



Predictive modeling of coral distribution in the Central Aleutian Islands, USA

Doug Woodby*, Dave Carlile, Lee Hulbert

Alaska Department of Fish and Game, 1255 West 8th Street, Juneau, Alaska 99811, USA

ABSTRACT: Logistic regression and generalized estimating equation (GEE) models were used to predict the probability of presence/absence for deep-sea corals in the central Aleutian Islands from 50 to 3000 m depth. Explanatory variables included in the models were depth, slope, and rugosity. Five of 15 logistic models were considered acceptable based on cross validation using a *Kappa* statistic threshold of 0.4. Models of occurrence north of the Aleutian chain were more successful than models for areas to the south of the chain. Based on the logistic regression predictive model, there are large swaths of sea floor between 100 and 1700 m depth with high rugosity and steep slope that have the potential to support coral gardens, particularly north of the Aleutian Islands arc and in Amchitka Pass. Improvements to the predictive models, including high resolution substrate and current data, as well as more *in situ* coral observations from a broader range of areas in the archipelago, would allow an improved evaluation of the adequacy of recent fishing gear restrictions in the Aleutian Islands to provide conservation benefits.

KEY WORDS: Coral · Aleutian Islands · Predictive modeling · Logistic regression · Generalized estimating equations · Gorgonians

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Cold-water corals are non-photosynthesizing colonies of anthozoans whose distribution in the world's oceans is not well known, largely because they occur in waters considerably deeper than photosynthesizing coral reefs. The Aleutian Islands provide habitat for what appears to be the highest diversity of cold-water corals in the world (Heifetz et al. 2005), based in part on collected specimens from incidental take of corals in commercial trawl, longline, and pot fisheries. Corals are widely distributed around the Aleutian Archipelago and appear to be important habitat for commercial fishery species (Stone 2006). Six small areas in the Aleutian Islands have been closed to all bottom contact gear to protect the most diverse and sensitive coral habitats (Witherell & Woodby 2005). These habitats were identified by observers in a submersible in 2002 as containing coral gardens, defined as highly diverse assemblages of corals and sponges with coverage approaching 100%

(Stone 2006). Outside of fishery bycatch data and the exploratory submersible dives in 2002, there has been very little quantitative information to inform decisions on areas most sensitive to disturbance by bottom contact fishing gear in the Aleutian Islands. To fill that need, the present study was designed to identify suitable habitat in the Central Aleutian Islands using predictive models of coral occurrence based on measurement of selected physical features.

We postulated that there are predictable relationships between the occurrence of corals and measurable environmental characteristics, including depth, slope angle, and seafloor roughness (rugosity). We mapped the seafloor in a representative sample of a large and significant section of the Aleutian Archipelago to directly observe the distribution and abundance of corals on the seafloor and to assess the relationships between major environmental factors and the abundance of corals. We then constructed predictive models of where corals are likely to occur for areas where

*Email: doug.woodby@alaska.gov

direct observations have not been made and where the available physical measurements would be limited to those derived from bathymetric data. For this reason, we did not include bottom temperature or chemistry data as dependent variables.

Our approach was tempered by the reality of geographic scale in the Aleutians. It would be ideal to have a map of the distribution of coral habitats for the entire Aleutian Islands chain, but the 1600 km long archipelago is too vast to attempt such a detailed mapping project all at once. Our solution to the problem of scale was 2-fold. First, we limited the study area to the central Aleutians from Seguam Pass on the east to Amchitka Pass on the west, a distance of about 500 km with a surface area of approximately 43 000 km² (Fig. 1). This study area was chosen because it is heavily fished, has substantial coral bycatch, and includes sites where coral gardens were observed from a submersible in 2002. Also, results from the present study in the central Aleutian Islands are likely to be helpful in guiding future efforts in the adjacent eastern and western sections. Second, the bathymetric mapping was conducted as a sampling, rather than as a complete coverage of the study area. The mapping technology with sufficient resolution for our purposes was too expensive and restricted in coverage to attempt a complete mapping; however, a representative sam-

pling protocol allowed us to describe the study area as a whole in a statistically valid manner.

Our predictive models use logistic regression and generalized estimating equations (GEE) to estimate the probability of presence and absence for corals. The logistic regression modeling approach is common in ecological modeling and these types of models are referred to as resource selection functions (Boyce et al. 2002), habitat suitability models (Tittensor et al. 2009), and species distribution models (Guisan & Thuiller 2005). GEE methods are not commonly used in ecological modeling, but have some advantageous statistical properties in cases where data have serial correlation, as is often the case when data are collected from contiguous or closely spaced sampling units, such as on long video transects. An ongoing concern with these approaches is how they are to be evaluated, which ultimately should depend on how the results are to be used. Our approach for the full logistic models was to use a cross-validation test to evaluate how well the models reliably predicted locations suitable for corals.

METHODS

Data sources. Observations of benthic fauna were made using the 'Delta' submersible in 2003 and 2004

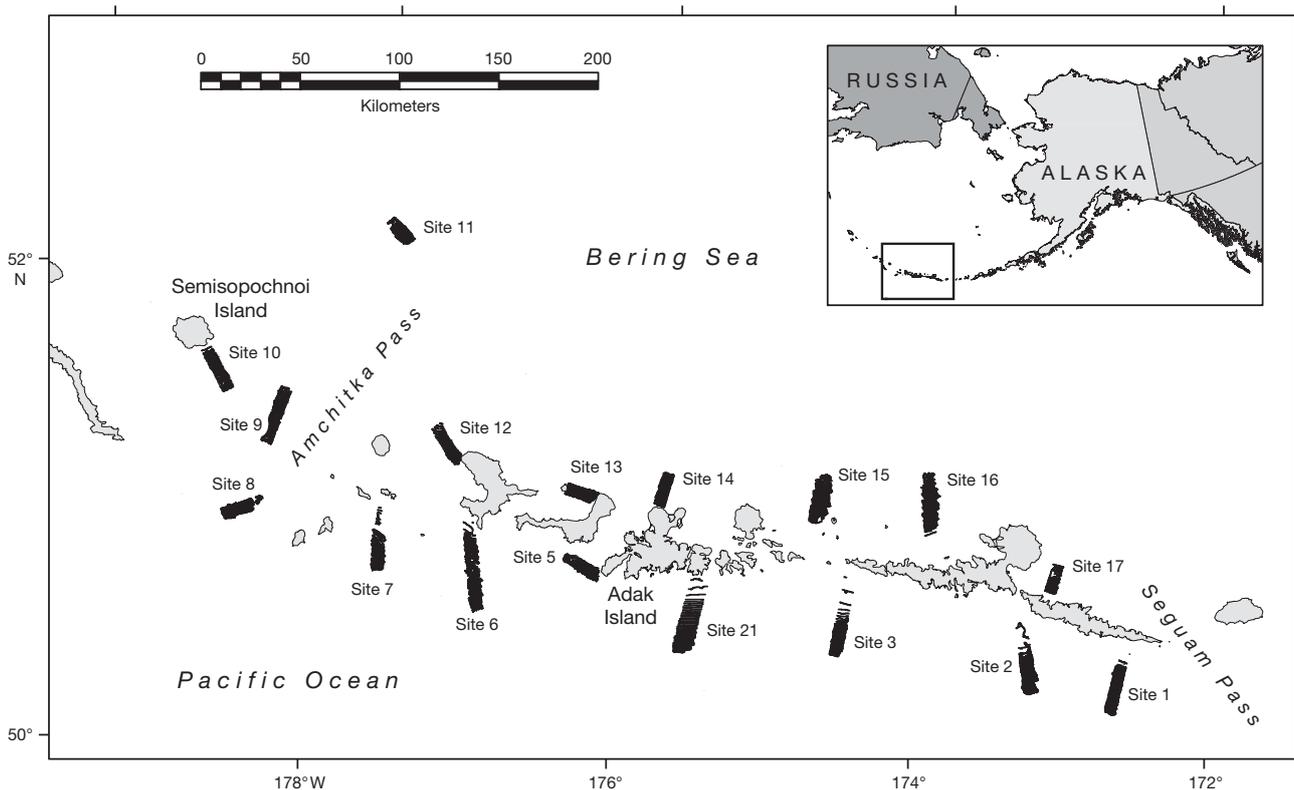


Fig. 1. Central Aleutian Islands coral study area. Black lines are ocean bottom mapping transects

