# Roadmap for demography/abundance project

I’ve tried to write down as many steps right now for this project. I have not yet gone and looked at every single script to create a really fine-grained outline but I think this captures many of the major steps.

### Workshops

SESYNC Bayesian Modeling for Socio-Environmental Data; workshop June 3-13, 2019; ([link](https://www.sesync.org/opportunities/research-short-courses/2019-bayesian-modeling-for-socio-environmental-data-short)) [taught by one of the co-authors of *Bayesian models: a statistical primer for ecologists*] [*apply by 3/15/19*]

Ecological Forecasting Summer Course; workshop July 28-August 2, 2019; ([link](https://ecoforecast.wordpress.com/summer-course/)) [taught by author of *Ecological Forecasting*] [*applied 2/15/19*]

Enhancing Linkages between Math and Ecology; 3 1 week-long workshops July 8-26, 2010 ([link](http://www.kbs.msu.edu/education/courses/#grad)) [teachers still TBD] [*application deadline?*]

### Contacts

Tom Miller (Rice) and/or Bret Elderd (LSU)

Janneke Hille Ris Lambers (UW)

Ian Breckheimer (Harvard)

Bob Shriver (USGS Southwest Biological Science Center)

### Step by step

1. Import demography data from excel to R to CSV. [*status*]
2. Identify things R is unhappy about (numbers as characters, etc.). [*status*]
3. Identify features of data to consider later in analysis (e.g. survival greater than 100%, uneven sampling across years). [*Monica Geber, Bill Morris, SESYNC*]
   1. Seedling survival > 100% - consider modeling this with an observation model in an HB framework. Bill suggests a state space model.
   2. Fruit per plant counts
   3. Seeds per fruit counts
   4. Seed bank
4. Make some decisions about what to do: filter data (toss out), correct data (e.g. adjust denominator in some years), or model observation error. [*status*]
5. Write script to clean data from csv file for analysis (incorporating the decisions about filtering or correcting data made above, if necessary). Requires knowledge of how you’re going to model the data - e.g. Stan won’t accept NAs so those need to be removed. [*status*]
6. Plot abundance time series for each site for each year in the dataset. [uses fruiting plant counts from survival dataset] [*status*]
   1. Calculate temporal coefficient of variation for each population, as well as spatial coefficient of variation. (Link and Nichols 1994, Royle on binomial mixture model?)
   2. Figure out whether temporal variability in abundance is different across the range.
7. Write script to put climate and abiotic data from excel into R to CSV. [*status*]
8. Plot climate time series for the dataset. Plot spatial patterns in climate/site variables. [*status*]
9. What are the possible questions to address/check at this point in the analysis? [*status*]
   1. Partition variation on each of the vital rate datasets. How much of the variance can be explained by the factors we are considering: time, space, within plot? [*cf. Adler & HRL 2008*]
   2. Can we identify particular vital rates where we expect low predictive ability a priori?
10. Write out each vital rate model. By this I mean write out the mathematical form of the regression model. Could also include drawing directed acyclic graphs. [*Hobbs & Hooten,* *Liz Duskey, SESYNC*]
    1. Vital rate models depend on decision in step 3.
    2. Fruit per plant and seed per fruit data is averaged in conversion
    3. Identify multiple ways that people write seed bank vital rate estimates?
11. Write stan script for seed bank. I include this as a separate step because it is not clear what steps you should take for this. This would integrate the germination data and both seed bank experiments using a latent variable of some sort. [*McElreath,* *Liz Duskey, SESYNC, Janneke HRL, Tom Miller*] [
    1. Not sure what best form to use.
    2. No time –
    3. Latent variable – (dependence of particular s2 vs. exponential decay)
    4. What are the data to validate this model?
12. Write script to join datasets (demography+climate+site) for analysis. [*status*]
13. Rescale input variables. Write script to calculate mean and standard deviation for each dataset for analysis so that parameter estimates for continuous and binary variables are directly comparable (http://www.stat.columbia.edu/ gelman/research/unpublished/standardizing.pdf). [*status*]
14. Write JAGS/stan script for , and as functions of climate, abiotic environment, and density. Use the mathematical forms you wrote out 3 steps back, and now add covariates. [*Tom Miller, Andrew Tredennick*]
15. Write R script to run stan code for , and . Also the seed bank parameters. [*status*]
16. Plot trace plot for parameter estimates to check whether chains are mixing during model fitting. [*status*]
17. Plot histogram of Brooks-Gelman-Rubin statistic to check model convergence. [*status*]
18. Save parameter estimates and posterior output to RData files. [*status*]
19. Use parameter estimates in the following ways: [*status*]
    1. Plot parameter estimates (mean + CI).
    2. Create table of parameter estimates (mean + CI).
20. Use posteriors to obtain posterior predictive for each line in the dataset for each vital rate, which you can use in the following ways: [*status*]
    1. Plot predicted values vs. residual of predicted value (scatter plot. histogram, sum of squares scatterplot).
    2. If trying out models with different priors, plot posterior densities to compare model output.
    3. Plot predicted vital rate (probability or count) in plot x at site y in year z to the observed value in that plot-site-year combination. A good model makes predictions that fall on a 1:1 line vs. observations.
    4. Plot predictions vs. some other variable. For example, look at the predicted mean probability of seedling survival at a site across years. This would be something like a plot of observed mean survival + mean predicted survival with 95% CI.
21. If able to get annual seed bank germination estimate, try to partition the portion of germinant each year that come from the seed bank vs. previous year to quantify relative importance of seed bank over time. [*status*]

[this is a stopping point] provide an updated look at the best way to look at lambda.

1. Write script to simulate climate covariation that matches the observed climate covariation. Either resample observed years, use copula to capture covariation in climate covariates by fitting multivariate probability distribution, or use Cholesky decomposition to model covariation. [*status*]
2. Import climate data and/or climate simulations to simulation. [*status*]
3. Import abiotic site covariates, and posterior distribution of parameter estimates from vital rate models to simulation. [*status*]
4. Write R script to simulate abundance dynamics. Need to think about initial conditions, and iterative process that prevents runaways (numeric overflow - R can’t deal with huge numbers that result if density dependence isn’t strong enough and shuts down). [*status*]
5. Discard initial transients, and save model output. [*status*]
6. Summarize simulated abundances from simulation to compare to observed abundances. [*Ecological Forecasting Short Course*]
   1. Comparison could be: abundance sites OR additional plots used to estimate fruits per plant
7. Plot predicted abundance vs. observed abundance at each site. [*status*]