**Read Me for estimating spatio-temporal models.**

goaTrawl.Rproj

#####

**inlaScript Estimation + Proj.R**

#####

This is the workhorse script for estimating the spatial model using the integrated nested Laplace approximation for the spatial field. This script can estimate either a binomial model for each species to provide the probability of occurrence (logit link, ar-1 spatial field) or a model for the abundance of each species (“positive” model conditional on the presence of each species: Gamma distribution with log link, ar-1 spatial field).

The very top of the script allows you to toggle between occurrence and positive models – the ‘model’ argument. It also determines whether you want to do a single model for each species totalBiomass = TRUE, or if you want to divide species into large and small categories totalBiomass = FALSE [this is described more below]. Finally, we can define whether we want to estimate a distinct intercept for each year of the model or a single intercept for all years – single.intercept = TRUE or FALSE.

This scripts reads in these data files to work:

goa\_projection\_points+temp.csv – provides the locations for predicted points

goa\_trawl\_final\_albers+temp.csv – provides the trimmed data for each trawl observation. Different species are represented in each column. Spatial locations are generated using an Albers projection. See “\_Processing GOA trawl survey data.docx” for a detailed description of the creation of this file.

goa\_trawl\_final\_size\_albers+temp.csv - provides the trimmed data for each trawl observation for six species of groundfish. The six species different species-size classes are represented in each column. Spatial locations are generated using an Albers projection. See “\_Processing GOA trawl survey data.docx” for a detailed description of the creation of this file and dependencies.

It creates a relatively fine mesh to estimate spatial the spatial field. At present, we use a mesh with 268 vertices. This seems to provide a balance between computational speed and predictive accuracy.

**Plotting Files:** This script calls other plotting files automatically. These two make preliminary plots of some of the marginal effects of either pres-abs data or the positive data.

trawl\_plot\_binom.r, trawl\_plot\_positive.r

These two make simple diagnostics of the fit of the models:

Plot model diagnostics pres.R, Plot model diagnostics pos.R

For each model run this script creates a few plots (see above) and a large data file of the form “SpeciesName\_pos\_;sing\_int=TRUE.Rdata” with “pos” indicating this is a positive model and “sing\_int” indicating that there is a single intercept. “Pos” can be replaced by “binomial” and sing\_int could be FALSE.

This is the model output that can be used to make projections and subsequently used to make predictions about various attributes of the groundfish portfolio.

#####

**Groundfish choose discrete areas.r**

#####

This file identifies areas in the Gulf of Alaska that can be used for index standardization for determining portfolio effect. It is largely a manual identification of areas based on depth and natural breaks in topography.

It uses goa\_projection\_points+temp.csv as data and produces: goa\_discrete\_areas\_for\_comparison(50\_to\_150m).csv. This file can then read in and used by the various projection files (e.g. processINLAOutput\_multispecies.r)

**Groundfish Methods.**

**Data:**

To assess spatial and temporal patterns in groundfish communities we used fisheries independent surveys conducted by the Alaska Fisheries Science Center (AFSC). between 1984 and 2011 (AFSC survey REF).

The average latitude and longitude of survey tow were used to represent the spatial locations of each sample, and these values were converted to ALBERS PROJECTION which is appropriate for the Gulf of Alaska. In addition to the spatial locations, we included information about the bottom depth (m). Both variables are thought to influence catchability, with depth as a proxy for light, and SST a function of upwelling (Pearcy 1970, Hannah 2011). The average depth of each trawl (m) was treated as a linear and quadratic predictor, following similar estimated relationships between depth and groundfish (Shelton et al 2014). Like other species distribution models, this framework is easily extended to include other static or time-varying habitat characteristics as predictors... BUT we didn’t include them.

This AFSE dataset has used the same methodology over the time series (stratified random sampling design, with the same bottom trawl sampling gear throughout the time series). As our interest was primarily in the waters likely most affected by the Exxon Valdez oil spill, we restricted our analysis to trawls conducted shallower that 600m. In total this dataset includes 7601 tows between 1984 and

We attempted to estimate model for 55 species groups. These groups represent species that were observed in at least 270 survey tows. XX represent individual species and YY represent species complexes that represent were not identified to species separately during field sampling throughout the survey extent (TABLE). For brevity, we refer to these as species and species groups simply as “species” subsequently. We provide the full species list in the supplement.

