

Defining EFH for Alaska Groundfish Species using Species Distribution Modeling

Principal Investigators: Chris Rooper, Ned Laman, Dan Cooper (RACE Division, AFSC)

Research Priority: This study will address Research Priority #1 – Characterize habitat utilization and productivity. We will elevate EFH information for all FMP groundfish species to level 2 using analysis of existing data sets (primarily trawl survey data and existing benthic habitat information). We will provide maps of essential fish habitat on a regional level for three Alaska regions (eastern Bering Sea, Aleutian Islands and Gulf of Alaska). Where possible, we will utilize available life history information to provide maps of essential fish habitat for the different life stages of each species.

Justification: In Alaska, most EFH descriptions for groundfish are limited to qualitative statements on the distribution of adult life stages. These are useful, but could be relatively easily refined both in terms of spatial extent and life history stage using species distribution models and available data. Distribution models have been widely used in conservation biology and terrestrial systems to define the potential habitat for organisms of interest (e.g. DeLong and Collie 2004, Lozier et al. 2009, Elith et al. 2011, Sagarese et al. 2014). Recently species distribution models have been developed for coral and sponge species in the eastern Bering Sea and Aleutian Islands (Rooper et al. 2014, Sigler et al. in review). The models themselves can take a number of forms, from relatively simple frameworks such as generalized linear or additive models to complex modeling frameworks such as boosted regression trees, maximum entropy models, two-stage models or other formulations. The models can be used to predict potential habitat, probability of presence or even abundance, but they all have some features in common.

- the underlying data consists of some type of independent variables (predictors) and a dependent response variable (presence, presence/absence or abundance)
- raster maps of independent variables are used to predict a response map
- confidence bounds on the predictions and partitioning of the data can produce test statistics useful for evaluating the model

Project Description: We are proposing to use a species distribution modeling framework to refine the descriptions (to level 2) of Essential Fish Habitat for Alaskan groundfish species. This will be attempted for each of the Alaska regions and for all groundfish species (see species list in Appendix 1). The independent variables will consist of those variables (such as depth, slope, bottom water temperature, current speeds, etc.) widely available from remote sensing or long-term monitoring programs at the AFSC. The dependent variables will be survey catches (primarily bottom trawl, but we will include pelagic surveys and ichthyoplankton surveys where available) of the Alaska FMP species. Where possible, the species will be divided by life history stage into egg/larval, juvenile and adult groups.

Because of the anticipation of variable data distributions (from log-normal to highly zero-inflated) a number of model frameworks will be considered, and the most appropriate for each species/life history stage will be used. An example preliminary analysis for Sablefish in the eastern Bering Sea is given below. Most of the habitat variables have already been accumulated for the GOA and AI as part of ongoing coral and sponge distribution modeling.

Example: Sablefish habitat model in the EBS slope

As an example of the modeling that will be performed for all species, we constructed a distribution model for sablefish (*Anoplopoma fimbria*) for the outer shelf and slope region of the eastern Bering Sea. The model is formulated as a two-stage (or hurdle) model, with the first stage producing estimates for probability of presence or absence and the second stage producing estimates of abundance. Bottom trawl survey data from EBS slope and shelf surveys (2002, 2004, 2008, 2010, and 2012) were used to construct the models.

Habitat Variables. Independent variables for modeling included the standard suite of habitat variables typically collected on the bottom trawl survey as well as a few derived and modeled variables. Haul positions (start and end) and depth were collected during each bottom trawl haul. The mid-point of the start and end positions of the net was used as the location variable in the modeling. The longitude and latitude data for each tow (and all other geographical data including the raster layers described below) were projected into Alaska Albers Equal Area Conic projection (center latitude = 50° N and center longitude = -154° W) and transformed into 100 m by 100 m square grids of eastings (longitude) and northings (latitude) for modeling.

The depth for each tow was estimated from a SeaBird SBE-39 microbathymograph attached to the headrope of the net plus the measured net height. Mean depth during the tow was calculated for inclusion as an explanatory variable in the modeling. A bathymetry raster for the entire Bering Sea region was also produced for this analysis. This raster was used for prediction, but not for parameterizing the models. Slope and rugosity were two habitat variables derived from the 100 m by 100 m bathymetry raster. Slope for each raster grid cell was computed as the maximum difference between the depth at a cell and its surrounding cells. The average summer water temperature at each site was estimated from data collected during EBS slope surveys. The temperature data used in our models were primarily designed to reflect long-term averages that could be compared spatially to the distribution of corals and sponges. Mean bottom temperature underneath each bottom trawl tow path was used as a habitat variable in the modeling. The 100 m by 100 m raster layers of average temperature were used for prediction.

