**Focal questions for groundfish**

1. **Have there been detectable changes in spatial distribution of groundfish related to oil spill (space/time)?** 
   1. **Different responses depending on life history characteristics?**
   2. **Shifts in diversity in groundfish assemblages.**

There are about 5 papers here. Basic thrust is descriptive. Do we see any difference

I’m going to outline the output we have in hand that allow a lot of flexibility with the kinds of analysis we can do. Then I’ll then sketch out one paper.

**DATA AND OUTPUT:**

We collected the AFSC groundfish trawl survey data from 1984-2011 (occurs every 3 years to 1999 then every 2 years after 1999) and estimated spatio-temporal models for each species independently. We estimated this as a hurdle model so there are two independent model components: one that describes the probability of the species occurring (presence-absence) and a second component that described the abundance conditional on the species being present. From these estimated models, we can produce predicted abundances (with uncertainty) for each species. At the moment, we have the ability to predict the abundance at the center of each 2x2km grid cell for the entire Gulf of Alaska. So this output is a summary of the species distribution for the 55 most abundant species in the depth range of AFSC trawl survey (approximately 50-800m). I put a list of the species at the end of this document (including 2 species complexes). There are a lot of different ways to summarize this output, as will be seen below.

**Paper 1 (there is a lot here right now. Prob need to trim.)**

Divide the gulf of Alaska in to a number of discrete chunks. Like so:



Each of the numbered circles is a shallow area that should have approximately similar fish communities based on physical parameters (50 to 150m) and they were all surveys before the exxon spill (1984 and 1987) and after the spill (1990 and on). They experienced different levels of direct exposure to the oil spill, though. We would expect that oil spill effects would be most intense near the oil (in space and time) and decline as a function of both distance from the spill and time since the spill. So we would expect to see the largest changes at the site number 3,4,5, somewhat less at 2,6,7, and much less at 1,8-11.

Need to do some thinking about how to best standardize among the areas (bottom area, other things).

So how do we measure changes among these areas to look at the effect of the spill? There are a bunch of ways.

For any given metric:

1. Treat each area as a univariate time-series. Ask if the metric changes notably after the spill.
   1. Do some metrics change while others do not.
2. Look at all areas simultaneously. Do the aggregate properties across all areas change with the spill
   1. Is there greater variance among sites
   2. Does the pairwise covariance between site change in some way?
   3. Are areas becoming more similar or more distinct post-spill?

This leads to the question- which metrics are the most interesting?

Here’s the list that we came up with in March:

Multi-species metrics

- Total Biomass (all species)

- Dominance

- Evenness

- Simple richness

Compare each of these metrics divisions dividing species up by life style:

* Benthic / Pelagic
* Feeing Guilds
  + Sarah Gaichas divisions
* Unresolved
  + treat species divided into multiple life-history categories as separate entities or not?

To DO:

Define prediction areas for AREAs 1-9. (DONE!)

Make examples maps for everyone to look

-two weeks before meeting, send out an update and example output.

Mary

**There are a bunch of other things that are described in the previous doc. I just pasted them below:**

**Hypotheses:**

1. Temporal comparisons in EVOS area  
   - Each species and aggregation

* Plot metric vs. time for a few metrics in oil spill area vs. non affected area.
* Proportion of change in species composition.
* Dominance - rate of change
* Diversity (Shannon-Weiner Index, Eveness)
* Come up with metric of oil spill intensity - where and when oil dispersed (ask Tammy)

Multiple “Control” Areas

- Decide which regions to compare

* Control areas in eastern and western GOA and the plume area. Try to have more than 1 reference area
* Compare at depths in and out of area, either control vs. impacted area or distance from source.

- Again, emphasis should be on species abundance

Potential model covariates for mixture model:

Temperature (Botttom) -> ROMS (Ben is working on this)  
 Fishing (Catch data – ask harvest portfolio group about this)

? Oil spill covariate (probably leave out of model)

? River runoff (freshwater – maybe the Royer index – need something discrete)

? Habitat (sediment information/grain size)

Metrics for comparison:

Multi-species metrics

- Total Biomass (all species)

- Dominance

- Evenness

Habitat divisions

- Benthic: Pelagic

Functional feeding group (guilds)

- Sarah Gaichas Divisions

- Include size info (immature vs. mature) for subset of species, including pollock, cod and halibut

Life-stage divisions

Size structure

**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

This AFSC 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; BEST REF?). 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 we included 7601 tows between 1984 and 2011.

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 6 abundant species, we used the size distribution data to divide our groundfish data into two components…. Big individuals and small individuals. We follow some ecopath models developed by Kerim and friends and use a 20cm cutoff to divide individuals into juvenile and adult size classes.

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 model outputs to generate a range of multi-species metrics of the groundfish communities. We first discuss the spatio-temporal statistical model before describing spatio-temporal metrics of community change.

