Dynamic occupancy modelling of temperate marine fish in area-based closures

Running title: Occupancy modelling for marine fish.

Authors

Jay Calvert[1]*, Chris McGonigle[1], Suresh Andrew Sethi[2], Bradley Harris[3], Rory Quinn[1], Jon Grabowski[4].

* Corresponding author: calvert-j3@ulster.ac.uk.

[1] School of Geography & Environmental Science, University of Ulster, Cromore Road, Coleraine, BT52 1SA, UK.

[2] U.S. Geological Survey, New York Cooperative Fish and Wildlife Research Unit, Cornell University, Ithaca, 14853, NY, USA.

[3] Fisheries, Aquatic Science and Technology Laboratory, Alaska Pacific University, Anchorage, 99508, Alaska, USA.

[4] Marine Science Center, Northeastern University, Nahant, 01908, Massachusetts, USA.

Abstract

1. Species distribution models (SDMs) are commonly used to model the spatial structure of species in the marine environment, however most fail to account for detectability of the target species. This can result in underestimates of occupancy, where non-detection is conflated with absence. The site occupancy model (SOM) overcomes this failure by treating occupancy as a latent variable of the model, and incorporates a detection sub-model to account for variability in detection rates. These have rarely been applied in the context of marine fish, and never for the multi-season dynamic occupancy model (DOM).
2. In this study, a DOM is developed for a designated species of concern, cusk (*Brosme brosme*), over a four season period. Making novel use of a high-resolution 3-dimensional hydrodynamic model, detectability of cusk is considered as a function of current speed and algae cover. Algal cover on the seabed is measured from video surveys to divide the study area into two distinct regions; those with canopy forming species of algae and those without (henceforth bottom types).
3. Modelled estimates of the proportion of sites occupied in each season are 0.88, 0.45, 0.74 and 0.83. These are significantly greater than the proportion of occupied sites measured from underwater video observations which are 0.57, 0.28, 0.43 and 0.57. Individual fish are detected more frequently with increasing current speed in areas lacking canopy, and less frequently with increasing current speed in areas with canopy.
4. The results indicate that, where possible, SDM studies for all marine species should take account of detectability to avoid underestimating the proportion of sites occupied at a given study area. Sampling closed areas or areas of conservation often requires the use of non-physical, low impact sampling methods like camera surveys. These methods inherently result in detection probabilities less than one, an issue compounded by time-varying features of the environment that are rarely accounted for in marine studies. This work highlights the use of modelled hydrodynamics as a tool to correct some of this imbalance.

Keywords: Detectability, dynamic occupancy model, hydrodynamics, temperate marine fish.

1 Introduction

The use of species distribution modelling as a tool for scientists and environmental managers has seen a substantial increase in the last three decades, driven by a growing demand for knowledge of species' ranges and facilitated by increasingly powerful computing resources (Barbosa & Schneck 2015). Species distribution models (SDMs) mathematically represent the relationship between species records and features of the environment, often with the intention of predicting suitable ranges for the target species (Franklin 2010). Depending on the aims of an investigation and the type of data collected, several approaches are available to researchers. For many applications where surveys have been planned in advance of statistical analysis, standard methods such as generalized linear models, generalized additive models and random forest are frequently used to model occurrence and abundance data (Elith *et al.* 2011).

However given the time and cost of performing a systematic survey, especially in the marine environment, researchers often use archived datasets to perform modelling investigations (Araújo & Guisan 2006). In many cases, these datasets only contain records on species presence, resulting in a situation where no absence data are available (Elith *et al.* 2011). For this reason, presence-only models such as MaxEnt (Phillips *et al.* 2006) have been developed that do not require absence data, but rather generate a large number of pseudo-absences from the study area (Phillips, Anderson & Schapire 2006).

While each approach outlined above has advantages, both fail to address the issue of detectability (Monk 2014). Both approaches assume that detection probability is invariant, i.e. that the target species is perfectly observed whenever it is present (Yackulic *et al.* 2013). This is often not the case

63 when studying cryptic species, and is an important consideration in the marine environment when
64 sampling methods often do not result in direct observation of the study environment, for example
65 when using trawl or camera surveys (Monk 2014). In a review of 108 articles that used MaxEnt,
66 Yackulic *et al.* (2013) found that only 14% mentioned detection probability. This failure to address
67 detectability introduces error to estimates of occurrence for the species being modelled and can
68 result in erroneous reporting of covariate effects (Guillera-Arroita *et al.* 2014).

69 Site occupancy modelling (SOM) allows occupancy and detectability to be analysed hierarchically as
70 two separate processes, accounting for the problem of imperfect detection (MacKenzie *et al.* 2002).
71 Monk (2014) provides a good overview of the need to adopt this class of model in the marine
72 environment; while available for a similar period as MaxEnt, the SOM has received much less
73 attention (e.g., Coggins *et al.* 2014) in marine ecology investigations. The multi-season occupancy
74 model (or dynamic occupancy model for its Bayesian counterpart (DOM)) allows occupancy,
75 detection, local colonisation and extinction to be accounted for across several sampling seasons
76 (MacKenzie *et al.* 2003). Seasons refer to primary sampling periods within which the population is
77 assumed closed but between which the population can be subject to local extinction and
78 colonisation. The model assumes that at least one site has been visited more than once within a
79 sampling period and that the true occupancy state is imperfectly observed, i.e., it is a latent variable
80 of the model. As such, the model can be thought of as a non-standard GLMM with a binary random
81 effect equal to 1 where the site is occupied by the target species and 0 where it is not (Kéry 2010).
82 Each of the four probabilities within the DOM (initial occupancy, colonisation, extinction and
83 detection) can be modelled as a function of a set of covariate data, or set as constant across sites
84 within a given sample period (MacKenzie *et al.* 2003; Royle & Kéry 2007). Where covariates are used
85 to model probabilities, these must represent variability in the environment along temporal scales
86 relevant to the phenomenon under study. Extinction and colonisation effects, for example, require
87 seasonally varying covariates. By contrast, detection effects require covariates that vary over much
88 shorter temporal scales, allowing differences in detectability to be discerned within relatively short

89 sampling periods. One possible reason why these models have not received as much attention in the
90 marine environment as they have for terrestrial studies is the prohibitive cost of marine sampling. In
91 order to satisfy the assumptions of the dynamic occupancy model, repeat visits are required within
92 each of a number of seasons to collect both response and environmental data.

93 Covariates used for modelling that have been collected *in-situ* at the time of making species
94 observations are often assumed to better describe observed patterns in species distributions
95 (Franklin 2010). Recent research has shown however that this is not always the case, and that
96 modelling studies can benefit from a combination of *in-situ* sampling and remote sensing data
97 (Niedballa *et al.* 2015). Moreover, Newton-Cross *et al.* (2007) demonstrated that remotely sensed or
98 computer generated data can be more effective than data collected *in-situ* for accurately predicting
99 the occurrence of some terrestrial based species. Again, this is important in the marine environment
100 where sampling is often expensive and time consuming compared to terrestrial studies. As such,
101 marine researchers often have to rely on remote sensing (Brown, Sameoto & Smith 2012) and
102 modelled data (Rattray, Ierodiaconou & Womersley 2015) to supplement *in-situ* sampling. These
103 latter datasets often come in the form of hydrodynamic models that mathematically represent tidal
104 and wave forcing of the marine environment (Gunn & Stock-Williams 2013). While widely used in
105 engineering and physical oceanography (e.g., McMillan & Lickley 2008; Chen *et al.* 2011), their use in
106 marine ecological investigations has been limited. This is likely due to the expense, in terms of time
107 and computational power, of setting up a hydrodynamic model that accurately reflects conditions at
108 spatial and temporal scales relevant to ecological processes and organism behaviour.

109 This investigation aims, for the first time, to create a dynamic occupancy model to demonstrate the
110 effectiveness of such an approach for a temperate marine fish species in a closed fishing area. The
111 study generates true, unbiased estimates of occupancy and compares these to occupancy estimates
112 obtained by survey methods alone. Inputs to the dynamic occupancy model comprise observed,
113 derived and simulated data. Observed data include observations of the target species, algae cover

and geomorphological complexity from video-surveys; derived data include depth measured using multibeam echosounder (MBES) data, and terrain attributes derived from the MBES depth data. Additionally, the study makes novel use of current velocities simulated using a high-resolution hydrodynamic model to demonstrate the utility of including these data in fisheries monitoring investigations.