Three measures of water movement and its potential interaction with the seafloor were used as habitat variables in modeling and prediction. The first variable was the maximum tidal speed on a 10 km² grid. Tidal speeds were estimated for 368 consecutive days (January 1st, 2009 to January 3rd, 2010) using a tidal inversion program parameterized for the eastern Bering Sea. This tidal prediction model was used to produce a time series of one year of tidal currents for spring and neap cycles at grid node. The mean values were then interpolated to a 100 m by 100 m grid using inverse distance weighting. The mean of the time series of predicted tidal current was then extracted for the position of each bottom trawl survey haul. This mean value was used as a habitat variable in the modeling.

The second water movement variable was the predicted bottom water layer current speed from ROM's model runs from 1970-2004. This long-term current speed and direction were available as points on a 10 km by 10 km grid. This regularly spaced data was interpolated to a 100 m by 100 m cell size raster covering the EBS slope using inverse distance weighting. Then the values

from this raster at each of the bottom trawl survey haul locations were extracted and the mean value computed for the path of each bottom trawl survey tow. The raster was also used for prediction.

To reflect average ocean productivity ($\text{g C m}^{-2} \text{ day}^{-1}$) at each of the bottom trawl survey sites, we used MODIS ocean color data for five spring-summer months (May-September) that encompass the spring and summer phytoplankton blooms over eight years (2003-2011) for the eastern Bering Sea region. These data were downloaded from the Oregon State University Ocean Productivity website. These data were averaged by cell and by month and then averaged again by cell and by year (to account for differences in the number of samples within each cell). The averages were then interpolated to 100 m by 100 m raster grids using inverse distance weighting. The mean value in this grid underlying each bottom trawl survey tow was extracted from this raster. The raster was used for prediction.

Eastings (longitude) and northings (latitude) were very strongly correlated because of the geographical shape of the eastern Bering Sea slope and as such were included as a bivariate term (location) in the model. The remaining habitat variables were used in a univariate form.

Model Fitting. Two-stage models were fit to the bottom trawl survey catches for sablefish. In the first stage, presence or absence was predicted using all bottom trawl survey hauls ($n=1357$) on the outer shelf and slope. In the second stage $\log(\text{CPUE})$ was predicted using only bottom trawl hauls that captured sablefish ($n = 483$). Generalized additive models (Hastie & Tibshirini 1990) using the *mgcv* package in R (Wood 2006) were used to predict the two dependent variables with the suite of untransformed habitat variables included, so that the full model was

$$y = s(\text{location}) + s(\text{depth}) + s(\text{temperature}) + s(\text{slope}) + s(\text{mean tidal current}) \\ + s(\text{mean current speed}) + s(\text{ocean color}) + s(\text{phi}) + s(\text{sort}) + \varepsilon$$

where y was the dependent variable presence or absence and abundance of sablefish in bottom trawl hauls and s indicates a thin plate regression spline smoothing function (Wood 2006). In each case the basis degrees of freedom used in the smoothing function was limited to ≤ 4 for univariate variables and ≤ 30 for the bivariate term (location). For presence or absence models a binomial distribution was used for the fitting. The Gaussian distribution with log-transformed CPUE data and a constant of half of the smallest positive value proved to best approximate normality for the CPUE data for sablefish.

A factorial analysis was used to reduce the number of variables in each model. Initially a full model containing the entire variable suite was fit to the data. Then the least significant variable was removed from the model, provided the GCV score for CPUE models or the UBRE score for binomial models was lower with the elimination of the variable, and then the reduced model was re-fit to the data. Stepwise variable removal was continued until a final best-fitting model was reached, where removal of additional variables did not result in a reduction in GCV or UBRE.

The best fitting presence-absence model for sablefish included the location variable (bivariate term of latitude*longitude), bottom depth, seafloor slope, bottom temperature, tidal current, ocean color, and the two measures of sediment character; sorting and phi (Figure 1). The model explained 68% of the deviation in the data set. The model fit very well, with an area under the receiver operating curve (AUC) of 0.97, which is considered “outstanding”. A threshold probability for determining presence or absence was set at 0.5 (i.e. for trawl hauls with a probability of presence of sablefish >0.5, presence was designated). Using this threshold resulted in about a 9% error rate for predicting presence or absence.

