marine spatial planning. Accordingly, both static (e.g. depth) and dynamic (e.g. sea surface temperature, SST) environmental variables are widely used in model construction (Michael et al. 2014). Remotely-sensed data fields are most often used as dynamic variables (Block et al. 2011, Nur et al. 2011, Becker et al. 2012, Pikesley et al. 2014), although oceanographic models (e.g. Regional Ocean Model System, ROMS) are becoming more widely used for this purpose because outputs provide valuable sub-surface information and are not cloud-affected (Becker et al. 2012). However, few studies investigate how the choice of environmental data fields can impact model performance. The comparative utility of environmental data fields contemporaneous to animal movement (e.g. daily, weekly) and longer-term averages (seasonal, climatological) is currently debated, as are the most appropriate levels of spatial resolution for

understanding mechanistic linkages between oceanographic conditions and animal responses.

Temporally-averaged environmental data fields of coarse spatial resolution are often used to contextualise movement or sightings datasets for wide-ranging marine predators (Kaschner et al. 2006, Louzao et al. 2011, Mannocci et al. 2014, 2015, Arrizabalaga et al. 2015). While climatological fields can provide a cloud-free, synoptic measure of the environment, of particular utility for modelling broad-scale movements of migratory animals, there is a clear mismatch in spatial and temporal scale between climatological data fields and limited duration (tracking) or snapshot (sightings) animal response datasets. To our knowledge, this is the first study to address the appropriate levels of spatial and temporal averaging in habitat-based models constructed using short-timespan movement data, despite the recent proliferation of tracking and biologging studies (Wilmers et al. 2015).

Here, we simulate movements of a hypothetical wide-ranging marine predator in a highly dynamic Eastern Boundary Upwelling Ecosystem (EBUE) to test the effects of spatial and temporal resolution on the accuracy of habitat-based models. We prescribe a habitat preference for contemporaneous conditions, based on high-resolution ROMS output. Recent work has revealed the value of including behavioural state (i.e. foraging, transit) in movement-based models (Morales et al. 2004, Patterson et al. 2009, Bestley et al. 2015). Accordingly, we create sets of movement simulations with environmental preferences underlying both presence-absence and behavioural state-switches, with three key aims; 1) to explore the effects of spatial resolution and temporal averaging of environmental data in habitat-based models; 2) to identify optimal levels of spatial and temporal resolution for oceanographic contextualisation of animal movements through a highly dynamic ocean region; 3) to examine variability in spatial predictions of habitat suitability resulting from models constructed using environmental fields of varying resolution.

## Material and methods

### Environmental data

Daily SST fields were obtained from an implementation of ROMS (Shchepetkin and McWilliams 2005) configured for the California Current System (CCS). The model was run from 1999 to 2013, within a spatial domain spanning 30–48°N and 115.5–134°W at 1/30° (~ 3 km) horizontal resolution and 42 terrain-following levels. Surface forcing (atmospheric influence such as wind stress, heat and fresh-water flux) was derived from the Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS; Hodur et al. 2002). Lateral boundary conditions (conditions at the edges of the model domain) were climatological, obtained from the World Ocean atlas (Antonov et al. 2010, Locarnini et al. 2010). The base model configuration and performance are described in Veneziani et al. (2009), and the higher resolution version used here has been applied to studies of larval dispersal (Drake et al. 2011, 2013, 2015) and quantifying resolution impacts on air-sea carbon exchange (Fiechter et al. 2014). Variability in (sub-

mesoscale (~ 1 km; 10 s–100 s of kilometres) oceanographic conditions within the modelled domain was assessed through calculation of decorrelation timescales per pixel (time lag in days such that no significant correlation remains, using an autocorrelation threshold of  $1/e$ , ~ 0.37; Supplementary material Appendix 1, Fig. A1).

In order to assess the confounding influence of cloud cover in remotely-sensed imagery, and compare models built using ROMS to those constructed using freely-available remotely-sensed data, we also acquired daily Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) SST imagery for a matching timespan at 0.0125 degree resolution (Local Area Coverage, US West Coast), via NOAA Environmental Research Division's Data Access Program (ERDDAP; <<http://coastwatch.pfeg.noaa.gov/erddap/>>). Cloud masks from each of these daily layers were used in subsequent analytical steps.

### Degradation of resolution

All daily environmental data fields were reprojected to an equal area projection (metres). We then degraded the spatial resolution of each SST field (ROMS, MODIS) to a series of progressively coarser resolutions (3, 4, 9, 25, 50, and 111 km). Temporal averaging was addressed through a similar process of progressive coarsening (daily, weekly, monthly, seasonal, climatological; mean SST over timespan) of each SST field at each spatial resolution (Fig. 1).

### Movement simulations

We constructed movement simulations for a hypothetical wide-ranging marine predator, representative of the scale of movement of a taxonomically diverse range of pelagic predator species that frequent the California Current Large Marine Ecosystem (CCLME; e.g. whales, albatrosses, sharks, cetaceans, turtles). To improve realism of these simulations, movement parameters were based on actual movements of 159 blue whales *Balaenoptera musculus* tracked through the CCLME using Argos satellite telemetry (Bailey et al. 2009).

We used travel speeds observed during transit and area-restricted search (ARS) behaviours (mean = 3.70 and 1.05 km h<sup>-1</sup> respectively) to parameterise step length distributions for each behavioural state, drawn from gamma distributions. The gamma distribution with parameters shape (a) and scale (s) has the probability density function:

$$f(x) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-\left(\frac{x}{s}\right)} \quad x \geq 0$$

For transit, we used  $a = 3.114$ ,  $s = 28900$ , calculated using a mean (SD) daily step length of  $90\,000 \pm 51\,000$  m and for ARS  $a = 1.181$ ,  $s = 21\,160$ , calculated from a mean (SD) of  $25\,000 \pm 23\,000$  (Fig. 2; Langrock et al. 2012; GammaDist() R function). Similarly, we used the observed distribution of turning angles between successive locations in each behavioural state to parameterise von Mises distributions of turning angles for simulations. The von Mises distribution with parameters  $\kappa$  and  $\mu$  has the probability density function:

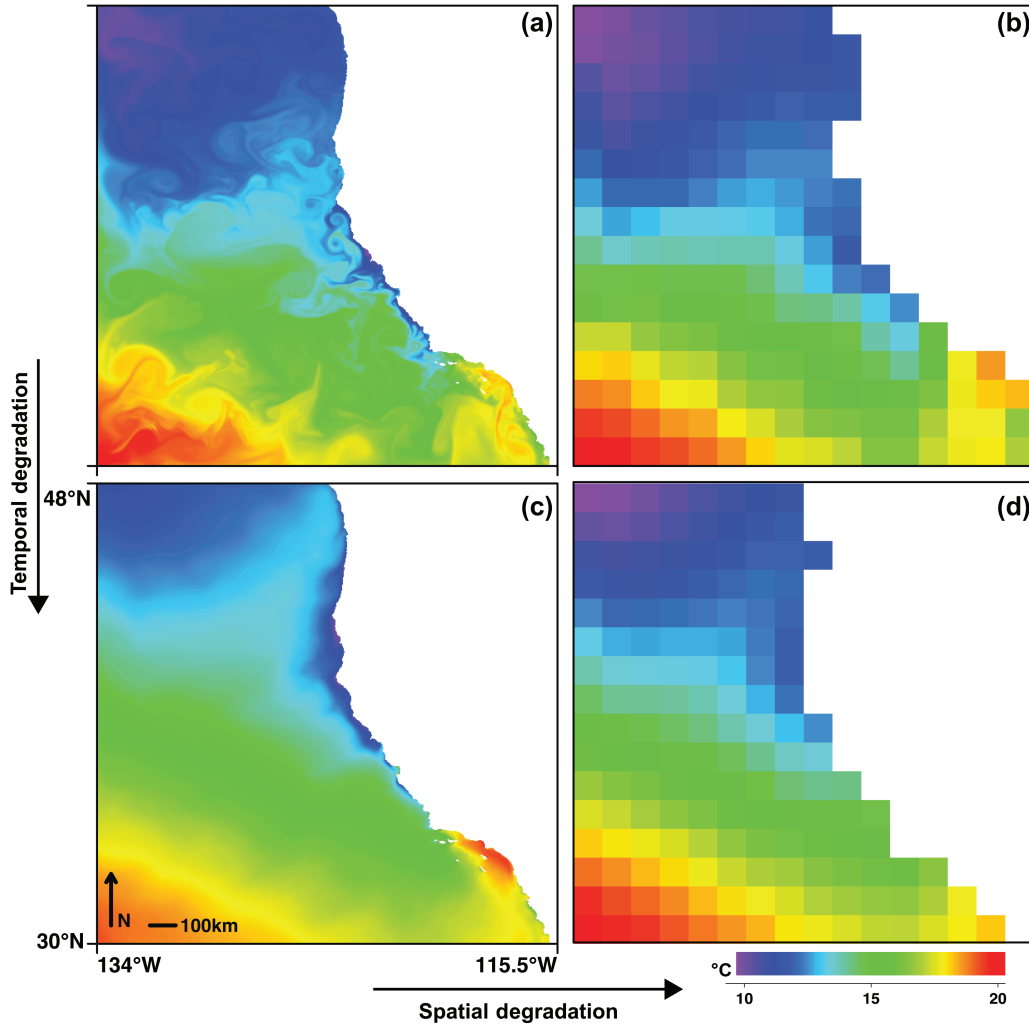


Figure 1. Degradation of environmental data fields. Degradation of spatial resolution from (a) daily 3 km ROMS SST fields to (b) daily 111 km ( $\sim 1$  degree) fields, and from (c) climatological 3 km ROMS SST field to (d) climatological 111 km ( $\sim 1$  degree) field, through bilinear interpolation. Temporal degradation from (a) daily 3 km ROMS SST fields to (c) climatological 3 km field, and from (b) daily 111 km ( $\sim 1$  degree) to (d) climatological 111 km ( $\sim 1$  degree), through median averaging.

$$f(x) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)} \quad 0 < x < 2\pi$$

We used von Mises parameters,  $\mu = 0$ ,  $\kappa = 10$  for transit and  $\mu = 0$ ,  $\kappa = -3$  for ARS, (Fig. 2; CircStats package for R).

In order to assess the effects of degrading resolution in both presence-availability and behavioural state model frameworks, we used three forms of movement simulation, 1) correlated random walks (CRWs) with an encoded environmental preference paired with, 2) CRWs with no environmental preference to generate pseudo-absences and 3) CRWs based on a two-state, discrete-time semi-Markovian state-switching process with an encoded environmental preference (Fig. 2) underlying behavioural state switches (similar to Hidden Markov Models, HMMs, applied to tracking data; Patterson et al. 2009, Langrock et al. 2012).

Daily ROMS SST fields for May–July 2010 were used to construct movement simulations. Each simulation generated a daily location over a three month duration (92 total). Starting locations were sampled at random from within a

bounding box enclosing the core CCLME (CRWs, latitude: 30°N–48°N, longitude: 126°W–116°W; HMM, latitude: 32°N–42°N, longitude: 122°W–125°W). Simulations roamed over a wider region of the Northeast Pacific (latitude: 30°N–48°N; longitude: 115.5°W–134°W) and were permitted to approach, but not to cross, land.

A total of 50 CRWs with an environmental preference, paired with 50 with no preference and matching starting locations, were constructed for presence-availability models. For CRWs, step length and turning angle distributions were combined in a 70:30 ratio of transit to ARS, matching that observed by Bailey et al. (2009). Step length–turning angle pairs were drawn at random from this combined dataset for each successive point location. For the CRWs with an environmental preference, a probability of presence at each successive location ( $P_i$ ;  $0 \leq P_i \leq 1$ ) was derived using a logistic link function:

$$\text{logit}(P_i) = \beta_0 + \beta_1 X_i$$

Where  $X$  is SST extracted at that location and  $\beta_0$ ,  $\beta_1$  describe the prescribed thermal preference (Fig. 3). We