2 Methods

2.1 Analysis overview

This study takes a multifaceted approach to producing a dynamic occupancy model (DOM) for a species of temperate marine fish at a remote rocky outcrop in the central Gulf of Maine. The model was specified in a Bayesian framework to account for a small amount of separation in the detection data and to allow the use of a finite sample estimator that generates more accurate estimates of occupancy for a small sample size (Royle & Kéry 2007). The analysis used data observed or derived from two primary datasets; a series of non-invasive underwater video surveys collected over four sampling seasons, and a MBES survey conducted for the Gulf of Maine mapping initiative (SAIC, 2005). Observations of the response variable (cusk) were recorded from the video footage. During each video survey detections of cusk were recorded along with the time of the observation. Simultaneous observations of algae cover and morphological complexity were made from the video footage, to be later used as explanatory variables. Separately, the digital elevation model extracted from the MBES data was used for three purposes; 1) to derive terrain attributes to assess morphological complexity over the study site; 2) to create a surface of algae cover over the study site based on empirical extinction depths and; 3) as an input to a standalone hydrodynamic model. The hydrodynamic model was used to estimate bottom current conditions at the study site, from which two outputs were generated; 1) point estimates of time-varying bottom current speed at each of the video locations surveyed for cusk and; 2) time-varying surfaces of current speed for the entire study area. Once all primary data had been processed, the DOM was created. Model coefficients for

139 the significant terms in the detection sub-model of the DOM (observed algae cover and bottom
140 current point estimates) were obtained and used to generate the final outputs. This was achieved by
141 predicting the model coefficients over the algae cover surface and bottom current surfaces created
142 above, creating spatio-temporally varying predictions of detection probability for the study area.
143 Each step of the analysis is described in full in the following sections.

144 **2.2 Candidate species and area**

145 The candidate species for this investigation is cusk (*Brosme brosme*, Lotidae); a cryptic, bottom
146 dwelling species found in the Eastern and Western Atlantic Ocean, and designated a species of
147 concern by the National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries
148 Service (NOAA 2009). In the Western Atlantic cusk are found from Nova Scotia in Canada to New
149 Jersey in the USA, and typically stay in deeper waters (>100 m) in these areas. Within the Gulf of
150 Maine, however, they are typically found in shallower water owing to the relatively shallow depths
151 of the internal Gulf (Bigelow & Schroeder 1953). While relatively little is known about their specific
152 life history and ecology (Davies & Jonsen 2011), it has been noted that cusk prefer structured habitat
153 and use kelp forests, boulder piles and rock crevices as refugia (Auster & Lindholm 2005; Hare *et al.*
154 2012). In addition they are considered to be weak swimmers (Bigelow & Schroeder 1953), so make
155 an ideal species for study on how their behaviour is affected by variability in movements of the
156 water column. The species also has a small home range (Dultz 2013), making it suited to the
157 assumptions of the dynamic occupancy model.

158 Data for the investigation were collected at Cashes Ledge in the central Gulf of Maine, approximately
159 170 km northeast of Boston (Figure 2a). The Ledge has been closed to bottom tending fishing gears
160 since 2002 (Sherwood & Grabowski 2016) and supports large resident populations of several
161 commercially important fish species (Grabowski, McGonigle & Brown 2010). The site displays a
162 tripartite zonation of macroalgae around the summit with each of the three zones reaching record
163 depths for boreal-subarctic waters; leathery macrophytes to 40 m, foliose red algae to 50 m and

crustose algae to 63 m (Vadas & Steneck 1988). The Ledge comprises morphologically complex granite shoaling at 10 m water depth, with sand and gravel deposits appearing around 60 m and silt dominated habitats below 80 m. To the east and west, the site is flanked by relatively deep basins (< 220 m) dominated by sands, fine silt and clay (Uchupi & Bolmer 2008).

2.3 Data collection and covariate generation.

Underwater video surveys were conducted at 14 sites on Cashes Ledge by the Gulf of Maine Research Institute using a drop camera in summer 2006 and spring, summer and autumn 2007. Sites are defined here as compact geographical areas wherein samples are less than 30 m apart and features of the abiotic environment are homogenous. Sampling was stratified by depth to include two shallow sites (< 20 m), six intermediate sites (20 – 40 m) and six deep sites (> 40 m). During each sampling season a maximum of three replicate surveys were conducted at each site, with an average of two. Camera units were deployed and left *in-situ* for up to 1.5 hours during each survey, recording the time in and position of the camera. The camera was mounted on the sampling equipment such that the field of view was parallel to the seabed; no directional controls were in place, so the azimuthal direction varied between samples. For full camera set up see Grabowski *et al.* (2010). Any samples where the camera equipment landed with the camera facing into the water column and therefore unable to view the seabed were discarded and not used in any further analyses. After checking each sample for positional accuracy, videos were examined noting the time on the video any cusk were observed. For videos where cusk were present, the video time was combined with the survey start time to obtain the exact time of the observation. Where cusk were absent in a video, a time was randomly sampled from the length of the video to obtain a time for the null observation.

In addition to the observations of cusk, the videos were used to qualitatively assess fine scale morphological complexity (high, moderate, low) and algae density. Algae density was used to classify the study area into areas of two different bottom types according to algae cover, henceforth bottom type; areas with canopy forming species, and areas with no canopy forming species . Sampling effort

189 was defined simply as the length of bottom time in each video. Depth for each observation was
190 obtained from 5 m resolution MBES data collected for the Gulf of Maine Mapping Initiative (SAIC
191 2005). Morphological complexity was derived from the MBES data using the relative deviation from
192 the mean value (RDMV) as recommended by Lecours *et al.* (2017) in a 3 x 3 cell moving window. A
193 breakdown of the mean, minimum and maximum values for each of the depth strata sampled is in
194 table 1.

195 **2.4 Hydrodynamic model**

196 Hydrodynamics were assessed using a coupled wave-current model produced using MIKE by DHI
197 (Danish Hydraulic Institute, 2014). The current model solves the three-dimensional incompressible
198 Reynolds averaged Navier-Stokes equations, while the wave model solves a fully spectral wind-swell
199 formulation. The models are coupled to include wave-current interactions and are solved using a
200 finite volume method over a flexible mesh that allows higher resolution in areas of interest (Danish
201 Hydraulic Institute, 2014).

202 The domain for the model (Figure 2b) incorporated the Gulf of Maine from Tor Bay to Rhode Island,
203 and extends seaward off the continental shelf to allow the Gulf to respond freely to tidal forcing
204 (McMillan and Lickley, 2008). Forcing was supplied to the model as spatially and temporally varying
205 surface elevation from the DTU10 0.125° global tidal model (Cheng & Andersen 2010). Calibration of
206 the model was achieved by adjusting the value of bed resistance over a series of 13 month
207 simulations (one month warmup, 12 month usable data). After each calibration simulation, harmonic
208 analysis was conducted for 67 sites within the Gulf of Maine for comparison against empirical data
209 from Moody *et al.* (1984). Where disagreement between known and modelled data was
210 unacceptably large, values of bed resistance were iteratively adjusted to fine-tune the harmonics.

211 Once the model harmonics were calibrated to maximally correspond to empirical data, the
212 computational mesh was refined to increase resolution at Cashes Ledge. Here, the maximum

horizontal resolution was 135 m with an average horizontal resolution of 220 m. In addition to refining the model mesh, atmospheric forcing was introduced to the model using the National Centres for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CSFR) 6-hourly, 0.5° global weather model (Saha et al., 2010; 2011). Model validation was subsequently conducted on the refined mesh model by hindcasting periods of time not included in the calibration models. Validation included the assessment of model outputs against measured wind, wave and current data. The model was considered validated when outputs were minimally different to measured data. All model runs were performed with a time step of 20 minutes to allow validation against measured data.

Full details of the model setup, calibration and validation can be found in the supplementary materials (S.1). Once the model had been validated, current speeds were extracted from the water-seabed interface layer to capture current variations on the bottom. Times for the observations of cusk were then matched to the temporally closest value of current speed. Visualisation of temporal variability in current magnitude and direction was achieved using a tidal ellipse created using MATLAB (Xu 2002). The ellipse was derived for a point in open water 100 m from the summit in order to capture the movement of water without influence from local topography.

2.5 Dynamic occupancy model

The model was specified using notation from Royle and Kéry (2007) where ψ_t is the probability of initial occupancy at time period t, ϕ_t is the probability that a site remains occupied between t and $t + 1$, γ_t is the probability that a site is colonised between t and $t + 1$ and p is the probability of detection. During data exploration boxplots revealed a high degree of collinearity between RDMV and fine scale complexity observed in the video data. Fine scale complexity was excluded from further analyses; it is a categorical variable so including it would require estimation of more model

parameters. Covariates tested for ψ_t were depth, fine-scale morphological complexity, RDMV, maximum current speed throughout the sampling season (CurMax) and CurMax^2. Depth, fine scale complexity and RDMV as several studies have shown the importance of depth and seabed roughness for cusk habitat selection (Knutsen *et al.* 2009; Davies & Jonsen 2011; Hare *et al.* 2012); CurMax and CurMax^2 to test for any limiting effect of hydrodynamic forcing on cusk habitat selection. Covariate effects were not included for ϕ_t or γ_t and were therefore assumed to be the same for all sites within each sampling period (MacKenzie *et al.* 2003; Royle & Kéry 2007). Covariates included for *p* in the detection sub-model were current speed, sampling effort and bottom type. Current speed as cusk are weak swimmers whose movement is likely to be affected by movements of the water column; sampling effort as the longer the camera is in the water, the more likely it is that a fish will be observed in any given sample; bottom type as canopy forming species are likely to obscure vision and therefore affect detectability of cusk.

