Bayesian hierarchical methods for sea turtle mixed stock analysis

Ben Bolker

Department of Zoology &

Archie Carr Center for Sea Turtle Research

University of Florida

9 December 2003

Outline

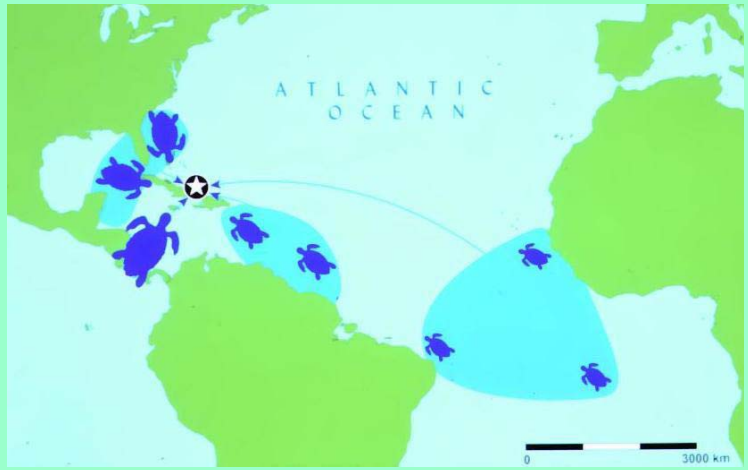
- Background
- Effects of sampling zeros on estimates and confidence limits
  unconditional maximum likelihood, Markov Chain Monte Carlo
- Ecological covariates and diffuse data
  hierarchical Bayesian models

Questions

- ultimate: how to integrate information/risks from rookeries (stocks) into long-term population projections?
- proximate: which rookeries do turtles in mixed stocks come from?
- technical: how can we solve statistical problems (sampling error, incorporating ecological covariates) to decrease bias/increase power of stock analysis?

Background

- Sea turtles (loggerhead, *Caretta caretta*, and green, *Chelonia mydas*): charismatic macrofauna
- Recovering populations, with the usual risks
- Direct tracking of individuals is difficult: the "lost year"
- Data: mitochondrial DNA samples from rookeries and mixed populations
- Basic methods: "beanbag genetics"



Rare and "missing" haplotypes

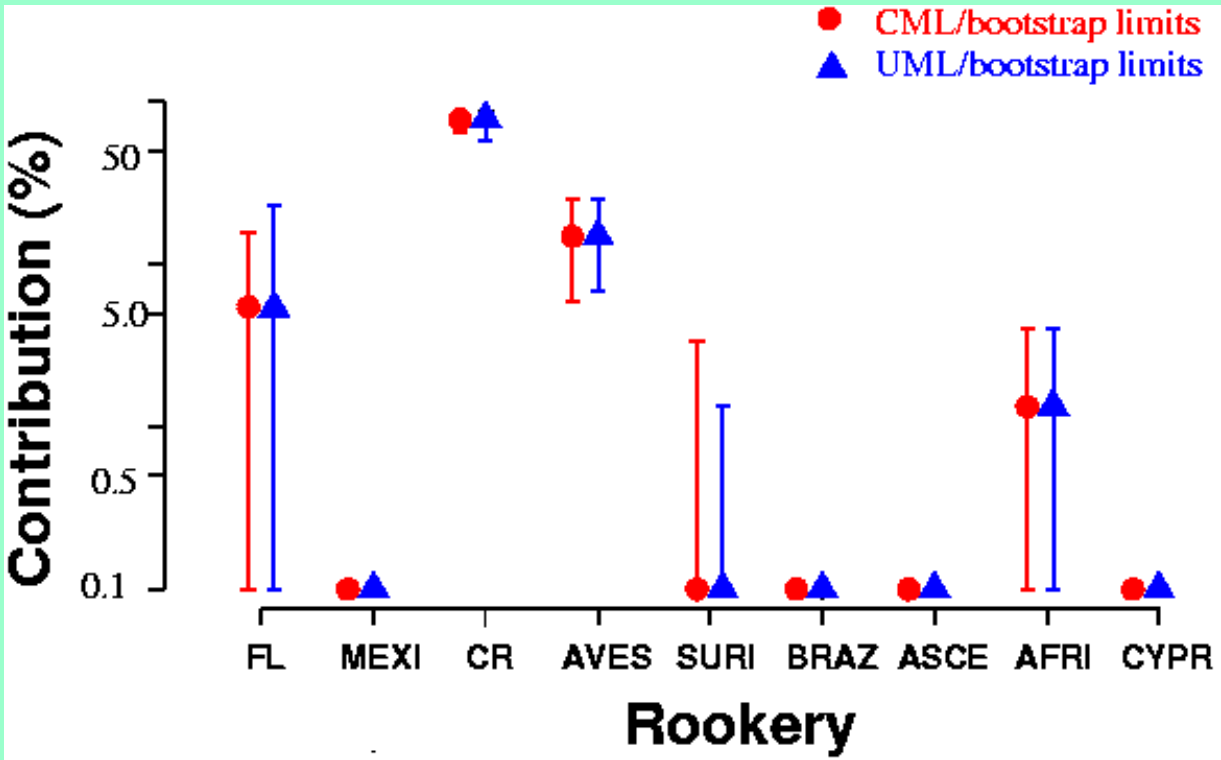
- Missing haplotypes skew estimates, even if they disagree with the rest of the data . . . and are really the result of sampling errors