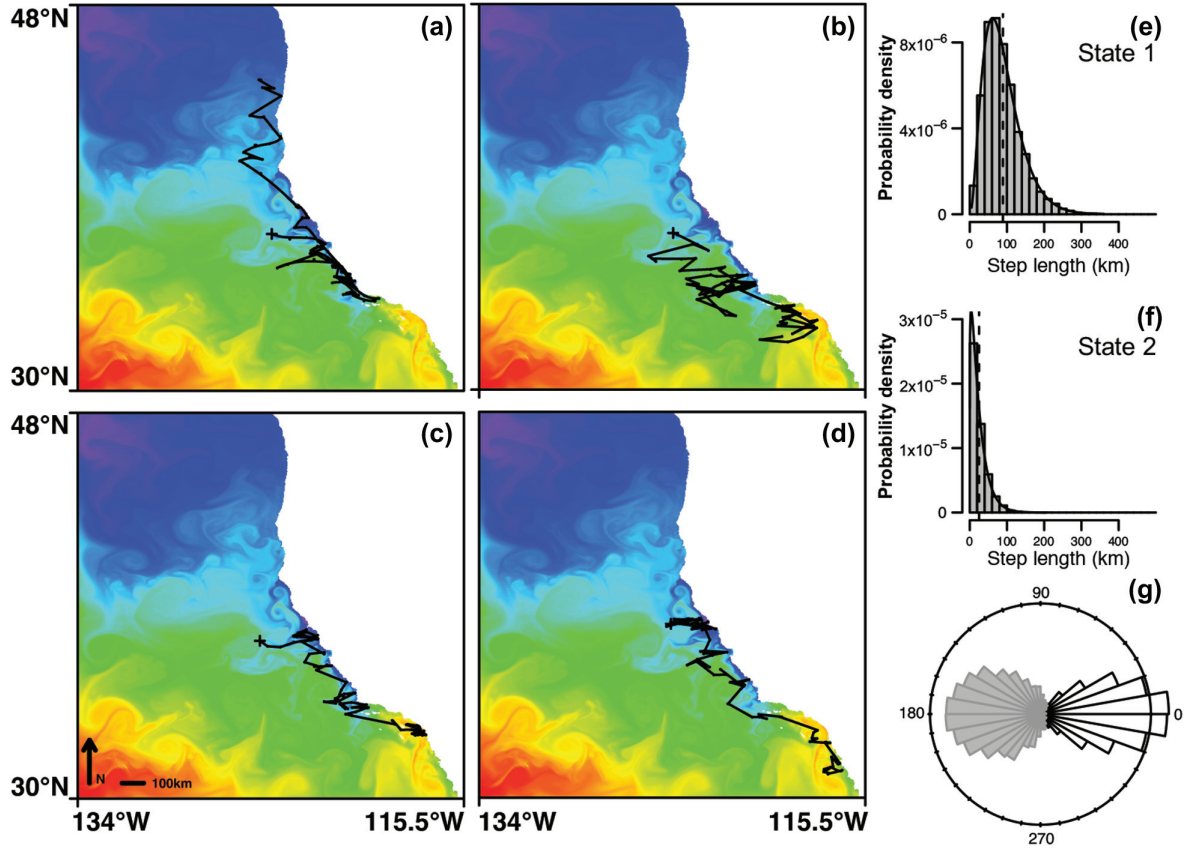


Figure 2. Movement simulations (a) correlated random walk (CRW) with encoded environmental preference, (b) correlated random walk (CRW) with no environmental preference but matching start location. (c), (d) Examples of movement paths resulting from two-state, discrete-time semi-Markovian state-switching process with environmental preference underlying behavioural state switches. (e), (f) Step lengths for behavioural states 1 (transit) and 2 (area-restricted search, ARS), used together with (g) turn angles for state 1 (black) and state 2 (grey) to build all movement simulations. Simulated paths plotted over one daily ROMS SST layer (1 June 2010).

conducted a sensitivity analysis to determine the optimal parameterisation of logistic functions, trialling both linear and quadratic functions. If the resultant probability of presence at each location was greater than 0.5, the location was retained in the series. If less than 0.5, the step length and turning angle distributions were resampled and a new location generated. This process was repeated to generate a complete movement path. Environmental preferences selected for use in simulations (Fig. 3) were those that were best detected by statistical models fitted at the same spatial and temporal resolution as the data fields upon which simulations were constructed, and were not designed to approximate true thermal preferences of blue whales. ROMS SST values for this period ranged from 7–25°C, so environmental preferences were simulated over this range.

For behavioural state models, we simulated a semi-Markovian state switching process, where the behavioural state at time  $t$  was conditional on both the contemporaneous thermal environment and the behavioural state at time  $t_{-1}$  (1st order Markov chain). Each simulation started in transit, state 1. A state-transition probability matrix conditioned on a logistic link to SST determined the probability of switching from transit (state 1,  $S_1$ ) to ARS ( $S_2$ , state 2) at each location:

$$\begin{aligned} \tau(X, \theta) &= \begin{pmatrix} S_2 \rightarrow S_2 & S_2 \rightarrow S_1 \\ S_1 \rightarrow S_2 & S_1 \rightarrow S_1 \end{pmatrix} \\ &= \begin{pmatrix} P(S_2 | S_2) & 1 - P(S_2 | S_2) \\ 1 - P(S_1 | S_1) & P(S_1 | S_1) \end{pmatrix} \\ &= P(S_t = j | s_{t-1} = k, X, \theta) \end{aligned}$$

where the elements of the transition matrix are  $\tau_{k,j} = P(S_t = j | s_{t-1} = k, X, \theta)$  and  $s_t = j$  denotes the behavioural state ( $S_1$  or  $S_2$ ) at time  $t$  (Patterson et al. 2009). This describes the probability of switching between states as a function of covariate  $X$  and parameters  $\theta$ . In our simulations, the covariate  $X$  is SST. The probability of switching from state 1 to state 2 is hence given by  $P(s_t = j | s_{t-1} = j) = \text{logit}^{-1}(\alpha_j + \beta_j X)$  and  $P(s_t = j | s_{t-1} = k) = 1 - P(s_t = j | s_{t-1} = j)$ . The probability of switching from state 2 to state 1 was set at a constant (0.7), obtained from the proportions of time tracked animals spent in each state in Bailey et al. 2009. Serial autocorrelation between states was implicitly driven by spatial and temporal autocorrelation in environmental data.

### Statistical models of habitat preference

Input datasets for model construction were derived through extraction of SST values from each of the series of ROMS



and MODIS fields (at different levels of spatial and temporal degradation), for all point locations along each simulated movement path. All input datasets were centred at 16°C to match logistic link functions describing the prescribed thermal preference. Any point locations falling within cloud-obscured areas of remotely-sensed imagery (i.e. within cloud-masks) were assigned a 'missing' value (NA).

We tested the effects of degrading environmental surfaces on model inference using generalised linear mixed modelling (GLMM; lme4 package for R, Bates et al. 2014). A series of logistic regression models were constructed using locations from the simulated tracks from both presence-absence and behavioural state models over each combination of spatial resolution and temporal averaging of SST fields. For all models, we used a binomial response, SST as a lone fixed effect and simulation replicate as a random effect. Pseudo-absences ('0' response) were derived from CRWs with no environmental preference for presence-availability models and from transit (state 1) locations for behavioural state models. We predicted response curves from each model, using an inverse-logit function, and compared with previously encoded temperature preferences. Absolute deviation from encoded preference parameters (intercept and slope,  $\beta_0$  and  $\beta_1$ ) was used as a scalar measure of model accuracy. We refer the reader to Aarts et al. (2008), Patterson et al. (2009) and Warton and Aarts (2013) for insights into the implications of using presence-absence and presence-only models in modelling habitat preference.

Model performance metrics that are commonly used for model selection from a series of candidates were derived for each output. We chose to use Akaike's information criterion, corrected for sample size (AICc; Sugiura 1978) and the area under the receiver operating characteristic curve (AUC; Bradley 1997) to compare among models (Supplementary material Appendix 1, Table A1–A4).

Spatial predictions of relative habitat suitability (scaled 0–1) were made over the entire study domain by extrapolating from presence-availability models. Predictions of presence-availability models fitted using daily 3 km ROMS data were projected over daily 3 km ROMS fields, and averaged to monthly predictive surfaces using the mean of habitat suitability per pixel. For comparison, models fitted using the climatological ROMS 111 km field were extrapolated over the same climatological field.

## Effects of cloud cover

To assess the effects of cloud cover in remotely-sensed imagery, we used cloud masks derived from daily MODIS fields to mask ROMS fields upon which movement simulations were constructed. MODIS cloud masks were used to ensure that spatial patterning in cloud cover was represented. We conducted a sensitivity analysis using the same cloud masks on all daily ROMS layers in succession (92 d), fitting a series of new models with progressively more cloud-masked datasets. Percentage cloud cover was calculated as the proportion of cloud-obscured pixels over the whole image extent. We also applied daily MODIS cloud masks to time-matched daily ROMS for a realistic representation of cloud cover variability.

