**Why less complexity produces better forecasts:
An independent data evaluation of kelp habitat models**

Edward J. Gregr¹, Daniel M. Palacios², Allison Thompson^{1,3}, Kai M.A. Chan¹

First published: 02 July 2018 <https://doi.org/10.1111/ecog.03470>

1. Institute for Resources, Environment, and Sustainability

University of British Columbia

429-2202 Main Mall, Vancouver, BC, V6T 1Z4, Canada

2. Marine Mammal Institute and Department of Fisheries and Wildlife

Oregon State University, Hatfield Marine Science Center

2030 SE Marine Science Drive, Newport, OR, 97365, USA

3. Wildlife Management Advisory Council (NWT)

107 Mackenzie Road, Inuvik, NT, X0E 0T0

Keywords: prediction, transferability, cross-validation, independence, uncertainty, habitat suitability, species distribution, sampling bias, ecological context, niche, initial conditions, climate change, giant kelp; bull kelp; canopy kelp

Why less complexity produce better forecasts:

An independent data evaluation of kelp habitat models

Abstract

Understanding how species are distributed in the environment is increasingly important for natural resource management, particularly for keystone and habitat forming species, and those of conservation concern. Habitat suitability models are fundamental to developing this understanding; however their use in management continues to be limited due to often-vague model objectives and inadequate evaluation methods. Along the Northeast Pacific coast, canopy kelps (*Macrocystis pyrifera* and *Nereocystis luetkeana*) provide biogenic habitat and considerable primary production to nearshore ecosystems. We investigated the distribution of these species by examining a series of increasingly complex habitat suitability models ranging from process-based models based on species' ecology to complex Generalised Additive Models applied to purpose-collected survey data. Seeking limits on model complexity, we explored the relationship between model complexity and forecast skill, measured using both cross-validation and independent data evaluation. Our analysis confirmed the importance of predictors used in models of coastal kelp distributions developed elsewhere (i.e., depth, bottom type, bottom slope, and exposure); it also identified additional important factors including salinity, and interactions between exposure and salinity, and slope and tidal energy. Comparative results showed that cross-validation can lead to over-fitting, while independent data evaluation clearly identified the appropriate model complexity for generating habitat forecasts. Our results also illustrate that, depending on the evaluation data, predictions from simpler models can outperform those from more complex models. Collectively, the insights from evaluating multiple

23 models with multiple data sets contribute to the holistic assessment of model forecast skill. The
24 continued development of methods and metrics for evaluating model forecasts with
25 independent data, and the explicit consideration of model objectives and assumptions, promise
26 to increase the utility of model forecasts to decision makers.

27

28 **Introduction**

29 Predictions of habitat, and ultimately species distributions, are recognised as integral to
30 ecosystem-based resource management {e.g., Tommasi et al. 2017) and marine spatial planning
31 {e.g., Robinson et al. 2011, Schmolke et al. 2010). Understanding how species may respond to
32 climate change is a growing part of this work {e.g., Araujo and Rahbek 2006, Sumaila et al.
33 2011, Tommasi et al. 2017). However, despite this clear management need, the use of habitat
34 forecasts for management continues to be hampered by a number of significant challenges
35 including a lack of clarity about model purpose {Araujo and Guisan 2006, Gregr and Chan 2014,
36 Houlahan et al. 2017), inadequate evaluation methods {Bahn and McGill 2013, Vaughan and
37 Ormerod 2005), and poor treatment of uncertainty {Beale and Lennon 2012, Gregr and Chan
38 2014, Planque et al. 2011). Using an emergent management need in Pacific Canada as a case
39 study, we confront these challenges by examining the utility of independent data for 1)
40 assessing model forecast skill, 2) identifying the appropriate complexity to maximise forecast
41 skill, and 3) exploring sampling bias and assumptions about ecological process, key components
42 of model uncertainty.

43 The ecological motivation for our study is the locally contentious issue of sea otter {*Enhydra*
44 *lutris*) recovery in the coastal waters of the Northeast Pacific. In this region, sea otters are
45 rapidly transforming the coastal ecosystem {Nichol et al. 2009) from one dominated by
46 invertebrates, to one dominated by kelp {Estes and Palmisano 1974). This is widely seen as an
47 ecological success story because canopy kelps {especially *Macrocystis pyrifera* and *Nereocystis*
48 *luetkeana*) provide biogenic habitat to species across several trophic levels, and serving as a
49 significant source of primary production, nourishing nearshore food webs {Duggins 1988,

Steneck et al. 2002). However, this ecosystem shift has led to conflict, as those economically dependent on invertebrate resources and frustrated by the competition from sea otters may sometimes resort to shooting sea otters (Nichol et al. 2009). Understanding the physical drivers of canopy kelp and the areas best suited to supporting kelp forests and their associated ecosystem services is an important piece of information for the spatial management of the region's natural resources. Thus, from a methodological perspective, the need is for a habitat suitability model that supports both an understanding of the mechanistic role of oceanographic drivers, particularly those susceptible to a changing climate, and also provides forecasts of habitat that are transferable in time.

Fortunately, habitat models are versatile, and can be built for description, understanding, or prediction (Araujo and Guisan 2006). Models for understanding how species distributions are influenced by environmental drivers should ideally test hypothesised relationships (Araujo and Guisan 2006), while models for predicting distributions require a greater focus on transferability (i.e., the skill of the model to forecast to different times or locations). In this study, we begin by proposing and testing hypotheses about the proximate drivers of canopy kelp distributions. We then examine the forecast skill of a range of increasingly complex kelp habitat suitability models, and use the results to highlight the consequences of current approaches to model evaluation, and to demonstrate how evaluation with independent data sets can provide guidance on the limits to model complexity.

The complexity of ecosystems, comprised as they are of many diverse biotic and abiotic components and often interacting processes, means there are myriad hypotheses that can be

considered in any particular habitat model. Finding the right balance between the complexity necessary to precisely describe patterns in a sample of data while preserving enough generality to make reasonably accurate predictions is one of the fundamental challenges for effective habitat suitability modelling, and emphasises the importance deliberate model design (Grg and Chan 2014, Guisan and Zimmermann 2000, Jrgensen 1986). This includes the articulation of clear model objectives, explicit hypotheses about ecological process, and a rationale for the modelling approach selected. Deliberate, explicit design is critically important because it can counter-balance the methodological focus on correlation between observations and abiotic drivers. This is important because understanding the ecological consequences of what is, and is not represented is critical to model interpretation and uncertainty assessment.

Machine-learning methods (especially MaxEnt - Phillips et al. 2006) have emerged as the dominant approach to habitat suitability modelling in the last decade because they are excellent at teasing out relationships between potential predictors and observations. They are typically among the top performers in model comparison studies (e.g., Elith and Graham 2009, Rodriguez-Rey et al. 2013). However, it has been clear for some time that they will quite happily fit to noise if left unconstrained, resulting in over-fit models that are difficult to interpret (e.g., Grg 2011) and exhibit poor transferability (Warren and Seifert 2011, Moreno-Amat et al. 2015, Merow et al. 2014).

Combined with increases in computing power and the now ubiquitous availability of global predictor data (e.g., WorldClim, World Ocean Atlas), machine learning methods now essentially apply a data mining approach, bypassing the process of considered variable selection,

particularly for studies over large spatial extents. This transition towards a 'big data' approach is concerning for several reasons. Not only is there a lack of theoretical justification for the added complexity machine learning methods (Bell and Schlaepfer 2016), it has also shifted the complexity conversation from an ecological one about model sufficiency (i.e., what complexity is needed to achieve the model objectives) to one about how to constrain over-fitting (e.g., Radosavljevic and Anderson 2014)

In this case study, we compared models across a spectrum of complexities, ranging from simple mechanistic envelope models to correlative models whose complexity is limited only by commonly used stopping conditions based on sample size. Since most methods exhibit a significant difference in performance between internal validation and tests of transferability (Heikkinen et al. 2012), our goal was to investigate how this difference relates to model complexity, and whether limits on model complexity can be identified using independent evaluation data.

How to best evaluate model predictions depends on goal of the model (Araujo and Guisan 2006, Rykiel Jr. 1996). If the objective is description or understanding, there are a variety of metrics available for choosing the model that best fits the response data (e.g., Guisan and Zimmermann 2000, Hirzel et al. 2006, Lawson et al. 2014). These statistics are typically calculated using a subset of the available data using a process called cross-validation, where the sample of species observations is divided into model training and model testing data sets. Assessments of how well the model describes the various subsets of the data used to build it can yield important internal verification of a model and provide confidence in the underlying

113 hypotheses about species-habitat relationships {Araujo and Guisan 2006}. However, such in-
114 sample verification generally reveals little about the quality of model predictions at other times
115 and places {Bahn and McGill 2013, Houlahan et al. 2017}, particularly when random cross-
116 validation is used. Rather, evaluating forecast skill {i.e., transferability} depends on how well a
117 model generalises in space or time, and it has long been recognised that this is best done using
118 independent distributional data {Araujo et al. 2005b, Goodall, 1972 #2280, Guisan and
119 Zimmermann 2000, Rykiel Jr. 1996, Vaughan and Ormerod 2005}.

120 However, examples of independent data evaluation {IDE} are rare because observations of
121 forecasted states {in space or time} typically do not exist. Thus, cross-validation is often
122 interpreted as a measure of predictive performance, even though it is known to over-estimate
123 model skill because the training and testing subsets are not independent {e.g., Araujo et al.
124 2005a, Jimenez-Valverde et al. 2011, Roberts et al. 2017}. This has led to the development of
125 approaches like block cross-validation {Roberts et al. 2017}, intended to divide collections of
126 observations into more independent samples.