and using the remotely operated vehicle (ROV) 'Jason II' in 2004. The sampling units used in the predictive models were closely spaced, or contiguous video fields of view, referred to as a frames, on long transects. Due to the very large quantity of video data, frames were systematically subsampled on many of the dives at rates of 10 to 50%. For the 'Delta' transects, sampling rates of 10 and 50% were achieved by sampling every 5th and every other frame, respectively. For the 'Jason II' transects, all frames were sampled at a 50% rate, achieved by sampling 5 adjacent frames and then skipping the next 5 frames, and so on. This method was used to expedite sampling, but may have introduced additional bias due to the potential spatial correlation of adjacent samples. In total, there were 21 885 frames from 27 transects distributed among 9 study sites for which data were transcribed to provide counts of taxa and other related information. The total area of the sampled frames was 3.2 km². The frequency of samples was highest between depths of 200 and 500 m, due in part to selection of sites for the 'Delta' submersible, which is limited to operating in depths less than 365 m, and also reflecting the proportion of sea-floor area in the region sampled.

Coral taxa were classified into 5 groups for the purposes of predictive modeling. Gorgonians (Group 1, including *Acanthogorgia*, *Alaskagorgia*, *Arthrogorgia*, *Calcigorgia*, *Euplexaura*, *Fanellia*, *Muriceides*, *Paragorgia*, *Plumarella*, *Primnoa*, *Swiftia*, and *Thouarella*, but not bamboo corals or *Radicipes*) comprised the largest group (41% of total coral colony counts). Scleractinians (stony cup corals, not differentiated to genera) were the second largest group (Group 2, 23%), followed by pennatulaceans (sea whips and sea pens, including *Anthoptilum*, *Halipterus*, *Ptilosarcus*, and *Umbellula*: Group 3, 16%), and stylasteriids (hydrocorals, not differentiated to genera: Group 4, 14%). Four other groups were judged to be too infrequent to provide sufficient observations for reliable predictive modeling. These included stoloniferans (stolon corals such as *Clavularia*) and alcyonarians (true soft corals such as *Anthomastus*, *Gersemia*, and *Nepheids*) in Group 5 (3%), Isididae (bamboo corals, not differentiated to genera: 2%), *Radicipes* (0.2%), and antipatharians (black corals, not differentiated to genera: 0.1%). A composite of coral Groups 1, 2, 4, and 5 was formed to represent those corals commonly found together in coral gardens.

Physical habitat parameter data came from several sources. Depth data were derived from depth gauge records from the ROV and submersible, multibeam bathymetry for each of the 17 mapped study sites, and regional bathymetry data compiled by the National Marine Fisheries Service (S. Lewis pers. comm.). Slope values were derived from the depth data with ESRI ArcGIS 9.2 (ArcInfo) software and Spatial Analyst

extension tools. Rugosity (roughness) rasters were produced with Benthic Terrain Modeler (BTM) Version 1.0 software (Wright et al. 2005) using the ArcMap application of ESRI's ArcGIS mapping software, where rugosity is defined as the ratio of surface area to planar area and is calculated using a nearest neighbor analysis. Locations were classified as north or south of the Aleutian Islands arc. From previous work (Heifetz 2003), we anticipated having estimates of currents across the study area, but reliable and high resolution current data were not available for model development due to the general lack of current observations in the central Aleutian Islands. Current data on a 15 km grid were initially incorporated in our modeling approach (T. Bryan pers. comm.), but were found to be at too low a resolution to allow differentiation of currents between many of the transects. Measures of other environmental factors such as concentrations of aragonite, dissolved oxygen, and chlorophyll *a* were not included because they were not available at sufficiently fine scales of resolution or depths in our study areas. Although temperature by depth was available for all dives, it was highly correlated with depth, and was not available by depth away from the transects.

Model development. We conducted 2 rounds of model development. For the first round, we used logistic regression to model the presence/absence of 5 coral groups (Groups 1 to 4 and the composite group) as functions of several candidate explanatory variables (Hosmer & Lemeshow 2000, SAS Institute 2004). We predicted presence/absence as opposed to densities after examining the data and finding many zero counts at the frame level. The linear logistic model is expressed as:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta'x \quad (1)$$

where $\pi = P(Y = 1|x)$, α is the intercept parameter, and β' is a vector of parameters.

The predicted probabilities of occurrence for the coral groups ($\pi = P[Y = 1|x]$, i.e. the probability of occurrence given the explanatory variables, x) were obtained by back-transforming the $\text{logit}(\pi)$ values from the logistic models as follows:

$$\Pr(Y = 1 | x) = \frac{1}{1 + e^{-(\alpha + \beta'x)}} \quad (2)$$

Response variables were presence/absence per frame of corals, with corals differentiated by taxonomic categories. Modeling was done using the logistic procedure in SAS (SAS Institute 2004). Final explanatory variables retained in models were those selected by a stepwise selection procedure, using an $\alpha = 0.05$ cutoff level for entry and retention in the model, and which differed significantly from zero at the 0.05 alpha level.

Candidate explanatory variables for modeling (i.e. x), included depth, slope, and rugosity. We included quadratic and cubic terms for these variables, as well as 2-way interactions among variables, in the logistic models from the first round of modeling. The purpose of including quadratic and cubic terms was to allow for greater sensitivity to changes in variable values. The first round logistic models are referred to as full logistic models to distinguish them from the limited logistic models conducted for comparison to the GEE during the second round of modeling.

Location, either north or south of the Aleutian Islands arc, was evaluated as a factor by creating subsidiary models based on data from northern sites and a separate set of models for the southern sites. These, in combination with models using all site data (combined north–south model), resulted in 3 sets of models, with each set consisting of 5 coral models (Groups 1 to 4 and the composite group).

For the full logistic models, in addition to the step-wise selection procedure and the tests for significance of candidate explanatory variables, 2 diagnostics were used to evaluate the predictive power of the models: the maximum-rescaled R^2 and Cohen's *Kappa* (Cohen 1960). The maximum-rescaled R^2 is a generalized coefficient of determination scaled to a maximum value of 1.

Our individual dive transects, comprised of hundreds to thousands of contiguous or closely spaced sampling frames, constitute large clusters of data. The clustered nature of the sampling frames can result in spatial autocorrelation of coral observations. Clustering and the potential for non-independence of observations, if unaccounted for in analyses such as regular logistic regression analysis, can produce underestimates of standard errors, overestimates of test statistics, and, consequently, underestimates of observed significance levels for tests (Allison 1999). Therefore, as an alternative to the full logistic regression, we applied GEE for the second round of model development to explore coral occurrence as a function of the same depth and depth-derived habitat (explanatory) variables used for the full logistic modeling (Eq. 1). The 2 approaches differ in the way the variances are estimated, with GEE models yielding standard errors of parameters and test statistics that are adjusted for spatial correlation (Allison 1999).

