

**RESEARCH ARTICLE**

# Comparison of spatial distribution models to predict subtidal burying habitat of the forage fish *Ammodytes personatus* in the Strait of Georgia, British Columbia, Canada

Clifford L.K. Robinson<sup>1</sup>  | Beatrice Proudfoot<sup>1</sup> | Christopher N. Rooper<sup>1</sup> | Douglas F. Bertram<sup>2</sup>

<sup>1</sup>Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada

<sup>2</sup>Institute of Ocean Sciences, Environment and Climate Change Canada, Wildlife Research Division, Sidney, British Columbia, Canada

**Correspondence**

Clifford L.K. Robinson, Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia V9T 6N7, Canada.  
Email: cliff.robinson@dfpo-mpo.gc.ca

**Funding information**

Fisheries and Oceans Canada, Grant/Award Number: Project: 91601 Strategic Program for Ecosystem-bas; Karen C. Drayer Wildlife Health Center, Grant/Award Number: SeaDoc Society Grant

**Abstract**

1. The Pacific sand lance (*Ammodytes personatus*) is a key forage species for many commercially important fish (e.g. salmon and groundfish), marine birds, and whales found in nearshore coastal waters of British Columbia, Canada.
2. Sand lance lack a swim bladder and have a requirement for low-silt, medium-coarse sandy sea-bed habitat for burying. Little information is available describing the distribution of burying habitat, partly because there are no commercial fisheries for *A. personatus* in British Columbia.
3. This information is required by habitat and wildlife managers to identify and protect uncommon patches of burying habitats from detrimental activities, including dredging, infilling, and oil spills.
4. In this study, habitat distribution results from five suitability modelling algorithms were evaluated: maximum entropy, generalized linear model, generalized additive model, random forest, and an ensemble model of the latter three.
5. The maximum entropy model had the highest performance score (area under the receiver operator characteristic curve was 0.78) and was selected as the model that most accurately identified the presence of suitable *A. personatus* burying habitat.
6. Model results indicate that suitable burying habitat is primarily influenced by derived sea-bed substrate, distance to estuary, distance to sand-gravel beaches, and bottom sea temperature.
7. Overall, the spatial modelling identified only 105 km<sup>2</sup> of highly suitable sand lance burying habitat, or 2.6% of the study area (0–150 m), primarily in Haro Strait, along the east coast of Vancouver Island, and in northern regions of the strait near Cortes, Savary, and Harwood islands.
8. Identification of this uncommon and patchy burying habitat will contribute to the ongoing conservation of an important coastal prey species.

**KEYWORDS**

*Ammodytes personatus*, burying habitat, ensemble modelling, forage fish, habitat suitability model, Salish Sea

## 1 | INTRODUCTION

The Pacific sand lance (*Ammodytes personatus*; sand lance) is an important forage fish for many species of marine birds (e.g. rhinoceros auklet, *Cerorhinca monocerata*), numerous coastal fish (e.g. salmon and groundfish), and baleen whales (e.g. humpback whale, *Megaptera novaeangliae*) found in the coastal waters of British Columbia, such as the Strait of Georgia, Haro Strait, Puget Sound, and Juan de Fuca Strait, collectively known as the Salish Sea. Compared with other small forage fish species, sand lance are unique in that they lack a swim bladder and rely upon sea-bed substrates to bury in when not foraging in the water column or to escape from vertebrate predators. Importantly, the life history of *A. personatus* also includes an extended period where fish remain buried in sea-bed sediments from late autumn until spawning in early winter while developing gonads. After spawning in mid-winter, the sand lance resume foraging excursions into the water column from spring to early autumn during daylight, returning to bury in sand patches at night (Robards et al., 1999; Haynes & Robinson, 2011).

*Ammodytes personatus* is not commercially fished on the Pacific coast of Canada, and hence there is little information available describing the potential locations of important subtidal burying habitats. This information is ultimately required to support the recovery of threatened marine species dependent on sand lance, such as the marbled murrelet (*Brachyramphus marmoratus*; Environment Canada, 2014). Moreover, information on the location of burying habitats is required in light of increased coastal development and shipping traffic in British Columbia and associated activities such as sea-bed dredging, infilling, and oiling. These activities would result in permanent changes in suitable sand lance burying habitat and might constitute loss of habitat under the Canadian Fisheries Act (R.S.C., 1985, c. F-14) Government of Canada, 1985. Overall, identifying factors that influence the distribution of sand lance burying habitat and mapping its location offers value for wildlife and fish habitat managers regulating marine activities, designing and establishing protected areas, and supporting other conservation and management activities.

The key attributes of the sea-bed substrate that make it a suitable habitat for sand lance burying include low silt content and a high percentage composition of medium-coarse sand (e.g., >70%). Earlier field work in coastal British Columbia validated these key sand lance burying habitat properties. For instance, Haynes, Robinson & Dearden (2008) evaluated the properties of sand patches where young sand lance buried themselves in intertidal and shallow subtidal habitats and found that patches consisted of coarse sands and shell hash. The importance of coarse-grained burying habitat was further emphasized by Haynes & Robinson (2011), in that sand lance were found to reuse the same sand patches during summer and appeared to remain within several kilometres of their subtidal burying habitats when foraging in the water column. Overall, shallow sea-bed burying habitat properties of sand lance in coastal British Columbia are very similar to the North Sea (e.g. Wright,

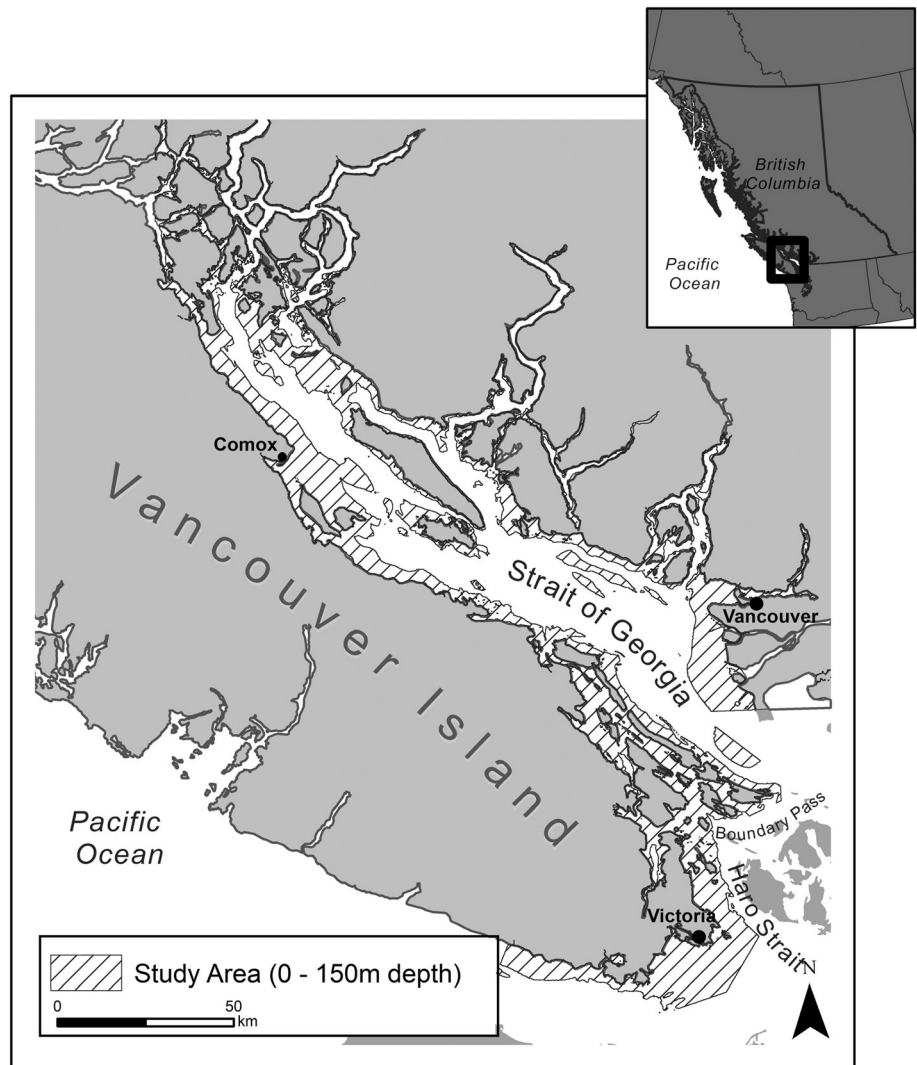
Jensen & Tuck, 2000; Holland et al., 2005; Greenstreet et al., 2010), southern Salish Sea (Bizzarro et al., 2016), and Alaska (Robards et al., 1999). It has been suggested that medium-coarse sands with relatively low silt content (<4%) due to specific oceanographic processes maintain relatively high surface dissolved oxygen, allowing *Ammodytes* to remain buried for extended non-foraging periods (Wright, Jensen & Tuck, 2000; Holland et al., 2005).