127 In this study, we use independently collected data to evaluate model forecast skill across a
128 range of increasingly complex habitat suitability models. We compare the results to those
129 obtained via both random, and block cross-validation to illustrate how IDE can help limit model
130 complexity. We also consider how IDE may inform sampling bias in spatially and temporally
131 dynamic systems, and the validity of model assumptions about ecological process.

132 The many questions around model objectives, assumptions, and hypotheses contribute to the
133 uncertainty in model predictions, and the associated complexities are central to the inadequate

treatment of uncertainty in both habitat modelling {Beale and Lennon 2012, Planque et al. 2011) and in ecosystem modelling generally {GREGG and Chan 2014, Link et al. 2012). In the few cases where uncertainty is examined, the focus is typically on parameterisation. Examination of uncertainties in source data, model structure, or model hypotheses and assumptions are much rarer {GREGG and Chan 2014, Planque et al. 2011). By using IDE to examine the fundamental question of model complexity, our analysis advances understanding of model forecast skill, and is relevant to any all ecological models seeking to generate predictions in other times or places.

Methods

We conducted our study in Kyuquot Sound on the west coast of Vancouver Island, British Columbia, where sea otters have been at or close to carrying capacity for over a decade {Nichol et al. 2009). This allowed us to assume that the system had reached a new equilibrium, where canopy kelps had re-occupied all of their potential habitat {within the limits of environmental variability). We restricted our study area {Fig. 1) to depths between 0 and 20 m, the observed limits of canopy kelp distribution. We defined habitat suitability models using canopy kelp percent cover data collected over 3 years as the dependent data, a set of abiotic predictor variables identified from a review of earlier models of kelp distribution, and a general understanding of the factors that influence kelp distribution and abundance. We developed a series of eight increasingly complex habitat models for canopy kelp, compared them using both random and block cross-validation, evaluated their forecasting skill using independent observations of canopy kelp distributions, and examined how the prevalence of the independent sample influenced measures of forecast skill. We elaborate on each of these components below.

156 *Kelp observations*

157 We collected survey data on the percent cover of the two canopy kelps using stratified random
158 boat-based surveys conducted in September {i.e., fall - the typical time of maximum canopy
159 cover) from 2010 to 2012, and in June {i.e., spring) of 2010. Percent cover was estimated
160 visually in four quadrats within a 100-m radius of the survey vessel, and averaged for each
161 sampling location. We also recorded the relative proportion of the two species. We sampled
162 percent cover a total of 375 times at 173 locations over the 3 years.

163 We used two types of independent data on kelp presence to evaluate the predictive models.
164 The first includes observations of canopy kelp presence compiled by Living Oceans Society {LOS)
165 from marine charts and assorted kelp surveys conducted over many years by a variety of
166 organizations {Ardron 2004). These observations were available as polygons indicating the
167 presence of canopy kelp, with no temporal information. The second set of independent test
168 data were based on opportunistically collected images taken from the SPOT-5 {Satellite Pour
169 l'Observation de la Terre) satellite-borne sensor archive {SPOT 2017). Three high resolution {10
170 x 10 m²) colour images were taken in mid summer {July/August) for the years 2005, 2006, and
171 2009. Visible bands 1 through 3 were classified to vegetation presence/absence using the
172 Support Vector Machine supervised clustering algorithm {Ge et al. 2008). Six control points
173 {areas where kelp cover was known to occur) were defined based on visual inspection and local
174 knowledge to calibrate the classification. Collectively, these data represent four independent
175 sets of canopy kelp observations {1 from LOS, and 3 from SPOT-5).

Several accommodations were necessary to make the survey data {used for kelp model development} comparable with the independent data {used for evaluation}. First, since the independent data did not include species, we developed a generic model of canopy kelp after examining the response of each species individually to the predictor variables. Second, we converted the boat-based percent cover estimates to presence-absence to match the independent data. Re-sighting data {repeat visits to selected sites on later days} indicated consistent values above 5% percent cover. We therefore selected 5% as the threshold for converting percent cover to presence or absence. Finally, we used only the fall data (n=240) for model development after noting a strong seasonal effect associated with the spring data.

Predictor variables

Canopy kelp in the Northeast Pacific are understood to be influenced by bottom type, sunlight {i.e., insolation}, temperature, water motion, salinity, sedimentation, and nutrient levels {Dayton 1985, Druehl 1978, Druehl 2001, Springer et al. 2007}. We developed a set of predictor variables {Table 1} based on this general understanding, and earlier models of kelp distribution {e.g., Bekkby et al. 2009, Gorman et al. 2013, Meleder et al. 2010}. The rationale for their inclusion and their derivation is described in Supplementary Appendix 1. In addition, given that the individual factors affecting kelp habitat are unlikely to act independently {Dayton 1985}, we included several interactions to test if they could improve our prediction of canopy kelp distributions. All predictors were scaled to align with a 20 x 20 m² bathymetric model of the coast, tested for collinearity, and centred before developing the models.

Model development

197 Habitat models can range in complexity from the straightforward habitat suitability index (HSI)
198 approach pioneered in the 1980s by the US Fish and Wildlife Service (Van Horne and Wiens
199 1991) to sophisticated regression-based machine learning methods that are very adept at
200 creating complex relationships to capture nuanced patterns in the data (reviewed by Merow et
201 al. 2013). Comparisons of how the various methods perform appear regularly (e.g., Elith et al.
202 2006, Guisan and Zimmermann 2000, Pearson et al. 2006, Segurado and Araujo 2004).

203 The "best" method depends on the model context and intent (Segurado and Araujo 2004),
204 which will also influence model complexity (Merow et al. 2014). Our intent here was to
205 evaluate habitat forecasts from a set of models with increasing complexity, with the complexity
206 driven in part by explicit hypotheses about ecological processes, but also by model structure,
207 and by methods intended to maximise fit to data. We began by reviewing both the ecological
208 understanding of kelp distributions and the available abiotic data (i.e., predictors), making
209 explicit hypotheses about how the predictors may influence kelp distributions. We used these
210 hypotheses to define our simplest models using the HSI approach, and reviewed the model
211 structure with local experts. We then defined increasingly complex binomial regression models
212 (GLMs and GAMs) with a logistic link function (the equivalent of a logistic model when the
213 response variable is dichotomous) to examine the relationships between model structure (i.e.,
214 complexity), variance explained (i.e., model fit), and forecast skill. This approach aligns with that
215 recommended by Merow et al. (2014), for exploring the functional limits of model complexity
216 based on the study objective, limitations of the data, and model comparison.

We began with an HSI model using *Depth*, bottom type, *Salt*, and exposure, representing habitat suitability for canopy kelp as: decreasing non-linearly with *Depth*; increasing linearly with the probability of rocky reefs and salinity {within a defined range}; and optimal across a range of exposures {Fig. 2}. We used the three variants of bottom type {*RF*, *RMSM*, and *BoP*} to assess the sensitivity of the HSI models to this predictor. We assumed all factors contributed equally, and that all were essential {i.e., low on any factor led to low habitat suitability}.

We then examined the strength of association between each potential predictor variable and our survey data using univariate GLMs. We examined correlations between linear, quadratic, and cubic forms of all predictors with the presence-absence of each species {i.e., giant and bull kelp} as well as the combined canopy to inform how to best to generalise the observations {see Appendix 2 in Supplemental Material for details}. We used the significant predictors from this univariate analysis in a structured variable selection approach to create four multivariate GLMs with increasing complexity. We began by defining a GLM to represent a parametric form of the HSI model, and a second using only the linear forms of the predictors identified as significant for the combined canopy kelp response variable. In the third model we increased model complexity by considering higher-order polynomial forms of the predictors. These polynomials always entered the models with all its terms. We used higher order polynomials to allow the GLMs to approach the complexity of GAMs and facilitate model comparison. Finally, for the fourth GLM, we considered interaction terms.

At each step, the best model was selected using Akaike's Information Criterion {AIC}, after considering both single term addition and single term removal. AIC also served as a stopping

condition to prevent increasing model complexity without a concomitant gain in predictive ability. A final review of the GLMs ensured all terms in the models were significant at the 0.05 level or better. These steps ensured the predictors with the highest explanatory power entered the models first, and that all predictors contributed significantly to model performance. This structured step-wise approach to variable selection is fundamentally different from the automated step-wise variable selection approach understood to have a range of methodological problems {Whittingham et al. 2006}.

We added two GAMs to our analysis: the first with predictors, interactions, and degrees of freedom equivalent to the most complex GLM, and the second with the same terms but with penalised smoothing terms. The degrees of freedom for GAM1 were specified {and fixed} using the k parameter to represent the non-parametric {i.e., smoothed} equivalent of the best performing GLM. GAM2 represented the most complex model in our analysis and was created by allowing the default GAM fitting algorithm to define smooths using as many degrees of freedom as penalization would admit. Tensor product smooths were used for both GAMs.

Data analyses and graphing were done in the R language for statistical analysis {R Core Team 2015} using the add-on packages *mgcv* {Wood 2004} and *AUC* {Ballings and Van den Poel 2013} for analysis, and *dplyr* {Wickham and Francois 2015}, and *ggplot2* {Wickham 2009} for presentation of results. Predictors were tested for correlation using the variance inflation factor tests in package *fmsb* {Nakazawa 2017}; residuals were examined using simulated residual plots generated using the *dHARMA* package {Hartig 2017}.