Covariate effects for each sub-model were transformed using a logit link function and all continuous covariates were standardised before analysis. In order to simplify model specification two models were assessed, one without (A) and one with (B) covariates for ψ, both of which included covariates for *p*.

Specification for ψ in model B:

$$\text{logit}(\psi) = \beta_1 + \beta_2.\text{Depth} + \beta_3.\text{RDMV} + \beta_4.\text{Current Max} + \beta_5.\text{Current Max}^2$$

where β_i are the regression coefficients, *RDMV* is the relative deviation from the mean value and *current max* is the maximum current speed at the site in any of the sampling seasons.

Specification for *p* in models A and B:

$$\text{logit}(p) = \alpha_1 + \alpha_2 \cdot \text{Current} + \alpha_3 \cdot \text{none} + \alpha_4 \cdot \text{Current}|\text{none}$$

where *current* is the current speed at the time of sampling for cusk, *none* is the bottom type with no canopy forming algae and *Current*|*none* is the interaction between current speed and the bottom type with no canopy forming algae.

2.6 Model validation and outputs

Goodness of fit (GOF) for each model was assessed using Bayesian p-values; for each draw of the MCMC algorithm, new data were simulated given the set of parameters estimated by the model, and Pearson residuals were calculated for each observed and simulated data point. The Bayesian p-value is the proportion of times the simulated residual is greater than the observed residual, with values closer to 0.5 indicating better fit (Kéry 2010).

For the k covariates in each model, importance was tested using binary inclusion variables $w_k \sim$ Bernoulli(0.5). Averaging over the posterior distribution of w_k gives the probability that term k belongs in the model, with values closer to 1 indicating a higher inclusion probability. The inclusion variable for interaction terms was defined as the product of itself and the component inclusion variables as recommended by Kruschke (2014), ensuring the interaction was only assessed when the lower order terms were included. Covariates and interactions were kept in the model when their inclusion variable had a posterior mean greater than 0.5 (Coggins, Bacheler & Gwinn 2014).

All logit scale coefficients were given weakly informative t-distribution priors ($\mu = 0$, $\sigma = 1.566$, $\nu = 7.763$) as recommended by (Dorazio *et al.* 2011), while all non-varying probabilities (ϕ, γ) were given uniform priors from 0 to 1. Sensitivity of the posterior distribution of parameter estimates was assessed following Dorazio *et al.* (2011) using three prior distributions recommended for logistic regression; Jeffreys prior (Firth 1993), t-distribution with $\mu = 0$, $\sigma = 2.5$, $\nu = 7$; those of Gelman *et al.* (2008), t-distribution with $\mu = 0$, $\sigma = 2.5$, $\nu = 1$; and those of Dorazio *et al.* (2011), as detailed above. Identifiability of parameter estimates was assessed by plotting and calculating the amount of overlap between the prior and posterior distributions. These were considered identifiable if the

283 overlap was below 35% (Garrett & Zeger 2000; Gimenez, Morgan & Brooks 2009). The MCMC
284 algorithm was set up with 5 chains, each sampling 50 000 draws. The first 10 000 draws discarded as
285 a burn-in period and every tenth sample was stored for analysis. Chains were assessed visually for
286 mixing and autocorrelation, and convergence was assessed using the Gelman-Ruben diagnostic
287 (Gelman & Rubin 1992) with values less than 1.1 considered converged. Occupancy was determined
288 using a finite sample estimator that is recommended for a small number of non-randomly selected
289 sites (Royle & Kéry 2007). This estimator allows a distinction to be made between estimates derived
290 for parameters of the population, and those derived for sites in the actual sample. The finite sample
291 estimator reduces the variance of the point estimates generated for the sampled sites, and is an
292 additional benefit of fitting the model in a Bayesian framework (Royle & Kéry 2007).

293 Continuous surfaces of modelled current speed were generated from the validated hydrodynamic
294 model to visualise spatio-temporal variability of current speed along a tidal cycle at Cashes Ledge.
295 Predictions of bottom type were made based on the extinction depths of various algae previously
296 reported at the Ledge (Vadas & Steneck 1988). Surfaces of detection probability were then
297 generated for the entire Ledge for the tidal cycle by predicting the occupancy model coefficients
298 over the current speed and bottom type surfaces.

299 All statistical analyses and predictive mapping were carried out in the R environment (R Core Team
300 2015), and Bayesian inference was conducted in JAGS (Plummer 2009).

301

302 **3 Results**

303 **3.1 Cusk observations and data exploration**

304 Of 112 replicate surveys analysed, 39 contained cusk and 73 did not. These correspond to 21 shallow
305 samples, 46 intermediate samples and 45 deep samples representing average depths of 17 m, 33 m

306 and 49 m respectively. Analysis showed 57 video surveys were conducted in the canopy forming
307 region and 55 in the non-canopy forming region.

308 Harmonic analysis of the hydrodynamic model reveals accuracy of 3.8 cm mean absolute deviation in
309 amplitude, and 4.6° in Greenwich phase lag for the M_2 tidal constituent. Bottom-current speeds at
310 Cashes Ledge range from 0.01 ms^{-1} to 0.31 ms^{-1} throughout the four sampling periods, with a
311 maximum of 0.21 ms^{-1} while the camera equipment was deployed. The tidal ellipse and current rose
312 show that maximum flow occurs in a North-South orientation along the semi-major axis of the
313 ellipse (Figure 3). A short time lag of around 1 hour is observed in the arrival time of the tidal signal
314 between the north and south of the Ledge (Figure 4).

315 **3.2 Model validation and outputs**

316 Bayesian p-values were 0.58 for model A (Table 2, top), and 0.62 for model B (Table 2, bottom).
317 Sensitivity analysis showed no significant effect of choice of prior on posterior distributions for all
318 parameters excluding those for the covariates for ψ in model B. The overlap of the posterior
319 distributions of parameters in model A with their prior distributions ranged from 20% to 31%, with
320 the highest amount of overlap for the estimate for sampling effort. All parameters for this model
321 were therefore considered identifiable. Inclusion variables for model A indicated that current speed,
322 bottom type and the interaction between current speed and bottom type should be included in the
323 detection sub-model (Table 2). Inclusion variables for model B indicate less than or equal to 50%
324 probability that any covariates belonged in the occupancy sub-model, and again that current speed,
325 bottom type and the interaction term between current speed and bottom type should be included in
326 the detection sub-model (Table 2).

327 Results for model A show that with increasing current speed, detection rates in the two bottom
328 types diverge; increasing in the non-canopy forming region and decreasing in the canopy forming
329 region (Figure 5a). At low current speeds (≈ 1 cm s^{-1}), the probability of detecting cusk is almost

identical in both bottom types, and there is no credible difference in detection probability between 0 cm s^{-1} and around 5 cm s^{-1} among bottom types (Figure 5b and inset).

Predictions of detection probability reveal similar rates of detectability in both bottom types at low and high tide (Figures 5.t1 & 5.t39), while varying significantly between these times (Figures 5.10, 5.20 & 5.30).

Estimates of the proportion of sites occupied by the finite sample estimator are 0.88 (95% CI: 0.64 – 1.0), 0.45 (95% CI: 0.29 – 0.79), 0.74 (95% CI: 0.50 – 1.0) and 0.83 (95% CI: 0.64 – 1.0) for seasons 1, 2, 3 and 4 respectively. This compares to observed proportions of 0.57, 0.28, 0.43 and 0.57 for season 1, 2, 3 and 4 respectively (Figure 7). Colonisation probabilities for all sites are 0.52, 0.77 and 0.55 for colonisation between seasons 1-2, 2-3 and 3-4 respectively. Extinction probabilities are 0.56, 0.44 and 0.15 for extinction between seasons 1-2, 2-3 and 3-4 respectively; these probabilities are the compliment of ϕ in the parametrisation described in the methods. Population growth is estimated at 0.56 between season 1 and 2, 1.74 between season 2 and 3, and 1.15 between season 3 and 4. This represents a decrease in the number of occupied sites from summer 2006 to spring 2007, an increase from spring to summer 2007, and an increase from summer to autumn 2007. JAGS code for the final model is given in the supplementary information (S.2).