As described by Horton & Lipsitz (1999) Eq. (1) implies:

$$E[Y_{ij}] = \frac{\exp(\alpha + \beta'x)}{1 + \exp(\alpha + \beta'x)} \quad (3)$$

where $E[Y_{ij}]$ is the expected (binary) value of outcome j for subject i .

Because the outcomes are binary (i.e. presence/absence of coral),

$$\text{var}(Y_{ij}) = \frac{\exp(\alpha + \beta'x)}{(1 + \exp(\alpha + \beta'x))^2} \quad (4)$$

Eq. (4) specifies the variance of the model, but because of the probable spatial autocorrelation in our data, the covariance structure of the observations also needs to be modeled. The covariance matrix of Y_i can be expressed as:

$$V_i = \phi \mathbf{A}_i^{1/2} \mathbf{R}(\alpha) \mathbf{A}_i^{1/2} \quad (5)$$

where ϕ is a dispersion parameter (McCullagh & Nelder 1989), \mathbf{A}_i is a diagonal matrix of variance functions (Eq. 4), and $\mathbf{R}(\alpha)$ is the working correlation matrix of Y_i . Because of the spatial longitudinal nature of our sampling frames along each dive transect, we specified an auto-regressive working correlation matrix for our GEE regression modeling.

To simplify interpretation of the GEE model results, we restricted the complexity of GEE model components to a maximum of 2-way interactions and quadratic terms. For comparison with the GEE models, we also re-ran some of the logistic models using the same simplifying model restrictions.

Modeling with GEE was done using the GENMOD procedure in SAS (SAS Institute 2004). In GENMOD, we used the TYPE=AR option to specify the probable lag-1 autoregressive structure of the working correlation matrix. Initially, we used $\alpha = 0.1$ for the significance level for retention of parameters in the model. However, this alpha level yielded very sparse models with few significant terms and implausible predicted probability maps. As a result we increased the power of the tests for identifying significant factors by changing the significance level to $\alpha = 0.2$.

Model evaluation. For the full logistic regression models, cross-validation was used to evaluate the reliability of the various predictive models of coral occurrence. For each hypothesized model, the appropriate set of video transect observations (n varied from 3147 to 21886) were resampled 200 times, where 90% of the observations in each resample were randomly chosen to serve as the training data set, and the remaining 10% served as the test data set. The randomization was repeated with each resampling. The same set of factors identified as significant in each hypothesized model was used for cross-validation, but the coefficients were newly determined with logistic regression for each of the 200 resamples of training data, and these new coefficients were used to score each of the observations in the test data set as present or absent (e.g. presence or absence of a taxon group), using a cutoff value of 0.5. The sum of predictions (counts) was tabulated for each of the 4 possible outcomes within a standard confusion matrix (Fielding & Bell 1997).

A variety of commonly used diagnostic measures of prediction accuracy may be calculated from a confusion matrix (Fielding & Bell 1997). Of these measures, we focused on *Kappa* as a generally robust measure with respect to common failings of the other measures (Cohen 1960, reviews in Fielding & Bell 1997, Manel et al. 2001).

An important concern with confusion matrix diagnostics, including the measure of correct classification rate, is a low prevalence of true events. This situation was frequent in our presence/absence data, given that 57% of the frames did not have corals. Manel et al. (2001) demonstrated that the effect of low prevalence on *Kappa* was apparently negligible. We adopted the following ranges of agreement for *Kappa* (K) as suggested by Landis & Koch (1977): $K < 0.4$ is poor; $4 \leq K \leq 0.75$ is good, and $K > 0.75$ is excellent. We rejected any of our full logistic prediction models where K was less than 0.4.

Unlike the SAS logistic procedure, which allows ready output of model results for the cross validation, the SAS GENMOD procedure used for the GEE models does not provide for easy, repeated output of results necessary for the resampling required for cross-validations as applied here. This situation precluded the use of cross-validation for the GEE and limited logistic modeling that was applied to the full logistic regressions in the first round of modeling.

Mapping products. Maps of predicted probabilities were made for the composite coral group using the models deemed acceptable (see below). The maps were generated from raster-based grid files using ArcGIS with cell sizes ranging from 15 to 50 m on a side. Cell sizes varied because of the depth-dependent resolution of the multi-beam sonar bathymetry, which was provided for this project by Fugro Pelagos. We did not make corrections for depth-dependence in cell size, but recognize the potential bias that might result, for example, when modeling effects of rugosity, given that rugosity is scale-dependent.

Based on results from the full logistic modeling, we made probability maps of each of the 17 multibeam bathymetry sites as well as 3 larger areas (termed 'off-sites') that were not part of the multibeam sample. These latter areas were selected for the purpose of applying the model predictions to areas without the high reso-

lution bathymetry available for the 17 study sites. The off-sites included 1 area north of the Aleutian Island chain, 1 area to the south, and another area on the west side of Amchitka Pass. Because of space limitations, here we present maps only for study site 10 and the off-site area west of Amchitka Pass. Probability maps of the other areas are available in the electronic supplement (see www.int-res.com/articles/suppl/m397p227_app.pdf; original source: Heifetz et al. 2007). For the GEE modeling, maps of predicted probabilities were made only for a small subset of the 17 multibeam bathymetry sites and only for the composite coral group, for the purpose of comparing the GEE results to the maps from the full logistic regression modeling. Model predictions were confined to depths within the 52 to 2987 m depths that were sampled.

RESULTS

Modeling

For the full logistic modeling, 10 of the 15 models (5 taxonomic groups \times 3 locations: north, south, combined) incorporated depth as a significant factor. All of the models incorporated quadratic and/or cubic depth factors as well as depth as an interactive factor with either slope or rugosity (Table 1). Rugosity entered as a significant factor in 12 of the 15 models and slope entered significantly in 8 of the models. Both rugosity and slope often entered as higher order terms and in interaction with the other variables. Depth, as a linear, quadratic, or cubic term, was generally the most important factor, as shown in the standardized parameters in Table 2 (except Group 3, rugosity quadratic).

Table 1. Model parameter estimates and predictive diagnostic statistics for full logistic models including data from north and south of the Aleutian Islands. Cor.comp. = Coral composite

Parameter/ Diagnostic	Taxonomic group				
	1	2	3	4	Cor.comp
Intercept	5.63×10^{-1}	-3.89×10^0	-9.30×10^{-1}	-5.26×10^0	4.75×10^{-1}
Depth		-5.09×10^{-3}	-2.34×10^{-2}	1.45×10^{-2}	
Depth \times Depth	-1.00×10^{-5}	-2.93×10^6	1.90×10^5	-2.00×10^{-5}	-1.00×10^{-5}
Depth \times Depth \times Depth	1.86×10^9		-4.44×10^9	3.86×10^9	2.16×10^9
Slope	-3.09×10^{-1}			-1.67×10^1	-2.56×10^{-1}
Slope \times Slope	-5.00×10^{-4}			-4.80×10^4	-5.20×10^{-4}
Slope \times Slope \times Slope	9.79×10^7			1.19×10^6	1.22×10^6
Depth \times Slope	-1.30×10^{-4}	5.81×10^6	-1.00×10^{-4}	-3.00×10^5	-1.10×10^{-4}
Rugosity	6.46×10^0	-9.61×10^0	-1.46×10^2		4.67×10^0
Rugosity \times Rugosity	-1.04×10^1	1.51×10^1	2.66×10^2		-8.12×10^0
Rugosity \times Rugosity \times Rugosity		-4.72×10^0	-1.20×10^2	-1.58×10^0	
Depth \times Rugosity	1.29×10^{-2}	-1.40×10^{-3}	4.00×10^{-3}	5.90×10^{-3}	1.14×10^{-2}
Slope \times Rugosity	3.86×10^{-1}			1.98×10^{-1}	3.33×10^{-1}
Max-rescaled R ²	4.12×10^{-1}	1.88×10^{-1}	4.80×10^{-1}	4.02×10^{-1}	4.19×10^{-1}