In all models, we used catch per unit effort (kg/ hectare) 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,

**(1)**

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:

**(2)**

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.

For the second groundfish sub-model (hereafter positive model), we assumed that groundfish CPU was gamma distributed and used a log-link,

**(3)**

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). Then represents the mean of the gamma distribution and represents the **shape** parameter of the gamma distribution. This model only uses the hauls in which the species was observed and so the positive model describes the CPUE of each species conditioned on the species presence.

For all models we used available haul level information about bottom depth (m) and included both linear and quadratic terms for log(depth) in the occurrence and positive models (Shelton et al. 2014). We centered log(depth) by it’s mean before estimating the model. We also considered two models for fixed year effects: in one model we estimated a single intercept while in the other we allowed for a distinct intercept for each year. As the intercept scales the occurrence (or CPUE, respectively) for the entire region, models with variable intercepts allow for spatially uniform, region wide changes in occurrence or CPUE.

For the spatial random effects in the occurrence 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 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; Ward et al 2015). 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) and 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. 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.

For each species, we estimated two occurrence models and two positive models. For each submodel, we estimate one model with a single intercept and one with a year-specific intercept. We compare single and year-specific models using posterior predictive plots and deviance information criterion (DIC) and then identified preferred models for each species.

**Generating predictive densities**

After the two sub-models were estimated for each species, we used these estimated model to generate predicted densities for the gulf of Alaska. We projected our model estimates to the center of 2 x 2 km grid created for the entire Gulf of Alaska (add details of the projection to a supplement). We generated 1,000 Markov chain Monte Carlo samples from the joint approximate posterior density for each species and used these samples to project a predicteddensity for each

Specifically, because the occurrence and positive models are estimated independently, the unconditional expectation for CPUE of species at time *t* and slocation *s* is in units kghectare-1.

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 distri

Species list (roughly in order of abundance (frequency of observation in the trawl survey))

|  |  |
| --- | --- |
| Common | Scientific |
| Arrowtoothflounder | Atheresthesstomias |
| Pacifichalibut | Hippoglossusstenolepis |
| Walleyepollock | Theragrachalcogramma |
| Pacificcod | Gadusmacrocephalus |
| rexsole | Glyptocephaluszachirus |
| flatheadsole | Hippoglossoideselassodon |
| Doversole | Microstomuspacificus |
| sablefish | Anoplopomafimbria |
| Pacificoceanperch | Sebastesalutus |
| eulachon | Thaleichthyspacificus |
| shortspinethornyhead | Sebastolobusalascanus |
| yellowIrishlord | Hemilepidotusjordani |
| southernrocksole | Lepidopsettabilineata |
| northernrocksole | Lepidopsettapolyxystra |
| rocksoleunident. | Lepidopsettasp. |
| Tannercrab | Chionoecetesbairdi |
| spinydogfish | Squalusacanthias |
| northernrockfish | Sebastespolyspinis |
| magistratearmhooksquid | Berryteuthismagister |
| longnoseskate | Rajarhina |
| searcher | Bathymastersignatus |
| darkfinsculpin | Malacocottuszonurus |
| Pacificlyrecrab | Hyaslyratus |
| bigmouthsculpin | Hemitripterusbolini |
| capelin | Mallotusvillosus |
| spinyheadsculpin | Dasycottussetiger |
| shortfineelpout | Lycodesbrevipes |
| wattledeelpout | Lycodespalearis |
| lingcod | Ophiodonelongatus |
| redbandedrockfish | Sebastesbabcocki |
| bigskate | Rajabinoculata |
| buttersole | Isopsettaisolepis |
| greatsculpin | Myoxocephaluspolyacanthocephalus |
| shortrakerrockfish | Sebastesborealis |
| harlequinrockfish | Sebastesvariegatus |
| Beringskate | Bathyrajainterrupta |
| sharpchinrockfish | Sebasteszacentrus |
| Atkamackerel | Pleurogrammusmonopterygius |
| prowfish | Zaprorasilenus |
| Englishsole | Parophrysvetulus |
| Aleutianskate | Bathyrajaaleutica |
| giantgrenadier | Albatrossiapectoralis |
| silvergrayrockfish | Sebastesbrevispinis |
| kelpgreenling | Hexagrammosdecagrammus |
| yellowfinsole | Limandaaspera |
| Pacificherring | Clupeapallasi |
| starryflounder | Platichthysstellatus |
| slendersole | Lyopsettaexilis |
| spottedratfish | Hydrolaguscolliei |
| chinooksalmon | Oncorhynchustshawytscha |
| Alaskaplaice | Pleuronectesquadrituberculatus |
| lanternfishunident. | Myctophidae |
| rosethornrockfish | Sebasteshelvomaculatus |
| redstriperockfish | Sebastesproriger |
| rougheyeandblackspottedrockfishunid. |  |
| duskyanddarkrockfishesunid |  |
| Pacifichake | Merlucciusproductus |