Habitat suitability models (HSMs) are valuable tools that have become increasingly important for supporting the conservation and management of species and their habitats. By exploring relationships between a biological entity of interest (e.g. community, biome, species, gene, habitat) and environmental variables, the presence of species and habitats can be predicted. Additionally, HSMs can be used to investigate environmental conditions that meet a given species' habitat requirements. Here, the more general term 'habitat suitability modelling' is used because the intent of the study is to predict the presence of suitable sand lance burying habitat in the Strait of Georgia rather than predict the distribution of sand lance. The choice of HSM method can considerably influence predictive outputs and performance of modelled outputs (Elith et al., 2006; Pearson et al., 2006). Additionally, selecting the best modelling approach can be challenging, as no single method is consistently superior in terms of predictive performance (Segurado & Araujo, 2004; Shabani, Kumar & Ahmadi, 2016). One approach that can help find consensus among the varying predictive outputs and performances of HSMs is ensemble modelling (Thuiller et al., 2009; Opper et al., 2012). Ensemble modelling involves combining predictive outputs from multiple different modelling algorithms into a single model (Araújo & New, 2007; Robinson et al., 2017). The objective of this study is to compare the predictive performance of individual and ensemble HSMs to identify the modelling approach that best represents the distribution of potential sand lance burying habitat in the Strait of Georgia. The results will ultimately be used to inform conservation and management decisions related to sand lance and their habitats.

## 2 | METHODS

The Strait of Georgia consists of three main oceanographic basins that are defined by a submarine sill in Boundary Pass and a second sill south of Victoria (Figure 1; Thomson, 1981). The region to the north of the Boundary Pass sill is referred to as the Strait of Georgia, and the region between the two sills is known as Haro Strait. Not considered in this study is the region seaward of the Victoria sill, namely Juan de Fuca Strait. The study area encompassed the Canadian portion of Haro Strait and the Strait of Georgia for depths ranging between chart datum and 150 m—depths that previous studies have indicated would most likely contain sand lance burying habitat (e.g. Jensen et al., 2011; Robinson et al., 2013). The shallow study area consisted of about 4,100 km<sup>2</sup> (Figure 1), which is approximately 60% of the sea surface area of the Strait of Georgia and Haro Strait.

**FIGURE 1** Map of the Strait of Georgia study area. Hatched lines indicate the extent of the Pacific sand lance burying habitat modelling (maximum depth 150 m)



## 2.1 | Species data

Information on the presence of sand lance in sea-bed sediments was derived from a 15 L Van Veen benthic grab sampler. Grab samples of sea-bed sediments and fish were collected from 500 shallow subtidal stations during late spring–early summer 2017. At each station, sediment from the grab sampler was placed in a 40 L tray, photographed, and a 250 ml representative subsample removed and frozen for later laboratory analysis (see later). Any sand lances caught in the Van Veen grabs were placed in plastic bags, labelled (e.g. date, time, location), and frozen for later laboratory analysis. Fish were collected under Fisheries and Oceans Canada Scientific Research Permit XR 2352017.

The 500 sediment grab samples collected by the field programme were analysed in a Parks Canada Agency laboratory in Vancouver, British Columbia, for grain size analysis. Sediment samples were first oven-dried at 60°C for 48 hr. Each sediment sample was then shaken over a series of graded, stacked sieves from largest to smallest mesh size (4.0, 2.0, 1.0, 0.5, 0.125, 0.063, and <0.063 mm) using a Retsch AS200 vibratory sieve shaker. After shaking, sediment fractions were then removed from each sieve and weighed to the nearest 0.01 g.

Percentages of the fractions of grain size diameters in each sample were calculated and reported according to the Udden–Wentworth scale of sediment/rock size classification: silt (<0.0625 mm), very fine sand (0.0625–0.125 mm), fine sand (0.125–0.25 mm), silts and fine sands ( $\leq 0.25$  mm), medium sand (0.25–0.5 mm), coarse sand (0.5–1.0 mm), very coarse sand (1.0–2.0 mm), medium-coarse sands (0.25–1.0 mm), and very fine gravel (>2.00 mm).

Additional grain size data from 177 grab samples were obtained from previous studies conducted within the Strait of Georgia. Sixty-nine subtidal grab samples were obtained from a cooperative student project in the southern Strait of Georgia in 2010 in an examination of sand lance shallow subtidal spawning habitats (Robinson CLK, unpublished data, 2021), while 108 grab samples with grain size data were assembled from Environment and Climate Change Canada for a project in the southern Gulf Islands (Bertram D, Environment and Climate Change Canada, unpublished data, 2021).

To define whether a grab sample might contain sand with suitable sand lance burying properties, the sediment properties from grab samples containing both sand lance and grain size data were evaluated. The grain size properties associated with sand lance presence

were then used to generate criteria that were applied to the remaining grab sediment samples (not containing fish) and hence classify the sediment as to its suitability as burying habitat. The importance of other environmental factors in determining sand lance burying habitat were to be identified during the habitat suitability modelling process.

## 2.2 | Environmental variables

A suite of environmental variables (Table 1) that may potentially influence the presence of buried sand lance was obtained at 100 m × 100 m resolution from chart datum to −150 m depth. Additional bathymetric derivatives (slope, bathymetric position index, vector ruggedness) were calculated using the Benthic Terrain Modeler toolbox (Walbridge et al., 2018) in ArcGIS. A substrate layer was developed by combining existing Fisheries and Oceans Canada grain size (Gregg et al., 2016) and bottom patch (Gregg, Lessard & Harper, 2013) models to produce a substrate classification composed of nine classes. However, owing to the sample size constraints (i.e. not all sediment substrate classes contained sufficient observations), the nine classes were reclassified into four classes ranging from high to low suitability for sand lance. The reclassification scheme (Table 2) was based on published literature of sand lance habitat suitability (Haynes & Robinson, 2011). The sediment composition of the four reclassified classes ranged from hard, consolidated substrates (low suitability) to soft/sandy substrates (high suitability) with the exception of silt, which was reclassified as low suitability because sand lances require a low silt content to remain buried and able to respire during extended non-foraging periods.

To produce a higher resolution 50 m × 50 m data set that aligned with the bathymetric and bathymetric derivative data, the oceanographic variables were downscaled using ArcGIS Spatial Analyst Tools. Though downscaling the oceanographic variables does not provide any additional information, given that pixel values remain the same

after downscaling, general patterns remain and the species observations can be aggregated to the scale of the higher resolution geomorphic variables.

## 2.3 | Data preparation

Species observations were aggregated to the resolution of the 50 m × 50 m environmental data. If more than one grab observation fell within a single raster cell, the cell was assigned a presence if at least one observation was present (Guinotte & Davies, 2014; Nephin et al., 2020). This step standardizes the observations and predictions to the same spatial resolution. However, aggregating presence observations within a large cell grid can often lead to a reduction in sample size, particularly if observations are in close proximity to each other. After aggregating, 402 sample points, 225 with suitable substrate (HS) and 177 with no suitable substrate

**TABLE 2** Reclassification of the derived sea-bed substrate layer

Original classification <sup>a</sup>	Reclassification
Hard	1
Bedrock dominant	1
Boulder dominant	1
Mixed hard and soft	2
Soft surface with patchy distribution of boulder and cobble	2
Soft surface overlaying hard substrate	2
Soft	3
Sand/shell	4
Silt/mud	1

<sup>a</sup>Refer to Gregg, Lessard & Harper (2013) for details of classification.

**TABLE 1** Environmental variables used to model Pacific sand lance burying habitat in the Strait of Georgia

Variable	Units	Native resolution	Reference (calculation tool)
Bathymetry	m	50 m × 50 m	DFO Pacific Region
Medium BPI <sup>a</sup>	—	50 m × 50 m	Bathymetric derivative (BTM Toolbox)
Rugosity	—	50 m × 50 m	Bathymetric derivative (BTM Toolbox)
Slope	degrees	50 m × 50 m	Bathymetric derivative (BTM Toolbox)
Bottom salinity max (1998–2007)	PSU	3 km × 3 km	Masson & Fine (2012)
Bottom temperature max (1998–2007)	Celsius	3 km × 3 km	Masson & Fine (2012)
Tidal current	cm/s	Variable	Foreman et al. (2008)
Distance to estuaries	km	100 m × 100 m	DFO Pacific Region
Distance to sand-gravel source	km	100 m × 100 m	DFO Pacific Region
Distance to terrestrial sand	km	100 m × 100 m	DFO Pacific Region
Derived sea-bed substrate	—	—	Gregg, Lessard & Harper (2013); Gregg et al. (2016)

Abbreviation: BPI, benthic position index; PSU, practical salinity unit; DFO, Fisheries and Oceans Canada.

<sup>a</sup>Inner radius of 200m, outer radius of 2,000m.

(NS) were used to model the presence of suitable sand lance burying habitat. After aggregating, 102 presence points were used as an additional measure of performance by determining the proportion of sand lance presence observations that were correctly classified when binary presence/absence habitat maps were produced for each model (refer to Section 2.4.7).

### 2.3.1 | Cross-validation

Data were partitioned into training and testing data sets using spatial blocking fivefold cross-validation (CV). Spatial blocking CV is considered best practice when partitioning data for model fitting and testing because spatial blocking improves the spatial independence of the training and testing data, resulting in more accurate estimates of model performance (Roberts et al., 2017; Fourcade, Besnard & Secondi, 2018; Nephin et al., 2020). The range of spatial autocorrelation in the environmental predictor data was used to determine the optimal block size. This was done using the blockCV R package (Valavi et al., 2019). The approach involves automatically fitting variogram models to each continuous predictor layer and finding the effective range of spatial autocorrelation; the median range of spatial autocorrelation of the predictors defines the block size. An even dispersal of presence and absence observations across folds was obtained by iteratively ( $n = 2,000$ ) and randomly assigning blocks to folds and selecting the configuration with the most even distribution of HS and NS points. Model fitting and evaluation were performed for each CV run. Individual CV runs were then combined to build the final models.