Table 2. Standardized parameter estimates from the full logistic models using data from north and south of the Aleutian Islands. Cor.comp. = Coral composite

Parameter/Diagnostic	Taxonomic group				Cor.comp.
	1	2	3	4	
Depth		2.75	-12.63	7.83	
Depth × Depth	-17.46	-4.53	29.71	-28.11	-16.24
Depth × Depth × Depth	8.04		-19.18	16.64	9.34
Slope	-4.49			-2.42	-3.72
Slope × Slope	-1.27			-1.34	-1.45
Slope × Slope × Slope	0.75			0.91	0.93
Depth × Slope	-0.93	0.04	-0.53	-0.23	-0.81
Rugosity	0.39	-0.59	-8.93		0.29
Rugosity × Rugosity	-1.21	1.76	30.91		-0.94
Rugosity × Rugosity × Rugosity		-1.02	-25.98	-0.34	
Depth × Rugosity	7.14	-0.77	2.24	3.29	6.33
Slope × Rugosity	5.87			3.02	5.08

The GEE models had much fewer significant factors than the full logistic models. Of the 15 GEE models (3 data combinations × 5 taxonomic groups), 14 incorporated depth as a significant linear, quadratic, or interaction factor with one or more of the other factors. Twelve of the 15 models incorporated a quadratic factor, and 8 of the models had at least one significant interaction term. Rugosity entered as a significant factor—either as a linear or more complex term—in 8 of the 15 models, whereas slope entered significantly in only 6 of the models. Focusing on the 5 models in which north and south data were combined, depth was included in some form in 3 of the 5 taxonomic group models (Table 3). Also, for the combined (north/south) models, 3 of 5 models had significant interactions (Table 3). Rugosity entered more often as a higher order term and/or in interaction with the other variables than as a linear term (e.g. Table 3).

Only a third of the full logistic models (5 of 15) resulted in *Kappa* values greater than or equal to the 0.4 cutoff for admissible models (Table 4). The north site models yielded the highest predictive successes, with 4 of the models yielding *Kappa* values between 0.698 and 0.776. The south site models performed poorly, with only one of the taxon-specific models having a *Kappa* value >0.4. The combined north–south area models also performed poorly. Based on the acceptance criteria, the only acceptable model of the composite coral group was for locations north of the Aleutian Islands chain (*Kappa* = 0.704).

Probability maps

Multibeam mapped sites:
full logistic models

The pattern of mapped probabilities from the full logistic model of the composite coral group across the north sites illustrates the importance of depth and slope (e.g. Fig. 2 for Site 10). Relatively high probabilities of occurrence (orange to red tones) were predicted for depths from 550 to 1500 m, and lower probabilities (green tones) occurred in the shallowest (200 to 500 m) and

deepest (1550 to 1650 m) areas. Mapped probabilities increased (to red tones) on steep slopes within the high probability depth range (Fig. 2). Where steeper slopes occurred above 200 m, high probabilities were predicted in waters as shallow as 100 m, as at Site 14 north of Adak Island (see Fig. S6 in the electronic supplement). Site 17 is a mostly shallow water site and the predicted probabilities of occurrence for the coral composite group were very low even in those areas with greater slopes and/or rugosity (e.g. a scarp that runs horizontally through the middle of the site; see Fig. 10 in the electronic supplement).

Table 3. Model parameter estimates for generalized estimating equation models including data from north and south of the Aleutian Islands. Cor.comp. = Coral composite

Parameter/ Diagnostic	Taxonomic group				Cor.comp.
	1	2	3	4	
Intercept		-2.33×10 ⁰		-4.28×10 ⁰	
Depth					
Depth × Depth	-6.58×10 ⁶			-8.63×10 ⁶	-4.74×10 ⁶
Slope	-6.69×10 ⁻¹				
Slope × Slope					
Depth × Slope					
Rugosity	1.55×10 ¹				
Rugosity × Rugosity	-1.90×10 ¹				
Depth × Rugosity	9.50×10 ⁻³		2.84×10 ⁻²	1.24×10 ⁻²	
Slope × Rugosity	8.51×10 ⁻¹				

Table 4. Cross-validation *Kappa* diagnostic measure for the 5 full logistic models. **Bold** values are those that exceed the 0.4 cutoff for acceptability. Cor.comp. = Coral composite

Geographic source of data for model	<i>Kappa</i> values for taxonomic group				
	1	2	3	4	Cor.comp.
North and south of the Aleutian Islands	0.252	0.010	0.360	0.110	0.298
North of the Aleutian Islands	0.698	0.346	0.734	0.776	0.704
South of the Aleutian Islands	0.310	0.086	0.585	0.016	0.353