Evaluating model performance

During the development of the statistical models, we used deviance explained (ΔAIC) to complement AIC in the model selection process. This statistic is typically used for GLMs and GAMs and is analogous to unadjusted variance explained (i.e., R^2) in linear models (Guisan and Zimmermann 2000).

We tested the in-sample performance of the completed models using cross-validation. We compared randomised 3-fold cross-validation (Fielding and Bell 1997) to block-crossvalidation (Roberts et al. 2017) using the survey year to block the data. For the 3-fold test we divided the observations randomly into 3 equal sized sub-samples (i.e., folds) parameterised each fold, and evaluated the resulting prediction using the observations in the other folds. We repeated this process 100 times to obtain a distribution of Area Under the Curve (AUC) scores. AUC is a widely used, threshold-independent measure of model performance (Fielding and Bell 1997), where a score of 1.0 implies perfect prediction while 0.5 implies a predictive power no better than random. For the block cross-validation, we fit the models using data from each of the 3 years individually, and tested how well it predicted the other two years, generating 3 AUC scores for each model. To evaluate the forecast skill of the different models, we mapped the independent test data and the predictor variables to a regular hexagonal (100 m diameter) grid of the study area (Fig. 1). Hexagons were assigned presence-absence values for each of the four independent test data sets using the 5% rule developed for the survey data: if kelp occupied less than or equal to 5% of the hexagon, it was assigned an absence, otherwise the hexagon was defined as a presence cell. To avoid spatial processing errors on boundary hexagons, only those hexagons fully within the extents of the independent test data set were used in the evaluation.

281 To increase the comparability of the independent data validation with the cross-validation
282 results, we simulated a field sampling program by drawing 500 observations from the extents of
283 each remote sensing layer (Fig. 1). We ensured each sample had the same prevalence as our
284 original field survey data (used to build the models). The statistical models were re-
285 parameterised for each pseudo-survey, and a kelp prediction was generated for the entire
286 hexagonal grid. We compared predictions from all models to each of the four independent data
287 sets using AUC. We repeated the process 100 times for each model-independent data set pair.

288 A variety of statistics are available to compare predictions to observations (e.g., Fielding and
289 Bell 1997), and a considerable body of literature has explored how species prevalence (the
290 proportion of sampled sites where a species is present) can be a complicating factor (Allouche
291 et al. 2006, Lawson et al. 2014, Mouton et al. 2010, Santika 2011) in part because of its
292 relationship to threshold selection (Manel et al. 2001). A threshold allows the continuous
293 predictions typically generated by habitat models to be classified into presences and absences
294 so that a broader set of performance metrics can be derived from the resulting contingency
295 table (also, for some reason, called a confusion matrix). However, threshold selection is not a
296 trivial matter (Lawson et al. 2014, Manel et al. 2001). So to avoid complicating our analysis with
297 the question of threshold selection and focus on cross-validation and IDE, we use the AUC (Area
298 Under the Curve) statistic – a widely used, threshold-independent metric of agreement
299 between predictions and observations.

300 **Results**

301 *Regression model development*

Our univariate analysis of the individual predictors formed the basis of the subsequent model development. These results largely mirrored earlier findings from the literature, with *Depth*, *Slope*, bottom type, *Salt*, temperature, and fetch all identified as significant (Supplementary material Appendix B, Table B1). Notable differences between the kelp species including the greater significance of *Salt* and temperature to giant kelp. Other differences were largely a matter of degree, with one species or the other showing significance at higher-order predictors, or different representations of the same predictor (e.g., *SST* vs. *SSTQ3*). The similarities between the species response variables were taken to indicate that a combined canopy model would adequately represent both for the purposes of this study.

Predictors were tested for collinearity using the variance inflation factor and values were well below the recommended threshold of 10.0 (Dormann et al. 2013) for all predictors indicating no collinearity concerns among the predictors. Residuals from all final models showed no heteroscedasticity based on simulated residual plots.

Deviance explained increased from 0.3 in the simplest GLM (GLM0), to 0.79 in the most complex model (GAM2; Table 2). All predictors in GLM0 were significant at $p < 0.001$ except *Salt* which was significant at $p < 0.1$ (Table 3). GLM1 explained 10% more deviance in the data from the addition of *Slope*, and *MaxTidal* (Table 2). Six of 7 predictors in GLM1 were significant at the 0.05 level or better. Only the bottom type (*RF*) term was not significant. In GLM2, exposure (as *FetchTot*) was the only predictor to enter as a polynomial term. Its addition explained an additional 2% of the deviance. We found that centring the predictor data notably reduced the number of higher order terms that entered the model at this step.

323 The biggest improvements in model fit were from the inclusion of interaction terms {in GLM3}.
 324 We defined interactions using the variable complexity from GLM2, and began by examining the
 325 change in both deviance explained and the significance of model terms for each interaction
 326 separately. We found no effect from the temperature-exposure interaction, but both the
 327 salinity-exposure $\{FetchTot^2:Salt\}$ and the slope-tidal current $\{Slope^2:MaxTidal\}$ interactions
 328 significantly improved model fit to data, explaining an additional 17% of the deviance {Table 3}.
 329 Adding the salinity interaction eliminated the need for the separate *Salt* predictor, which we
 330 removed and did not include in the more complex models. The remaining predictors were all
 331 significant at the 0.01 level, except for the first order term of the slope-tidal current interaction.
 332 At this stage, we also tested the explanatory power of the insolation model before and after
 333 adding the interactions, but in no case did either insolation variable reduce the AIC.

334 *Model performance*

335 The model fit statistics {AIC and Φ^2 } changed predictably as model complexity was increased
 336 {Table 2}, and the AUC score from both cross-validation tests showed a similar pattern of
 337 increasing complexity generating higher scores. Results for the 3-fold cross-validation ranged
 338 between 0.67 and 0.89 for the HSI models, with the statistical models increasing steadily from
 339 GLM0 {0.77} to GAM2 {0.97}, and a short plateau in performance at GLM3 {Fig. 3}. The AUC
 340 score based on block cross-validation {using survey years} was also similar, but plateaued at the
 341 complexity of GLM3.

342 In contrast to this internal verification, forecast skill was more variable when evaluated using
 343 the independent data. The AUC scores shows a high variability both among the independent

data and between models, but generally trended upwards from below 0.6 for the HSI models to just over 0.8 for GLM2. Forecast skill dropped dramatically to below 0.6 for the remaining, more complex models {Fig. 3}. GLM2 was best at forecasting the independent remote sensing data, while the IDE models out-performed all statistical models when evaluated against the independent LOS data.

The distribution of predicted values ranged from almost uniform in GLM0 {Fig. 4a} to the extremely bimodal GAM2 {Fig. 4c}. This bimodality appeared in all the regression models {except GLM0}, and increased with model complexity. This increasing bimodality is expressed spatially as increasingly concentrated and homogeneous patches of suitable and unsuitable habitat {Fig. 4}.

Alternate predictor representations

As part of model development, we considered different representations of temperature {*BTemp*, *SST*, *SSTQ2*, *SSTQ3*}, insolation {*SumSol*, *Slope*, *Aspect*}, and bottom type {*RF*, *RMSM*, *BoP*}. Temperature only entered the models when seasonal observations were included and thus do not factor into the remainder of this analysis. However, it is worth noting that if more than one temperature variable was allowed to enter the model, temperature quickly became a key driver of the model, likely because of the strength of the seasonal effect. Our hypothesised insolation variable {*SumSol*} showed little correlation with canopy kelp {Supplementary Data Appendix B Table B1}, and only contributed to the statistical models in the presence of temperature.

Of the three different representations of hard substrate in the HSI models, *RMSM* performed marginally better than *RF*, and both significantly outperformed the *BoP* model. Comparing the HSI model spatial predictions {where the only difference was the form of this substrate variable, i.e., as per Fig. 2) shows the leverage of this predictor on the habitat predictions. Predictions from the *BoP* and *RMSM* models show the greatest difference, while both share similarities with the *RF* model {Fig. 5}. The distributions of predicted values show *RMSM* to have a bimodal distribution similar to the statistical models, while the other two models show roughly linearly decreasing probability as the prediction values increase to 1.0.

Discussion

Drivers of kelp distribution

Our models of canopy kelp distribution confirm the importance of abiotic drivers identified by studies of other kelp species, in other regions in the Northeast Pacific. In addition to depth, bottom type, bottom slope, exposure, tidal flow, and insolation, we further identified salinity as an important predictor, particularly for the distribution of giant kelp which is found in more sheltered waters than bull kelp. While not considered in earlier studies of kelp distributions, it is presumably relevant for broadly distributed species in regions such as the Northeast Pacific with complex coastlines, large embayments and fjords. It also appears to have an important interaction with exposure {see the parameter comparison below}.

While the processes underlying some of these drivers are clear {e.g., bottom type, exposure}, the role of others {e.g., depth} is more equivocal, in that they often serve as proxies for other, more proximate variables. How well the predictor variables characterise ecological function is

central to a model's forecasting skill {Austin 2002}. The desire to identify more proximate ecological predictors was the motivation for developing an explicit insolation predictor and exploring variable interactions. However, insolation did not enter any of the models, and post-hoc tests showed that *Slope* and *Depth* consistently explained more variability, suggesting these proxies do a very good job capturing this important ecological effect. Noting that *Depth* was a significant univariate predictor while *SumSol* was not strongly suggests that attenuation {driven primarily by depth and secondarily by slope and in some cases aspect} is more important than surface insolation, even when {as here} depth is restricted to the photic zone.