### 2.3.2 | Variable selection

Collinearity among predictor variables was assessed using the variance inflation factor (Zuur et al., 2009) and Pearson's correlation. Maximum salinity was removed, as it was highly correlated ( $>0.70$ ) with maximum bottom temperature, while slope was also removed due to a high VIF ( $> 10$ ), leaving nine environmental variables for modelling (Table 1).

## 2.4 | Modelling

Ensemble modelling is a technique that involves combining predictive outputs from multiple different modelling algorithms and is one option for incorporating the varying predictions and performance of multiple HSMs (Araújo & New, 2007; Robinson et al., 2017). For this study, three individual modelling algorithms were combined into an ensemble model to predict the presence of burying habitat in the Strait of Georgia: a generalized linear model (GLM), a generalized additive model (GAM), and a random forest (RF) model. Ensemble modelling was completed with the Biomod2 R package version 3.4.6 (Thuiller et al., 2016). A MaxEnt (Phillips, Anderson & Schapire, 2006) model

was also produced and compared with the individual and ensemble models, and the model with superior performance judged by area under the receiver operator characteristic curve (AUC) and true skill statistic (TSS) was selected as the model that best represents the location of burying habitat in the study area.

### 2.4.1 | Generalized linear model

GLMs are flexible and widely used distribution modelling methods used to relate species observations to environmental variables by calculating species occurrence as parametric functions of environmental predictors (McCullagh & Nelder, 2019). In this study, five GLMs (one for each CV run) were developed to predict the presence of suitable sand lance burying habitat in the Strait of Georgia. A binomial error distribution and logit link function were used (Guisan, Thuiller & Zimmermann, 2017), and variable selection was done by ranking models based on the Bayesian information criterion. A GLM prediction surface representing the probability of occurrence of sand lance burying habitat was derived by calculating the mean prediction from the five CV runs.

### 2.4.2 | Generalized additive model

GAMs are an extension of GLMs that are not restricted to linear relationships between species occurrences and environmental predictors (Hastie & Tibshirani, 1990). GAMs are flexible and utilize smoothing functions of predictor variables, allowing for non-linear relationships between the environmental covariates and the presence of sand lance burying habitat. The following default Biomod2 settings were used for model fitting: thin-plate regression spline smoothing function (which is more flexible than the cubic spline smoothing function), binomial error distribution, and logit link function. Variable selection was based on generalized CV by attributing low weights to the less important variables. A GAM prediction surface representing the probability of occurrence of sand lance burying habitat was derived by calculating the mean prediction from the five CV runs.

### 2.4.3 | Random forest model

RF (Breiman, 2001) is a machine-learning classification and regression-tree-based approach to species distribution modelling that involves creating multiple decision trees and combining the outputs. Default Biomod2 settings were applied: The square root of the number of explanatory values was used to determine the number of explanatory variables selected at each split, the number of trees was set to 500, and node size was set to 5. An RF prediction surface representing the probability of occurrence of sand lance burying habitat was derived by calculating the mean prediction from the five CV models.

## 2.4.4 | Maximum entropy model

In contrast to the regression-based GLM, GAM, and RF models, the MaxEnt model (Phillips, Anderson & Schapire, 2006) uses the principle of maximum entropy to model the distribution of a species or biological entity of interest. MaxEnt models assume that the optimal approach for addressing remaining uncertainty in a distribution within known environmental constraints is to maximize entropy (Elith et al., 2011; Georgian, Anderson & Rowden, 2019). Instead of using a random selection of background points to define the available environmental conditions (the default MaxEnt settings), absence data were specified as the NS grab locations for model parameterization and testing in a 'targeted background' approach (Phillips & Dudík, 2008). The same HS and NS observations divided into five CV folds (training and test data) as used for the GLM, GAM, and RF were also used for the MaxEnt modelling. MaxEnt modelling was carried out using the *dismo* package in R (<https://cran.r-project.org/web/packages/dismo/dismo.pdf>). During initial data exploration, the regularization parameter and feature classes used in the MaxEnt modelling were varied systematically to choose the best settings for analysis and minimize overfitting of the training data (Merow, Smith & Silander, 2013; Morales, Fernandez & Baca-Gonzalez, 2017). Models were evaluated based on the joint rank of AUC calculated for the testing data partition, the difference between the AUC for the training data and the testing data, and the Akaike information criterion calculated for small sample sizes (Rooper et al., 2019). A regularization parameter of 3.9 with exclusion of the product feature class was found to produce the best average rank. A final MaxEnt prediction surface representing the probability of occurrence of sand lance burying habitat was derived by calculating the mean prediction from the five CV models. It is important to note that the MaxEnt model did not predict the presence or absence of suitable habitat, but only those areas where habitats were relatively more or less suitable within the background. So, a probability of 0.8 in the MaxEnt output does not infer a probability of suitable habitat of 80%, but only that relative to the overall conditions there is an 80% chance that the habitat is better than an average background point.

## 2.4.5 | Model performance

Models derived from the GLM, GAM, RF, and MaxEnt CV folds were evaluated using the mean and *SD* of the following performance metrics (definitions obtained from Nephin et al. (2020)):

- **AUC**—A common metric for estimating the performance of HSMs. AUC values of >0.5 are estimated to be better than chance, values >0.7 are considered acceptable, and values between or higher than 0.8 and 0.9 are considered excellent (Pearce & Ferrier, 2000).
- **TSS**—Balances sensitivity and specificity and is independent of prevalence of observations. TSS ranges from -1 to +1, with values <0 indicating that models are no better than chance and values of 1 indicating perfect agreement (Allouche, Tsoar & Kadmon, 2006).

TSS values >0.6 are considered good, and values between 0.2 and 0.6 are considered moderate (Landis & Koch, 1977).

- **Accuracy**—The percentage of predictions that are correctly classified when the probability of presence of suitable habitat is converted to a binary presence-absence classification. The threshold probability was determined by maximizing model sensitivity and specificity (see Section 2.4.7 for details on how thresholds were calculated).

Model uncertainty was also illustrated using the multivariate environmental similarity index to quantify uncertainty associated with extrapolation (Elith, Kearny & Phillips, 2010; Guillaumot et al., 2020). The approach estimates areas where environmental conditions are outside the range of conditions used to build the models (i.e. observed in the species data). Extrapolated areas occur where at least one environmental descriptor value is outside the range of environmental data used in model calibration.

## 2.4.6 | Ensemble modelling

Ensemble model predictions were created by averaging the GLM, GAM, and RF model predictions for each CV run. One ensemble model was built for each CV run to ensure that all models could be accurately evaluated and compared. Predictions from individual CV runs were weighted based on their performance (AUC). The threshold for including models in the ensemble was set at 0.5 to ensure that a model performing worse or no better than random ( $AUC \leq 0.5$ ) does not contribute to the ensemble prediction (Nephin et al., 2020). The threshold was selected because any greater value (e.g. 0.7) would be arbitrary (Nephin et al., 2020). The mean prediction from the five ensemble models (one for each CV run) was used to produce the final ensemble output. The fivefold CV was used to compute a standard error for prediction at each raster grid cell on the map.

Because the MaxEnt algorithm models the density of environmental conditions used that are then back transformed to be interpreted as the probability of suitable conditions for substrate, rather than the presence or absence of suitable habitat (Guisan, Thuiller & Zimmermann, 2017), it is not appropriate to include the MaxEnt model within the ensemble. However, MaxEnt models are known to perform consistently well and can outperform other modelling approaches (Elith et al., 2006; Reiss et al., 2011), and hence including MaxEnt in the model comparisons is pertinent for determining the model algorithm that best represents sand lance burying habitat in the Strait of Georgia.

## 2.4.7 | Model comparison and selection

To determine the potential areas of sand lance burying habitat in the Strait of Georgia, a threshold probability of suitable habitat was generated for each model, and the proportion of sand lance presence observations from grab samples (Figure 2) that were correctly classified was calculated and used as an additional measure of performance. For the



MaxEnt model, a threshold probability was generated using the confusion matrix calculated by the PresenceAbsence package in R (Freeman & Moisen, 2008). For the GLM, GAM, RF, and ensemble models, the threshold was generated using the Find.Optim.Stat function in the Biomod2 R package (Thuiller et al., 2016). Both methods use a function to find the threshold at which a selected evaluation metric is optimized. In this study, a threshold that maximized TSS was determined by iterating through 1,000 fitted values to determine the optimal TSS score and the associated threshold cut-off for converting the continuous values to binary. The model with the highest average performance metrics was determined to be the best representation of sand lance burying habitat in the Strait of Georgia.