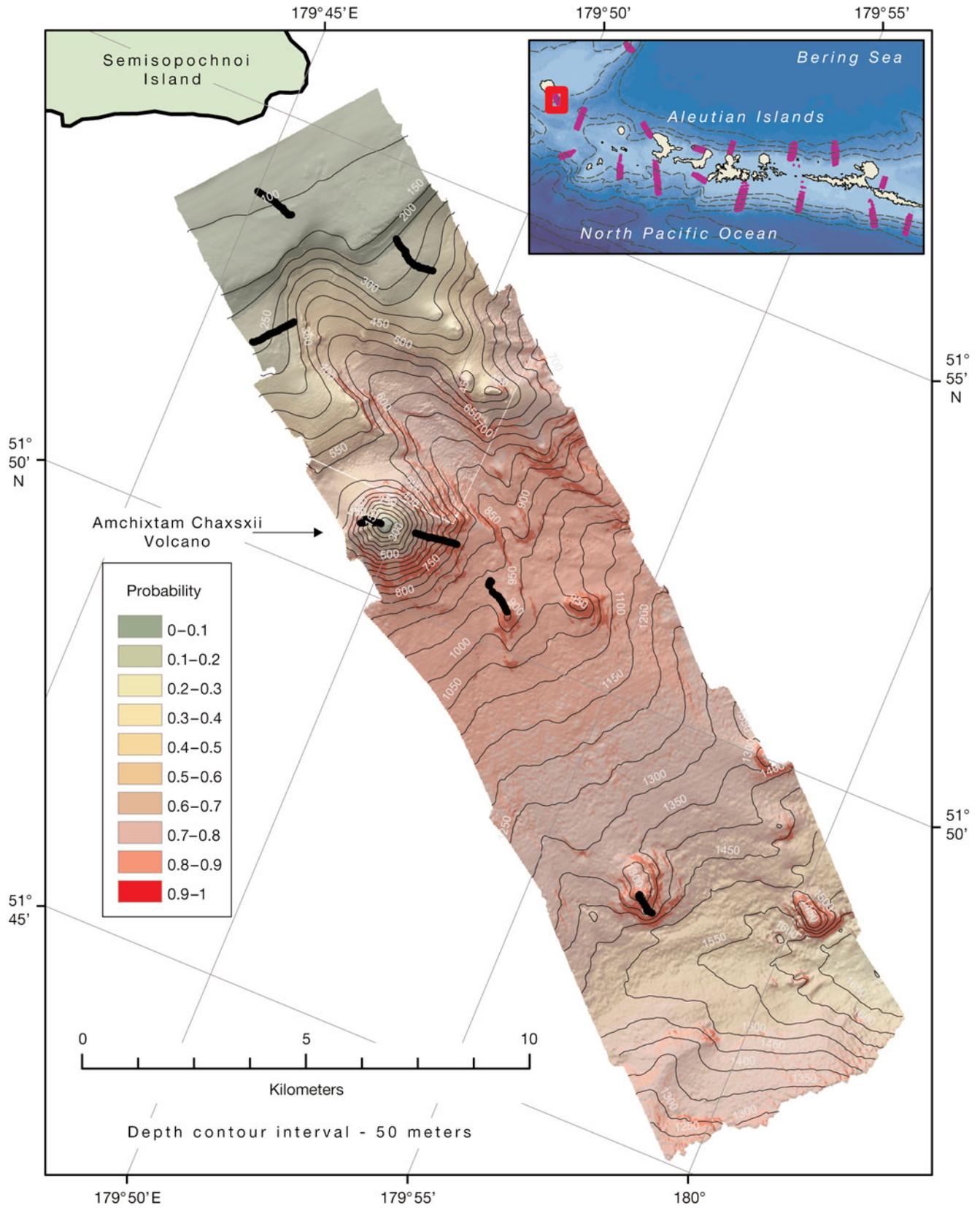


Fig. 2. Predicted probabilities of occurrence of the coral composite group, north full logistic model, at multibeam survey Site 10. Thick black lines are survey transect paths of the 'Delta' submersible and ROV 'Jason'

There was a noticeably lower predicted probability of occurrence of the composite coral group on the upper flanks of the Amchixtam Chaxsxii volcano at Site 10 (Fig. 2) above 350 m. That prediction is at odds with visual observations made from the submersible 'Delta' on Dive no. 6000, which revealed rich coral gardens up to the volcano's summit (120 m), but in concurrence with the overall distribution of coral densities for all transects, which showed peak densities from 500 to 1300 m. In effect, the north model is fit to a larger set of observations than those for Site 10 alone, and it is plausible that a model developed specifically for Site 10 would be more faithful to the observations made there.

Multibeam mapped sites: GEE models

Predicted probabilities of occurrence of the composite coral group from the GEE model, when applied to Site 10, suggested that depth is the factor of primary influence (Fig. 3). Predicted probabilities for the coral composite group from the GEE model were entirely in the 0.0 to 0.1 probability range; an erroneously low prediction, given the 7 'Delta' or 'Jason II' transects completed within this multibeam site and the direct observations of corals often in sufficient density and richness to qualify as coral gardens.

Multibeam mapped sites: limited logistic models

Probabilities of occurrence of the composite coral group at Site 10 under the limited logistic model exhibited the influence of depth, but also slope and rugosity (Fig. 4). Lowest probabilities of occurrence—consistently 0 to 0.1—were predicted for the greatest depths; those beyond about 1300 m. The shallowest depths, from 100 to 300 m, and intermediate depths, 1050 to 1300 m, had probabilities ranging between 0.1 and 0.3. Depths between 300 and 1050 m had greater variability in, and among the highest, mapped probabilities for Site 10. Probabilities in this depth range varied mostly between 0.2 and 0.8, with a few small spots of probabilities in the 0.9 to 1 range. Probabilities in this mid-depth range exhibited what appear to be the influence of slope and rugosity on mapped probabilities, with the highest probabilities (>0.6, and including the few spots with probabilities >0.9) being in steeper terrain, such as along the flanks of the Amchixtam Chaxsxii volcano. Like the mapped probabilities from the full logistic model, a known error from the limited logistic model is the predicted low (≤ 0.3) probability of coral occurrence near the summit of the volcano, when direct observation from the 'Delta' showed dense coral gardens.

Off-site predictions: corals

Fig. 5 shows the full logistic model predictions of coral distribution on a 50×60 km site that includes terrain beyond the multibeam-mapped Site 10 south of Semisopchnoi Island on the west side of Amchitka Pass. This figure is an example of extrapolation of the predictive models to an area with lower resolution bathymetry. Depths were interpolated from bathymetric data of varied but low resolution onto a 100 m grid using an inverse distance weighting algorithm. Predictions of probabilities of occurrence of the coral composite group in this area showed patterns very similar to those made for the northern multibeam-mapped sites. The dominant pattern was a depth distribution of the highest probabilities between 700 and 1700 m and where slopes were steepest.

DISCUSSION

Coral distribution patterns

Physical, especially topographical, features of the seafloor are likely to be important in the distribution of deep-water corals in general (Cimberg et al. 1981). In particular, areas of flow acceleration, as expected with strong currents in areas of high relief, are thought to be important in promoting coral growth by delivering more food particles per unit time (Genin et al. 1986, Mortensen et al. 2001, MacIsaac et al. 2001). With a few exceptions, many deep-water corals are expected to require a hard substrate for attachment following settlement, and near-bottom currents are likely to be important in keeping suspended sediments and organic particles from settling and smothering developing corals (White et al. 2005). For example, in the NE Atlantic the anthozoans *Paragorgia*, *Primnoa*, and *Acanthogorgia* were generally found to occur on cobbles and boulders (Mortensen & Buhl-Mortensen 2004). *Radicipes* and the pennatulaceans (sea whips and sea pens) appear to be clear exceptions to this hard-bottom requirement, as they have anchoring structures that allow them to hold fast in soft sediments.

Several recent attempts to model suitable habitat for deep-water corals have been undertaken at low resolution across large areas. These include a worldwide effort to predict the distribution of scleractinians (Titensor et al. 2009) using oceanographic factors (e.g. chemistry, temperature, depth, and regional current velocity), an assessment of suitable habitat for a single species of scleractinian (*Lophelia pertusis*; Davies et al. 2008) comparing a suite of factors (physical chemical, and biological) at both global and regional (NE At-

