Meeting with Krkosek (08-15-18)

Admin:

* Get form signed
* Figure out the logistical format for project – am I meeting with him? A grad student? How often? As needed or regularly?
* Should I do email follow-up’s after meetings? Should I do weekly check-in progress reports?

My updates:

* Been reading lots of papers, I think I have a decent handle on the ecology of the system
* A few questions:
  + Koch’s postulates: has that actually been shown in this system at all? Is it a good way to go about determining causality of one agent or another?
  + Can they infect as nauplii or just once they reach copepodid stage?
  + Are the smaller salmonids (i.e. pink salmon) show higher mortality than other species of bigger salmon?

\*marty emailing me papers rn

Notes here:

Some management changes were effective, and corresponded with signs of recovery, components of wild fish dynamics

Lots of fishery data from DFO and stock assessment data

One paper uses atlanti salmon data,

One of the main issues is that it’s controversial – remains controversial because the scale is effects on individual fish – lethal or sublethal effects on fish, and then taking that and making inference on a population level is hard, the epidemics are quite large, the fish naturally have very high mortality rates so if they’re going to die anyways, who cares if they get infected, but are the effects additive or synergistic – complicated analyses, all analyses are correlative in nature, and so ti remains hard to show definitive stuff, there’s a paper on atlantic salmon where they tag and release salmon and look at the differential return rates (the paper he sent is a meta analysis) – they found here an overall effect – differential survival confers a 35+% difference in survival, major correlate was baseline survival rate, if it was a bad year for salmon, the lice have huge effects, in good years it makes less of a difference

Predation and infection can be synergistic or compensatory depending on different factors

Pink and chum enter the ocean at much smaller sizes (don’t stay in freshwater for a year) – decline in pink salmon during years where epidemics have occurred, but the chum salmon don’t show a decline like the pink salmon – main predator of both is coho salmon smolts, there is an obvious pink salmon preference even when rare, because parasites are elevating the attack rate, the predation becomes more elevated on pink

Infection affects swimming speed, position within schools etc, and it definitely has an effect on attack rate

Dataset:

* Different species
* We think we see more of certain lice species on certain species
  + Even though they’re comigrating through the same environments
* First thing is to look at this quantitatively and do exploratory data analysis, comparative
  + Sockeye, pink and chum
* For some samples there will be correlates that have to do with stomach contents or potentially (followup) odolith analysis or coinfections (microparasites)
* First cut: look at differences among juveniles in ectoparasite communities
  + Comparative, field-based

Meeting with MK & Sean

* Hand lens in the boat, killing them and looking under a scope, and then under the scope and then just looking at the motiles
* Motiles are always consistent among the fish – so this can be accounted for across all species, but then in a subset, theres some juvenile ones
* Notes: summarizes the dataset
* pulled fish from collections that had a reasonable number of th three fish species we're interested
* at least five of each of the three species, (some years have more than others)
* D for site means discovery islands and then J means johnstone strait

Notes about dataset:

* each row is a fish, two columns (survey columns) aren't important
* siene ID is going to be important to go into the summary dataset
* GH and are the two copepidite for the two species and then an unkown one
* calimus life stage is grouped into younger and older (clemensi A is first two and B is second 2, lep A is first stage and B is second one)
* if I end up wanting to look at species differences between the two lice chalimus stages, and you could look at this for some of the fish (just the ones that have the fine scale lice protocol)
* this means looking primarily at pre-adult stages and then some you can look at the copepodite stages
* motile stages: bunch of columns cor caligus motiles and the lep motiles, male claigus, any preadult or adult caligus male, and then same with female, then a column with caligus motile that's preidentified but its not a gravid one - column I'll work with the most is the çlaigus motile that's not a gravid female, this is the sum of the previous four columns - column Q and R are not gravid and gravid female
* for leps, the data is the same for every single fish so that's self explanatory, two groupings of caligus and five for leps for all fish

(10/04/18)

Updates:

* Currently just looking at the data, starting to subset (with regards to site is what I’m paying attention to right now)
* Plotted a map of the sites to take a look at that, am going to make a few choropleth maps that will allow us to see if there are any interesting spatial patterns in terms of gradients or not
* Took a look at a couple plots that will be interesting once I get a bit further into it
* There’s pretty big differences between the different sites in terms of how many fish as well as lice per fish
* In terms of fish per lice, there seems to be a pretty normal distribution in terms of number of lice on a fish
* Notes:
  + Early collections weren’t really balanced between species but
  + You have to look at comparisons collections by collections (is one species more infected with one species of sea lice etc)
  + eventually:
    - species id as a response variable and to pool the data, we need hierarchical structure (mixed effects models) random effect (identity of sample collection) so that any variation among collections is represented through a random variable (random normal variable) and then variation among species that’s systematic is estimated at the within collection level
  + Collection: individual space time combination in the dataset – when the boat arrives at a particular site at a particular date (often a single employment of the net, sometimes a few)
    - So aggregate so a collection is all the fish from a particularl location on a particular date
    - If theres a lot of times when there’s multiple deployments of the net we might be able to pool data to that level but it’ll make it quite unbalanced (you’ll end up having a bunch of pinks in the first deployment and sockeye in the second) so treat it as an aggregation of fish collected on a particular day at a particular day
  + Write a loop that goes through the dataframe collection by collection and calculates summary stats:
    - Relative abundance of parasites by fish species (total number of paratsites on average, per fish, per species) – store this in an array
  + and then start looking at patterns in the collections in that array
  + think of a way to standardize the data by collection – so average number of parasites per fish per species is important, then also calculate the total number of parasites in the collection and then calculate the proportion of parasites on each of the species (that’ll be six quantities per collection to calculate) – therefore that’ll put the stats on the same scale for each collection – for the proportion quantities calculate a mean and a variance for that and see where we’re at from there

10/11/18

Updates:

* created collections based on unique time/location pairings – there are 52 collections in total
* then, according to last weeks instructions, I calculated the average number of parasites per fish per species, as well as the total number of parasites in that collection, as well as calculated the proportion of lice for each of the species
  + then I calculated the mean and variance for the proportions
  + then plotted the proportions and the mean number of lice per fish per species
* it appears that Sockeye are the worst affected, and this is controlled for (it’s not just because there’s more sockeye) because of the way it was calculated

Questions:

* Proposal
  + What is it you’re looking for? An intro, more of a methods proposal?

Partition lice for leps and caligus and do the same meanlice per fish plot (exclude the ones that don’t have species ids for the parasites)

For proposal, put in some proposed methods, and also have an introduction showing that I know what I’m talking about and places the project within the literature (specific literature on salmon and sealice and broader literature on host parasite communities and host specificity and generalist vs specialist parasites etc.) and so talk about the biological background for both the parasite and host species – then there’s conceptual stuff about parasite life histories and tradeoffs between specialists vs generalists.

Generalized linear mixed effects model – negative binomial distribution likely (count data is the parasites per fish) and count data (which can be low integer values), simplest distribution is poisson, but probably negative binomial – parasite individuals will aggregate on individual fish – more fish who have zero just by chance alone 0 could use a zero inflated poisson distribution, and negative binomial, and zero-inflated negative binomial. Random effects for collection (normal distribution) and that will allow us to account for variation among samples, which we’re not very interested in, and account for variation among parasite or host species which will be fixed effects – might fit one model for caligus, one for leps, and if there’s differene in host specificity, then the fixed effects in each of the models for the host species will differ. So that’s what we’d be looking at, we could even have parasite abundance as response variable, NOT carved out by the species it is, fixed effects for the host (general specificity) and then have general interaction term between fish species id and parasite species id, which would be a more direct test of whether parasites have host specificity

10/18/18

Updates:

* Created the new graphs
  + Essentially just repeated the same process but for each species of louse
* Working on proposal currently, presenting some of my ideas for it at lab meeting on the 29th

Model will look similar, but since the distribution won’t be gaussian, but function will be glmer(), and you can specify what the error distribution was, and if it was a binary response variable, you choose a binomial distribution, and you use this to specify the family. The syntax is the same.