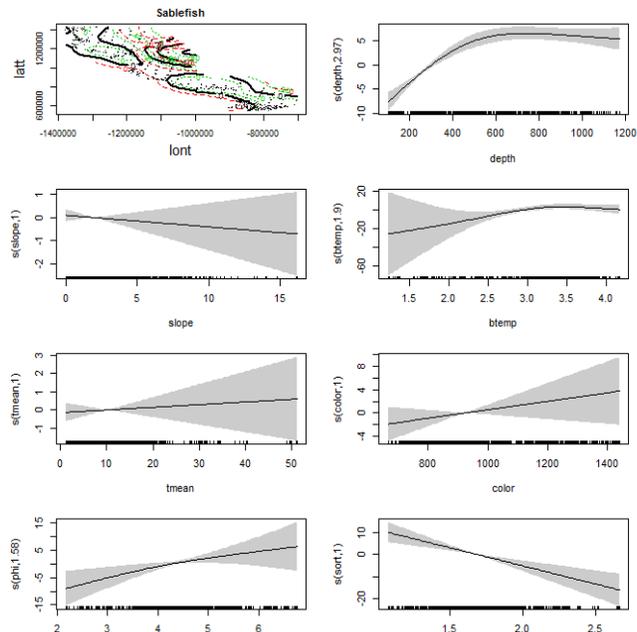


Figure 1. Response of sablefish probability of presence with significant variables in the best fitting model of sablefish presence or absence.

The best fitting model of abundance (trawl survey log(CPUE)) included; location, depth, slope, bottom temperature, ocean color and phi (Figure 2). This model also fit the data well, explaining 43% of the deviation in the data set. A map of abundance showed that occurred at medium depths (500-800 m) along the slope (Figure 4). Almost no sablefish were predicted to occur on the eastern Bering Sea outer shelf.

Interestingly, most of the areas of high abundance of sablefish were predicted to occur in canyons (Bering, Pribilof and Zhemchug).

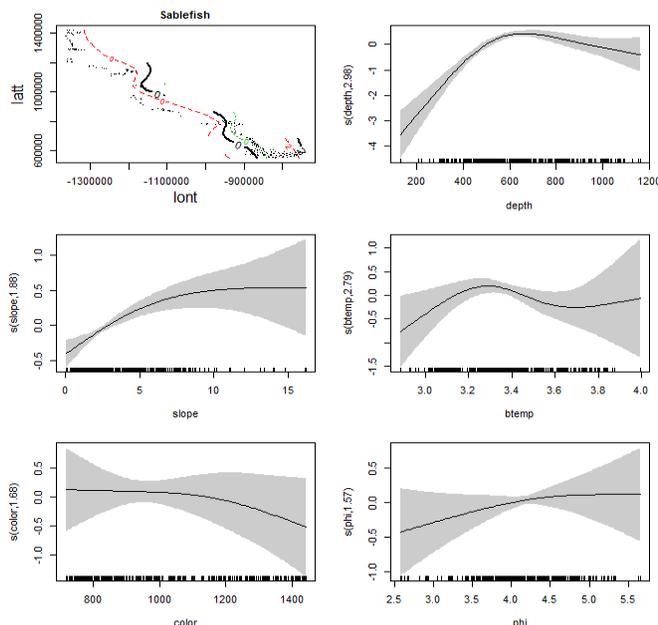


Figure 2. Response of sablefish log-transformed CPUE to significant variables in the best fitting model. Data were only bottom trawl hauls with sablefish present.