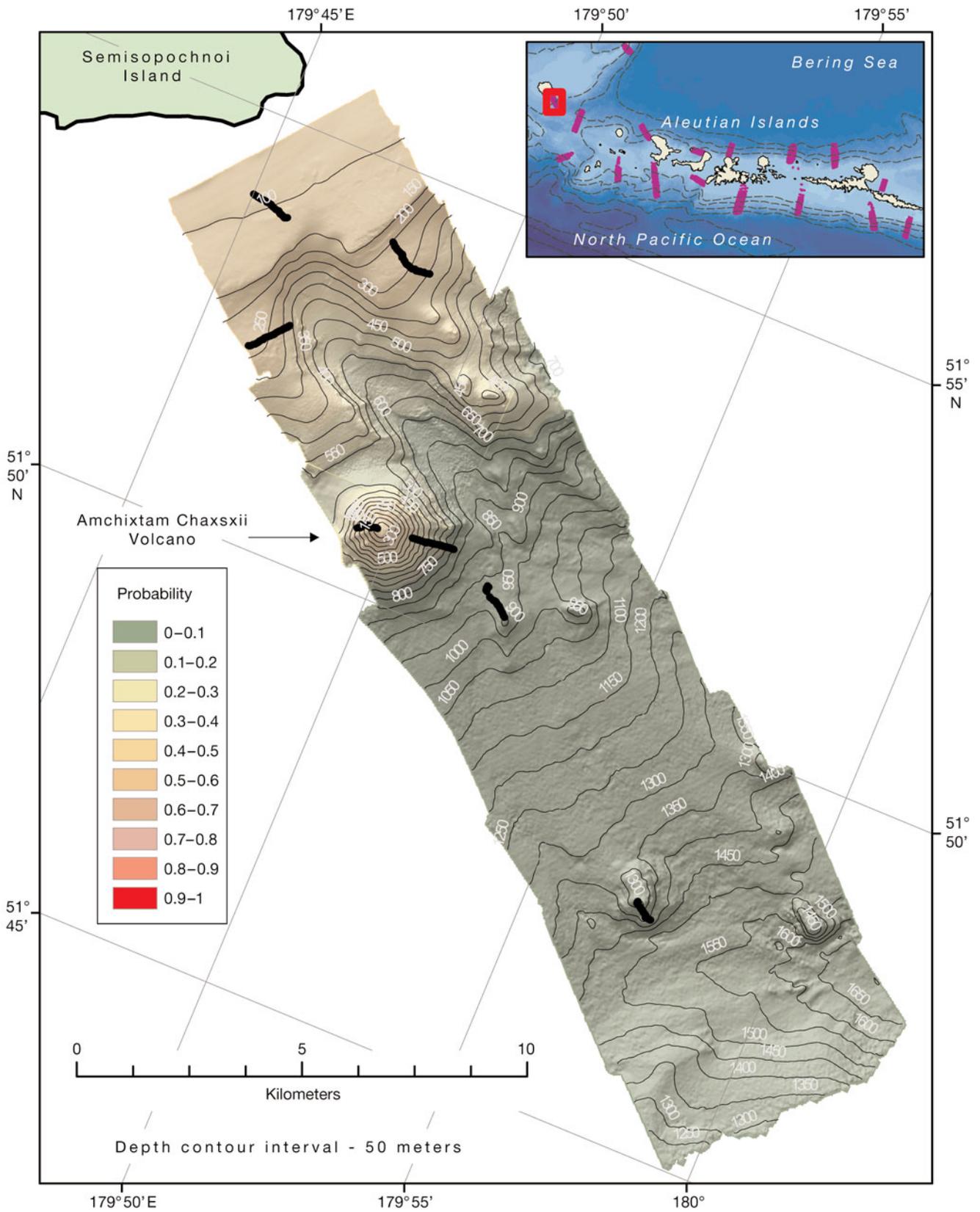


Fig. 3. Predicted probabilities of occurrence of the coral composite group, north generalized estimating equation model, at multibeam survey Site 10. Thick, black, lines are survey transect paths of the 'Delta' submersible and ROV 'Jason'

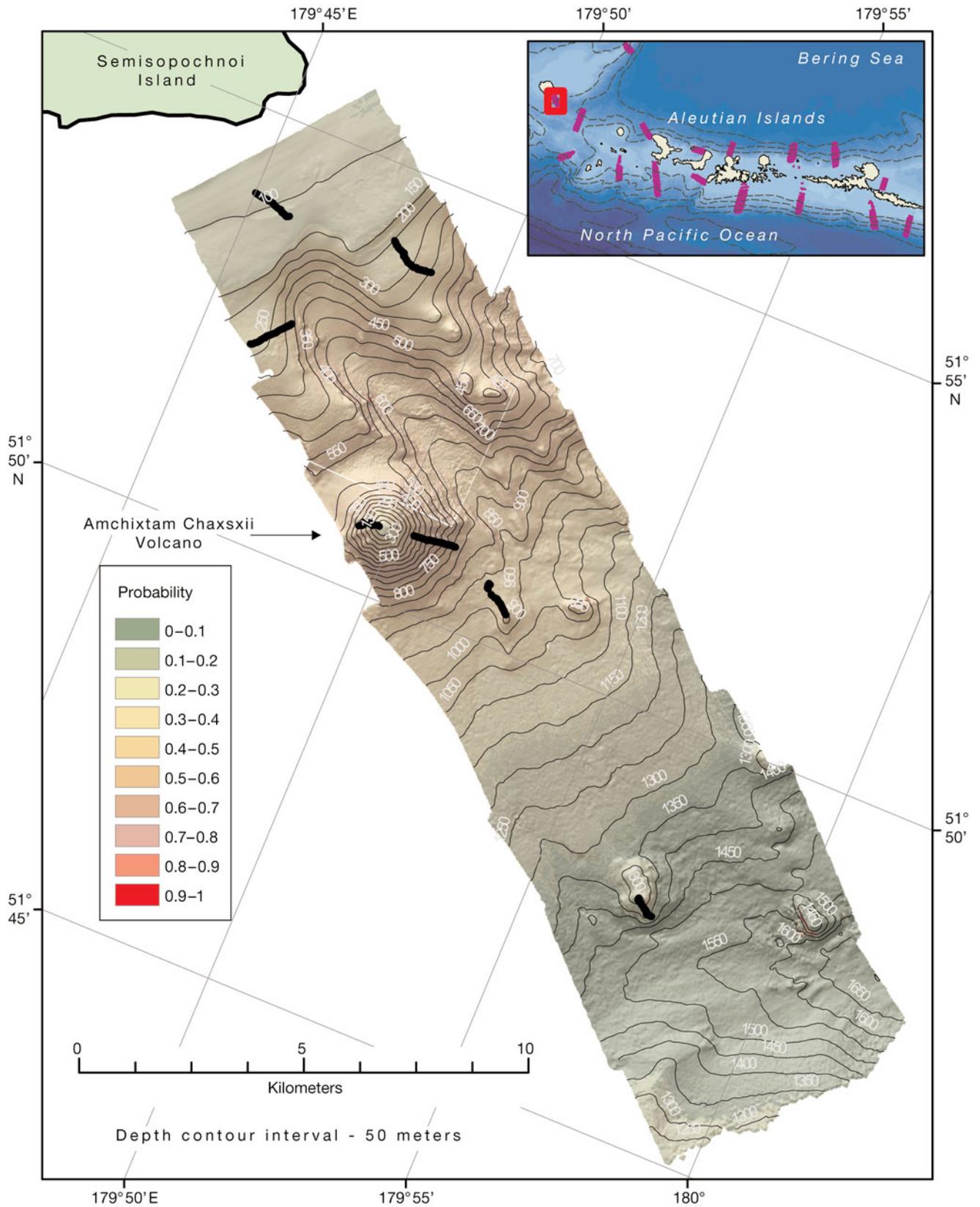


Fig. 4. Predicted probabilities of occurrence of the coral composite group, north limited logistic model, at multibeam survey Site 10. Thick black lines are survey transect paths of the 'Delta' submersible and ROV 'Jason'