4 Discussion

The problem of bias introduced by failing to account for detectability of mobile fish species when estimating their occupancy was addressed in this study. Using a dynamic occupancy model and treating the true occupancy state as a latent variable of the model, occupancy estimates are significantly greater than occupancy assessed from observation alone. Taking account of imperfect detection, the difference in estimated and observed proportion of sites occupied by cusk range from 0.17 to 0.31. This highlights the need to incorporate detectability of the target species into species distribution modelling efforts (Rota *et al.* 2011).

354 **4.1 Cusk behaviour**

355 Cusk are slow moving weak swimmers (Bigelow & Schroeder 1953; COSEWIC 2003) and it is
356 therefore not unreasonable to expect them to be influenced by hydrodynamic conditions. Outputs
357 of model A show that, indeed, cusk detectability is affected by changes in current speed along a tidal
358 cycle. Within the non-canopy forming regions of Cashes Ledge, the increase in detectability with
359 increasing current speed can be thought of as a proxy for increased activity of cusk. This increase in
360 activity is interpreted in one of two ways, either as cusk searching for morphologically complex
361 environments as refuge, or as cusk using the movement of water as an opportunity to forage for
362 food. Cusk have been observed to prey on cunner (*Tautogolabrus adspersus*) (Auster & Lindholm
363 2005), and cunner in turn have been observed to forage more on exposed surfaces near refugia with
364 increasing current velocity (Auster 1988, 1989).

365 Other than these prey, little is known about cusk diet in the Western Atlantic. In European waters
366 however, stomach contents analysis shows that cusk will eat a range of crustaceans and molluscs,
367 both of which have been found in abundance in the shallower kelp dominated regions at Cashes
368 Ledge (Vadas & Steneck 1988; Witman & Sebens 1992). Within this canopy forming algae region, as
369 current speed begins to increase, detection probability decreases quite rapidly. While marine fish
370 have been shown to use kelp habitats as both refuge and foraging grounds (Holbrook *et al.* 1990;
371 Uhl, Bartsch & Oppelt 2016), the presence of canopy forming species of algae will encumber
372 detectability, especially in relatively high flows or with high energy wave conditions (Rattray,
373 Ierodiaconou & Womersley 2015).

374 In previous studies of cusk habitat usage, the species has been recorded at much greater depths
375 than those observed in this study (Nye *et al.* 2009; Davies & Jonsen 2011; Hare *et al.* 2012). One
376 explanation of the depths observed in this study is that the fish in this region are year round
377 residents that have become accustomed to living at comparatively shallower depths. Bigelow and
378 Schroeder (1953) note that cusk do not often move from bank to bank, and no seasonal spawning

migrations have been noted for the species (COSEWIC 2003). The spawning season for cusk in the Gulf of Maine extends from April to July (Bigelow and Schroeder 1953), and this might explain the large growth rate in the number of sites occupied between Spring 2007 and Summer 2007 (sampling seasons 2 and 3 respectively) as fish become more active in search of a mating partner. These two time periods together take in portions of the spawning season for cusk; it is therefore possible that the later it is in the season, the more active the fish are in their search.

An assumption of the dynamic occupancy model is that within a primary sampling season the population remains closed, but can be open to local extinction and colonisation between seasons (Kéry 2010). Cusk are described as a 'station keeping bottom' or 'station keeping cover' species (Auster & Lindholm 2005), and evidence suggests that the home range of cusk is small (Dultz 2013). Furthermore, cusk show strong affinity for complex habitats while avoiding substrata with no structure (Bigelow & Schroeder 1953; COSEWIC 2003). Cashes Ledge is surrounded on all sides by a number of deep basins, most notably Ammen Basin to the east and Cashes Basin to the west (Figure 2), which consist of unconsolidated substrata (Uchupi 1966; Uchupi & Bolmer 2008). As such, the assumptions of the model are satisfied; cusk remain on station and are kept on Cashes Ledge by expanses of non-favourable habitat on all sides.

Throughout the four sampling seasons within this study, the same camera setup was used. No variability in detection probability should therefore be expected due to gear differences. In investigations using camera surveys combined with fish traps to assess detectability using simultaneous data collection methods (Coggins, Bacheler & Gwinn 2014), the approach of adding cameras to other gear types has been recommended (Bacheler *et al.* 2014). In the current study, the camera system was used in isolation, and given the limited field of view and a lack of control over orientation once on the seabed, detection probabilities are expected to be and are less than one.

No significant effects were observed for any of the covariates tested in the initial occupancy sub-model of model B. While it is likely that some covariates are missing from this sub-model, due to the

small sample size, it is not possible to add more terms without overfitting the model. The covariates that were included were of importance in two ways. First, they were important in terms of what has previously been reported as driving cusk habitat; depth and surface complexity (Hare *et al.,* 2012). Secondly, maximum current speed and maximum current speed squared were included to test whether hydrodynamics play a role not only in determining cusk behaviour, but also in limiting cusk habitat choice. Nevertheless, the main purpose of this study is to give consideration to the detectability issue in marine fish occupancy modelling. Given that the dynamic occupancy model is able to handle constant initial occupancy probabilities across all sites in the study domain, this lack of fit for the initial occupancy state does not present problems for inference about detectability.

4.2 Recommendations for future studies

In this study, detection probability ranged from 0.59 to 0.88 in the non-canopy forming region, and from 0.03 to 0.64 in the canopy forming algae region. These results are broadly comparable to detection probabilities from other studies using camera surveys with other modelling approaches to detect marine fish (Bacheler *et al.* 2014; Coggins, Bacheler & Gwinn 2014). These detection probabilities are conditional on cusk being present at the site being observed, and also need to be considered in light of the fact that bottom time for the camera was relatively high in this investigation. It should also be noted that, for some species of fish, the presence of camera equipment on the seabed may encourage more curious individuals to investigate so may introduce some bias to the results (Stoner *et al.* 2008). In both models assessed in this study, the relationship between sampling effort and cusk detection was found to be insignificant. The term was left in the models however due to its theoretical importance; one of the most important factors affecting the detection of any organism is the amount of effort put into trying to detect it. Especially when dealing with small sample sizes, insignificant covariates should be left in models when they are theoretically important (Schuenemeyer & Drew 2010).

While most of the spatial variability in detection probability can be explained by tidal phase, some small differences persist within each of the bottom types in Figure 6. These can be explained by local differences in current speed caused by water movement around topographic features on the seabed. Differences in detection probability created by these local variations in current speed in the deeper regions, and to a lesser extent in the shallow regions, are better understood when flow direction and topography are considered. Given that the strongest flows indicated by the tidal ellipse occur in a northerly and southerly direction (Figure 3) and that the orientation of the Ledge is approximately SE – NW (Figure 4c), it follows that different areas of the Ledge will experience maximal flow at different stages of the tidal cycle. Similarly, minimum flow occurs at different times throughout the 12.4 hour M_2 tidal period (Figure 3). This has an effect on the arrival time of the increase in current speed at different locations throughout the study area (Figure 4b). Failing to take these local differences into consideration can have consequences when trying to plan similar surveys when the target species may be affected by flow rates. Such camera surveys are often the most suitable method for monitoring marine reserves and areas closed to mobile fishing gears as they are considered both cost effective and non-invasive (Bouchet & Meeuwig 2015; DeCelles *et al.* 2017). The need to consider detectability is therefore paramount to obtaining true, unbiased results.

It is recognised that organisms living in marine environments exist in a multidimensional space, where even though they may live on the seabed, the constant flux of the water column plays an important role in shaping their distributions (Brown *et al.* 2011). While the same is true in the terrestrial environment (Jung *et al.* 2012), the data needed to describe the nature of the ocean are often harder to collect or generate. Broad scale covariates used in marine investigations, such as temperature, current speed and direction, often come from depth averaged or surface values of the covariate of concern. Additionally, if these covariates are not collected at the time of survey as is often the case, they need to be found as records elsewhere, or modelled using an appropriate method. Temperature, for example, is known to affect the metabolism and behaviour of marine fish (Biro, Beckmann & Stamps 2010; Remen *et al.* 2015), which can affect the detection probability.

Including temperature in the hydrodynamic model used in the current study can only be achieved by specifying water densities as a function of temperature and salinity. This presents significant challenges in coastal environments where highly variable surface conditions due to fresh water inputs cause errors in modelled hydrodynamics. As a result, temperature data were available only as low resolution depth averaged values, providing information important only to seasonal variations in temperature. Any variability in cusk occupancy due to seasonal temperature variations would already have been captured in the occupancy model by the latent variables for gamma and phi. Nevertheless, the inclusion of temperature data in future studies could potentially provide more insight into variability in fish behaviour and detection as a result of thermal stresses. Hydrodynamic information in this investigation came from a 3-dimensional hindcast model of wave and current conditions at Cashes Ledge, and explained ecologically relevant phenomena that may otherwise have been overlooked. While forecasting these types of data may not be a viable option for many researchers planning future studies, it is a recommendation of this study that an effort be made to consider the fine scale variability of any feature of the environment that may impede detection of their target species.