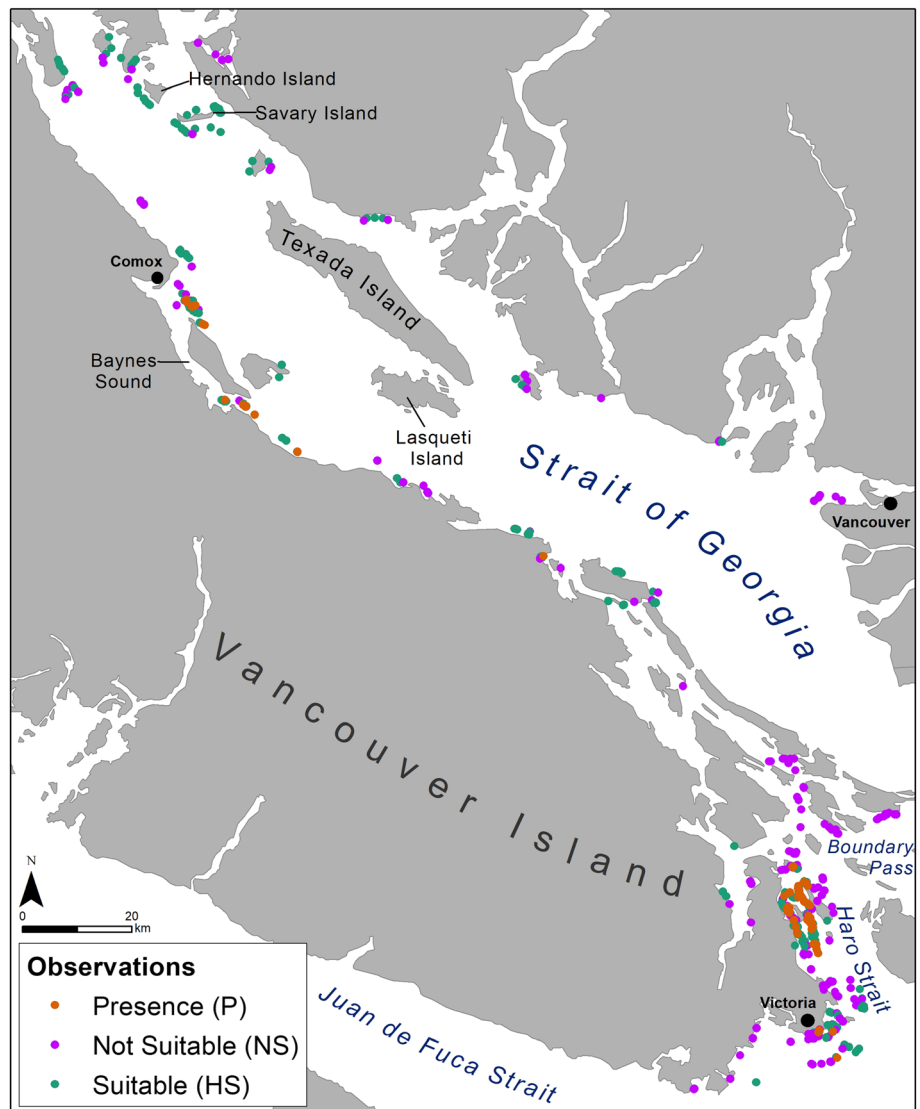
### 3 | RESULTS

#### 3.1 | Model performance

CV testing suggested that the GAM, MaxEnt model, and ensemble model performed reasonably well, with AUC scores of 0.78 (MaxEnt),

0.76 (GAM), and 0.75 (ensemble; Table 3). The GLM and RF model performed the poorest, with AUC scores of 0.73 and 0.63 respectively. TSS scores for all models were moderate to good, with the ensemble model having the highest TSS score (0.64; Table 3).

Using training data thresholds that optimize the balance of sensitivity and specificity, the MaxEnt, ensemble, and GAMs correctly predicted higher probability of suitable habitat or the presence of suitable habitat at places where sand lance were captured in 88% (threshold 0.54), 85% (threshold 0.57), and 86% (threshold 0.47) of occurrences respectively. Presence observations used to evaluate the models are generally located in areas of high density of HS/NS sampling points, suggesting that these are areas where the model is predicting sand lance habitat quite well, resulting in high accuracy scores when evaluating using the presence observations. Additionally, the high accuracy and associated high threshold values for the MaxEnt and ensemble confirm high probabilities of occurrence of suitable habitat associated with presence observations. Across all models, the results show that the presence of suitable burying habitat in the Strait of Georgia is strongly influenced by derived substrate, distance to estuary, distance to sand-gravel beach, and bottom temperature (Figure 3).



**FIGURE 2** Distribution of suitable and non-suitable burying habitat grabs used for model training and testing. Orange symbols indicate Pacifics and lance presence observations used for model evaluation

Response curves for all models (Figure 4) are similar, showing that the probability of the presence of sand lance burying habitat decreases as distance to estuaries and distance to sand-gravel beaches increase and as bottom temperature decreases. In contrast, the relationship between depth and the presence of suitable sand lance burying habitat is less clear, with lower probabilities observed at mid-depths and no apparent relationship in the RF model (Figure 4c). Not surprisingly, across all models, the probability of the presence of suitable burying habitat increases as substrate suitability increases.

**TABLE 3** Performance metrics for generalized linear model (GLM), generalized additive model (GAM), random forest (RF) model, ensemble model, and MaxEnt model when evaluated using fivefold cross-validation

Model	Cross-validation evaluation			
	AUC	TSS	Sensitivity	Specificity
GLM	0.73	0.47	0.75	0.75
GAM	0.76	0.52	0.76	0.76
RF	0.63	0.31	0.66	0.63
Ensemble	0.75	0.64	0.82	0.82
MaxEnt	0.78	0.45	0.76	0.69

Note: Higher AUC scores (closer to 1) indicate superior model performance.

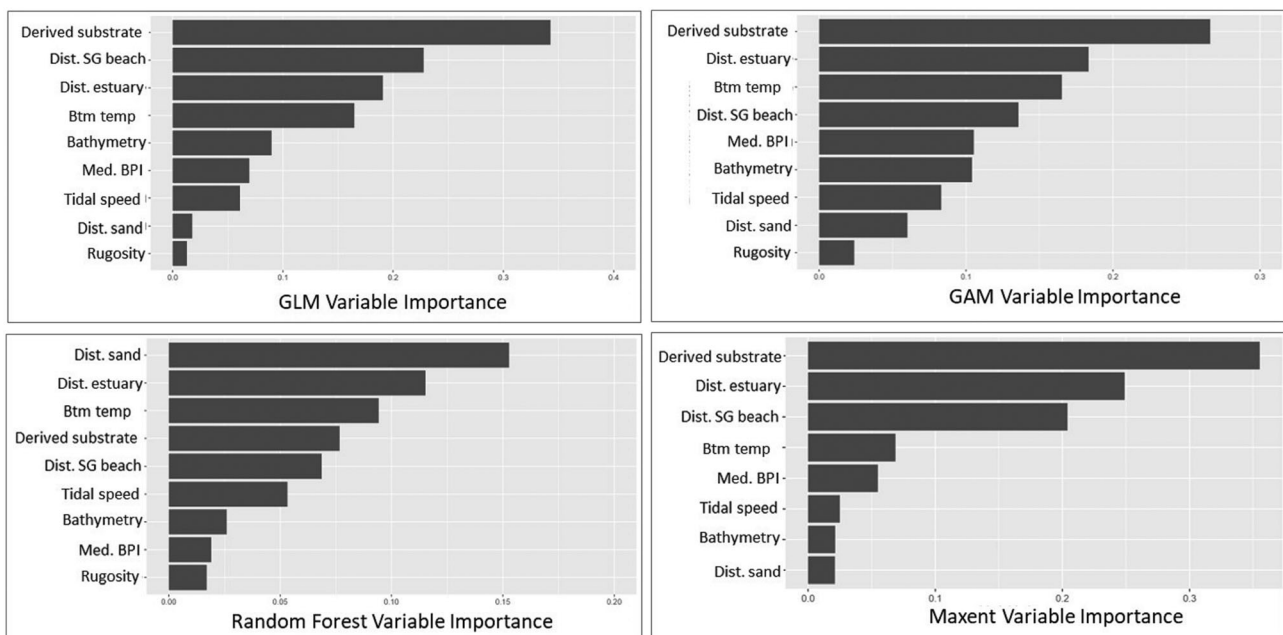
AUC, area under the receiver operator characteristic curve; TSS, true skill statistic

## 3.2 | Spatial distribution

Across all models, predictions were generally consistent with the data on presences and absences. All models produced similar patterns, predicting high probabilities of the presence of suitable sand lance burying habitat (>0.8) in Baynes Sound, Haro Strait, and near Savary, Hernando, and Marina islands (near the entrance to Desolation Sound; Figure 5). In general, low suitability is predicted in deeper regions of the study area (>70 m) with the exception of the GLM and GAM predictions east of Baynes Sound, which is reflected in the ensemble model built on the GLM, GAM, and RF model. Low suitability is also predicted at the Fraser River Delta and near Texada and Lasqueti islands (Figure 5).