Given that ecosystems have been described as a prototypical, complex adaptive systems {Levin 1998}, it would be surprising if interactions were ever truly absent from species-habitat associations. We considered three interactions {Temperature-Exposure, Salinity-Exposure, and Slope-Tidal energy}. The last two explained a significant part of the deviance in our data. This could be an indication of important ecological processes useful for refining descriptions of kelp habitat in coastal ecosystems. However, the dramatic drop in skill observed in the IDE evaluation {but not evident in the cross-validation analysis} suggests those interactions are fitting something very specific in the training data. Further investigation is necessary to determine whether this fit is to ecological process or model noise.

Comparing parameter values across the different models {Table 3} provides a potentially rich source of insight into the ecological relevance of the predictors. For example, consistent parameter values across models {e.g., *Depth*, *RF*, *MaxTidal*} strongly imply high ecological relevance, while unexpected values {e.g., a negative correlation with salinity} may imply

406 unanticipated proxies - in this case, the interaction between salinity and exposure may be
407 indicative of a more proximate process that dominates or is in addition to the hypothesised role
408 of salinity in capturing freshwater influence. Testing such interpretations of correlative model
409 results is necessary to improve our understanding of process.

410 Finally, the high variability in the significance of the bottom type predictor (R_{Reef} in Table 3)
411 suggests its spatial representation is poor because it is necessary for canopy kelp occurrence. In
412 fact, we understand predictions of bottom type to be unreliable because the true spatial
413 variability of substrate can be very high, making characterisation challenging. This variability in
414 variable significance and perhaps parameter value across multiple models may be a useful
415 indicator of uncertainty in habitat predictions, and a generic means of identifying poorly
416 characterised predictors.

417 *Challenges to evaluating model transferability*

418 Our analysis of model forecast skill illustrates the value of IDE and the potential shortcomings of
419 cross-validation. Using IDE, all statistical models had a mean AUC score considerably lower than
420 when assessed with cross-validation, whether random or blocked, contributing more evidence
421 that cross-validation yields an inflated view of forecast skill. These results corroborate similar
422 findings on the landscape (Bahn and McGill 2013, Petitpierre et al. 2017), and from simulation
423 studies (Garcia-Callejas and Araujo 2016) showing how cross-validation can inflate model
424 performance scores. Vaughan and Ormerod (2005) attributed this to what they called
425 transportability. This is effectively the degree to which processes in the training data (and
426 represented by the statistical model) persist in the testing data. This is essentially the

assumption of stationarity (i.e., invariance in process across time and space) and it is unlikely to be true for highly heterogeneous habitats (Wagner and Fortin 2005), or for distributions influenced by temporal variability or biological interactions. Non-stationarity is perhaps the biggest challenge facing accurate ecological prediction. Various methods for addressing non-stationarity in correlative models have been proposed including spatial statistics (Wagner and Fortin 2005), geographically weighted regression (e.g., Austin 2007), and hierarchical models (Beale and Lennon 2012). However, such approaches have yet to be widely adopted, perhaps because current modelling methods are focused less on understanding and more on correlation (Bell and Schlaepfer 2016, Houlahan et al. 2017, Palacios et al. 2013b).

The interpretation of statistics based on cross-validation as estimates of model forecast skill requires the additional assumption of representativity (see discussion below). Since the interpretation of cross-validation as a measure of forecast skill routinely violates both stationarity and representativity, it is fair to conclude that models relying on cross-validation for estimates of forecast skill are over-fit for accurate prediction. Models are now built to maximise their description of the data with which they were built (Bell and Schlaepfer 2016). Maximising internal validation in this way can lead to unnecessary model complexity, particularly for non-parametric methods (Merow et al. 2013). This is somewhat troubling given what should be a general understanding that high complexity is no guarantee of high forecast skill, and may even reduce it (Bahn and McGill 2013, Costanza and Sklar 1985, Fulton et al. 2003). We argue that a reliance on cross-validation undermines the essential need to balance complexity and generality, and that this balance will be achieved more effectively with IDE and a greater focus on process (and thus on non-stationarity).

449 *Representativity*

450 The assumption of representativity (i.e., that the data used to build the model are
451 representative of the place and time of the desired forecast) is how predictive models treat the
452 challenge of sampling bias and non-independence, and applies to both the observations of
453 species distributions and the predictor data. Habitat suitability models continue to rely
454 primarily on opportunistic observations of species distributions, and these data are understood
455 to usually contain spatial sampling bias (Jarnevich et al. 2015, Phillips et al. 2009). Since cross-
456 validation measures model performance by subsampling from a larger sample using a single
457 model structure, it essentially provides a description of parametric uncertainty within a single
458 ecological context (i.e., a particular time, place, and ecosystem). This is appropriate when
459 developing a descriptive model or evaluating process hypotheses, but is clearly inadequate for
460 assessing model forecast skill which by definition must be measured by how well it performs in
461 different contexts (Houlahan et al. 2017).

462 While questions around the validity of random cross-validation are not new, they have focused
463 largely on the question of independence. In this case, independence can be seen as the aspect
464 of representativity that addresses the biases in observational data set - that is, how well they
465 describes a particular context. Block cross-validation (e.g., Pearson et al. 2006, Roberts et al.
466 2017) provides a means of dividing data sets into pseudo-independent samples, essentially
467 striving to restore some independence by dividing large collections of observations according to
468 the ecological contexts they represent. While the approach appears promising for reducing
469 over-inflated metrics of performance, it requires a sample size large enough to capture the full

environmental range of the predictor variables, making it less effective (as shown in this study) for relatively small collections of largely independent field observations.

The inter-annual variability observed in model forecast skill (Fig. 4) led us to also consider the role of temporal sampling bias, something not as widely considered as spatial bias. Canopy kelp distributions change seasonally during the growing season, and annually in response to winter storms. Edwards and Estes (2006) linked inter-annual variability in giant kelp abundance to El Nino events. In Pacific Canada, cooler (La Nina) years generally lead to higher winter storm frequency on the coast (NOAA 2015b), suggesting a higher likelihood of kelp forest destruction, and thus lower prevalence the following summer. This explicit, inter-annual effect suggests our model may agree more with SPOT-5 data from a similar ocean climate (i.e., context), than from a different context. Using the ONI (Oceanic Nino Index - NOAA 2015a) we found our survey data (2010 to 2012) were collected during a cooler La Nina period, implying the observations were biased towards kelp in less exposed areas. The SPOT-5 data included years from both a mild El Nino (2005) and mild La Nina (2006, 2009). While our results appear to contradict this hypothesis (the 2005 SPOT-5 data consistently evaluated better than the other years), this nevertheless illustrates the potential for temporal bias in observational data, especially if data are collected during an anomalous period.

Temporal variability is reduced as data are aggregated over time. However, representing potential habitat using time-invariant observations assumes these training data are representative of the long-term average (GREGG 2011). This can have a strong effect on model accuracy when evaluated with data from particular years (Scales et al. 2017). The actual

491 difference between the training sample context and the long-term average can be termed the
492 temporal sampling bias, and its magnitude can be inferred from how well the model forecasts
493 evaluate against different time steps (e.g., years) of independent observations. We argue that
494 temporal sampling bias can be a critical component of the forecast uncertainty, analogous to
495 the role played by initial conditions in simulation models (Fulton et al. 2003, Gibson and Spitz
496 2011). Understanding temporal sampling bias is also likely to help estimate forecast horizons –
497 the distance (either spatial or temporal) to which a prediction can be expected to be reliable
498 (Petchey et al. 2015).

499 The idea of temporal sampling bias also applies to the averaging of dynamic predictors.
500 Typically, habitat models use dynamic climate variables as either climatologies (i.e., long-term
501 climate averages) or temporally concurrent data. Using contemporaneous physical data can
502 lead to high correlations and insights into process at short time scales. While challenges remain
503 (Mannocci et al. 2017), this approach to dynamic ocean management is proving effective for
504 short-term, tactical management of, for example, vessel-based impacts on endangered large
505 pelagics such as turtles and cetaceans (Maxwell et al. 2015).

506 In contrast, climatologies remove inter-annual variability, and Gregr (2011) noted that using
507 climatological predictors with multiple years of observations leads to models that capture the
508 persistent, long term signals indicative of a species' potential habitat. However, accurately
509 describing these signals depends critically on long-term observations – using a climatology to
510 predict a single year of observations is subject to the temporal bias described above. This is
511 illustrated in our study by the significant improvement in performance observed for the HSI

models when evaluated with the aggregated LOS data. As a proxy for long-term observations of canopy kelp presence, these data are more appropriately scaled to the averaged predictors. This emphasises the importance of matching the resolutions of observations and predictors {sensu Wiens 1989} to accurately characterise the underlying ecological processes.

Since habitat suitability models often rely on predictions from other types of models {e.g., simulations of ocean climate and global climate change, or processed remote sensing products}, we examined how the accuracy of one such modelled predictor can influence habitat predictions. Bottom type is critical to canopy kelp. However, this predictor can be derived in a number of ways {Fig. 2} and these differences can manifest both statistically {Fig. 3} and spatially {Fig. 5}. Our methods led us to using *RF* as the predictor of bottom type in our multi-model comparison, but we know this predictor is biased towards hard substrate {Haggarty 2015}. This over-representation of hard bottom influences all the habitat predictions and emphasises the need to consider the representativity and potential biases in modelled predictor variables, something that will also be facilitated through IDE.