5 Conclusion

This study demonstrates a novel, multifaceted approach to produce a dynamic occupancy model for a species of concern in a closed area. It has generated estimates of occupancy that are considerably greater than occupancy measured from observation alone, for the first time using outputs from a high-resolution 3-dimensional hydrodynamic model in such a modelling framework. While the need for species distribution models to consider the 3-dimensional nature of the marine environment has been documented previously (Duffy & Chown 2017), this study reinforces it using modern statistical methods. Using the outputs, this investigation has shown how the behaviour of cusk changes in two different environments as a function of current speed. This behaviour has implications for the detectability of the species, which in turn has implications for the occupancy estimates. This

imperative to consider detectability in marine SDM studies is true not only for cusk, but for all species surveyed using non-invasive sampling methods. It holds especially true for areas where managers must use these methods to monitor stocks. Failing to recognise the limitations of models that do not account for imperfect detection will impact future estimates of abundance, potentially for many species. It is imperative that practitioners of future marine SDM applications consider the detectability of the species under study. In order to do this, they must first understand the processes that govern the fine-scale, time-varying features of the environment that may affect detectability, not just for the species in question but also for the specific habitat type being observed (Bacheler *et al.* 2014) in order to obtain true, unbiased estimates of occupancy in the marine environment.

Acknowledgements

This research was part funded by a Fulbright-Marine Institute (2014/15) grant, and supported by the Royal Society International Exchanges Scheme (Award IE131346).

Author Contributions

J.C., C.M. and B.H. conceived of and designed the investigation, J.G. collected data, S.A.S guided the analytical process, R.Q. advised on technical aspects of manuscript. All authors contributed to preparation of final manuscript and gave approval for submission.

Data Accessibility

Data will be archived on Dryad.

References

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Auster, P.J. (1988) A Review of the Present State of Understanding of Marine Fish Communities. *Journal of Northwest Atlantic Fishery Science*, **8**, 67–75.

502 Auster, P.J. (1989) *Species Profiles : Life Histories and Environmental Requirements of Coastal Fishes*.
503 Groton, CT.

504 Auster, P.J. & Lindholm, J. (2005) The Ecology of Fishes on Deep Boulder Reefs in the Western Gulf of
505 Maine (NW Atlantic). *Diving for Science 2005, Proceedings of the American Academy of*
506 *Underwater Sciences*, 89–127.

507 Bacheler, N.M., Berrane, D.J., Mitchell, W.A., Schobernd, C.M., Schobernd, Z.H., Teer, B.Z. &
508 Ballenger, J.C. (2014) Environmental conditions and habitat characteristics influence trap and
509 video detection probabilities for reef fish species. *Marine Ecology Progress Series*, **517**, 1–14.

510 Barbosa, F.G. & Schneck, F. (2015) Characteristics of the top-cited papers in species distribution
511 predictive models. *Ecological Modelling*, **313**, 77–83.

512 Bigelow, H.B. & Schroeder, A.C. (1953) Fishes of the Gulf of Maine,
513 http://www.gma.org/fogm/Brosme_brosme.htm

514 Biro, P.A., Beckmann, C. & Stamps, J.A. (2010) Small within-day increases in temperature affects
515 boldness and alters personality in coral reef fish. *Proceedings of the Royal Society B: Biological*
516 *Sciences*, **277**, 71–77.

517 Bouchet, P.J. & Meeuwig, J.J. (2015) Drifting baited stereo-videography: a novel sampling tool for
518 surveying pelagic wildlife in offshore marine reserves. *Ecosphere*, **6**, art137.

519 Brown, C.J., Sameoto, J.A. & Smith, S.J. (2012) Multiple methods, maps, and management
520 applications: Purpose made seafloor maps in support of ocean management. *Journal of Sea*
521 *Research*, **72**, 1–13.

522 Brown, C.J., Smith, S.J., Lawton, P. & Anderson, J.T. (2011) Benthic habitat mapping: A review of
523 progress towards improved understanding of the spatial ecology of the seafloor using acoustic
524 techniques. *Estuarine, Coastal and Shelf Science*, **92**, 502–520.

525 Chen, C., Huang, H., Beardsley, R.C., Xu, Q., Limeburner, R., Cowles, G.W., Sun, Y., Qi, J. & Lin, H.
526 (2011) Tidal dynamics in the Gulf of Maine and New England Shelf: An application of FVCOM.
527 *Journal of Geophysical Research: Oceans*, **116**, 1–14.

528 Cheng, Y. & Andersen, O.B. (2010) Improvement in global ocean tide model in shallow water regions.
529 Poster, SV. *Ostst*, pp. 1–68. Lisbon.

530 Coggins, L.G., Bacheler, N.M. & Gwinn, D.C. (2014) Occupancy Models for Monitoring Marine Fish: A
531 Bayesian Hierarchical Approach to Model Imperfect Detection with a Novel Gear Combination
532 ed I. Corsi. *PLoS ONE*, **9**, e108302.

533 COSEWIC. (2003) *Assessment and Status Report on the Cusk in Canada*. Ottawa.

534 Davies, T.D. & Jonsen, I.D. (2011) Identifying nonproportionality of fishery-independent survey data
535 to estimate population trends and assess recovery potential for cusk (Brosme brosme).
536 *Canadian Journal of Fisheries and Aquatic Sciences*, **68**, 413–425.

537 DeCelles, G.R., Keiley, E.F., Lowery, T.M., Calabrese, N.M. & Stokesbury, K.D.E. (2017) Development
538 of a Video Trawl Survey System for New England Groundfish. *Transactions of the American*
539 *Fisheries Society*, **146**, 462–477.

540 Dorazio, R., Gotelli, N., Ellison, A. & Editors. (2011) Modern methods of estimating biodiversity from
541 presence-absence surveys. *Biodiversity Loss in a Changing Planet*, 277–302.

542 Duffy, G.A. & Chown, S.L. (2017) Explicitly integrating a third dimension in marine species
543 distribution modelling. *Marine Ecology Progress Series*, **564**, 1–8.

544 Dultz, E. (2013) Brosme brosme (On-line), http://animaldiversity.org/accounts/Brosme_brosme/

545 Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of
546 MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

547 Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

548 Franklin, J. (2010) *Mapping Species Distributions. Spatial Inference and Prediction*. Cambridge
549 University Press, Cambridge.

550 Garrett, E.S. & Zeger, S.L. (2000) Latent class model diagnosis. *Biometrics*, **56**, 1055–1067.

551 Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y.S. (2008) A weakly informative default prior distribution
552 for logistic and other regression models. *Annals of Applied Statistics*, **2**, 1360–1383.

553 Gelman, A. & Rubin, D.B. (1992) Inference from Iterative Simulation Using Multiple Sequences.
554 *Statistics*, **7**, 457–511.

555 Gimenez, O., Morgan, B.J.T. & Brooks, S.P. (2009) Weak Identifiability in Models for Mark-Recapture-
556 Recovery Data. *Modeling Demographic Processes in Marked Populations*, 3rd ed (eds D.L.
557 Thompson, E.G. Cooch, & M.J. Conroy), pp. 1055–1067. Springer, Boston, MA.

558 Grabowski, J.H., McGonigle, C. & Brown, C.J. (2010) Evaluation of closed areas: Cashes Ledge as
559 juvenile cod habitat. A final report submitted to the Gulf of Maine Council. , 1–38.

560 Guillera-Arroita, G., Lahoz-Monfort, J.J., MacKenzie, D.I., Wintle, B.A. & McCarthy, M.A. (2014)
561 Ignoring Imperfect Detection in Biological Surveys Is Dangerous: A Response to 'Fitting and
562 Interpreting Occupancy Models''' ed E.P. White. *PLoS ONE*, **9**, e99571.

563 Gunn, K. & Stock-Williams, C. (2013) On validating numerical hydrodynamic models of complex tidal
564 flow. *International Journal of Marine Energy*, **3**–**4**, e82–e97.

565 Hare, J.A., Manderson, J.P., Nye, J.A., Alexander, M.A., Auster, P.J., Borggaard, D.L., Capotondi, A.M.,
566 Damon-Randall, K.B., Heupel, E., Mateo, I., O'Brien, L., Richardson, D.E., Stock, C.A. & Biegel,
567 S.T. (2012) Cusk (Brosme brosme) and climate change: assessing the threat to a candidate
568 marine fish species under the US Endangered Species Act. *ICES Journal of Marine Science*, **69**,
569 1753–1768.