R code used to create simulations is available in Supplemental material Appendix 1.

## Results

### Model performance

Our simulations reveal a propensity for the introduction of bias in modelled habitat preferences where temporally-averaged environmental fields (i.e. seasonal, climatological) are used together with movement data to define preferred habitats of wide-ranging marine species in highly dynamic domains (Fig. 3a–b, 4a–d). For presence-availability models based on ROMS fields, the greatest rate of deviation from encoded parameters was observed when moving from monthly to seasonal averaging (Supplementary material Appendix 1, Fig. A2; Table A1).

Additionally, our models' capacity to detect the environmental preferences underlying behavioural state switches was eroded when moving from daily, weekly and monthly environmental data to seasonal or climatological fields, with the greatest rate of information loss occurring between seasonal and climatological (Fig. 4e–h). However, the rate of increase in deviation from encoded parameters was lower for HMM than within the presence-availability model framework (Supplementary material Appendix 1, Table A1–A4), indicating that behavioural state models were less adversely affected by increased temporal averaging.

Importantly, we also observed a divergence between those models selected as the best performing using common measures of model performance (i.e. AICc, AUC) and the accuracy of those models in reproducing a known environmental preference. For example, within the presence-availability framework using ROMS SST, coarse-scale climatological models were identified as the best performing despite notable deviation from the prescribed habitat preference (Fig. 3a–b, 4a–b; Supplementary material Appendix 1, Table A1–A4).

Decorrelation timescales of ROMS SST fields show clear evidence of a dominance of sub-monthly and sub-seasonal variability in (sub-)mesoscale oceanographic conditions within the CCLME domain (Supplementary material Appendix 1, Fig. A1). The high temporal variability of SSTs in the coastal CCLME, where most of the simulations occurred, suggest a mechanism by which temporal averaging over successively longer periods produced the greatest deviation from true environmental preferences (Fig. 4; Supplementary material Appendix 1, Fig. A2; Table A1–A4).

### Predicting presence

The choice of spatial and temporal resolution of environmental data used to fit and project presence-availability models strongly influences spatial predictions of habitat suitability (Fig. 5). Our predictions of monthly means of habitat suitability differ markedly among months, and in comparison to predictions of coarse climatological models. For example, finer-resolution models identified mesoscale features of potential biological importance in this system, such as eddies and upwelling filaments (Fig. 5a–c; Checkley

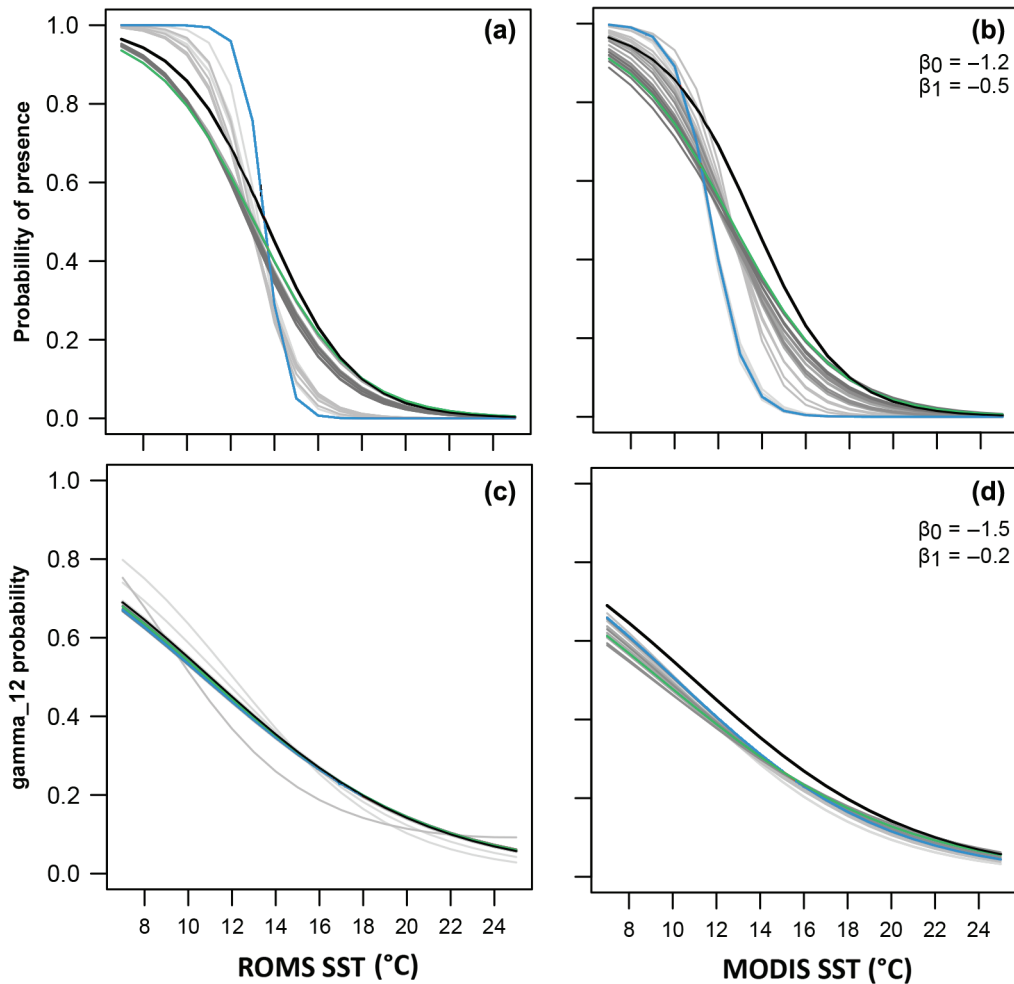


Figure 3. Environmental preference curves. Encoded environmental preferences for (a), (b) presence-availability models and (c), (d) behavioural switching models shown as bold black line, described by parameters  $\beta_0$  and  $\beta_1$ . Model-generated preference curves for spatial and temporal degradation process shown in grey, with AUC-identified ‘best’ model identified in blue and most accurate model (lowest cumulative absolute deviation from encoded parameters,  $\beta_0 + \beta_1$ ) in green. (a), (c) Models fitted using ROMS SST. (b), (d) Models fitted using MODIS SST.

and Barth 2009), whereas climatological model predictions undervalued the importance of the entire coastal upwelling region (Fig. 5d).

### Effects of cloud cover

Cloud cover in daily MODIS fields affected the performance of both presence-availability and behavioural state models (Fig. 4, Fig. 6, Supplementary material Appendix 1, Table A2, A4), particularly for behavioural state models (Fig. 4h) owing to a higher proportion of data loss. In fact, behavioural state models using daily 4 km MODIS SST failed to converge owing to missing environmental data (Supplementary material Appendix 1, Table A4). For presence-availability models fitted with daily 4 km MODIS SST, model error increased rapidly above 70% data loss (Fig. 6). Time-matched cloud masks led to the loss of approximately 67% of environmental data from point extractions, indicating that the use of daily MODIS imagery in cloudy regions can generate inaccurate models.