Fundamentally, representativity is a question of combining correctly scaled observations and predictors to allow the presumed ecological process(es) to be appropriately captured. This is essentially a re-phrasing of the scaling challenge articulated by Wiens {1989}. Whereas Wiens emphasised the importance of correctly matching spatial and temporal resolutions before inferring process, we emphasise the consequential need for predictors to properly represent the ecological context of the observations {Greegr 2011, Montalto et al. 2014}.

The case for reduced model complexity

533 We found that the GAMs, by providing flexible non-linear relationships {similar to machine
534 learning methods), tended to emphasise predictors {e.g., depth, bottom type) that could be
535 closely fit to data. Over-fitting predictors with high leverage while overlooking other, potentially
536 more ecological predictors is one way flexible non-linear pattern matching tools may produce
537 less generalisable results than simpler models for an equivalent number of degrees of freedom
538 {a result also observed by Randin et al. 2006). This also explains why pattern-matching
539 approaches often create ecologically implausible relationships {GREGG 2011, Merow et al. 2013,
540 Warren and Seifert 2011). In contrast, GLMs had less precise fits to individual drivers, resulting
541 in a poorer fit to training data but better forecasts. GLMs also provide a coherent way to
542 rationally construct higher-order terms and interactions to represent the full range of
543 ecologically reasonable processes. Considering ecological relevance is central to integrating
544 ecological theory and modelling practice {Austin 2007). By requiring the analyst to be explicit,
545 and consider ecological processes a priori, we argue GLMs are better suited to this integration.
546 They are also easier to explain {Wintle et al. 2005), and thus more likely to be ecologically
547 defensible.

548 When ecological context is well understood, the need for complex models to predict potential
549 habitat distributions may be even less. Our analysis shows what can be accomplished with a
550 simple habitat suitability model if the ecology is reasonably well understood, re-enforcing the
551 importance of process to generating accurate model predictions. While poor at predicting any
552 particular instance of the kelp data {i.e., a realized habitat) our HSI model proved generalisable
553 enough to not only sometimes outscore the statistical models across the different evaluation
554 data sets, but to far outperform those models when predicting potential habitat {Fig. 3). While

no substitute for analyses grounded in both ecological understanding and statistical analysis, HSI models can provide a reasonable starting point for estimating habitat, particularly in data-poor contexts. Equally important is that they may well outperform statistical models when training data are biased, or when models are fit without ecological consideration, which can lead to model mis-specification if relevant biotic or anthropogenic drivers are overlooked {Palacios et al. 2013a, Sober6n and Nakamura 2009}.

The distribution of predicted values may also provide some insight into whether a simpler model is preferred, particularly when forecast skill is variable. The increased bimodality we observed as model complexity increased {Fig. 4} means that at high complexity, threshold values used to partition the predictions into 'habitat' and 'not habitat' have little effect at intermediate values. This can lead to lower AUC scores because if the model forces potentially equivocal locations into one state or the other {e.g., Fig. 4c}, incorrect observations in those locations will contribute less to the AUC than for models with a more uniform distribution of predictions. Thus, the distribution of predictions may provide a useful continuous metric for assessing model performance as recommended by Lawson et al. {2014}, and also inform the well-understood challenges of threshold selection and the trade-off between false positives and false negatives.

Objectives, assumptions and uncertainties

Appropriate interpretation of model performance fundamentally depends on model intent {Jimenez-Valverde et al. 2008, Rykiel Jr. 1996}. The vast majority of models continue to predict potential habitat {i.e., by considering abiotic predictors only}, yet many continue to be

576 characterised as models of *species distribution*, {Araujo and Peterson 2012}. This
577 misrepresentation of potential habitat models not only confounds model assessment, but
578 downplays the importance of realised habitat and the challenge of accurately predicting
579 realised speciesdistributions. This challenge is illustrated by our evaluation of potential habitat
580 predictions with observed {i.e., realised} distributions, which clearly shows inter-annual
581 variability. However, given the high potential for temporal bias discussed above, predictions of
582 realised habitat will likely require a representation of temporal variability in addition to other
583 drivers of species' realised distributions such as barriers to movement and biological
584 interactions {see Sober6n and Nakamura 2009}. The necessary increases in model complexity
585 will add both parametric and structural uncertainties, emphasising the trade-off between
586 model complexity and generality.

587 Model intent also informs underlying model design, including decisions about what predictors
588 to include, and how to scale them to match the observed distributional data. While such
589 decisions relate directly to the structural uncertainty and representativity of the model, they
590 are most often implicit, and rarely evaluated after the initial model design {Gregn and Chan
591 2014}. However, because such decisions typically comprise a significant portion of the
592 uncertainty in any model prediction, a holistic sense of model uncertainty will not be achieved
593 without articulating model objectives, the relevant ecological contexts, the processes that are
594 {and are not} represented, and any known biasesin the data. Such transparency is essential for
595 conveying the overall confidence that can be placed in any model results and should ultimately
596 be rooted in ecological plausibility, particularly for data poor species . However, the integration

597 of these diverse uncertainties remains beyond the reach of most current quantitative
598 uncertainty assessments {but see Gelfand et al. 2006}.

599 **Conclusions**

600 Our models of canopy kelp habitat, the first developed for the Northeast Pacific, provide insight
601 into the role of various habitat predictors, and emphasise the critical role ecological
602 understanding plays in model development through both the consideration of interactions {a
603 potentially important, under-studied aspect of habitat) and the construction of process-based
604 models.

605 Our tests of cross-validation show that it can lead to models with high complexity, closely fit to
606 a single ecological context. Thus, any associated performance metrics will typically inflate
607 measures of model forecast skill. This is because cross-validation relies on the generally false
608 assumptions of process stationarity and data representativity. Our results show that by
609 confronting model predictions with data from different contexts, IDE provides a more accurate
610 test of model transferability. Our collective understanding of model forecast skill would thus be
611 greatly improved by avoiding claims that cross-validation is a reliable measure of forecast skill.

612 In our methods, we emphasised the importance of deliberate model design, focusing on
613 ecological process rather than relying strictly on correlation. The performance of our HSI
614 models, and the simple GLMs, shows that when ecologically sound hypotheses about process
615 can be described, simpler, process-based models can provide better predictions of potential
616 habitat suitability than more complex models. This has significant implications for data-poor

situations, or when observational data contain significant spatial or temporal bias, as when they are only snapshots of a temporally or spatially variable pattern.

Down the road, to be truly useful for decision-making, we will need to be clear on what our models are predicting, and offer an assessment of the confidence that can be placed in their predictions. We therefore encourage the modelling community to clearly distinguish between predictions of potential and realised habitat, and to articulate assumptions about data representativity, the suitability of the predictor variables, and the processes linking the data to the model context from the outset. Only by explicitly considering the uncertainties associated with our assumptions can a comprehensive view of model confidence be obtained. Without this holistic treatment of uncertainty, our models will {and should} remain largely outside the policy and decision-making spheres.

Acknowledgements

Thanks to Dr. Bill Crawford for helpful conversations regarding oceanography and model scaling. We are also grateful to Dr. Louis Druehl for sharing his considerable expertise on kelp habitat in his review of the Habitat Suitability Index model, and to Joanne Lessard from Fisheries and Oceans Canada who has been instrumental in the development of nearshore habitat predictors in Pacific Canada. We also thank Alan Smith of Kyuqout, BC and Wayne Markel of Bamfield, BC for their support in the field, and acknowledge the Nuuchahnulth First Nations, within whose unceded territories this field work was conducted. We are grateful for the thoughtful and detailed comments provided by a number of anonymous reviewers which led to a much more concise and substantive article. This research was funded by NSERC through Strategic Projects Grant UBC # 17R43675, Discovery Grant UBC # 17R82930, and a Canadian

639 Graduate Scholarship to EJG. KMAC was also funded by the Canada Research Chairs program,
640 and the CFI Leaders Opportunity Fund {UBC # 17R39196}.

641

642 **References**

- 643 Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence,
644 kappa and the true skill statistic {TSS}. — J. Appl. Ecol. 43: 1223-1232.
- 645 Araujo, M. B. and Guisan, A. 2006. Five {or so} challenges for species distribution modelling. -
646 J. Biogeogr. 33: 1677-1688.
- 647 Araujo, M. B. et al. 2005a. Validation of species-climate impact models under climate change.
648 — Global Change Biol. 11: 1504-1513.
- 649 Araujo, M. B. and Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modeling. -
650 Ecology 93: 1527-1539.
- 651 Araujo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? - Science
652 313: 1396.
- 653 Araujo, M. B. et al. 2005b. Reducing uncertainty in projections of extinction risk from climate
654 change. — Global Ecol. Biogeogr. 14: 529-538.
- 655 Ardron, J. 2004. Kelp Notes. Living Oceans Society, p 2.
- 656 Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and
657 some possible new approaches. — Ecol. Model. 200: 1-19.
- 658 Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological
659 theory and statistical modelling. — Ecol. Model. 157: 101-118.
- 660 Bahn, V. and McGill, B. J. 2013. Testing the predictive performance of distribution models. -
661 Oikos 122: 321-331.
- 662 Ballings, M. and Van den Poel, D. 2013. AUC: Threshold independent performance measures for
663 probabilistic classifiers.
- 664 Beale, C. M. and Lennon, J. J. 2012. Incorporating uncertainty in predictive species distribution
665 modelling. — Philosophical Transactions of the Royal Society B: Biological Sciences 367:
666 247-258.
- 667 Bekkby, T. et al. 2009. Spatial predictive distribution modelling of the kelp species *Laminaria*
668 *hyperborea*. — ICES J. Mar. Sci. 66: 2106-2115.