570 Holbrook, S.J., Carr, M.H., Schmitt, R.J. & Coyer, J. a. (1990) EFFECT OF GIANT KELP ON LOCAL
571 ABUNDANCE OF REEF FISHES : THE IMPORTANCE OF ONTOGENETIC RESOURCE
572 REQUIREMENTS. *Direct*, **47**, 104–114.

573 Jung, K., Kaiser, S., Böhm, S., Nieschulze, J. & Kalko, E.K. V. (2012) Moving in three dimensions:
574 Effects of structural complexity on occurrence and activity of insectivorous bats in managed
575 forest stands. *Journal of Applied Ecology*, **49**, 523–531.

576 Kéry, M. (2010) *Introduction to Winbugs for Ecologists: Bayesian Approach to Regression, Anova,*
577 *Mixed Models and Related Analyses*.

578 Knutsen, H., Jorde, P.E., SannÆs, H., Rus Hoelzel, A., Bergstad, O.A., Stefanni, S., Johansen, T. &
579 Stenseth, N.C. (2009) Bathymetric barriers promoting genetic structure in the deepwater
580 demersal fish tusk (Brosme brosme). *Molecular Ecology*, **18**, 3151–3162.

581 Kruschke, J.K. (2014) *Doing Bayesian Data Analysis. A Tutorial with R, JAGS, and Stan*, 2nd ed.
582 Elsevier, London.

583 Lecours, V., Devillers, R., Simms, A.E., Lucieer, V.L. & Brown, C.J. (2017) Towards a framework for
584 terrain attribute selection in environmental studies. *Environmental Modelling & Software*, **89**,
585 19–30.

586 MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003) Estimating site
587 occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*,
588 **84**, 2200–2207.

589 MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C. a. (2002)
590 ESTIMATING SITE OCCUPANCY RATES WHEN DETECTION PROBABILITIES ARE LESS THAN ONE.
591 *Ecology*, **83**, 2248–2255.

592 McMillan, J. & Lickley, M. (2008) The Potential of Tidal Power from the Bay of Fundy. *SIAM*

593 *Undergraduate Research Online*, **1**, 20–37.

594 Monk, J. (2014) How long should we ignore imperfect detection of species in the marine
595 environment when modelling their distribution? *Fish and Fisheries*, **15**, 352–358.

596 Moody, J.A., Butman, B., Beardsley, R.C., Brown, W.S., Daifuku, P., Irish, J.D., Mayer, D.A., Mofjeld,
597 H.O., Petrie, B.D., Ramp, S., Smith, P.C. & Wright, W.R. (1984) *Atlas of Tidal Elevation and*
598 *Current Observation on the Northeast American Continental Shelf and Slope*. Alexandria VA.

599 Newton-Cross, G., White, P.C.L. & Harris, S. (2007) Modelling the distribution of badgers Meles
600 meles: Comparing predictions from field-based and remotely derived habitat data. *Mammal*
601 *Review*, **37**, 54–70.

602 Niedballa, J., Sollmann, R., Mohamed, A. bin, Bender, J. & Wilting, A. (2015) Defining habitat
603 covariates in camera-trap based occupancy studies. *Scientific Reports*, **5**, 17041.

604 NOAA. (2009) *Species of Concern: Cusk Brosme Brosme*. Silver Spring MD.

605 Nye, J.A., Link, J.S., Hare, J.A. & Overholtz, W.J. (2009) Changing spatial distribution of fish stocks in
606 relation to climate and population size on the Northeast United States continental shelf.
607 *Marine Ecology Progress Series*, **393**, 111–129.

608 Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic
609 distributions. *Ecological Modelling*, **190**, 231–259.

610 Phillips, S.J. & Dudík, M. (2008) Modeling of species distribution with Maxent: new extensions and a
611 comprehensive evalutation. *Ecograpy*, **31**, 161–175.

612 Plummer, M. (2009) JAGS: A program for analysis of Bayesian graphical models using Gibbs samping.
613 *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (eds F.
614 Leisch & A. Zeileis), p. Vienna.

615 R Core Team. (2015) R: a language and environment for statistical computing.

616 Rattray, A., Ierodiaconou, D. & Womersley, T. (2015) Wave exposure as a predictor of benthic
617 habitat distribution on high energy temperate reefs. *Frontiers in Marine Science*, **2**, 1–14.

618 Remen, M., Nederlof, M.A.J., Folkedal, O., Thorsheim, G., Sitjà-Bobadilla, A., Pérez-Sánchez, J.,
619 Oppedal, F. & Olsen, R.E. (2015) Effect of temperature on the metabolism, behaviour and
620 oxygen requirements of Sparus aurata. *Aquaculture Environment Interactions*, **7**, 115–123.

621 Rota, C.T., Fletcher, R.J., Evans, J.M. & Hutto, R.L. (2011) Does accounting for imperfect detection
622 improve species distribution models? *Ecography*, **34**, 659–670.

623 Royle, J.A. & Kéry, M. (2007) A BAYESIAN STATE-SPACE FORMULATION OF DYNAMIC OCCUPANCY
624 MODELS. *Ecology*, **88**, 1813–1823.

625 SAIC. (2005) *Gulf of Maine Mapping Initiative , Priority 1 Area Survey Report (Doc. 05-TR-017)*. Rhode
626 Island.

627 Schuenemeyer, J.H. & Drew, L.J. (2010) *Statistics for Earth and Environmental Scientists*. John Wiley
628 & Sons, Inc., Hoboken, NJ, USA.

629 Sherwood, G.D. & Grabowski, J.H. (2016) A comparison of cod life-history parameters inside and
630 outside of four year-round groundfish closed areas in New England, USA. *ICES Journal of*
631 *Marine Science: Journal du Conseil*, **73**, 316–328.

632 Stoner, A.W., Ryer, C.H., Parker, S.J., Auster, P.J. & Wakefield, W.W. (2008) Evaluating the role of fish
633 behavior in surveys conducted with underwater vehicles. *Canadian Journal of Fisheries and*
634 *Aquatic Sciences*, **65**, 1230–1243.

635 Uchupi, E. (1966) Topography and structure of Cashes Ledge, Gulf of Maine. *Atlantic Geology*, **2**,
636 117–120.

637 Uchupi, E. & Bolmer, S.T. (2008) Geologic evolution of the Gulf of Maine region. *Earth-Science*
638 *Reviews*, **91**, 27–76.

639 Uhl, F., Bartsch, I. & Oppelt, N. (2016) Submerged kelp detection with hyperspectral data. *Remote*
640 *Sensing*, **8**.

641 Vadas, R.L. & Steneck, R.S. (1988) Zonation of Deep Water Benthic Algae in the Gulf of Maine.
642 *Journal of Phycology*, **24**, 338–346.

643 Witman, J.D. & Sebens, K.P. (1992) Regional variation in fish predation intensity: a historical
644 perspective in the Gulf of Maine. *Oecologia*, **90**, 305–315.

645 Xu, Z. (2002) *Ellipse Parameters Conversion and Vertical Velocity Profilesfor Tidal Currents*. Quebec.

646 Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S.
647 (2013) Presence-only modelling using MAXENT: When can we trust the inferences? *Methods in*
648 *Ecology and Evolution*, **4**, 236–243.

649

Table 1. Mean, maximum and minimum for each of the covariates falling within the depth strata sampled. Values for algae cover are reported as counts, as these are categorical variables. "-" indicates no relevant data available.

Covariate	Stratum	Min	Mean	Max	Number
Depth	Shallow	-23	-17	-11	-
	Medium	-43	-33	-25	-
	Deep	-57	-49	-39	-
Current speed	Shallow	0.05	0.11	0.17	-
	Medium	0.03	0.09	0.19	-
	Deep	0.01	0.09	0.2	-
CurMax	Shallow	0.21	0.22	0.24	-
	Medium	0.16	0.19	0.23	-
	Deep	0.16	0.19	0.22	-
RDMV	Shallow	0.14	0.56	1.39	-
	Medium	0.06	0.31	0.36	-
	Deep	0.08	0.25	0.67	-
Algae cover: none	Shallow	-	-	-	0
	Medium	-	-	-	10
	Deep	-	-	-	45
Algae cover: canopy	Shallow	-	-	-	21
	Medium	-	-	-	36
	Deep	-	-	-	0

Table 2: Parameters estimated for coefficients for occupancy model A (without covariates for ψ) and model B (with covariates for ψ). GOF is the Bayesian p-value, where values closer to 0.5 indicate better fit. Gelman-Ruben MV is the multivariate Gelman-Ruben convergence statistic, where values close to 1 mean the model has successfully converged. Current is the current speed at the time of observation of cusk, effort is sampling effort, none is regions with no canopy forming species of algae, current | none is the interaction between current and none, complexity is RDMV, current max is the maximum current speed during the sampling season.