haplotype	rookery A	rookery B	mixed
I	50	10	10
II	10	50	50
III	1	0	5

conditional maximum likelihood (CML) estimates contributions of {A=20%, B=80%}

Unconditional maximum likelihood (green turtles)

- Incorporate estimates of rookery genotype frequencies



little difference in estimates for green turtle data …

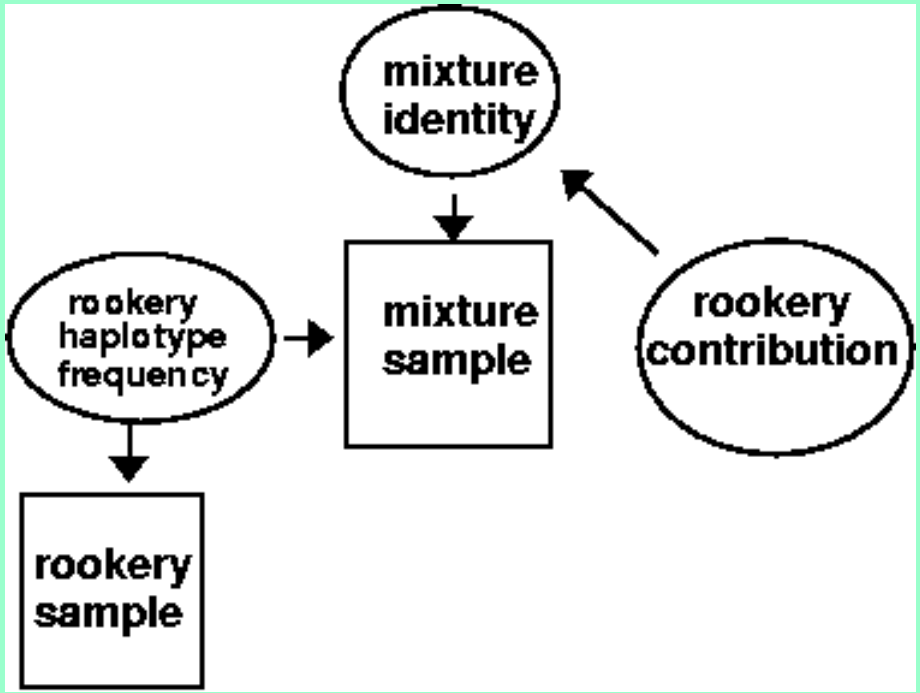
UML most valuable when source populations are undersampled