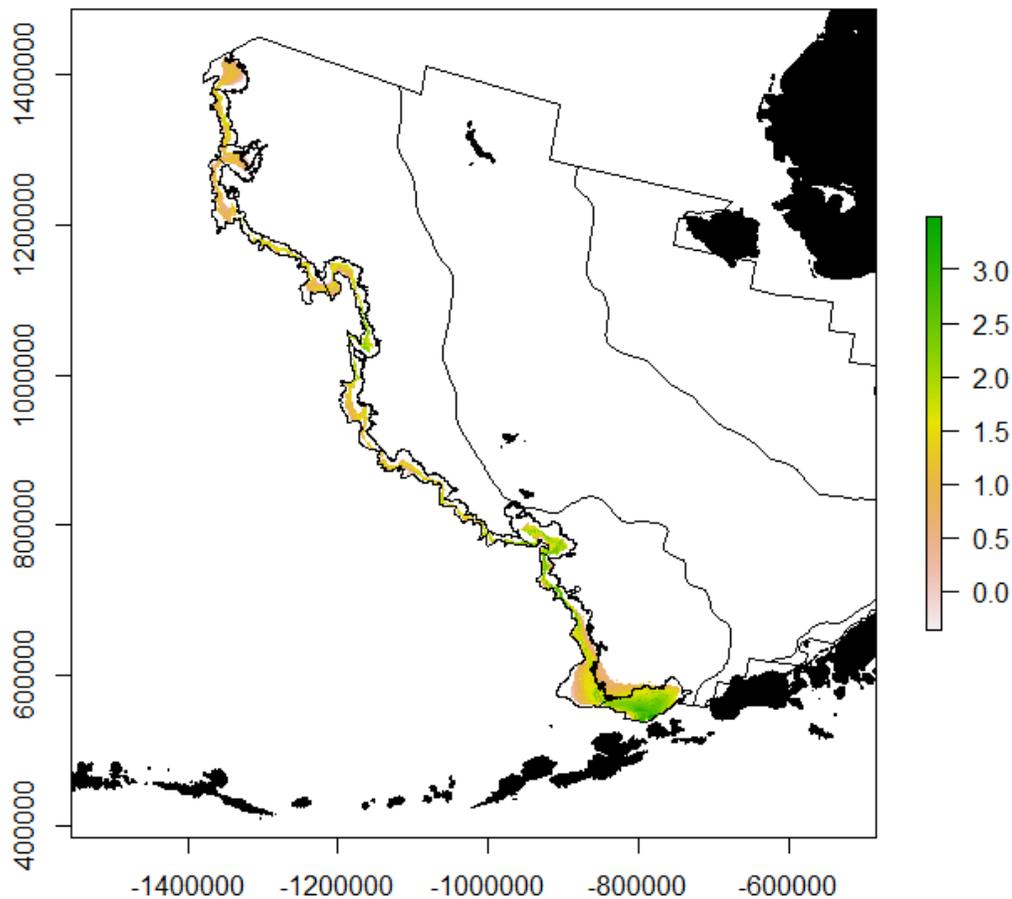


Figure 3. Predicted sablefish abundance (log-transformed) from the hurdle model using a threshold value of 0.5, meaning that at grid cells where the probability of presence is predicted to be >0.5 , sablefish abundance is predicted using the two-stage model.

Expected Products: A NOAA Technical Memorandum will be delivered that describes the individual species modeling results. The document will produce maps of the distribution of each species in each region for all life history stages where modeling can be accomplished. ArcGIS coverages will also be provided for each species, region and life history stage that can be incorporated into SAFE documents and used for further analyses. Finally, a manuscript describing the general methodology and results will be produced for a peer-reviewed journal.

Timeline: We anticipate beginning preliminary work on this project in November 2014. A complete set of models for 2-5 example species (i.e. Pacific ocean perch, sablefish, arrowtooth flounder and Pacific cod) will be presented to the NPFMC SSC in February 2015. The contractor would begin work on the additional species using the template developed on the example species in March 2015 (assuming funds become available in March). Data compilation should take 2-4 months, modeling should be completed in 4-6 months and the document could be produced in October 2015. The final manuscript would be completed by March 2016.

References:

- DeLong, A.K. and J.S. Collie. 2004. Defining Essential Fish Habitat: A Model-Based Approach. Rhode Island Sea Grant, Narragansett, R.I. 4pp.
- Elith, J. Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., and Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Biodiversity Research* 17:43-57.
- Hastie TJ, Tibshirani RJ (1990) *Generalized Additive Models*. Chapman & Hall, London
- Lozier, J.D., Aniello, P. and Hickerson, M.J. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modeling. *J. Biogeogr.* 36:1623-1627.
- Rooper, C.N., M. Zimmermann, M. Prescott, A. Hermann. 2014. Predictive models of coral and sponge distribution, abundance and diversity in bottom trawl surveys of the Aleutian Islands, Alaska. *Mar. Ecol. Prog. Ser.* 503:157-176.
- Sagarese, S.R, Frisk, M.G., Cerrato, R.M., Sosebee, K.A., Musick, J.A., and Rago, P.J. 2014. Application of generalized additive models to examine ontogenetic and seasonal distributions of spiny dogfish (*Squalis acanthias*) in the Northeast (US) shelf large marine ecosystem. *Can. J. Fish. Aquat. Sci.* 71:847-877.
- Sigler, M.S., Rooper, C.N., Hoff, G.R, Stone, R.P., McConnaughey, R.A. and T. K. Wilderbeur. In review. Are Bering Sea canyons unique habitats within the eastern Bering Sea? *Mar. Ecol. Prog. Ser.*
- Wood SN (2006) *Generalized Additive Models: an introduction with R*. Chapman and Hall/CRC. 391 p