Model A	GOF	Gelman-Ruben MV					
	0.58	1.01			Credible Interval		
Sub-model	Parameter		Mean	s.d.	2.5%	97.5%	Inclusion probability
Detection	Intercept		-1.57	0.79	-3.27	-0.04	
	Current		-1.26	0.59	-2.58	-0.19	0.66
	Effort		-0.37	0.75	-1.94	1.05	0.32
	none		3.01	0.88	1.52	5.19	0.99
	Current | none		1.77	0.71	0.44	3.32	0.79

Model B	GOF	Gelman-Ruben MV					
	0.62	1.01			Credible interval		
Sub-model	Parameter		Mean	s.d	2.5%	97.5%	Inclusion probability
Occupancy	Intercept		1.41	1.4	-0.93	4.55	
	Depth		-1.66	1.55	-5.08	1.24	0.50
	Complexity		0.99	1.66	-2.01	4.63	0.49
	Current max		0.04	1.31	-2.53	2.64	0.40
	Current max^2		0.17	1.31	-2.42	2.78	0.39
Detection	Intercept		-1.43	0.83	-3.17	0.2	
	Current		-1.31	0.63	-2.67	-0.21	0.7
	Effort		-0.35	0.79	-1.97	1.09	0.34
	none		2.85	0.93	1.3	4.98	0.99
	Current | none		1.85	0.75	0.47	3.44	0.82

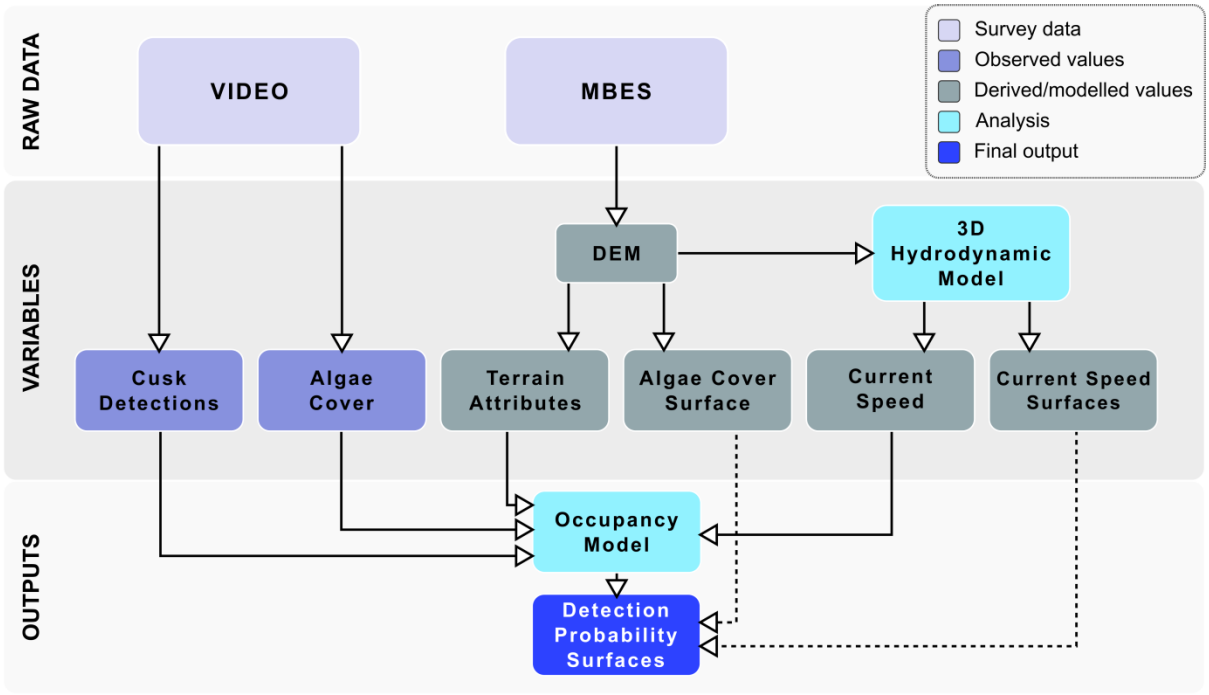

Figure 1. Flow diagram for analyses described in the methods. Solid lines represent the flow of information to create the dynamic occupancy model, and dashed lines represent the flow of information for producing the output probability surfaces in Figure 6.

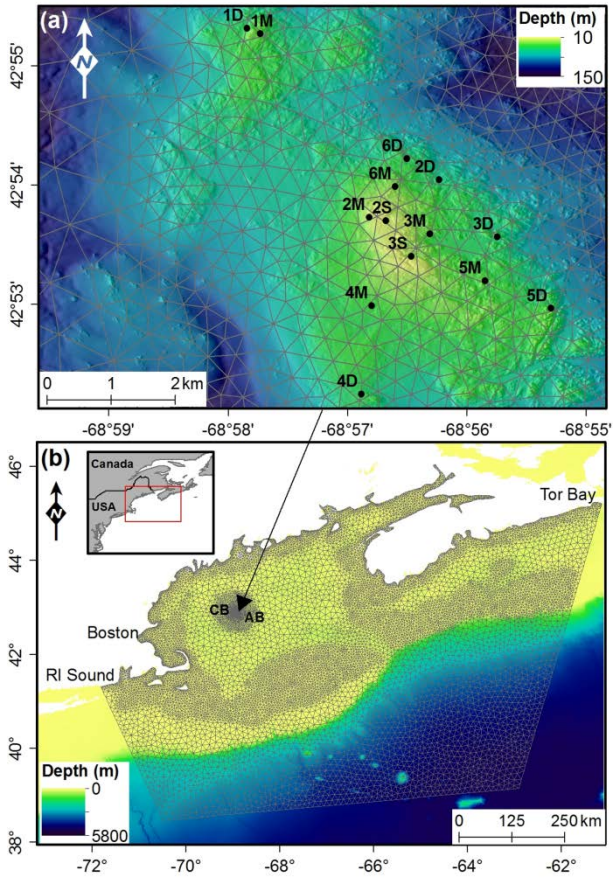

Figure 2. (a) Location of 14 sites sampled for cusk at Cashes Ledge. Grey triangles are from the hydrodynamic model computational mesh, the full domain of which is in (b) along the location of Cashes Ledge in the Gulf of Maine. CB is Cashes Basin, AB is Ammen Basin.

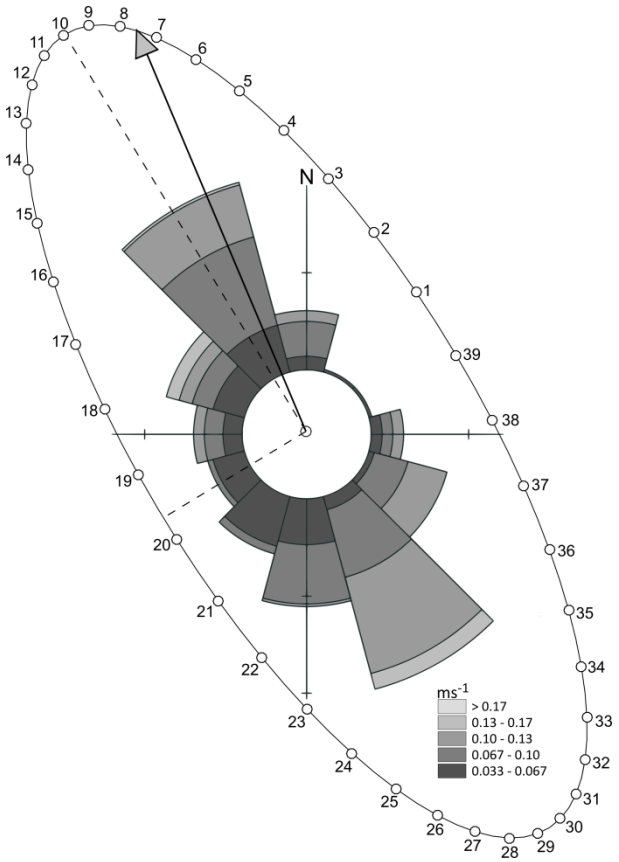

Figure 3: Tidal ellipse for a point in open water 100 m away from Cashes Ledge. The distance from the centroid to the arc of the ellipse is proportional to the current speed, and the direction of a vector radiating from the centroid to any point on the arc represents the direction any that point. The number on the arc correspond to the numbers on the tidal curve and in Figure 6.