## Discussion

Our simulations reveal that the use of coarse resolution climatological data fields in habitat-based models for wide-ranging marine predators can lead to inaccuracy in predictions of environmental preferences. This is particularly pronounced where animals exhibit a strong preference for contemporaneous conditions in highly dynamic oceanographic domains.

### Model performance

Regional oceanographic characteristics are a crucial consideration in the construction of habitat-based models – the complexity and dynamism inherent in oceanic conditions interacts with the scale of movement of tracked animals and the resolution of environmental data to affect model inference. For example, the California Current is an intensely dynamic and productive eastern-boundary upwelling ecosystem, in which oceanographic variability influences marine predator behaviour and space use over a continuum

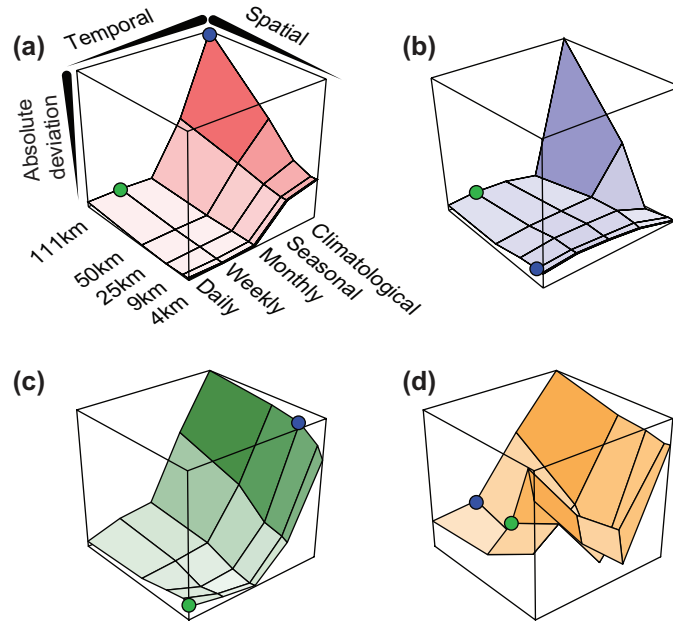


Figure 4. Effects of spatial and temporal degradation on model accuracy, as absolute deviation from parameters describing prescribed thermal preference. (a) Presence-availability model with ROMS SST,  $\beta_1$  parameter. (b) Presence-availability model with MODIS SST,  $\beta_1$  parameter, (c) behavioural state model with ROMS SST,  $\beta_1$  parameter, (d) behavioural state model with MODIS SST,  $\beta_1$  parameter. Best model identified using AUC highlighted with blue circle and best performing model (lowest cumulative absolute deviation from encoded parameters,  $\beta_0 + \beta_1$ ) with green circle.

of timescales, from second-to-second dynamics to multi-decadal periodicity (Checkley and Barth 2009). Capturing this variability in model predictions is crucial to their predictive capacity.

Some studies have selected coarse-scale environmental data in preference to contemporaneous surfaces when using limited duration sightings or movement datasets to predict the responses of wide-ranging marine predators to

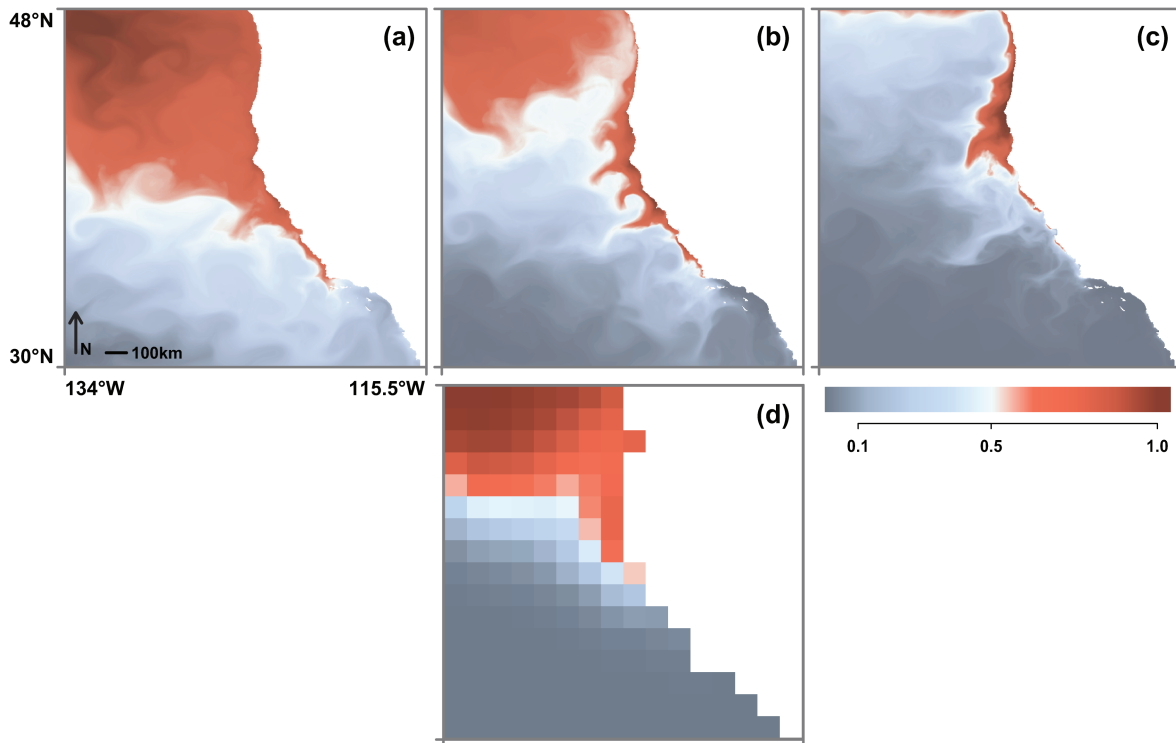


Figure 5. Spatial predictions of presence-availability models. (a), (b), (c) Monthly mean probability of presence per pixel for May, June, July 2010 respectively, generated using presence-availability models fitted with daily ROMS 3 km SST. (d) Probability of presence per pixel as a climatological layer at 111 km ( $\sim 1$  degree) resolution, generated using presence-availability model fitted with climatological ROMS data spatially degraded to 111 km resolution. Same CRW datasets used to fit both sets of models.