- 669 Bell, D. M. and Schlaepfer, D. R. 2016. On the dangers of model complexity without ecological
670 justification in species distribution modeling. — Ecol. Model. 330: 50-59.
- 671 Costanza, R. and Sklar, F. H. 1985. Articulation, accuracy and effectiveness of mathematical
672 models: A review of freshwater applications. — Ecol. Model. 27: 45-68.
- 673 Dayton, P. K. 1985. Ecology of kelp communities. — Annu. Rev. Ecol. Syst. 215-245.
- 674 Dormann, C. F. et al. 2013. Collinearity: a review of methods to deal with it and a simulation
675 study evaluating their performance. — Ecography 36: 27-46.
- 676 Druehl, L. D. 1978. The distribution of *Macrocystis integrifolia* in British Columbia as related to
677 environmental parameters. — Canadian Journal of Botany 56: 69-79.
- 678 Druehl, L. D. 2001. Pacific Seaweeds: A guide to common seaweeds of the West Coast. -
679 Harbour Publishing.
- 680 Duggins, D. 1988. The effects of kelp forests on nearshore environments: biomass, detritus, and
681 altered flow. — In: Estes, J. A. and VanBlaricom, G. R. {eds), The community ecology of
682 sea otters. Springer-Verlag, pp. 192-201.
- 683 Edwards, M. S. and Estes, J. A. 2006. Catastrophe, recovery and range limitation in NE Pacific
684 kelp forests: a large-scale perspective. — Mar. Ecol. Prog. Ser. 320: 79-87.
- 685 Elith, J. and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons
686 for differing performances of species distribution models. — Ecography 32: 66-77.
- 687 Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence
688 data. — Ecography 29: 129-151.
- 689 Estes, J. A. and Palmisano, J. F. 1974. Sea otters: Their role in structuring nearshore
690 communities. — Science 185: 1058-1060.
- 691 Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors
692 in conservation presence/absence models. — Environ. Conserv. 24: 38-49.
- 693 Fulton, E. A. et al. 2003. Effect of complexity on marine ecosystem models. - Mar. Ecol. Prog.
694 Ser. 253: 1-16.
- 695 Garcia-Callejas, D. and Araujo, M. B. 2016. The effects of model and data complexity on
696 predictions from species distributions models. — Ecol. Model. 326: 4-12.

- 697 Ge, Q. Z. et al. 2008. High efficient classification on remote sensing images based on SVM. –
 698 Proc. of the International Archives of the Photogrammetry, Remote Sensing and Spatial
 699 Information Sciences 37:
- 700 Gelfand, A. E. et al. 2006. Explaining species distribution patterns through hierarchical
 701 modeling. – Bayesian Analysis 1: 41-92.
- 702 Gibson, G. A. and Spitz, Y. H. 2011. Impacts of biological parameterization, initial conditions,
 703 and environmental forcing on parameter sensitivity and uncertainty in a marine
 704 ecosystem model for the Bering Sea. – J. Mar. Syst. 88: 214-231.
- 705 Gorman, D. et al. 2013. Modeling kelp forest distribution and biomass along temperate rocky
 706 coastlines. – Mar. Biol. 160: 309-325.
- 707 Gregr, E. J. 2011. Insights into North Pacific right whale *Eubalaena japonica* habitat from
 708 historic whaling records. – Endangered Species Research 15: 223-239.
- 709 Gregr, E. J. and Chan, K. M. A. 2014. Leaps of Faith: How implicit assumptions compromise the
 710 utility of ecosystem models for decision-making. – Bioscience 65: 43-54.
- 711 Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. -
 712 Ecol. Model. 135: 147-186.
- 713 Hartig, F. 2017. DHARMA: Residual Diagnostics for Hierarchical {Multi-Level / Mixed} Regression
 714 Models. R package version 0.1.5.
- 715 Heikkinen, R. K. et al. 2012. Does the interpolation accuracy of species distribution models
 716 come at the expense of transferability? – Ecography 35: 276-288.
- 717 Hirzel, A. H. et al. 2006. Evaluating the ability of habitat suitability models to predict species
 718 presences. – Ecol. Model. 199: 142-152.
- 719 Houlahan, J. E. et al. 2017. The priority of prediction in ecological understanding. - Oikos 126:
 720 1-7.
- 721 Jarnevich, C. S. et al. 2015. Caveats for correlative species distribution modeling. - Ecological
 722 Informatics 29: 6-15.
- 723 Jimenez-Valverde, A. et al. 2011. Use of niche models in invasive species risk assessments. -
 724 Biol. Invasions 13: 2785-2797.

- 725 Jimenez-Valverde, A. et al. 2008. Not as good as they seem: the importance of concepts in
726 species distribution modelling. — *Divers. Distrib.* 14: 885-890.
- 727 Jørgensen, S. E. 1986. *Fundamentals of Ecological Modelling*. — Elsevier.
- 728 Lawson, C. R. et al. 2014. Prevalence, thresholds and the performance of presence-absence
729 models. — *Methods in Ecology and Evolution* 5: 54-64.
- 730 Levin, S. A. 1998. Ecosystems and the biosphere as complex adaptive systems. - *Ecosystems* 1:
731 431-436.
- 732 Link, J. S. et al. 2012. Dealing with uncertainty in ecosystem models: The paradox of use for
733 living marine resource management. — *Prog. Oceanogr.* 102: 102-114.
- 734 Manel, S. et al. 2001. Evaluating presence-absence models in ecology: the need to account for
735 prevalence. — *J. Appl. Ecol.* 38: 921-931.
- 736 Mannocci, L. et al. 2017. Temporal resolutions in species distribution models of highly mobile
737 marine animals: recommendations for ecologists and managers. — *Divers. Distrib.* 23:
738 1098-1109.
- 739 Maxwell, S. M. et al. 2015. Dynamic ocean management: Defining and conceptualizing real-time
740 management of the ocean. — *Mar. Policy* 58: 42-50.
- 741 Meleder, V. et al. 2010. Predictive modelling of seabed habitats: case study of subtidal kelp
742 forests on the coast of Brittany, France. — *Mar. Biol.* 157: 1525-1541.
- 743 Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution
744 models? — *Ecography* 37: 1267-1281.
- 745 Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it
746 does, and why inputs and settings matter. — *Ecography* 36: 1058-1069.
- 747 Montalto, V. et al. 2014. Testing the effects of temporal data resolution on predictions of the
748 effects of climate change on bivalves. — *Ecol. Model.* 278: 1-8.
- 749 Mouton, A. M. et al. 2010. Ecological relevance of performance criteria for species distribution
750 models. — *Ecol. Model.* 221: 1995-2002.
- 751 Nakazawa, M. 2017. fmsb: Functions for Medical Statistics Book with some Demographic Data.
752 R package version 0.6.1.

- 753 Nichol, L. M. et al. 2009. Recent trends in the abundance and distribution of sea otters (*Enhydra*
754 *lutris*) in British Columbia. — DFO Can. Sci. Advis. Sec. Res. Doc. 2009/016 iv + 16 p.
- 755 NOAA. 2015a. Cold and Warm Episodes by Season. — Accessed: May 30, 2017
756 http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml.
- 757 NOAA. 2015b. Storm Tracks. — Accessed: May 30, 2017
758 http://www.cpc.ncep.noaa.gov/products/precip/CWlink/stormtracks/strack_NH.shtml.
- 759 Palacios, D. M. et al. 2013a. Beyond correlation: integrating environmentally and behaviourally
760 mediated processes in models of marine mammal distributions. — *Endangered Species*
761 *Research* 22: 191-203.
- 762 Palacios, D. M. et al. 2013b. Modeling the temperature-nitrate relationship in the coastal
763 upwelling domain of the California Current. - *Journal of Geophysical Research: Oceans*
764 118: 3223-3239.
- 765 Pearson, R. G. et al. 2006. Model-based uncertainty in species range prediction. - *J. Biogeogr.*
766 33: 1704-1711.
- 767 Petchey, O. L. et al. 2015. The Ecological Forecast Horizon, and examples of its uses and
768 determinants. — *Ecol. Lett.* 18: 597-671.
- 769 Petitpierre, B. et al. 2017. Selecting predictors to maximize the transferability of species
770 distribution models: lessons from cross-continental plant invasions. — *Global Ecol.*
771 *Biogeogr.* 26: 275-287.
- 772 Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models:
773 implications for background and pseudo-absence data. — *Ecol. Appl.* 19: 181-197.
- 774 Planque, B. et al. 2011. Understanding what controls the spatial distribution of fish populations
775 using a multi-model approach. — *Fish. Oceanogr.* 20: 1-17.
- 776 R Core Team 2015. R: A language and environment for statistical computing. R Foundation for
777 Statistical Computing.
- 778 Radosavljevic, A. and Anderson, R. P. 2014. Making better Maxent models of species
779 distributions: complexity, overfitting and evaluation. — *J. Biogeogr.* 41: 629-643.
- 780 Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? -
781 *J. Biogeogr.* 33: 1689-1703.