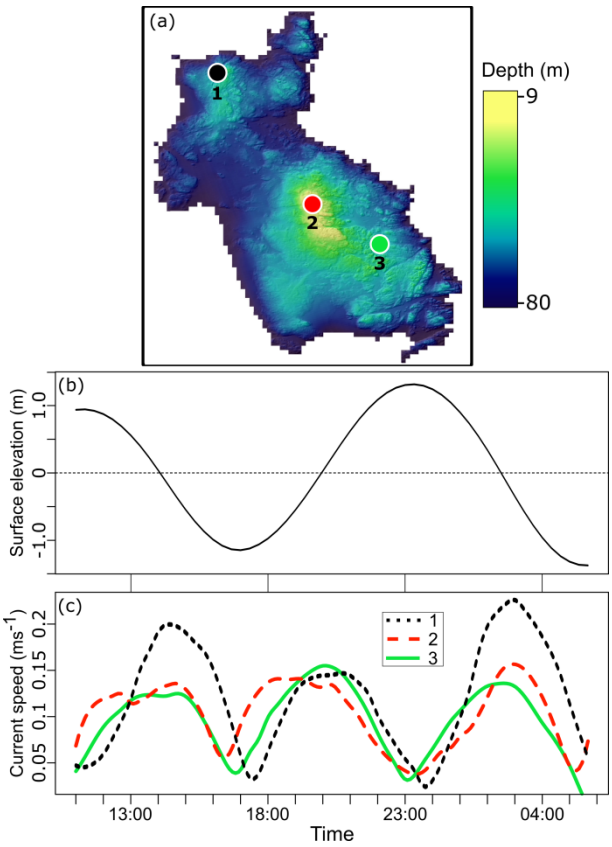

Figure 4: Example of difference in arrival time of tidal signal for three sites, one deep (1), one intermediate (3) and one shallow (2) at Cashes Ledge over one and a half tidal cycles throughout the summer 2006 sampling season. Time on the x-axis of the tidal curve (b) and current speed plot (c) are the same. The locations of the three sites are marked on the map (a).

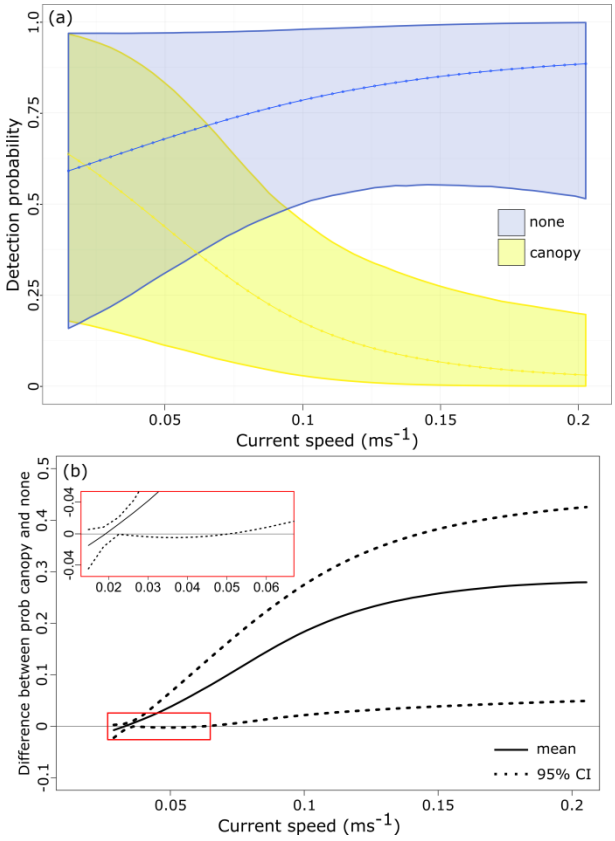

Figure 5: Model outputs showing (a) final model for detectability of cusk at Cashes Ledge based on current speed and bottom type, coloured bands are 95 % credible interval, (b) difference in detection probability between canopy forming and non-canopy forming regions. Where the 95% credible interval includes zero, there is no significant difference in detection probability.

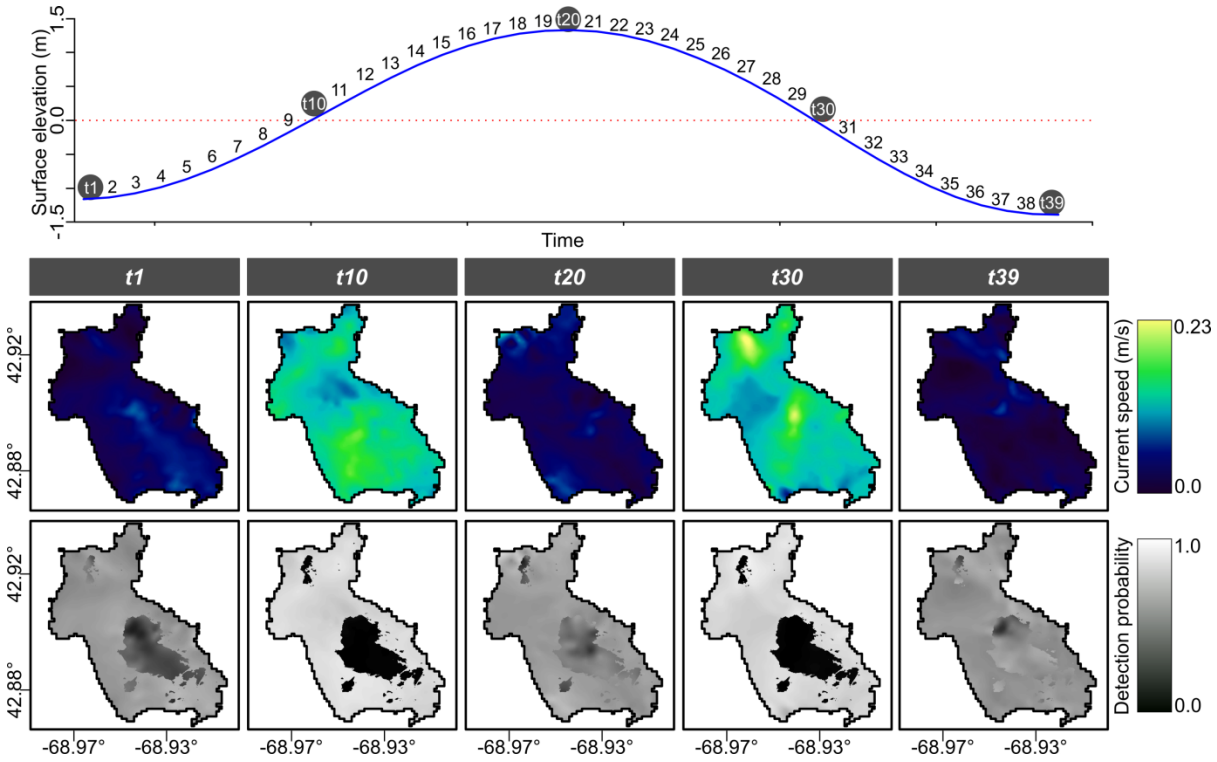

Figure 6: Tidal curve (top), current speeds (middle) and predicted probability surfaces (bottom) along a full tidal cycle during the summer 2006 sampling season. Surfaces are predicted from the detection model outputs. Numbers of each panel in the grey boxes (t1 – t39) correspond to the numbers on the tidal curve. The darker regions in t10 and t30 indicate the areas with canopy forming species of algae based on extinction depth for those species at Cashes Ledge, while the lighter regions are the areas with no canopy forming species of algae.

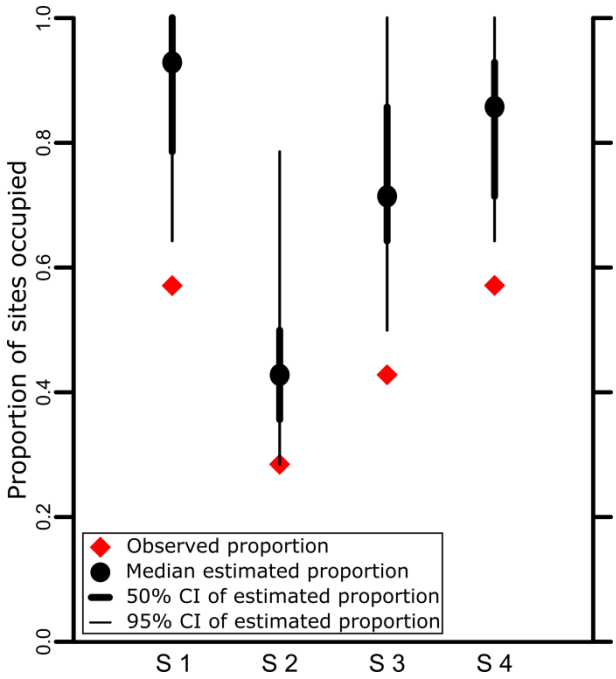

Figure 7. Observed and median estimated proportions of sites occupied by cusk at Cashes Ledge during the four sampling periods; S1 (Summer 2006), S2 (Spring 2007), S3 (Summer 2007 and S4 (Autumn 2007). Also shown are the 50% and 95% credible intervals for the posterior estimates of the proportion of occupied sites.