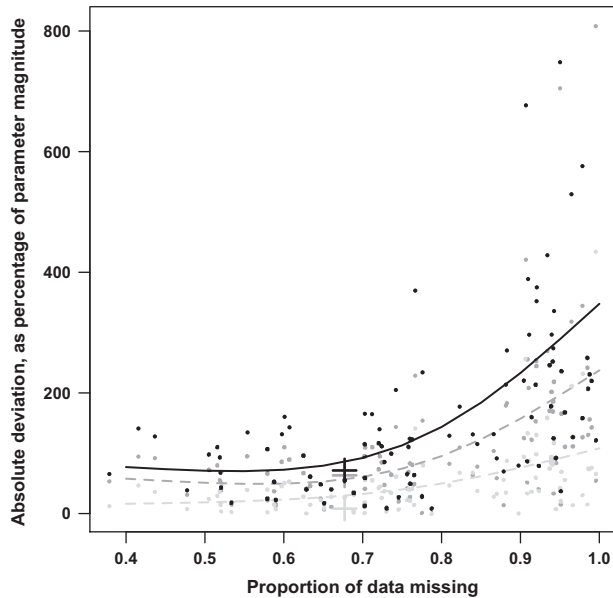


Figure 6. Effects of cloud cover in MODIS SST fields on presence-availability model performance. Proportion of data missing from combined CRW simulations dataset after extraction of SST from cloud-masked daily ROMS 3 km SST fields. Absolute deviation (normalised to magnitude of parameter) from encoded  $\beta_1$  parameter in models fitted using resultant reduced datasets shown as light grey points;  $\beta_0$  as dark grey and cumulative deviation as black points. Regression lines highlight trend of decreasing model performance with increasing cloud cover, of particular significance above 70% missing data. Crosses show parameters derived from model fitted using daily MODIS cloud masks superimposed on time-matched ROMS SST fields.

oceanographic conditions (e.g. climatological, Kaschner et al. 2006, Mannocci et al. 2014, 2015; seasonal, Praca et al. 2009, Arcos et al. 2012). While this approach may be appropriate for highly stable systems, we urge caution when using environmental data fields in this way. Common model performance metrics (AICc, AUC) may select coarse-scale climatological models from a set of candidates, yet our analyses show that these models can fail to accurately reproduce known environmental preferences. Seasonal or climatological fields can smooth over (sub-)mesoscale oceanographic variability that is critical to capture if we are to understand the mechanisms underlying animal–environment interactions in dynamic marine systems.

We found that spatial resolution had a lesser effect than temporal averaging on the magnitude of inaccuracy, over the scales at which spatial and temporal averaging are commonly performed. A progressive coarsening of spatial resolution from 3 km (ROMS) or 4 km (MODIS) to 111 km (~ 1 degree) had little effect on error in models constructed using daily, weekly or monthly environmental fields, for both presence-availability and behavioural state models. This result concurs with previous analyses of the effects of spatial resolution in other modelling applications (Guisan et al. 2007, Redfern et al. 2008, Becker et al. 2010), but see (Storch 2002, Gottschalk et al. 2011, Bean et al. 2014). However, we found that spatial degradation of seasonal and climatological fields increased model error substantially. Bias in environmental preferences was more pronounced

at coarse spatial resolutions, detracting from the reliability of coarse-scale climatological models as predictive tools in marine ecology.

## Predicting presence, predicting behaviour

Variation in the spatial predictions of presence-availability models fitted using environmental fields of differing resolutions highlight important pitfalls of habitat-based modelling. Our results suggests that, in highly dynamic systems, habitat suitability predictions should be based upon presence-availability models constructed using contemporaneous environmental data, particularly in highly dynamic systems (daily, weekly or monthly at most; Nordstrom et al. 2013, Howell et al. 2015). Predicting on the finest temporal scale and averaging those predictions rather than averaging environmental data fields may improve predictive capacity (Becker et al. 2012, 2014, Forney et al. 2015). Coarse-scale climatological models projected onto coarse-scale climatological fields appear likely to lead to more severe prediction error than averaging predictions made on contemporaneous environmental data fields, particularly where animals respond strongly to the contemporaneous physical environment.

Moreover, non-stationarity in animal–environmental interactions through space and time, including behavioural plasticity in movement strategies, diminishes the predictive capacity of presence-availability models constructed using climatological fields (Hidalgo et al. 2012, Dodge et al. 2014, Schmidt et al. 2014). Many species distribution models for wide-ranging species show evidence of poor extrapolation through space and time (Elith and Leathwick 2009, Torres et al. 2015). The modelling of ‘generic’ animal–environment relationships over vast ecological provinces ignores this issue of non-stationarity, particularly where coarse-scale predictions are made using coarse-scale models.

Furthermore, movement data are often obtained from individual animals at frequent intervals via tracking technologies, and so a scale mismatch can manifest when climatological variables are used to model preferences of animals tracked over timescales of days-weeks-months. The effects of this scale mismatch are particularly pronounced where climatological fields smooth out (sub-)mesoscale variability in highly dynamic systems such as the CCLME. Where environmental data fields degrade across specific scales of temporal and spatial variability (Kelly 1985, Abbott and Letelier 1998), effects on model accuracy are likely to be particularly pronounced. In summary, we surmise that the relative value of contemporaneous and climatological oceanographic data fields is likely a function of regional oceanographic character – contemporaneous variables are more informative in highly dynamic regions, whereas climatological fields may suffice in more stable systems.

Wide-ranging marine predators are known to interact with the environment over multiple, nested spatial and temporal scales (Kenney et al. 2001, Fauchald et al. 2011, Hazen et al. 2013b). Including this complexity in habitat-based models may require a range of environmental predictors, for example large-scale migratory movements may be best matched with climatological datasets, whereas understanding fine-scale foraging movements is likely to require



high-resolution data and derived variables (e.g. fronts, eddies; Scales et al. 2014, Cotté et al. 2015). Where mechanistic linkages between seasonal or climatological fields and animal responses can be clearly inferred, synoptic environmental data fields can be useful (Block et al. 2011, Louzao et al. 2011, Hazen et al. 2013a, b, Scales et al. 2014, Pardo et al. 2015). Moreover, the inclusion of a suite of contemporaneous and climatological data fields in the same models may be necessary to capture animal–environment interactions over scales relevant to conservation and management. However, our results indicate that explicit and careful consideration of scale-match is critical in constructing accurate habitat-based models.

Incorporating behavioural state information has been shown to be highly informative in identifying preferred environmental conditions under which animals perform certain activities (e.g. foraging, migration, reproduction; Bailey et al. 2009, Patterson et al. 2009, Roever et al. 2014, Bestley et al. 2015). Consequently, we assessed the influence of spatial resolution and temporal averaging in behavioural state models in addition to the more standard techniques used in presence-availability designs. We found that predictions were less affected by spatial or temporal degradation than presence-availability models but were more adversely affected by data loss owing to cloud cover. Data loss was responsible for complete lack of convergence in behavioural state models using daily MODIS imagery, and poor performance of models using weekly imagery. For these reasons, we recognise a trade-off between the proportion of data missing and the effects of temporal averaging. We found that, in a cloudy region, the best performing behavioural state models were built with daily 3 km ROMS fields, whereas seasonal layers with a fine spatial resolution were preferable when using MODIS (Kinlan et al. 2012).