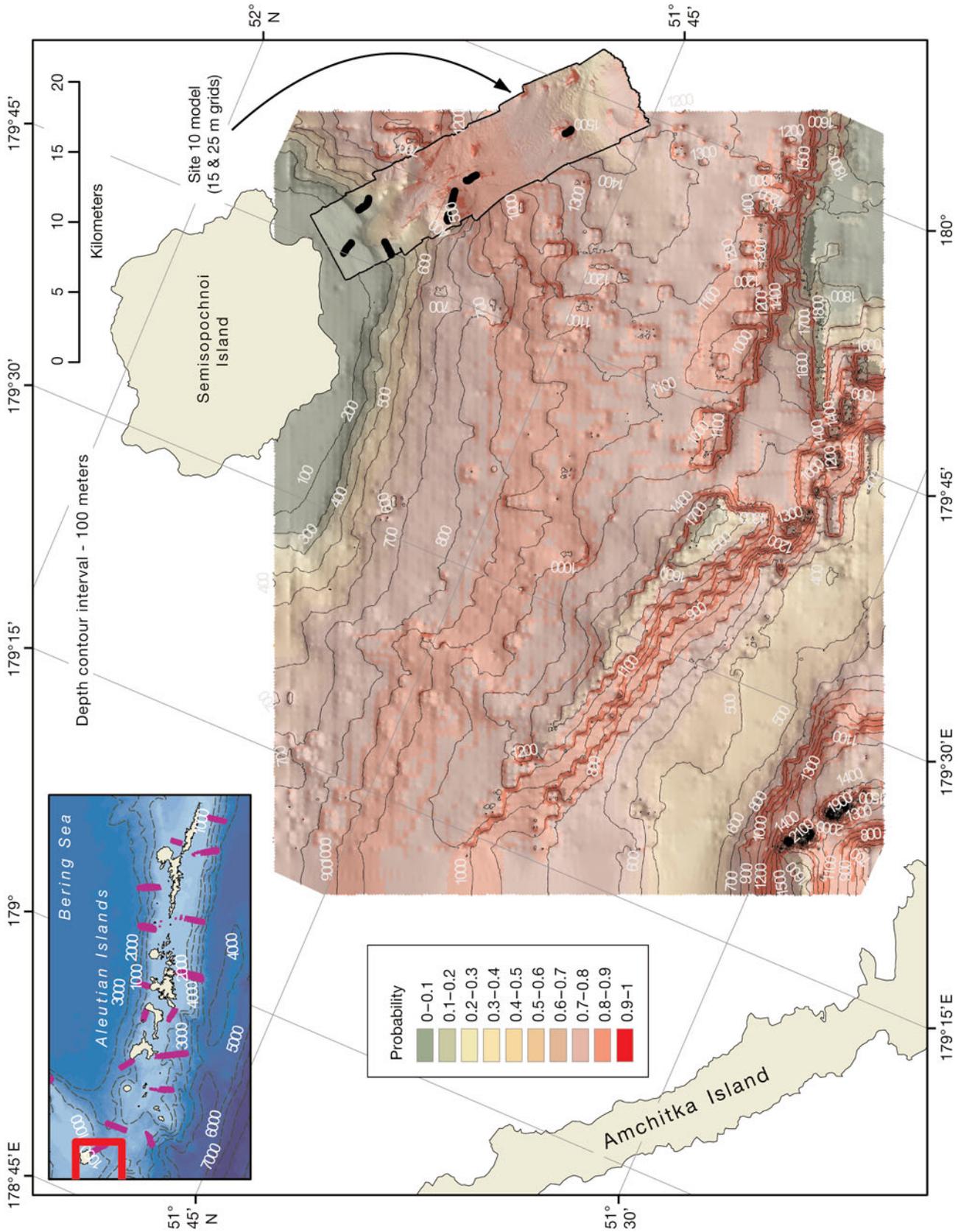


Fig. 5. Predicted probabilities of occurrence of the coral composite group, north full logistic model, for an 'off-site' area near multibeam survey Site 10 (note Site 10 overlay, outlined in black). Thick black lines are survey transect paths of the 'Delta' submersible and ROV 'Jason'

lantic) scales, and habitat suitability modeling for Paragorgiidae and Primnoidae (Bryan & Metaxas 2006, 2007) on the Atlantic and Pacific coastal margins of North America, based on depth/slope, temperature, current, chlorophyll *a*, and substrate data. All 3 of these large-scale studies employed presence-only data, which is an appropriate choice where absence data may not be reliable or available (Hirzel et al. 2002), as may be true for large-scale environmental data sets. In the Pacific coast model by Bryan & Metaxas (2006, 2007), which overlaps our study area, their most accurate model for Paragorgiidae combined temperature, slope, current, and chlorophyll *a*. Their model for Primnoidae was similar, except it included depth instead of temperature.

Our findings predicted presence and absence of corals using only bathymetrically derived factors, including linear effects of depth, slope, and rugosity, as well as higher order and interaction terms for those factors. Based on these factors alone, the mapped probabilities for corals are characterized by relatively large areas of uniformly high probabilities. It is not possible at this time to say whether this pattern of distribution is realistic; however, it is likely, given the heterogeneity of substrate in the Aleutians Islands area, that the actual distribution of coral garden habitat is patchy, and that the maps of model predictions are most appropriately viewed as indicating the potential for occurrence.

Further modeling efforts may benefit substantially from addition of other factors, especially reliable substrate data and high resolution current data. In lieu of current data, it may be possible to identify proxy measures that influence current, such as designations of ridge tops or summits of seamount features where currents would be expected to accelerate (Genin et al. 1986). Although depth is an easily measured factor, it may chiefly serve as a surrogate for factors strongly correlated with depth, such as temperature and concentrations of oxygen and aragonite. Inclusion of factors such as these would be expected to improve the predictive power of the models, and provide more realistic indications of where we might expect to find as yet unidentified coral gardens.

Model evaluation

The measure of success of the present models is based on a moderately rigorous cross-validation, using test data from the same visual transects from which the randomly chosen training data were taken. A more rigorous evaluation would use test observations taken from sites other than those from which the training data were derived. This is a potential next step, but is

marginally feasible given the limited number of transects and sites to choose from.

Each modeling approach we investigated has advantages and disadvantages. For the full logistic regression, available software (e.g. SAS) has stepwise, best subsets modeling capability which is currently not available in software that allows application of GEE. In addition, there are readily estimated goodness of fit test statistics, and relatively easily calculated diagnostic statistics for model checking applicable to logistic regressions available in statistical software. For GEE methods, these model building and evaluation capabilities are lacking, or not as well developed, in the current crop of statistical software. However, a primary advantage of habitat modeling based on methods such as GEE is that the spatial correlation often inherent in data collected along transects from contiguous or closely spaced sampling frames is accounted for in the modeling process. This accounting is in the form of an adjustment of the standard errors of parameters and the test statistics to more accurately reflect underlying relationships between coral occurrence and habitat factors, such as depth, while reducing the potential bias introduced by spatial proximity of the sampling frames. Failure to account for this spatial correlation in the modeling process can yield erroneous results, such as inclusion of a habitat factor as being a significant predictor of coral occurrence when it may not truly exercise significant influence.

While the predictive models based on the full logistic regression yielded predicted probabilities of occurrence that were more consistent with gross observations of coral distribution—primarily the often patchy nature of the distributions—to some degree, the statistical significance of depth or depth-derived factors in some of the full logistic models may have been overestimated because of the failure of the logistic regression to account for the covariance component of the overall variances. Nevertheless, the full logistic models based on data from north of the Aleutian Islands yielded *Kappa* values which suggest fairly strong capabilities of these models to predict probabilities of occurrence of coral north of the Aleutians.

Application of GEE is based on asymptotic theory, so a large number of clusters are needed to yield consistent results. The number of clusters (in our case, the number of transects), rather than sampling units within clusters (i.e. video frames within transects), determines adequate sample size (Stokes et al. 2000). In modeling situations with just a few candidate explanatory variables, 25 clusters may be sufficient to avoid misleading results. With more than 5 explanatory variables, a minimum of 100 clusters is needed and around 200 clusters are needed for the GEE test statistics to yield assessments of parameter significance with confidence

(Stokes et al. 2000). Despite the large number of sampling frames (100+) within most of our transects, the emphasis on the number of clusters for effective application of GEE regression modeling, and the relatively small numbers of clusters in our surveys ($n = 27$), may have contributed to the relatively poor performance of the GEE models.