As a first step, fit some models using poisson family in glmer, and that’s gonna be good practice, but it’s not actually correct because the data are probably overdispersed and the poisson doesn’t account for this (overdispersed means more zeros than you would have by chanve and way more parasites than you would have just by chance).

As step 2, in order to deal with overdispersion, so we’re going to use a different package which is called ‘glmmadmb’ (it might not be available yet as a download in r, but you can still download and install manually in R) – marty to send a link – info about the package will be there, as well as how to install. Jags and Rjags are also really good for fitting these types of models

When you use the family, put in poisson and zeroInlfation = FALSE to start with

Start with analyzing the caligus and leps separately and we’ll wanna have essentially just mean abundance by host species (whatever the header label is for fish species)

All.cal ~ spp -1 + (1|Collection)

For glmer …. , family = ‘poisson’

For glmmadmb …., family = ‘nbinom’, zeroinflation = FALSE

The minus 1, if you don’t have it, it takes the first level in the categorical variable and treats it as an intercept, a then your estimates for the other levels of the categorical variable are additive, so the minus 1 estimates it directly for every level of the categorical variable

For the output, we want to see what we’re doing, and then we might see the results as being simple or not. Interested in coefficient for species level effects

Make a plot with parameter estimates and two standard errors – for two species of the parasite (two sd is 95% confidence interval) and you can see if they overlap the point estimates and that’ll make it really easy to see if there’s differences – for the plot, take the estimates, exponentiate them, along with the two standard errors

Take the estimate values, plus or minus two standard errors and raise each of them e^(x) and that’ll be our estimate of parasite abundance per species per fish

Yaxis = average abundance, on the x, have pink, chum, and sockeye (whatever order) and for each of these, there’s going to be an estimate (two points, one symbol for cal, one for leps) and

Could potentially in the future look at body size covariates, etc.

Tell marty what other interesting stuff is in the data, i.e. is some of the variation also due to environmental and body size data – so see if there’s environmental data, sean might have body size data

So we need to find the estimate of average abundance of parasite sp per fish species

10/25/18

Not a ton to go over today – showed Marty results and stuff yesterday

* Put some plots and model outputs in a document and sent it to Sean and Dylan
* Have to wait for feedback from Sean before anything else really moves forwards

Spend some time reading up on the field-work methods that have gone into generating the dataset – Look at the Hakai institute website (Salmon Program) locations, frequency of sampling, how sampling is done etec.

For lab meeting: include the collection stuff as a part of the methods, the dataset, part of the program etc. – Sean has a video of them doing the purseineing – and then the preliminary results

11/01/18

* Not a ton of updates
  + Speaking with sean tomorrow
  + Working on getting the environmental stuff all shifted over, but still not quite there yet

One dataset has measurements for fish length stuff – some are measurements from lab some from field, some fish have both, and they’re often not the same number (dead fish who go through the freeze thaw, they change a bit) – for each fish that has the body size info, if it has both, take an average, and there will be a handful of fish that have no body size data, and those will be excluded from the next set of models

When I put together a data frame, because it will be two columns that I’ll have to put it into one.

Pull out the year as well as the Julian day and make those two separate columns in the dataset (try and convert julian day by a function) – this will allow us to look at using the year as a fixed effect in the model

Another thing, would be to run the models again with just the copepodite data

The infectious stage of caligus has much higher preference for sockeye, the leps can respond to chemical cues and respond preferentially to the hosts they want

Alternative explanation: initial infection is random among the fish stages, and the ones in the older life stages would actually be choosing.

There could be a lot of movement between fish in the later stages of the lice life stages – the parasites might sort preferentially on one fish species vs the other, but this requires an ability to do this.

So fix it so that column is only pre-adult and adult – and if there’s NAs in the adult and preadult stages should be removed from that group

Redo plots first, then work on merging the data, then run the models using year and body size as fixed effects