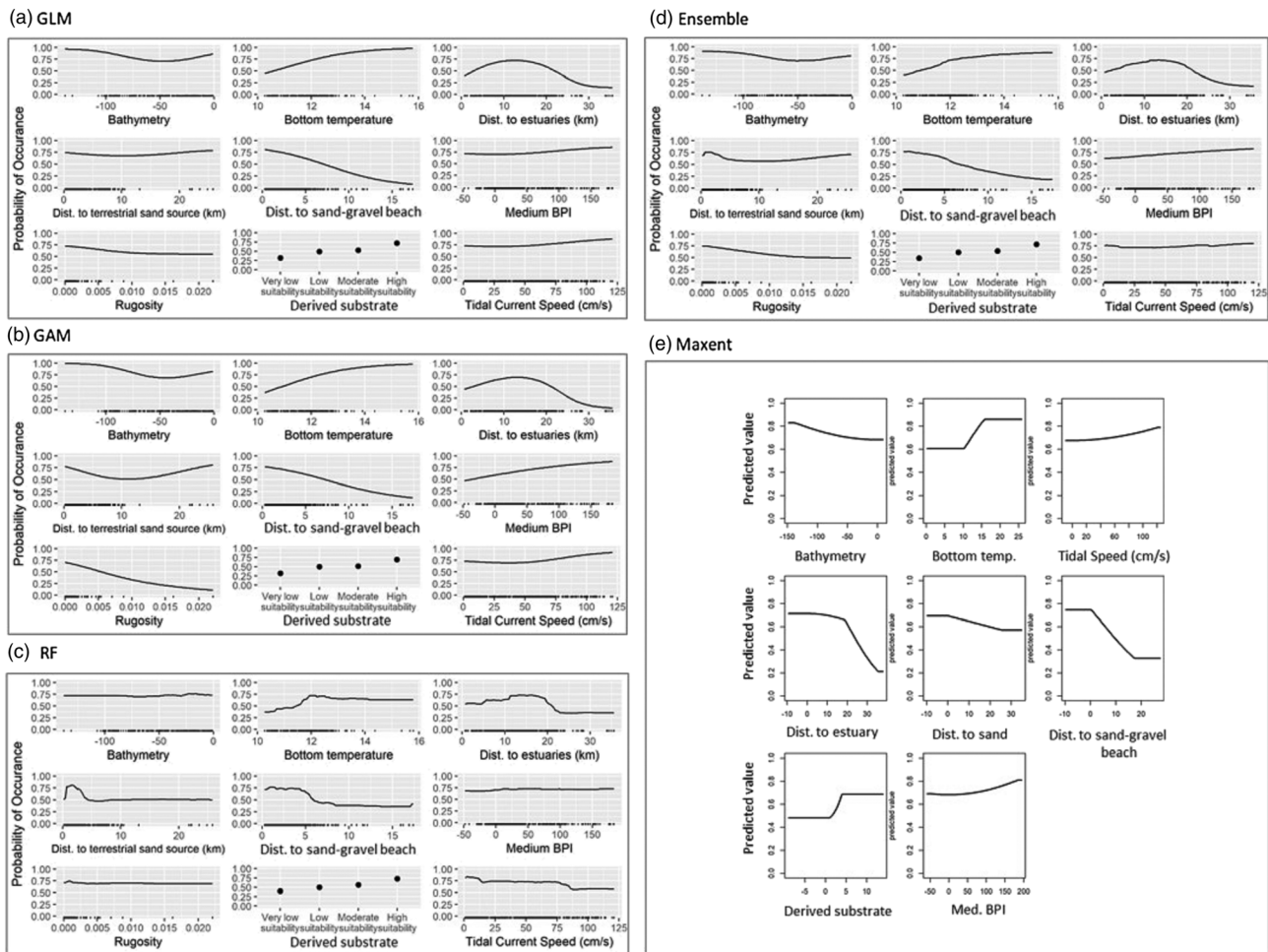
### 3.2.1 | Model selection

When both the CV and threshold/accuracy performance metrics are considered, the MaxEnt model performs consistently well, with an AUC score of 0.78 and an accuracy score of 0.88 when evaluated against independent sand lance presence data. As such, the MaxEnt model (Figure 6) was elected as the model that most accurately predicts the presence of suitable sand lance burying habitat in the Strait of Georgia. The ensemble and GAM models also likely represent reasonable predictions of the probability of presence of sand lance burying habitat in the Strait of Georgia, as both the ensemble model and GAM have higher sensitivity and specificity than the MaxEnt



**FIGURE 3** Variable importance plots for generalized linear (GLM), generalized additive (GAM), random forest (RF) and maximum entropy models predicting the distribution of suitable Pacific sand lance burying habitat in the Strait of Georgia. Variable importance is measured as the mean increase in accuracy based on the mean square error of predictions. A higher percentage increase indicates a higher variable importance. SG = sand-gravel; BPI = bathymetric position index





**FIGURE 4** Response Curves from the generalized linear model (GLM; a), generalized additive model (GAM; b), Random Forest (RF; c), Ensemble (d) and MaxEnt (e) models predicting Pacific sand lance burying habitat in the Strait of Georgia. Dist = distance. Note that rugosity in the Maxent was eliminated from further modelling

model. However, additional field validation is required to confirm the presence of suitable burying habitat (e.g. low silt, medium-coarse sand) in deeper regions of the study area as identified in the GAM and ensemble model.

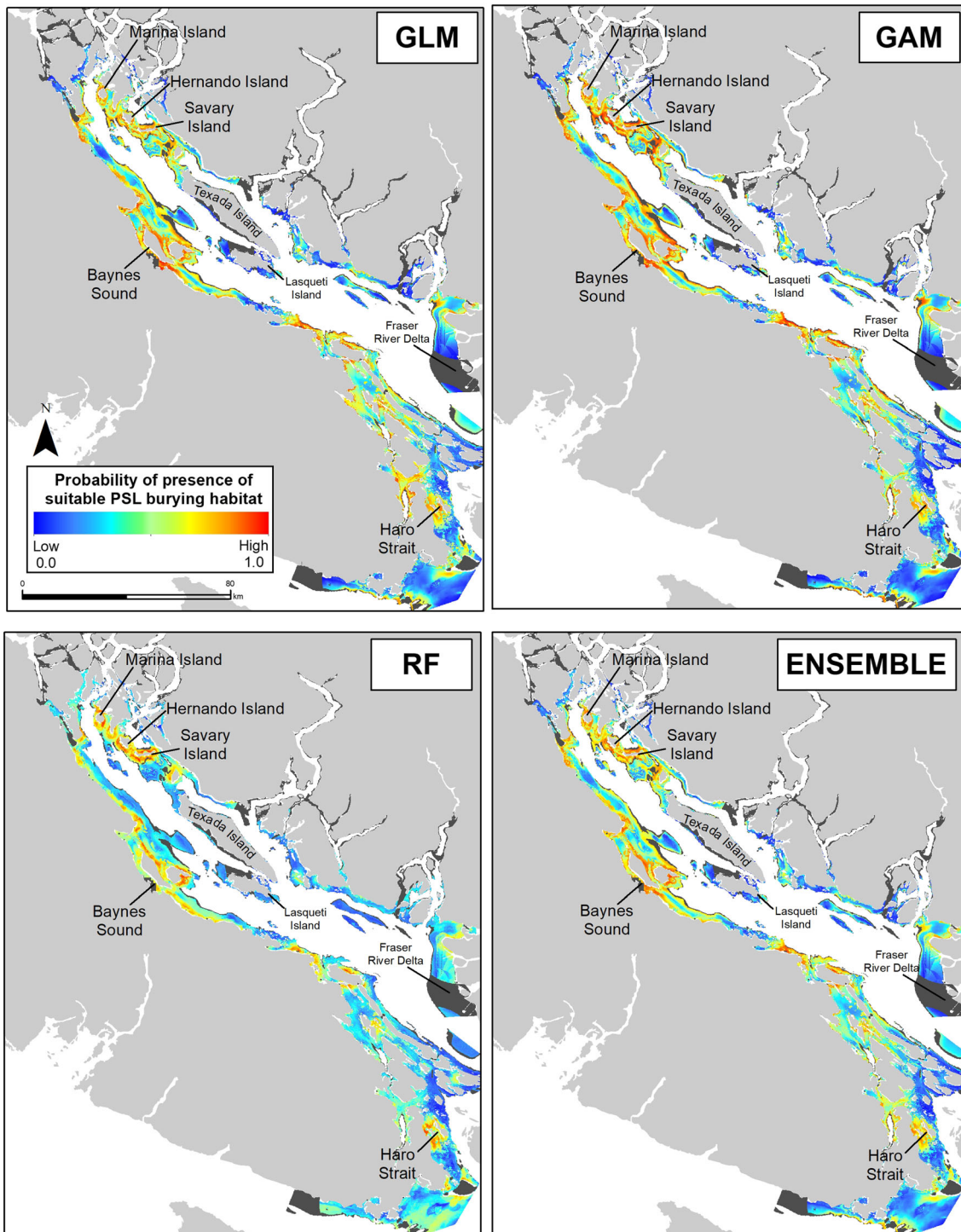
### 3.3 | Spatial uncertainty

Models generally exhibited similar patterns of spatial uncertainty (measured as the standard error of averaging predictions from CV folds) compared with the individual models they were built on. Areas with the highest uncertainty values include areas at greater depths and along the margins of areas predicted to be highly suitable. Uncertainty was generally low in areas where highly suitable sand lance burying habitat is predicted. Figure 6 illustrates the spatial distribution of the mean and standard deviation of the MaxEnt model. Extrapolated areas identified by the multivariate environmental similarity calculation are located throughout the study area (Figures 5 and 6),

particularly at the deep margin of the study area, in undersampled regions near Texada and Lasqueti islands and in the Fraser Delta.