For 5 abundant species, we used the size distribution data to divide our groundfish data into two components…. Big individuals and small individuals. Something about how the cutoff size was identified (chat with Mary H.)

**Things to do for Data:**

Make table of sample sizes by year for GOA surveys.

Make map of trawl locations for each year.

* Write detailed methods for how I separated out the different sizes
  + Dealing with no observed size data.
  + Converting observed length to weight using allometric equations.
  + Determining which fraction are “big” versus “small” and applying that to the CPUE data to generate CPUE for each size class.
  + What is the “UE” in CPUE (per km^2? Hectare?)

**Statistical modeling**

We constructed separate models for each groundfish species to understand the spatiotemporal patterns of occurrence and abundance. We estimated a model for each species independently and subsequently combined the models to produce a range of multi-species metrics of the groundfish communities.

In all models, we used weight (kg) observed on each trawl as the response variable. Because most species were absent from a large number of observed trawls, we separately modeled the presence and distribution of eulachon density, adopting a delta-GLM approach with two sub-models (Pennington 1983, Maunder and Punt 2004, Shelton et al. 2014). Probability of occurrence of the *i*th species in year *t* for a set of locations ***s,*** *,* was modeled using a binomial GLMM with logit link,

where represents a matrix of fixed effect covariates at locations ***s*** in year *t*, represents a vector of coefficients for species *i*, and represents a vector of spatial random effects that follow a first-order autoregressive process:

where *ρ* represents the degree of autoregression in encounter probabilities and Σ represents spatial covariation in random effects (discussed below). Spatial random effects were assumed to be autoregressive to account for variation not explicitly included in our model (variation due to the environment or population processes such as density dependence). Note that because trawl surveys are not conducted annually but triennially (1984-1999) or biennially (1999-2011) the autoregressive term refers to the date of the previous survey year, not the previous calendar year.

To model the distribution of groundfish as the second sub-model, we assumed that groundfish CPU was gamma distributed and used a log-link and modeled as

(4)

where is a matrix of covariates corresponding to each haul location (similar to ), represents the estimated species-specific coefficients, represent spatial random effects that again follow an autoregressive process (similar to eqn 2, but with a independently derived covariance matrix), and represents the **shape** parameter of the gamma distribution.

For both covariance matrices of the spatial random effects in the presence-absence and positive models, we used the Matern function to model covariance as a function of Euclidian distance, so , is the covariance between location and , is the spatial variance, and represent the gamma and Bessel functions, respectively, is the Euclidian distance between locations and , and is an estimated scaling parameter (Lindgren et al. 2011). The parameter controls the smoothness of the Matern function and is usually fixed rather than estimated from data (when = 0.5, the Matern reduces to the simpler exponential covariance function). Following previous work, we chose = 3/2; this allows the Matern to be more flexible than the exponential, but also allows the function to be differentiable (Rasmussen & Williams 2006). The covariance matrices for the presence-absence and positive models for each species have separate parameters , and , reflecting the assumptions that each model component may have a different variance or rate at which correlations decline as a function of distance. Further details can be found in (Ono et al. 2015, Ward et al. 2015).

Estimation of latent Gaussian Markov random fields is difficult, and can be

challenging in a Bayesian framework if the dimensionality of the knots or locations is large (Shelton et al. 2014). Recent advances have allowed the spatial covariance matrix to be approximated via stochastic partial differential equations (SPDE) as calculated within INLA (Rue et al. 2009, Ruiz-Cardenas et al. 2012). More specifically, INLA approximates the inverse of the spatial variance-covariance matrix of fixed locations using three large sparse matrices (see Thorson et al. In press for more details). Estimation of the fixed effects is then done via maximum marginal likelihood using the Laplace approximation to approximate the integral across random effects, and random effects are estimated via Empirical Bayes. Using these estimates (and Bayesian priors on fixed effects), INLA allows Monte Carlo samples to be generated from the posterior distribution, as the Laplace approximation to the marginal likelihood.

Write:

Generating MCMC samples from the estimated model objects.

To produce estimates of occurrence and density and We generated 1,000 samples from the approximate posterior distribution