Global climate models (GCM), increasingly used to generate predictive surfaces on which to forward-project habitat-based models (Hazen et al. 2013a, b, Willis-Norton et al. 2015) typically take a coarse spatial (~ 1 degree) but high- to medium-frequency temporal (daily-monthly) form. Encouragingly, our results indicate that GCM fields may be adequate for forward-projection of models built using movement datasets, providing that non-stationarity and behavioural plasticity in the habitat preferences of focal species can be adequately addressed. In combination with the proliferation of animal movements datasets obtained through tracking and biologging (Wilmers et al. 2015), this increases our capacity to understand, and possibly predict, responses to climate change.

Our findings lead us to recommend the averaging of daily habitat suitability predictions into seasonal or climatological projections in preference to projecting models over seasonal or climatological fields. Where seasonal, annual or climatological GCM products are all that is available, statistical downscaling may be necessary for the preservation of accuracy in habitat-based models (Araújo et al. 2005). Moreover, there is a clear need for a better understanding of the mechanistic linkages between the behaviour of marine predators and dynamic biophysical conditions in pelagic systems, and of how these linkages scale through space and time.

## Conclusions and recommendations

Modelling the environmental preferences of wide-ranging marine predators is non-trivial and constrained by the limitations of current methods for data acquisition, yet models must be accurate if they are to provide realistic predictions for conservation and management. We assert that the use of coarse-scale climatological fields in habitat-based models risks significant predictive inaccuracy, particularly in presence-availability designs that use movement data to detect animal responses to dynamic contemporaneous conditions. Our simulations suggest that the use of contemporaneous environmental data, such as daily or weekly fields, is preferable to fitting and projecting models on coarse-scale climatological fields, particularly in highly dynamic domains.

Spatial management of anthropogenic threats to marine predator populations can only be effective where model predictions correctly identify key habitats and threat hotspots (Maxwell et al. 2015). If inaccurate habitat-based models are used to guide the implementation of spatial conservation measures, we risk focusing on the wrong areas, and the intensification of threat in unidentified key habitats. This consideration is likely to become increasingly relevant as habitat-based models are integrated into dynamic, adaptive ocean management frameworks (Hobday et al. 2011). Full consideration of scale-match between regional oceanographic dynamics and animal responses in habitat-based models is critical if model predictions are to make a meaningful contribution to solutions to the global marine biodiversity crisis.

*Acknowledgements* – The authors gratefully acknowledge the contributions of Roland Langrock in advising HMM development, and participants of a December 2014 workshop at Duke Univ., North Carolina ‘Contemporaneous vs climatological variable use in habitat models’. We also thank Toby Garfield and Ian Jonsen for their contributions to the manuscript. This work was funded under the Cooperative Inst. for Marine Ecosystems and Climate (NOAA SWFSC/UC Santa Cruz).

## References

- Aarts, G. et al. 2008. Estimating space-use and habitat preference from wildlife telemetry data. – *Ecography* 31: 140–160.
- Abbott, M. R. and Letelier, R. M. 1998. Decorrelation scales of chlorophyll as observed from bio-optical drifters in the California Current. – *Deep Sea Res. Part II* 45: 1639–1667.
- Antonov, J. I. et al. 2010. World Ocean atlas 2009, volume 2: salinity. – NOAA Atlas NESDIS 69, U.S. Government Printing Office, Washington, DC.
- Araújo, M. B. et al. 2005. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. – *Global Ecol. Biogeogr.* 14: 17–30.
- Arcos, J. M. et al. 2012. Assessing the location and stability of foraging hotspots for pelagic seabirds: an approach to identify marine Important Bird Areas (IBAs) in Spain. – *Biol. Conserv.* 156: 30–42.
- Arrizabalaga, H. et al. 2015. Global habitat preferences of commercially valuable tuna. – *Deep Sea Res. Part II* 113: 102–112.