- 782 Roberts, D. R. et al. 2017. Cross-validation strategies for data with temporal, spatial,
783 hierarchical, or phylogenetic structure. — *Ecography* 40: 913-929.
- 784 Robinson, L. M. et al. 2011. Pushing the limits in marine species distribution modelling: lessons
785 from the land present challenges and opportunities. — *Global Ecol. Biogeogr.* 20: 789-
786 802.
- 787 Rodriguez-Rey, M. et al. 2013. Species distribution models predict range expansion better than
788 chance but not better than a simple dispersal model. — *Ecol. Model.* 256: 1-5.
- 789 Rykiel Jr., E. J. 1996. Testing ecological models: the meaning of validation. - *Ecol. Model.* 90:
790 229-244.
- 791 Santika, T. 2011. Assessing the effect of prevalence on the predictive performance of species
792 distribution models using simulated data. — *Global Ecol. Biogeogr.* 20: 181-192.
- 793 Scales, K. L. et al. 2017. Scale of inference: on the sensitivity of habitat models for wide-ranging
794 marine predators to the resolution of environmental data. — *Ecography* 40: 210-220.
- 795 Schmolke, A. et al. 2010. Ecological models supporting environmental decision making: a
796 strategy for the future. — *Trends in Ecology and Evolution* 25: 479-486.
- 797 Segurado, P. and Araujo, M. B. 2004. An evaluation of methods for modelling species
798 distributions. — *J. Biogeogr.* 31: 1555-1568.
- 799 Soberón, J. and Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and
800 assumptions. — *Proceedings of the National Academy of Sciences* 106: 19644-19650.
- 801 SPOT. 2017. SPOT {Satellite Pour l'Observation de la Terre}. - Accessed: October 22 2017.
802 <http://eoedu.belspo.be/en/satellites/spot.htm>.
- 803 Springer, Y. et al. 2007. Ecology and management of the bull kelp, *Nereocystis luetkeana*.
804 Lenfest Ocean Program, p 48.
- 805 Steneck, R. S. et al. 2002. Kelp forest ecosystems: biodiversity, stability, resilience and future. -
806 *Environ. Conserv.* 29: 436-459.
- 807 Sumaila, U. R. et al. 2011. Climate change impacts on the biophysics and economics of world
808 fisheries. — *Nature Climate Change* 1: 449-456.

- 809 Tommasi, D. et al. 2017. Managing living marine resources in a dynamic environment: The role
810 of seasonal to decadal climate forecasts. — *Prog. Oceanogr.* 152: 15-49.
- 811 Van Horne, B. and Wiens, J. A. 1991. Forest bird habitat suitability models and the development
812 of general habitat models. — *U.S. Fish Wild. Serv. U.S. Fish Wild. Serv.* 8. 31 p.
- 813 Vaughan, I. P. and Ormerod, S. J. 2005. The continuing challenges of testing species distribution
814 models. — *J. Appl. Ecol.* 42: 720-730.
- 815 Wagner, H. H. and Fortin, M.-J. 2005. Spatial analysis of landscapes: concepts and statistics. -
816 *Ecology* 86: 1975-1987.
- 817 Warren, D. L. and Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of
818 model complexity and the performance of model selection criteria. — *Ecol. Appl.* 21:
819 335-342.
- 820 Whittingham, M. J. et al. 2006. Why do we still use stepwise modelling in ecology and
821 behaviour? — *J. Anim. Ecol.* 75: 1182-1189.
- 822 Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. - Springer Science & Business
823 Media.
- 824 Wickham, H. and Francois, R. 2015. *dplyr: A Grammar of Data Manipulation*.
- 825 Wiens, J. A. 1989. Spatial scaling in ecology. — *Funct. Ecol.* 3: 385-397.
- 826 Wintle, B. A. et al. 2005. Fauna habitat modelling and mapping: a review and case study in the
827 Lower Hunter Central Coast region of NSW. — *Austral Ecol.* 30: 719-738.
- 828 Wood, S. N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized
829 additive models. — *Journal of the American Statistical Association* 99: 673-686.
- 830

831 Table 1: Summary of variables and interactions considered in the development of a generic canopy kelp model for coastal British
832 Columbia, Canada. Variable names are used in the text, figures, and subsequent tables. Details on the data sources are provided in
833 Supplemental Material Appendix 1.

Variable name	Description	Units
Depth	Water column depth below the low water line.	Metres {m}
Slope	The steepness of the grid cell, based on the elevation of its neighbours.	
Aspect	The orientation of the grid cell, with 0 degrees indicating North	{0, 359}
RF	Predicted bottom type based on Random Forest analysis.	{0, 1}
RMSM	Random forest bottom type grouped into 4 classes.	[Rock, Mixed, Sand, Mud]
BoP	Estimate of bottom type extrapolated from observations.	[Rock, Mixed, Soft]
Salt	Average predicted bottom salinity.	
BTemp	Average predicted bottom temperature. Mean	C
SST	sea surface temperature {2010-2012}.	C
SSTQ2	Mean sea surface temperature {2010-2012} for April to June.	C
SSTQ3	Mean sea surface temperature {2010-2012} for July to September.	C
SumSol	Predicted summer insolation at the surface	
FetchTot	Cumulative potential exposure to wave energy based on fetch.	km
FetchSE	Cumulative potential exposure from the Southeast.	km
FetchNW	Cumulative potential exposure from the Northwest.	km
MaxTidal	Maximum predicted bottom tidal speed.	cm/s
Slope x Aspect	A potential proxy for sunlight.	--
Slope x MaxTidal	The potential for the effect of tidal energy to be a function of slope.	--
Fetch x BTemp	The potential for the effect of temperature to be a function of exposure.	--
Fetch x Salt	The potential for the effect of salinity to be a function of exposure.	--

834

835

Table 2: Results of model development showing model terms, AIC scores, and deviance explained (R^2) with increasing model complexity. First column shows the model names used in the comparative analysis. See text for a description of the variables. Lower order terms were included with all higher order GLM polynomials. In GAM1, degrees of freedom (k) were fixed. In GAM2, k was initially set to 10 individual smooths and 20 for interactions; the reported k is a result of penalization to allow AIC to drive the number of degrees of freedom.

Model Scope		AIC	R^2
<i>GLMs</i>			
GLM0	Depth + RReef + Salt + FetchTot	171	0.30
GLM1	Depth + RReef + Salt + FetchTot + Slope + MaxTidal	153.8	0.40
GLM2	Depth + RReef + Salt + FetchTot ² + Slope + MaxTidal	150	0.42
GLM3	Depth + RReef + Slope + MaxTidal + FetchTot ² + FetchTot ² :Salt + Slope ² :MaxTidal	116.6	0.59
<i>GAMs</i>			
GAM1	Depth + RReef + Salt + Slope + MaxTidal + te{FetchTot, $k = 3$ } + ti{Salt by FetchTot, $k = 5$ } + ti{MaxTidal by Slope, $k = 5$ }	120.6	0.59
GAM2	Depth + RReef + Salt + Slope + MaxTidal + te{FetchTot, $k = 5$ } + ti{Salt by FetchTot, $k = 6$ } + ti{MaxTidal by Slope, $k = 6$ }	81.1	0.79

852

8554

Figure 1: The location of the study area on the west coast of Vancouver Island, British Columbia (A) in the Northeast Pacific (Canada in dark grey; USA in light grey); (B) the depth limits of the study area (0-20 m, pale blue) and the survey locations (black dots), along with the boundaries of the available independent remote sensing data (dark blue line = 2005, grey line = 2006, red line = 2009); and (C) the resolution of the study with respect to land (grey) and the density of survey locations.

Figure 2: The definition of the three habitat suitability index (HSI) models, each using a different representation of bottom type (from top to bottom: Bottom Patches (BoP; Gregr et al. 2013) coded as Hard, Mixed, and Soft; a continuous 'hardness' index based on Random Forest analysis (RF, Haggarty 2015), and a classification of RF into Rock, Mixed, Sand, and Mud classes (termed RMSM in the text).

Figure 3: Correspondence scores (Area Under the Curve, AUC) showing cross-validation (CV) scores (black points = 3-fold; grey points = 3-year cross-block) and the independent data validation using four different independent data sets for nine increasingly complex models across the x-axis. Models include three Habitat Suitability Index models (BoP, RMSM, and RF), four Generalized Linear Models (GLMs), and two Generalized Additive Models (GAMs). Regression models were built using 3 years of survey data. Variability (± 1 standard deviation) for 3-fold CV is from 100 randomly sampled folds; for cross-block CV it is based on the 3 survey years. For independent data evaluation, points are shown with 95% confidence intervals, based on 100 simulated samples of 500 observations. Independent data included three years of classified remote sensing data (2005, 2006, 2009) and a compilation of long-term kelp observations (LOS).

Figure 4: Heat maps (black = low, red = high) showing habitat predictions from A) the simplest generalised linear model (GLM0), B) the simplest GLM (GLM1), and C) the most complex generalised additive model (GAM2) for the central portion of the study area showing how, with increasing model complexity, predictions are increasingly concentrated in shallow, high relief areas while reducing the amount of intermediate quality habitat predicted. Histograms above the predictions show how the predictions become increasingly bimodal with model complexity, an indicator of how closely the models are fit to the data.

Figure 5: Heat maps (black = low, red = high) showing habitat suitability index predictions based on the three representations of bottom type (A - Bottom Patches (BoP), B - Random Forest (RF) prediction, C - Random Forest classified into Rock, Mixed, Sand Mud (RMSM) categories). Histograms above each map show the distribution of predicted values. The difference between these predictions illustrates the sensitivity of identical model representations to differences in a key underlying predictor.

