JIMAR, PFRP ANNUAL PROGRESS REPORT: FY 2003

18 April 2003

P.I. Name: Benjamin Bolker

Project Proposal Title: Development of a hierarchical model to estimate sea turtle rookery contributions to mixed stocks in foraging habitats

Funding Agency: JIMAR

1 Purpose of the project and indicative results:

The purpose of the project is to develop general methods for incorporating ecological covariates in genetic stock analysis models. Stock analysis attempts to estimate the proportion of the individuals in a mixed population that come from each of a number of possible source populations: for example, comparing data from breeding grounds and an open-ocean population that combines individuals from many breeding grounds to figure out the importance of particular breeding grounds to the overall population. In the past, stock analysis has been based only on individual morphological or genetic measurements, such as the mitochondrial DNA haplotypes of individuals found in rookeries and in mixed-stock foraging grounds. Other ecological information such as the size of the breeding population or the distance from the breeding population to the foraging ground is often available (and ignored). We are using stock analysis of Atlantic sea turtle populations (loggerhead and green turtles) to test and develop models that include ecological covariates such as rookery size and location, and drawing initial conclusions about the more powerful or different conclusions that come from incorporating this information. In particular, we are developing hierarchical Bayesian models, which are a flexible but rigorous way to add rookery size and geographic location to stock analysis methods that have traditionally used only genetic data to try to infer the contributions from each rookery. We are also developing important auxiliary statistical tools, such as model selection methods that can determine whether adding particular ecological covariates to an analysis actually increases the precision and accuracy of our estimates, or whether (if we mistakenly try to add irrelevant information to the model) it actually dilutes the power of the analysis; these tools are necessary before one can confidently start using hierarchical Bayesian methods as a general tool to add information to stock analyses. We are building software tools that implement these methods and that can be used by a broader audience of researchers. Finally, we hope to apply these general methods to some broader questions in stock analysis: for example, where should we define boundaries between populations for the purpose of stock analysis? How do we know when we have enough information to justify analysis at a very fine spatial scale or using very detailed genetic differences, and when should we be satisfied with analyses on a coarser scale?

2 Progress during FY 2003.

We have only been funded since January (a pre-award from the University of Florida was necessary to cover salaries for the first few months of the year). In that time, we have prepared and submitted a manuscript to Ecological Applications ("Combining genetic and ecological data to estimate sea turtle origins", by Okuyama and Bolker) that details the first step of our grant: the construction of hierarchical Bayesian models for sea turtle stock analysis, the testing of such models with a broad range of simulated data to see what conditions favor the use of such models over other (non-hierarchical) stock analysis tools, and the application of hierarchical Bayesian models to existing data on mitochondrial DNA haplotypes of green and loggerhead turtles in the Atlantic Ocean. This is one of the first uses of hierarchical Bayesian models in an ecological context, and is unique in its emphasis on using these models to incorporate additional ecological covariates in a flexible way. While hierarchical Bayesian models are well described in the statistical literature, our paper gives a clear description of our particular method and of the general approach for ecologists. We then discuss a simulation scheme that generates random data sets given a series of parameters (the number of source populations, number of distinct haplotypes, sizes of rookeries, correlations between rookery size and contribution, range of contributions, etc.). We show that for sample sizes, numbers of rookeries and haplotypes, and variances in rookery size as observed in Atlantic sea turtle populations, the hierarchical Bayesian method often gives more accurate results. We then apply the model to Atlantic sea turtle data; the results do not suggest that previous analyses were generally correct, with the exception of some studies using small data sets. Importantly, the addition of ecological covariates is able to narrow confidence intervals sharply and to show statistically that some populations are definitely contributing to the mixed populations (where confidence intervals derived from non-hierarchical models were too broad to be able to make this conclusion).