Implications for conservation

In recognition of the uniqueness of deep-water coral habitats in the Aleutian Islands, and in response to a federal mandate as well as scientific and public interest for conserving essential fish habitat, the North Pacific Fishery Management Council established the Aleutian Island Habitat Conservation Area (AIHCA) in 2005, closing 950 463 km² to bottom trawl gear (Witherell & Woodby 2005). Traditional bottom trawl areas totaling 42 611 km² were left open, essentially 'freezing the footprint' of areas already impacted by trawling. Additional protection was afforded to 6 sites with especially high densities of corals, encompassing coral gardens discovered during the 2002 submersible explorations (Stone 2006). These areas, totaling 377.3 km², are closed to all bottom contact fishing gear, including trawls, longlines, and pots. Based on the visual display of model predictions presented here (and in the supplement), there are large swaths of seafloor with potential for supporting coral gardens, as seen for 8 of the 9 northern study sites and 2 off-site areas. These areas predicted to have potential coral gardens include large areas within the AIHCA.

CONCLUSIONS

The successful models demonstrated the potential for using relatively few physical factors, which in this case were all derived from bathymetry, to predict deep-water coral distribution. The rejection of slightly more than one-half of the models suggests that there are some potentially serious limitations of the input data and/or the models. Significant among these is the tremendous heterogeneity between transects and between sites, as well as the sparse distribution of transects across sites.

It is possible that the inclusion of more parameters in logistic models, compared to the GEE models, is at least partly an artifact of spatial correlation and not completely indicative of an inherent underlying relationship between coral occurrence and depth, or the depth-derived habitat factors of slope and rugosity.

Improvements in our ability to more precisely predict locations of coral gardens will require additional model

development, potentially including the addition of substrate and current data as well as additional transect observations from existing video collections to better represent coral distribution across the diversity of sites and substrates. Additional surveys focused on ground-truthing model predictions could also help refine and/or substantiate the predictive models.

Acknowledgements. This study is a product of undersea investigations in the central Aleutian Islands undertaken as collaboration between NOAA Fisheries, the Alaska Department of Fish and Game (ADF&G), the University of Alaska Fairbanks (UAF), and Moss Landing Marine Laboratories. The project was funded by NOAA Fisheries, the North Pacific Research Board, and ADF&G, and was led by Drs. J. Heifetz (NOAA), J. Reynolds (UAF), and D. Woodby (ADF&G), and Mr. R. Stone (NOAA). We thank the multi-agency field teams and the crews of the RV 'Velero IV', the RV 'Roger Revelle', Delta Oceanographics, and Woods Hole Oceanographic Institution for their support during fieldwork. Thanks also go to R. Stone, E. Brown, C. Martinez, and D. Alcorn for reviewing video tapes for data collection. Two anonymous reviewers provided thorough and thoughtful reviews that were especially helpful. This is contribution no. PP-279 of the ADF&G.

LITERATURE CITED

- Allison PD (1999) Logistic regression using SAS: theory and application. SAS Institute, Cary, NC
- Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. *Ecol Model* 157:281–300
- Bryan TL, Metaxas A (2006) Distribution of deep-water corals along the North American continental margins: relationships with environmental factors. *Deep-Sea Res I* 53: 1865–1879
- Bryan TL, Metaxas A (2007) Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific continental margins of North America. *Mar Ecol Prog Ser* 330:113–126
- Cimberg RL, Gerrodette T, Muzik K (1981) Habitat requirements and expected distribution of Alaska coral. Final Report 54. NOAA, OCSEAP, Silver Spring, MD, p 207–308
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
- Davies AJ, Wisshak M, Orr JC, Roberts JM (2008) Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). *Deep-Sea Res I* 55:1048–1062
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24:38–49
- Genin A, Dayton PK, Lonsdale PF, Speiss FN (1986) Corals on seamount peaks provide evidence of current acceleration over deep-sea topography. *Nature* 322:59–61
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8: 993–1009
- Heifetz J (2003) Deep sea coral distribution and habitat in the Aleutian Archipelago. Proposal to the North Pacific Research Board, Anchorage, Alaska. Available at http://doc.nprb.org/web/03_prjs/R0304_Heifetz_all_final.pdf
- Heifetz J, Wing BL, Stone RP, Malecha PW, Courtney DL (2005) Coral of the Aleutian Islands. *Fish Oceanogr* 14(Suppl):131–138

- Heifetz J, Woodby D, Reynolds J, Stone RP (2007) Deep sea coral distribution and habitat in the Aleutian Archipelago. North Pacific Research Board Final Report 304. Available at http://doc.nprb.org/web/03_prjs/r0304_final.pdf
- Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83:2027–2036
- Horton NJ, Lipsitz SR (1999) Review of software to fit generalized estimating equation regression models. *Am Stat* 53:160–169
- Hosmer DW, Lemeshow S (2000) *Applied logistic regression*, 2nd edn. John Wiley, New York
- Landis JR, Koch GG (1977) The measurements of observer agreement for categorical data. *Biometrics* 33:159–174
- MacIsaac K, Bourbomnais C, Kenchington E, Gordon D Jr, Gass S (2001) Observations on the occurrence and habitat preference of corals in Atlantic Canada. In: Willison JHM, Hall J, Gass SE, Kenchington ELR, Butler M, Doherty P (eds) *Proceedings of the first international symposium on deep-sea corals*, 30 Jul to 3 Aug 2000. Ecology Action Center and the Nova Scotia Museum, Halifax, p 58–75
- Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *J Appl Ecol* 38:921–931
- McCullagh P, Nelder JA (1989) *Generalized linear models*. Chapman & Hall, New York
- Mortensen PB, Buhl-Mortensen L (2004) Distribution of deep-water gorgonian corals in relation to benthic habitat features in the Northeast Channel (Atlantic Canada). *Mar Biol* 144:1223–1238
- Mortensen PB, Hovland MT, Fossa JH, Furevik DM (2001) Distribution, abundance and size of *Lophelia pertusa* coral reefs in mid-Norway in relation to seabed characteristics. *J Mar Biol Assoc UK* 81:581–597
- SAS Institute (2004) *SAS/STAT® 9.1 user's guide*. SAS Institute, Cary, NC
- Stokes M, Davis CS, Koch GG (2000) *Categorical data analysis using the SAS system*, 2nd edn. SAS Institute, Cary, NC
- Stone R (2006) Coral habitat in the Aleutian Islands of Alaska: depth distribution, fine-scale species associations, and fisheries interactions. *Coral Reefs* 25:229–238
- Tittensor DP, Baco AR, Brewin PE, Clark MR and others (2009) Predicting global habitat suitability for stony corals on seamounts. *J Biogeogr* 36:1111–1128
- White M, Mohn C, de Stigter H, Mottram G (2005) Deep-water coral development as a function of hydrodynamics and surface productivity around submarine banks of the Rockall Trough, NE Atlantic. In: Freiwald A, Roberts JM (eds) *Cold-water corals and ecosystems*. Springer-Verlag, Berlin, p 503–514
- Witherell D, Woodby D (2005) Application of marine protected areas for sustainable production and marine biodiversity off Alaska. *Mar Fish Rev* 67:1–27
- Wright DJ, Lundblad ER, Larkin EM, Rinehart RW, Murphy J, Cary-Kothera L, Draganov K (2005) ArcGIS Benthic Terrain Modeler. Davey Jones' Locker Seafloor Mapping/Marine GIS Laboratory, Oregon State University, Corvallis, OR

Submitted: February 27, 2009; Accepted: September 7, 2009

Proofs received from author(s): December 11, 2009