## 4 | DISCUSSION

This study compared results from GLM, GAM, RF, MaxEnt, and an ensemble model to determine the modelling approach that might best predict the probability of presence of highly suitable sand lance burying habitat in the Strait of Georgia. The MaxEnt model was selected as the best model based on its accuracy in predicting an independent set of presence observations and on its higher AUC (0.78). The results indicate that suitable burying habitat is primarily influenced by derived sea-bed substrate, distance to estuary, distance to sand-gravel beaches. The low performance of the RF model is likely due to overfitting of the training (CV) data, which is fairly common in RF models (Rooper, Zimmermann & Prescott, 2017). Additionally, using spatial-CV blocking with a relatively small sample size is also a likely

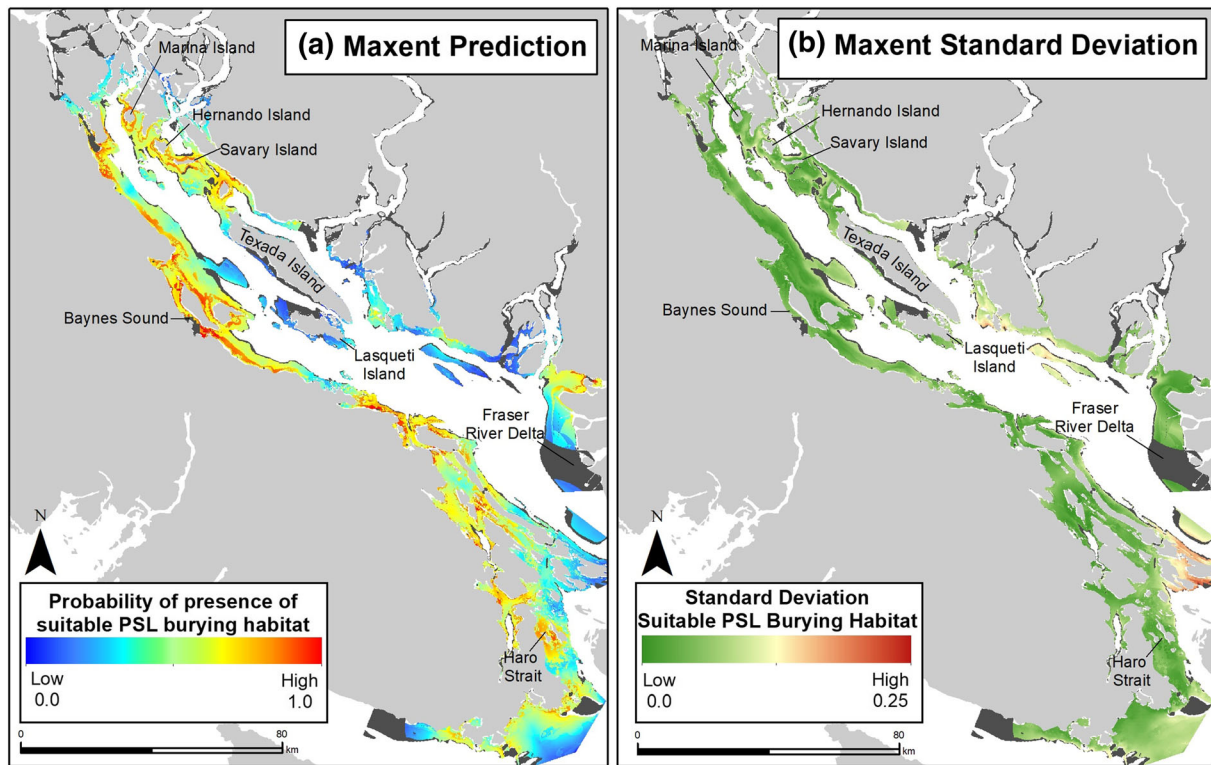


**FIGURE 5** Generalized linear (GLM), generalized additive (GAM) and random forest (RF) models showing the predicted probability of suitable Pacific sand lance (PSL) burying habitat in the Strait of Georgia. The ensemble model represents the mean probability calculated from the GLM, GAM and RF models for each cross validation fold. Warmer colours indicate a higher probability of suitable sand lance burying habitat. Extrapolation areas identified using the Multivariate Environmental Similarity Surface (MESS) approach are displayed in dark grey

contributor to the overfitting pattern observed in the RF model. The lower performance of the ensemble model compared with the MaxEnt model is also consistent with other studies, in which ensemble modelling has also been shown to perform no better than individual modelling approaches (Hao et al., 2019; Hao et al., 2020). The results

illustrate the importance of exploring multiple HSM algorithms to determine the model that is most suitable for a given study.

Across all models, the results indicate that the distribution and configuration of burying patches varied markedly between the southern Strait of Georgia (e.g. Haro Strait) and the central or



**FIGURE 6** (a) MaxEnt prediction and (b) standard deviation of the probability of occurrence of suitable Pacific sand lance burying habitat in the Strait of Georgia. Prediction and standard deviation surfaces derived from the five-fold cross validation models. Warmer colours indicate a higher probability of suitable sand lance burying habitat. Extrapolation areas identified using the Multivariate Environmental Similarity Surface (MESS) approach are displayed in dark grey

northern Strait of Georgia, likely because of their underlying oceanography and differences in geological processes. In the central/northern Strait of Georgia, there is generally a widespread along-shore distribution of narrow shallow sand lance burying habitat that is highly fragmented. In contrast, the majority of burying habitat predicted in the Haro Strait region is associated with large patches of deeper sea bed dominated by bathymetric restrictions that are enhanced by current flow and erosion of the underlying Quaternary sediments into subaqueous sand-to-fine gravel sand waves (Barrie et al., 2009). Another noticeable difference between regions is the smaller proportion of modelled sand lance burying habitat in deeper waters of the Strait of Georgia and adjoining inlets. It is expected, however, that given the widespread modelled low bottom-current speeds (<25 cm/s) for more than 75% of the Strait of Georgia, and the sediment starvation due to major rivers flowing into inlets (Hill et al., 2008), it is unlikely that many patches of suitable sand lance burying sediment will be confirmed in deeper, less exposed waters of the strait. Overall, the observed patchiness and wide-ranging patch size might be characteristic of *Ammodytes* spp. burying habitats in general. For example, in the North Sea, 217 elongated sandeel patches ranged in size from 1 km<sup>2</sup> to 4,023 km<sup>2</sup>, and they were broadly distributed over 5% (33,566 km<sup>2</sup>) of the North Sea and Skagerrak—see Fig. 1 in Jensen et al. (2011). Models indicate similar patchiness in sand lance habitat that will make it challenging for Salish Sea resource managers,

and this highlights the importance of adequate field sampling and model validation before potentially detrimental activities to sea-bed sediments (e.g. dredging or dumping) are allowed to proceed.

The approach used to identify burying habitat in this model study focused on physical environmental factors that are most likely stable over time, because it is expected that suitable sand lance burying habitat is spatially persistent. Barrie et al. (2009) indicated that from more than 6 years of repetitive multi-beam measurements there was no clear net directional movement or migration of the subaqueous dunes in Boundary Pass. The dune field maintained its overall geometry and volume, and it was suggested that sediments were derived from in situ erosion of the underlying sea bed and that sediments were trapped by a tidally oscillatory system. More recently, Greene, Cacchione & Hampton (2017) also found no evidence of net sediment transport in their sand bed fields that were composed of glacially derived sediment being winnowed and well oxygenated by strong tidal currents.

Though relatively stable sea bed conditions may exist in the deeper tidally driven Haro Strait, the nearshore environments in the central and northern Strait of Georgia might be more influenced by present-day sediment source and wind exposure. For instance, the two main sources of sediment supplying nearshore areas are riverine-derived bed-load transport and mass wasting from terrestrial sand bluffs, and ultimately redistribution via longshore wind-induced transport (Parks, Shaffer & Barry, 2013). Furthermore, well-known seasonal



beach dynamics on the British Columbia coast include the movement of shallow nearshore sands to deeper offshore bars parallel to the shoreline during winter storms, with spring and summer onshore sediment movement (Thomson, 1981). Hence, some portion of the burying habitats available to sand lance in the shallow alongshore subtidal strip in the northern Strait of Georgia in winter might not be available in spring–summer, and vice versa. Seasonal and interannual nearshore sediment dynamics have not been considered with respect to sand lance shallow nearshore burying habitat, and hence should be investigated more fully.

#### 4.1 | Study limitations

This study focused on modelling methods for identifying sand lance subtidal burying habitat. As with any modelling approach, there are several limitations to keep in mind. First, the mainland inlets were excluded from the modelling because the oceanographic and sea-bed properties of the inlets are very dissimilar to the majority of the Strait of Georgia (Thomson, 1981). Furthermore, there are major differences between Haro Strait sea-bed properties (e.g. relic glacier sediments) and more modern-day fluvial and longshore transport processes in the central and northern Strait of Georgia. This wide range of environmental conditions makes it challenging for any one spatial model to identify a specific suite of factors that can accurately represent the distribution of sand lance burying habitat. As with any spatial model, the resolution of grid cells (e.g. 50 m × 50 m) may present a mismatch in spatial resolution with the actual field patch size of sand lance burying habitat. Model results highlight for habitat managers general areas in the Strait of Georgia where potential burying habitat likely exists and where additional field sampling, and ultimately conservation initiatives, should be undertaken.

In terms of limitations of the models, a broader distribution of observations across predictor space could increase performance of the models. In particular, a broader distribution of observations would permit the use of an existing substrate data set that contains nine classes. This could produce predictions that incorporate a higher degree of substrate variability in the region as opposed to predictions that utilize the simplified substrate data set (four classes) used in this study. A larger sample size with broader coverage across geographic space (particularly for buried sand lance presence observations used for external model evaluation) could also increase the performance scores and better predict the probability of buried sand lance in more northern regions of the Strait of Georgia. Suitable sand lance burying habitat is predicted in northern parts of the study area (near Savary, Hernando, and Marina islands); however, there are no observations of sand lance present in grabs in the region (Figure 2); most observations are from fishery bycatch and predator diet analyses. Additionally, the presence observations used in model evaluation are not independent from the training data used to build the models, which could lead to overly optimistic accuracy metrics. Best practices indicate that model evaluation

should be done with independently collected, well-sampled data (Araújo et al., 2019); however, an independent data set was not available.