Since submitting the paper, we have made significant advances in the model selection problem, which is the next one we had set for ourselves. We have determined that the *Deviance Information Criterion* (DIC), a metric recently developed by researchers in Bayesian statistics, has the desired properties of testing model fit while penalizing complexity. (Simple rules such as the well-known Akaike Information Criterion are hard to apply to hierarchical models where the number of parameters is hard to define precisely.) Hierarchical models are essentially compromises between classical stock analysis models that contain a

fixed contribution parameter for every source population and regression models that base the contribution for each source population solely on its ecological covariates, without allowing for any variation from the contribution predicted by the covariates. As such, their effective number of parameters is hard to estimate; the DIC provides a method. We have implemented the DIC for our models and run extensive tests both for our simulated sea turtle scenarios and for simpler "strategic" models, to convince ourselves that the DIC really does reliably determine whether a model with or without ecological covariates gives more accurate answers under particular circumstances. We have compared DIC with other, more traditional metrics such as the Bayes factor, and found that it works better for our applications: Bayes factors do not penalize additional model complexity sufficiently to determine the most accurate model. We are now finalizing our studies on this topic and preparing a manuscript.

3 Plans for the next twelve months:

We plan to finish preparing the manuscript on model selection for Bayesian hierarchical models, and to move on to the central goal of the project, which is to use these methodologies to study the value of a wide range of ecological covariates for stock analysis. In our work so far, we have only incorporated rookery size and a very crude measure of rookery location—the major ocean current or gyre in which the rookery is situated, which serves only to divide rookeries into a few distinct blocks. We now want to explore different possible model frameworks (hierarchical regression, conditional autoregression) and different metrics for distance (geographic, hydrographic) on the accuracy of stock analysis. We will use the tools we have developed to date — simulations and the DIC — to assess the value of different models and different distance covariates, and will apply our models to existing data and to new data that is available from the Archie Carr Sea Turtle Research Center.

We will prepare a manuscript for peer-reviewed publication from this work. We also plan to prepare a manuscript that uses our new tools to survey available data more widely (some specifically relevant to Pacific fisheries), a more conservation-oriented and less technically-oriented paper. We will also work to make the procedures we have developed robust and to incorporate them into a relatively user-friendly package running on top of the R programming environment (and possibly tying in the BUGS statistical estimation package as well).

Finally, we will use these tools to tackle the broader question of model aggregation. Hierarchical models were designed with complex spatial data sets in mind, and should be ideally suited to asking questions about the spatial and genetic resolution at which we should characterize our data to get the most accurate estimates of rookery contributions to stocks. If time permits, we will explore coalescent methods for estimating spatial migration to incorporate the evolutionary and straying processes in our ecological model of relationships among rookeries and foraging grounds.

4 Papers published in refereed journals

None

5 Other papers and presentations

None

6 Students graduating

None

7 Budget for next fiscal year:

Much the same as this year's budget, except for the absence of computer equipment; movement of \$250 from computer maintenance to publication costs; and minor changes (PI summer salary halved) because of UF increases in graduate tuition .

Item	Cost
PI summer salary	\$3152
(1 month, including fringe ¹)	
Graduate RA salary	21526
(full year, 1/2 time)	
Graduate tuition ²	6325
(spring/summer/fall)	
Computer supplies &	250
maintenance	
Publication costs	250
Total direct costs	31503
UF modified indirect costs ³	\$14176
UH/JIMAR indirect costs ⁴	\$5150
Total	\$50929

Notes:

- 1. **fringe:** 18.73% fringe is charged on PI salary. Fringe *has been included* in the figure given here!
- 2. UF graduate salary has increased markedly (from estimated \$3648 in our original budget), this has been compensated by halving PI summer salary.
- 3. **UF indirect costs:** charged on all costs *except* graduate tuition and equipment. Rate is 45% through 6/30/04, 45.5% thereafter (applies only to PI salary summer 2004 and graduate student salary summer-fall 2004).
- 4. UH/JIMAR indirect costs: 20.6% charged on first \$25,000.