- Bailey, H. et al. 2009. Behavioural estimation of blue whale movements in the Northeast Pacific from state-space model analysis of satellite tracks. – *Endang. Species Res.* 10: 93–106.
- Bates, D. et al. 2014. lme4: linear mixed-effects models using Eigen and S4. – R package ver. 1.1-7, <<http://CRAN.R-project.org/package=lme4>>.
- Bean, W. T. et al. 2014. Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. – *J. Appl. Ecol.* 51: 1116–1125.
- Becker, E. et al. 2010. Comparing California Current cetacean – habitat models developed using in situ and remotely sensed sea surface temperature data. – *Mar. Ecol. Prog. Ser.* 413: 163–183.
- Becker, E. et al. 2012. Forecasting cetacean abundance patterns to enhance management decisions. – *Endang. Species Res.* 16: 97–112.
- Becker, E. A. et al. 2014. Predicting seasonal density patterns of California cetaceans based on habitat models. – *Endang. Species Res.* 23: 1–22.
- Bestley, S. et al. 2015. Taking animal tracking to new depths: synthesizing horizontal–vertical movement relationships for four marine predators. – *Ecology* 96: 417–427.
- Block, B. A. et al. 2011. Tracking apex marine predator movements in a dynamic ocean. – *Nature* 475: 86–90.
- Bradley, A. P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. – *Pattern Recog.* 30: 1145–1159.
- Checkley, D. M. and Barth, J. A. 2009. Patterns and processes in the California Current System. – *Prog. Oceanogr.* 83: 49–64.
- Cotté, C. et al. 2015. Flexible preference of southern elephant seals for distinct mesoscale features within the Antarctic Circumpolar Current. – *Prog. Oceanogr.* 131: 46–58.
- Dodge, S. et al. 2014. Environmental drivers of variability in the movement ecology of turkey vultures (*Cathartes aura*) in North and South America. – *Phil. Trans. R. Soc. B* 369: 20130195.
- Drake, P. T. et al. 2011. Dispersion and connectivity estimates along the U.S. west coast from a realistic numerical model. – *J. Mar. Res.* 69: 1–37.
- Drake, P. T. et al. 2013. Influence of larval behavior on transport and population connectivity in a realistic simulation of the California Current System. – *J. Mar. Res.* 71: 317–350.
- Drake, P. T. et al. 2015. The relationship between larval settlement and upwelling-related metrics in a numerical model of the central California coastal circulation. – *Mar. Ecol. Prog. Ser.* in press.
- Durner, G. M. et al. 2009. Predicting 21st-century polar bear habitat distribution from global climate models. – *Ecol. Monogr.* 79: 25–58.
- Eliith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677.
- Fauchald, P. et al. 2011. Scale-dependent response diversity of seabirds to prey in the North Sea. – *Ecology* 92: 228–239.
- Fiechter, J. et al. 2014. Air–sea CO<sub>2</sub> fluxes in the California Current: impacts of model resolution and coastal topography. – *Global Biogeochem. Cycles* 28: 371–385.
- Forney, K. A. et al. 2015. Habitat-based models of cetacean density and distribution in the central North Pacific. – *Endang. Species Res.* 27: 1–20.
- Gottschalk, T. K. et al. 2011. Influence of grain size on species–habitat models. – *Ecol. Model.* 222: 3403–3412.
- Guisan, A. et al. 2007. Sensitivity of predictive species distribution models to change in grain size. – *Divers. Distrib.* 13: 332–340.
- Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. – *Ecol. Lett.* 16: 1424–1435.
- Hannah, L. et al. 2014. Fine-grain modeling of species’ response to climate change: holdouts, stepping-stones, and microrefugia. – *Trends Ecol. Evol.* 29: 390–397.
- Hazen, E. L. et al. 2013a. Predicted habitat shifts of Pacific top predators in a changing climate. – *Nat. Clim. Change* 3: 234–238.
- Hazen, E. L. et al. 2013b. Scales and mechanisms of marine hotspot formation. – *Mar. Ecol. Prog. Ser.* 487: 177–183.
- Hidalgo, M. et al. 2012. A combination of hydrodynamical and statistical modelling reveals non-stationary climate effects on fish larvae distributions. – *Proc. R. Soc. B* 279: 275–283.
- Hobday, A. J. et al. 2011. Seasonal forecasting of tuna habitat for dynamic spatial management. – *Can. J. Fish. Aquat. Sci.* 68: 898–911.
- Hodur, R. M. et al. 2002. The Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). – *Oceanography* 15: 88–98.
- Howell, E. A. et al. 2015. Enhancing the TurtleWatch product for leatherback sea turtles, a dynamic habitat model for ecosystem-based management. – *Fish. Oceanogr.* 24: 57–68.
- Kaschner, K. et al. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. – *Mar. Ecol. Prog. Ser.* 316: 285–310.
- Kelly, K. A. 1985. The influence of winds and topography on the sea surface temperature patterns over the northern California slope. – *J. Geophys. Res. Oceans* 90: 11783–11798.
- Kenney, R. D. et al. 2001. Migration and foraging strategies at varying spatial scales in western North Atlantic right whales: a review of hypotheses. – *J. Cetacean. Res. Manage.* 2: 251–260.
- Kinlan, B. et al. 2012. Predictive modeling of seabird distribution patterns in the New York Bight. A biogeographic assessment of seabirds, deep sea corals and ocean habitats of the New York Bight: science to support offshore spatial planning. – NOAA Technical Memorandum NOS NCCOS, 141.
- Langrock, R. et al. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. – *Ecology* 93: 2336–2342.
- Locarnini, R. A. et al. 2010. World Ocean atlas 2009, volume 1: temperature. – NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington, DC.
- Louzao, M. et al. 2011. Conserving pelagic habitats: seascape modelling of an oceanic top predator. – *J. Appl. Ecol.* 48: 121–132.
- Mannocci, L. et al. 2014. Predicting top predator habitats in the southwest Indian Ocean. – *Ecography* 37: 261–278.
- Mannocci, L. et al. 2015. Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. – *J. Biogeogr.* 42: 1267–1280.
- Maxwell, S. M. et al. 2015. Dynamic ocean management: defining and conceptualizing real-time management of the ocean. – *Mar. Policy* 58: 42–50.
- Michael, P. E. et al. 2014. Relative influence of static and dynamic features on black-footed albatross (*Phoebastria nigripes*) habitat use in central California Sanctuaries. – *Fish. Oceanogr.* 23: 18–31.
- Morales, J. M. et al. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. – *Ecology* 85: 2436–2445.
- Nordstrom, C. A. et al. 2013. Foraging habitats of lactating northern fur seals are structured by thermocline depths and submesoscale fronts in the eastern Bering Sea. – *Deep Sea Res. Part II* 88: 78–96.
- Nur, N. et al. 2011. Where the wild things are: predicting hotspots of seabird aggregations in the California Current System. – *Ecol. Appl.* 21: 2241–2257.
- Pardo, M. A. et al. 2015. Inferring cetacean population densities from the absolute dynamic topography of the Ocean in a hierarchical Bayesian framework. – *PLoS One* 10: e0120727.
- Patterson, T. A. et al. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. – *J. Anim. Ecol.* 78: 1113–1123.

- Pikesley, S. K. et al. 2014. Modelling the niche for a marine vertebrate: a case study incorporating behavioural plasticity, proximate threats and climate change. – *Ecography* 38: 803–812.
- Praca, E. et al. 2009. Modelling the habitat suitability of cetaceans: example of the sperm whale in the northwestern Mediterranean Sea. – *Deep Sea Res. Part I* 56: 648–657.
- Redfern, J. V. et al. 2008. Absence of scale dependence in dolphin-habitat models for the eastern tropical Pacific Ocean. – *Mar. Ecol. Prog. Ser.* 363: 1–14.
- Redfern, J. et al. 2013. Assessing the risk of ships striking large whales in marine spatial planning. – *Conserv. Biol.* 27: 292–302.
- Robinson, L. M. et al. 2015. Trailing edges projected to move faster than leading edges for large pelagic fish habitats under climate change. – *Deep Sea Res. Part II* 113: 225–234.
- Roeber, C. et al. 2014. The pitfalls of ignoring behaviour when quantifying habitat selection. – *Divers. Distrib.* 20: 322–333.
- Scales, K. L. et al. 2014. Mesoscale fronts as foraging habitats: composite front mapping reveals oceanographic drivers of habitat use for a pelagic seabird. – *J. R. Soc. Int.* 11: 20140679.
- Schmidt, A. E. et al. 2014. Non-stationary seabird responses reveal shifting ENSO dynamics in the northeast Pacific. – *Mar. Ecol. Prog. Ser.* 499: 249–258.
- Shchepetkin, A. F. and McWilliams, J. C. 2005. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. – *Ocean Model.* 9: 347–404.
- Storch, I. 2002. On spatial resolution in habitat models: can small-scale forest structure explain capercaillie numbers? – *Conserv. Ecol.* 6: 6.
- Sugiura, N. 1978. Further analysts of the data by Akaike's information criterion and the finite corrections. – *Comm. Stat. Theor. Methods* 7: 13–26.
- Torres, L. G. et al. 2015. Poor transferability of species distribution models for a pelagic predator, the grey petrel, indicates contrasting habitat preferences across ocean basins. – *PLoS One* 10: e0120014.
- Veneziani, M. et al. 2009. A central California coastal ocean modeling study. Part I: the forward model and the influence of realistic versus climatological forcing. – *J. Geophys. Res.* 114: C04 015.
- Warton, D. and Aarts, G. 2013. Advancing our thinking in presence-only and used-available analysis. – *J. Anim. Ecol.* 82: 1125–1134.
- Willis-Norton, E. et al. 2015. Climate change impacts on leatherback turtle pelagic habitat in the Southeast Pacific. – *Deep Sea Res. Part II* 113: 260–267.
- Wilmers, C. C. et al. 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. – *Ecology* 96: 1741–1753.

Supplementary material (Appendix ECOG-02272 at <[www.ecography.org/appendix/ecog-02272](http://www.ecography.org/appendix/ecog-02272)>). Appendix 1.