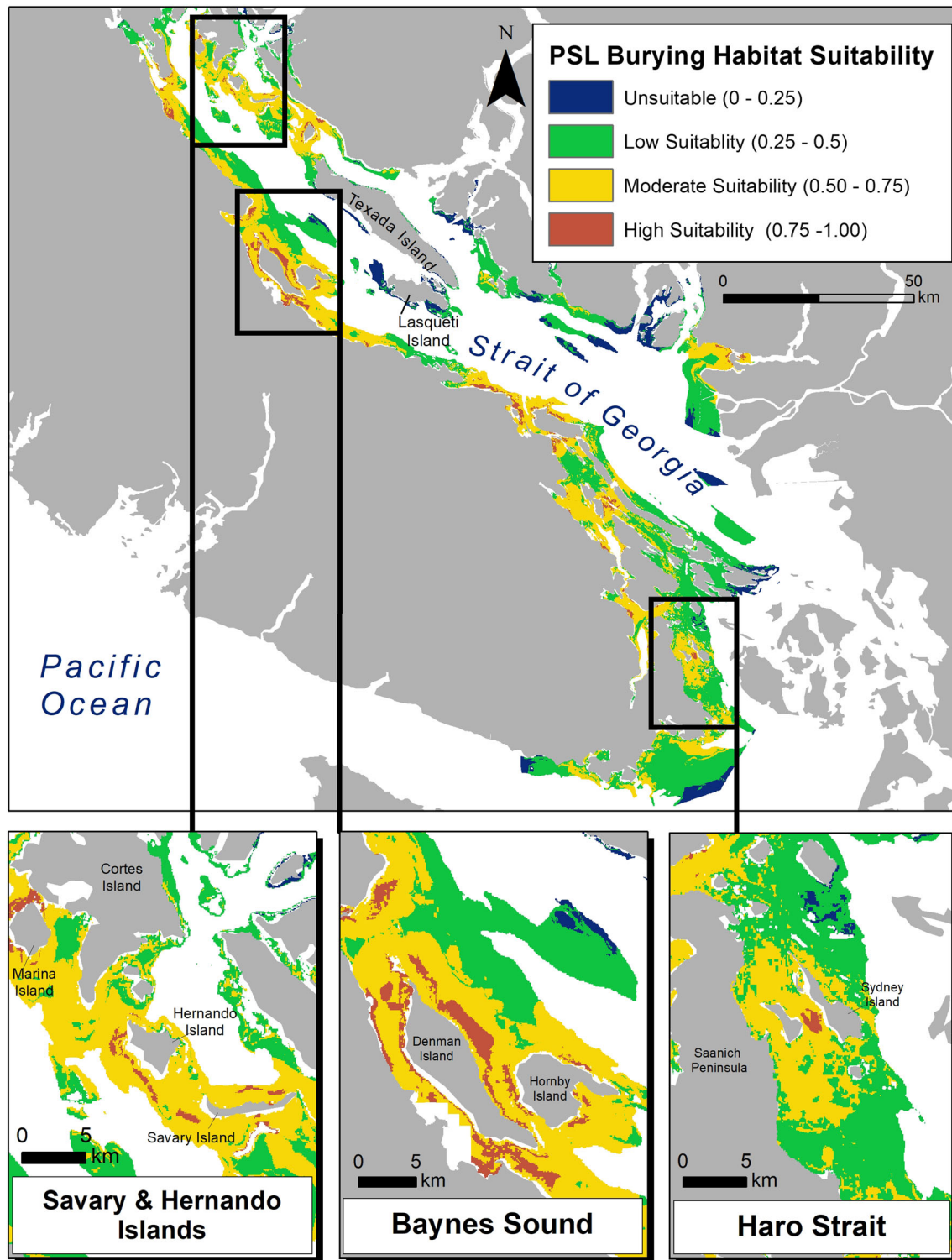
Examining the variable importance plots (Figure 3) and response curves (Figure 4) and the structures of the relationships between the model outputs and environmental predictors can also provide insights into the performance of the models. Substrate, distance to estuaries, distance to sand-gravel beaches, and bottom temperature were important variables. In contrast, medium bathymetric position index bathymetry and rugosity were both moderate to low contributors in all models and exhibited more variable structures in the response plots, suggesting that the relationship between these variables and the presence of sand lance burying habitat is less clear. Proximity to suitable spawning beaches may also be an important variable for predicting the presence of buried sand lance in the nearshore. Future sampling efforts could target subtidal areas in close proximity to known spawning beaches.

#### 4.2 | Burying habitat conservation and management recommendations

This study provides an improved quantitative estimate of sand lance burying habitat and its distribution in the Strait of Georgia and Haro Strait, British Columbia, compared with a study by Robinson et al. (2013) that relied mainly upon qualitative ecological information. The best available sea-bed bottom-type data and a relatively large grab data set used in this study helped model sand lance burying habitat distribution. From the results of the MaxEnt modelling, it is estimated that about 105 km<sup>2</sup> of sea bed in the Strait of Georgia and Haro Strait study area (2.6%) has a high probability (>0.75) of containing suitable sand lance burying habitat (Figure 7), compared with estimates of 6% by Robinson et al. (2013) using a simple expert-driven habitat suitability index approach. It is highly recommended that wildlife and habitat managers use the model maps in combination with enhanced field sampling to reduce model uncertainty and to increase confidence in identifying discrete nearshore subtidal burying patches, and ultimately to increase the conservation of uncommon burying habitat from harmful anthropogenic activities, such as dredging, infilling, dumping, or oiling.

It is also recommended that managers incorporate model results to inform evaluations of sand lance population size fluctuations, which might be an important component of the recovery of marbled murrelet and Chinook salmon populations, among other species. For example, sand lance found in modelled subtidal burying habitats in Haro Strait should be assessed and monitored over time to better understand how sand lance populations fluctuate in response to ocean warming.

Finally, a key recommendation from this research is that the modelling approach described should be incorporated into a decision-making framework and its uncertainty addressed with additional field sampling and then applied to the remainder of the British Columbia



**FIGURE 7** Reclassified MaxEnt model output into quartiles showing the patchy distribution of highly suitable Pacific sand lance (PSL) burying habitat in the Strait of Georgia. Inset maps illustrate the sub-areas where patches of predicted suitable burying habitat are concentrated. Extrapolation areas are removed

coast. Hence, the modelling approach should be used to better support management and conservation strategies aimed at reducing anthropogenic impacts to subtidal burying habitats of a key coastal forage species on the Pacific coast of Canada.

#### ACKNOWLEDGEMENTS

The field component of this project, which formed the basis of the spatial modelling, was critical to project success and was funded by the SeaDoc Society through the Karen C. Drayer Wildlife Health

Center, School of Veterinary Medicine, University of California, Davis, USA. Additional financial support for field sampling was received from the Canadian Wildlife Service (K. Woo). Captain Bruce Evans provided extensive logistical support and advice during field grab sampling in the Strait of Georgia aboard the FV *Misty Lady*. Project technical support was provided by L. McCartney, W. Peters, and H. Shannon. A Fisheries and Oceans Canada Strategic Program for Ecosystem-based Research and Advice to CLKR provided critical financial support for conducting the spatial modelling and analyses.