Ecography

ECOG-03470.

Supplementary material

Contents

Appendix A - Ecological relationships between predictors and kelp habitat

Appendix B - Univariate tests of significance for each of the independent variables

References

Appendix A

Ecological relationships between predictors and kelp habitat

Hard substrate is critical to provide kelp a strong attachment point. However, unequivocal descriptions of bottom type, particularly near shore, are not broadly available since acoustic bottom-mapping surveys (e.g., Gorman et al. 2013) are time consuming and spatially limited (i.e., suitable acoustic surveys were unavailable across our study area). Instead, we assess the utility of three different predictive models of bottom type. Two were derived from depth according to Haggarty (2015): *RF*, the score of a random forest analysis applied to the bathymetry, and *RMSM*, *RF* classified to Rock-Mixed-Sand-Mud. The third (bottom patch - *BoP*) classifies bottom type into Hard, Mixed, and Soft using the best available observational data following (Grega et al. 2013).

Insolation (i.e., sunlight received) is important to all life history stages of kelp (Dayton 1985). Gorman et al. (2013) used mean light levels derived from satellite data, while Bekkby et al. (2009) used a proxy based on slope and aspect of the bottom. We hypothesised that along coasts with high topographic relief like the Northeast Pacific, shading may play an important

role and so constructed an insolation model using the Solar Radiation module from ArcGIS 10.2 {ESRI 2015}. We estimated monthly potential sunlight for each depth pixel in the study area, and created predictors to represent mean summer {*SumSol*} and winter {*WinSol*} insolation. For comparison, we calculated the proxies *Slope* and *Aspect* proposed by Bekkby et al. {2009} from the bathymetry.

In California, spring growth of giant kelp has been negatively correlated with sea surface temperature {SST}. This has been interpreted as a proxy for nutrient availability {Cavanaugh et al. 2010} since cold, upwelled waters are also typically nutrient rich. However, temperature may also be physiologically limiting for canopy kelp, as Springer et al. {2007} reported an upper thermal limit for bull kelp, and Buschmann et al. {2004} reported improved early growth of giant kelp at lower temperatures.

We downloaded satellite-derived SST data using the Marine Geospatial Ecology Tool {version 0.8z40, Roberts et al. 2010} as 3-month averages from the 1-km resolution GLOB-G1SST data set for each quarter for the years 2010 to 2012. We derived three potential predictor variables from these remotely sensed data: *SST* represented the overall mean temperature across all quarters in all years, while *SSTQ2* and *SSTQ3* represented the mean temperature for the second and third quarters of the year {the peak growing period}, averaged across the 3 years of sampling. Because many kelp survey points were near shore, some did not fall in a 1 km grid cell. For these points, the closest SST values were used.

Water motion, particularly through wave action, is important because of its ability to dislodge kelp from its substrate {Springer et al. 2007} and was a significant factor in all the studies on

kelp distribution we reviewed {e.g., Bekkby et al. 2009, Cavanaugh et al. 2010, Gorman et al. 2013, Pedersen et al. 2012}. We used fetch {an index of exposure - Lessard et al. 2007} as a proxy for wave energy. We calculated total fetch {from 120 equally spaced bearings} and directional fetch for the dominant summer {Southeast} and winter {Northwest} wind directions for points spaced 100 m apart across our study area, and interpolated these variables {*FetchTot*, *FetchSE*, *FetchNW*} to the depth grid.

While excessive wave action may be responsible for dislodging plants, we hypothesised that less-disruptive water movement may play a critical role in the circulation of nutrient-rich waters. We therefore considered the role of tidal energy using maximum bottom tidal speed {*MaxTidal*} obtained from Foreman et al. {2008}.

Although not well studied, salinity may also affect kelp distribution, as few kelp species can tolerate low salinities {Dayton 1985}. This was clearly observed during the kelp survey results, where extensive sampling closer to freshwater inputs {i.e., away from the open ocean} yielded very little kelp. There is also evidence of a positive correlation between giant kelp spore production and salinity {Buschmann et al. 2004}. In the absence of contemporaneous salinity data, we defined our salinity predictor {*Salt*} using long-term average summer bottom salinities {Grega et al. 2016}.

Finally, while sedimentation has also been identified as important for kelp habitat {e.g., Springer et al. 2007}, we lacked suitable proxies for this factor and thus could not include it in the analysis. However, it is reasonable to assume that sedimentation is correlated with bottom

type, meaning that rocky reefs by definition receive little sedimentation. We therefore assumed that sedimentation was sufficiently represented by bottom type in the models.

We used an equidistant projection {BC Albers Conic} to map all predictors to the depth grid. Lower resolution predictors {*BTemp*, *Salt*, *MaxTidal*} were re-sampled to this working resolution.

Interactions

While difficult to quantify, interactions may be critical to understanding habitat suitability and the derivation of more process-based models. We tested a number of possible interactions defined using hypotheses describing how different abiotic variables could combine to improve kelp habitat quality:

- 1) Slope and aspect. We elaborated on Bekkby et al.'s {2009} use of these two variables as a proxy for sunlight by testing if interaction terms were more effective predictors of kelp.
- 2) Exposure and temperature. These variables are inversely correlated with distance from shore and also appear to have opposite effects on kelp habitat suitability. Cavanaugh et al. {2010} found the response of giant kelp biomass was dominated by wave disturbance in exposed regions, while a temperature effect was more dominant in sheltered regions. Such regional heterogeneity strongly suggests an interaction between temperature and exposure.
- 3) Exposure and salinity. In contrast to temperature, salinity is positively correlated with distance from shore, and with exposure. We hypothesised that the importance of elevated

salinity would vary with exposure, perhaps playing a more significant role in low exposure areas.

4) Slope and tidal energy. Tidal energy circulates and mixes water quality characteristics; slope can serve as both a proxy for bottom type and a component of insolation. We hypothesised that the role of tidal energy differed in regions with low and high slope.

Appendix B

Univariate tests of significance for each of the independent variables

References

- Bekkby, T. et al. 2009. Spatial predictive distribution modelling of the kelp species *Laminaria hyperborea*. — ICES J. Mar. Sci. 66: 2106-2115.
- Buschmann, A. et al. 2004. The effect of water movement, temperature and salinity on abundance and reproductive patterns of *Macrocystis* spp. (Phaeophyta) at different latitudes in Chile. — Mar. Biol. 145: 849-862.
- Cavanaugh, K. C. et al. 2010. Scaling giant kelp field measurements to regional scales using satellite observations. — Mar. Ecol. Prog. Ser. 403: 13-27.
- ESRI 2015. ArcGIS. Environmental Systems Research Institute.
- Foreman, M. G. G. et al. 2008. Dynamic ocean topography for the Northeast Pacific and its continental margins. — Geophys. Res. Lett. 35: L22606.
- Gorman, D. et al. 2013. Modeling kelp forest distribution and biomass along temperate rocky coastlines. — Mar. Biol. 160: 309-325.
- Gregr, E. J. et al. 2016. A benthic habitat template for Pacific Canada's continental shelf. Canadian Technical Report of the Fisheries and Aquatic Sciences. p ix+33p.
- Gregr, E. J. et al. 2013. A spatial framework for representing nearshore ecosystems. — Prog. Oceanogr. 115: 189-201.
- Haggarty, D. R. 2015. An evaluation of the effectiveness of rockfish Conservation areas in British Columbia, Canada. Zoology. University of British Columbia, p xxii + 218.
- Lessard, J. et al. 2007. Recovery Potential Assessment for the northern abalone (*Haliotis kamtschatkana*) in Canada. Canadian Science Advisory Secretariat, p 101.
- Pedersen, M. F. et al. 2012. Effects of wave exposure on population structure, demography, biomass and productivity of the kelp *Laminaria hyperborea*. — Mar. Ecol. Prog. Ser. 451: 45-60.
- Roberts, J. J. et al. 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. — Environ. Model. Software 25: 1197-1207.
- Springer, Y. et al. 2007. Ecology and management of the bull kelp, *Nereocystis luetkeana*. Lenfest Ocean Program, p 48.

Table B1: Significance levels {p values: "****"< 0.001; "***" < 0.01; "*" < 0.05; "." < 0.1} for the univariate relationships between linear, quadratic, and cubic polynomial forms of the proposed predictor variables and the presence/absence of kelp canopy and its two component species giant and bull kelp. See text for variable descriptions.

Variable	Canopy			Giant kelp			Bull kelp		
	Linear	2nd order	3rd order	Linear	2nd order	3rd order	Linear	2nd order	3rd order
Depth	***	--	--	***	***	**	--	.	--
Slope	**	***	**	--	--	**	**	.	*
Aspect	--	--	--	**	--	--	*	--	--
RF	***	--	--	.	*	--	***	*	--
RMSM ¹	***	--	--	**	--	--	*	--	--
BoP ¹	***	--	--	*	--	--	.	--	--
Salt	--	--	--	--	***	*	--	*	--
BTemp	--	--	.	*	*	--	**	**	.
SST	--	--	--	*	--	--	**	--	--
SSTQ2	.	--	--	.	--	--	--	--	*
SSTQ3	--	.	--	**	*	--	--	--	--
SumSol	--	--	--	--	--	--	--	--	--
FetchTot	--	*	***	***	***	--	***	--	--
FetchSE	--	--	*	***	*	*	--	***	.
FetchNW	--	--	**	***	--	--	***	--	--
MaxTidal	**	--	--	--	**	--	**	**	--

1. Higher order values not valid for categorical variables.