## ORCID

Clifford L.K. Robinson  <https://orcid.org/0000-0001-9199-2233>

## REFERENCES

- Allouche, O., Tsoar, A. & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R. et al. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M.B. & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Barrie, J.V., Conway, K.W., Picard, K. & Greene, H.G. (2009). Large-scale sedimentary bedforms and sediment dynamics on a glaciated tectonic continental shelf: Examples from the Pacific margin of Canada. *Continental Shelf Research*, 29(5–6), 796–806. <https://doi.org/10.1016/j.csr.2008.12.007>
- Bizzarro, J.J., Peterson, A.N., Blaine, J.M., Balaban, J.P., Greene, H.G. & Summers, A.P. (2016). Burrowing behavior, habitat, and functional morphology of the Pacific sand lance (*Ammodytes personatus*). *Fishery Bulletin*, 114(4), 445–460. <https://doi.org/10.7755/FB.114.4.7>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A. et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2), 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith, J., Kearney, M. & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Environment Canada. (2014). *Recovery strategy for the marbled murrelet (Brachyramphus marmoratus) in Canada*. Species at Risk Recovery Strategies Series. Environment Canada: Ottawa, ON, Canada. Available at: [https://sararegistry.gc.ca/virtual\\_sara/files/plans/rs\\_guillemot\\_marbre\\_marbled\\_murrelet\\_0614\\_e.pdf](https://sararegistry.gc.ca/virtual_sara/files/plans/rs_guillemot_marbre_marbled_murrelet_0614_e.pdf)
- Foreman, M., Crawford, W., Cherniawsky, J. & Galbraith, J. (2008). Dynamic ocean topography for the northeast Pacific and its continental margins. *Geophysical Research Letters*, 35(22), L22606. <https://doi.org/10.1029/2008GL035152>
- Fourcade, Y., Besnard, A.G. & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2), 245–256. <https://doi.org/10.1111/geb.12684>
- Freeman, E.A. & Moisen, G. (2008). PresenceAbsence: An R package for presence absence analysis. *Journal of Statistical Software*, 23(11), 1–31. <http://www.jstatsoft.org/v23/i11>
- Georgian, S.E., Anderson, O.F. & Rowden, A.A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries Research*, 211, 256–274. <https://doi.org/10.1016/j.fishres.2018.11.020>
- Government of Canada. (1985). Fisheries Act (R.S.C., 1985, c. F-14). Available at: <https://laws-lois.justice.gc.ca/eng/acts/F-14/>
- Greene, H.G., Cacchione, D.A. & Hampton, M.A. (2017). Characteristics and dynamics of a large sub-tidal sand wave field—Habitat for Pacific sand lance (*Ammodytes personatus*), Salish Sea, Washington, USA. *Geosciences*, 7(4), 107. <https://doi.org/10.3390/geosciences7040107>
- Greenstreet, S.P., Holland, G.J., Guirey, E.J., Armstrong, E., Fraser, H.M. & Gibb, I.M. (2010). Combining hydroacoustic seabed survey and grab sampling techniques to assess “local” sandeel population abundance. *ICES Journal of Marine Science*, 67(5), 971–984. <https://doi.org/10.1093/icesjms/fsp292>
- Gregg, E.J., Gryba, R., Li, M.Z., Alidina, H., Kostylev, V. & Hannah, C.G. (2016). *A benthic habitat template for Pacific Canada's continental shelf*. Canadian Technical Report of Hydrography and Oceanography Sciences, 312. Sidney, BC, Canada: Fisheries and Oceans Canada.
- Gregg, E.J., Lessard, J. & Harper, J. (2013). A spatial framework for representing nearshore ecosystems. *Progress in Oceanography*, 115, 189–201. <https://doi.org/10.1016/j.pocean.2013.05.028>
- Guillaumot, C., Moreau, C., Danis, B. & Saucède, T. (2020). Extrapolation in species distribution modelling. Application to Southern Ocean marine species. *Progress in Oceanography*, 188, 102438. <https://doi.org/10.1016/j.ocean.2020.102438>
- Guinotte, J.M. & Davies, A.J. (2014). Predicted deep-sea coral habitat suitability for the US West Coast. *PLoS ONE*, 9(4), e93918. <https://doi.org/10.1371/journal.pone.0093918>
- Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge, UK: Cambridge University Press.
- Hao, T., Elith, J., Guillera-Aroita, G. & Lahoz-Monfort, J.J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25(5), 839–852. <https://doi.org/10.1111/ddi.12892>
- Hao, T., Elith, J., Lahoz-Monfort, J.J. & Guillera-Aroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43(4), 549–558. <https://doi.org/10.1111/ecog.04890>
- Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. London, UK: Chapman & Hall, p. 335.
- Haynes, T.B. & Robinson, C.L.K. (2011). Re-use of shallow sediment patches by Pacific sand lance (*Ammodytes hexapterus*) in Barkley Sound, British Columbia, Canada. *Environmental Biology of Fishes*, 92(1), 1–12. <https://doi.org/10.1007/s10641-011-9809-z>
- Haynes, T.B., Robinson, C.L.K. & Dearden, P. (2008). Modelling nearshore intertidal habitat use of young-of-the-year Pacific sand lance (*Ammodytes hexapterus*) in Barkley Sound, British Columbia, Canada. *Environmental Biology of Fishes*, 83(4), 473–484. <https://doi.org/10.1007/s10641-008-9374-2>
- Hill, P.R., Conway, K., Lintern, D.G., Meulé, S., Picard, K. & Barrie, J.V. (2008). Sedimentary processes and sediment dispersal in the southern Strait of Georgia, BC, Canada. *Marine Environmental Research*, 66(Supplement), S39–S48. <https://doi.org/10.1016/j.marenvres.2008.09.003>
- Holland, G.J., Greenstreet, S.P., Gibb, I.M., Fraser, H.M. & Robertson, M.R. (2005). Identifying sandeel *Ammodytes marinus* sediment habitat preferences in the marine environment. *Marine Ecology Progress Series*, 303, 269–282. <https://doi.org/10.3354/meps303269>



- Jensen, H., Rindorf, A., Wright, P.J. & Mosegaard, H. (2011). Inferring the location and scale of mixing between habitat areas of lesser sandeel through information from the fishery. *ICES Journal of Marine Science*, 68(1), 43–51. <https://doi.org/10.1093/icesjms/fsq154>
- Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Masson, D. & Fine, I. (2012). Modeling seasonal to interannual ocean variability of coastal British Columbia. *Journal of Geophysical Research: Oceans*, 117(C10), C10019. <https://doi.org/10.1029/2012JC008151>
- McCullagh, P. & Nelder, J. A. (2019). *Generalized linear models*. London UK: Routledge.
- Merow, C., Smith, M.J. & Silander, J.A., Jr. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, 36(10), 1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Morales, N.S., Fernández, I.C. & Baca-González, V. (2017). MaxEnt's parameter configuration and small samples: Are we paying attention to recommendations? A systematic review. *PeerJ*, 5, e3093. <https://doi.org/10.7717/peerj.3093>
- Nephtin, J., Gregr, E.J., St. Germain, C., Fields, C. & Finney, J.L. (2020). Development of a species distribution modelling framework and its application to twelve species on Canada's Pacific coast. DFO Canadian Science Advisory Research Document, 2020/004. Available at: [http://publications.gc.ca/collections/collection\\_2020/mpo-dfo/fs70-5/Fs70-5-2020-004-eng.pdf](http://publications.gc.ca/collections/collection_2020/mpo-dfo/fs70-5/Fs70-5-2020-004-eng.pdf)
- Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I. et al. (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation*, 156, 94–104. <https://doi.org/10.1016/j.biocon.2011.11.013>
- Parks, D., Shaffer, A. & Barry, D. (2013). Nearshore drift-cell sediment processes and ecological function for forage fish: Implications for ecological restoration of impaired Pacific Northwest marine ecosystems. *Journal of Coastal Research*, 29(4), 984–997. <https://doi.org/10.2112/JCOASTRES-D-12-00264.1>
- Pearce, J. & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133(3), 225–245. [https://doi.org/10.1016/S0304-3800\(00\)00322-7](https://doi.org/10.1016/S0304-3800(00)00322-7)
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C. et al. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33(10), 1704–1711. <https://doi.org/10.1111/j.1365-2699.2006.01482.x>
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips, S.J. & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- Reiss, H., Cunze, S., König, K., Neumann, H. & Kröncke, I. (2011). Species distribution modelling of marine benthos: A North Sea case study. *Marine Ecology Progress Series*, 442, 71–86. <https://doi.org/10.3354/meps09391>
- Robards, M.D., Willson, M.F., Armstrong, R.H. & Piatt, J.F. (1999). *Sand lance: A review of biology and predator relations and annotated bibliography*. Research Paper PNW-RP-521. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. Available at: [https://www.fs.fed.us/pnw/pubs/pnw\\_rp521.pdf](https://www.fs.fed.us/pnw/pubs/pnw_rp521.pdf)
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Aroita, G. et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Robinson, C.L.K., Hrynyk, D., Barrie, J.V. & Schweigert, J. (2013). Identifying subtidal burying habitat of Pacific sand lance (*Ammodytes hexapterus*) in the Strait of Georgia, British Columbia, Canada. *Progress in Oceanography*, 115, 119–128. <https://doi.org/10.1016/j.pocean.2013.05.029>
- Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E. & Lundquist, C.J. (2017). A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, 4, 421. <https://doi.org/10.3389/fmars.2017.00421>
- Rooper, C.N., Hoff, G.R., Stevenson, D.E., Orr, J.W. & Spies, I.B. (2019). Skate egg nursery habitat in the eastern Bering Sea: A predictive model. *Marine Ecology Progress Series*, 609, 163–178. <https://doi.org/10.3354/meps12809>
- Rooper, C.N., Zimmermann, M. & Prescott, M.M. (2017). Comparison of modeling methods to predict the spatial distribution of deep-sea coral and sponge in the Gulf of Alaska. *Deep Sea Research Part I: Oceanographic Research Papers*, 126, 148–161. <https://doi.org/10.1016/j.dsr.2017.07.002>
- Segurado, P. & Araújo, M.B. (2004). An evaluation of methods for modeling species distributions. *Journal of Biogeography*, 31(10), 1555–1568. <https://doi.org/10.1111/j.1365-2699.2004.01076.x>
- Shabani, F., Kumar, L. & Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and Evolution*, 6(16), 5973–5986. <https://doi.org/10.1002/ece3.2332>
- Thomson, R. (1981). *Oceanography of the British Columbia coast*. Canadian Special Publication of Fisheries and Aquatic Sciences 56. Ottawa, ON, Canada: Department of Fisheries and Oceans.
- Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M.D. & Thuiller, C.W. (2016). 'biomod2': Ensemble platform for species distribution modeling. Available at: <https://cran.r-project.org/web/packages/biomod2/index.html>
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M. (2009). BIOMOD—A platform for ensemble forecasting of species distributions. *Ecography*, 32(3), 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Aroita, G. (2019). Block CV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225–232. <https://doi.org/10.1111/2041-210X.13107>
- Walbridge, S., Slocum, N., Pobuda, M. & Wright, D.J. (2018). Benthic Terrain Modeler (BTM) 3.0, tools for understanding and classifying the benthic environment [Software].
- Wright, P., Jensen, H. & Tuck, I. (2000). The influence of sediment type on the distribution of the lesser sandeel, *Ammodytes marinus*. *Journal of Sea Research*, 44(3–4), 243–256. [https://doi.org/10.1016/S1385-1101\(00\)00050-2](https://doi.org/10.1016/S1385-1101(00)00050-2)
- Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009). *Mixed effects models and extensions in ecology with R*. New York, NY: Springer Science & Business Media.

**How to cite this article:** Robinson CLK, Proudfoot B, Rooper CN, Bertram DF. Comparison of spatial distribution models to predict subtidal burying habitat of the forage fish *Ammodytes personatus* in the Strait of Georgia, British Columbia, Canada. *Aquatic Conserv: Mar Freshw Ecosyst*. 2021; 1–15. <https://doi.org/10.1002/aqc.3593>