Confidence limits

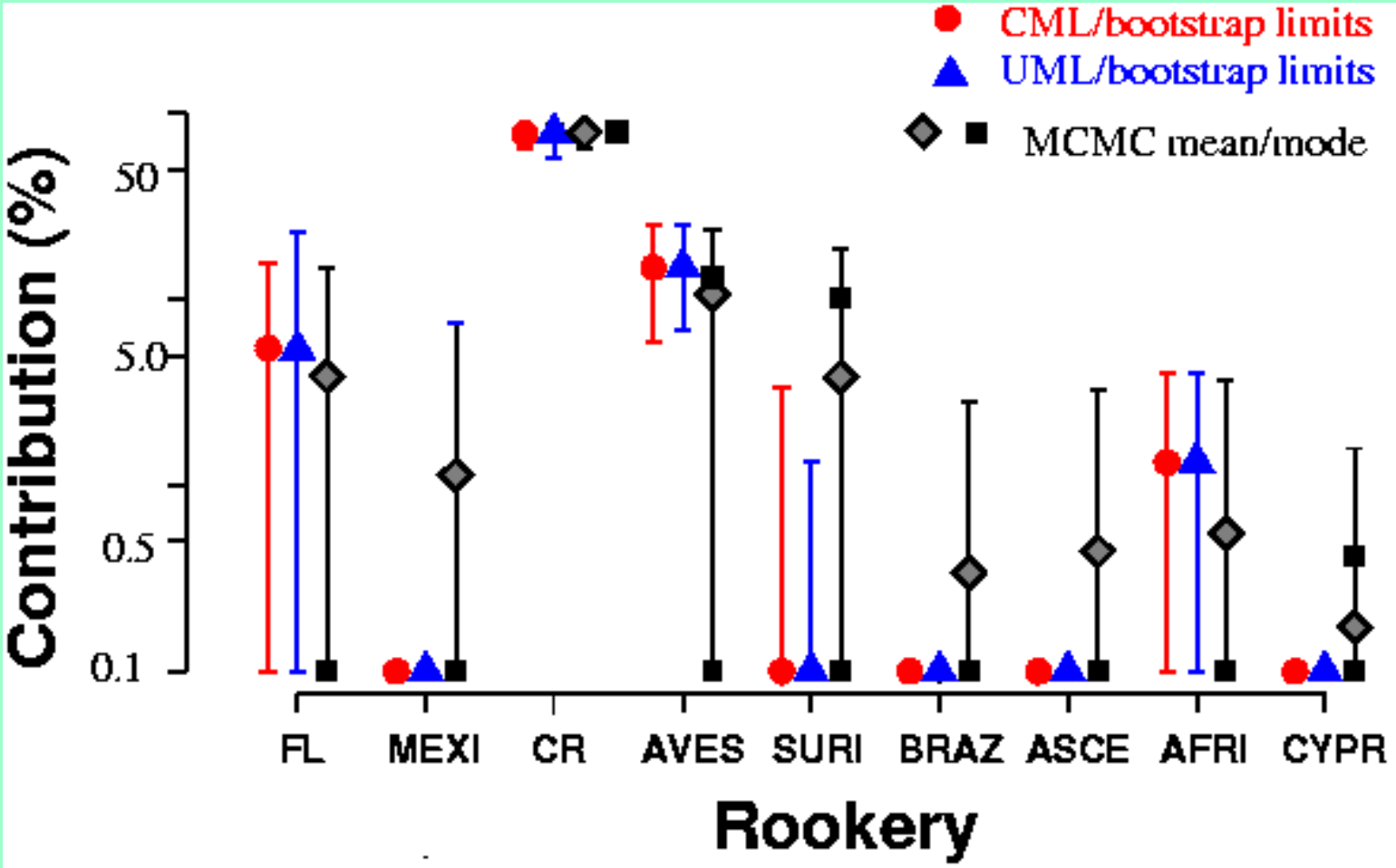
- nonparametric bootstrapping **underestimates** effects of sampling error (never samples missing haplotypes)
- **parametric bootstrapping:** resamples *frequencies* as well as haplotype numbers
- **Markov Chain Monte Carlo (MCMC):** alternative resampling method; alternates parameter estimation and stock imputation

MCMC basics

- Solves difficult estimation problems in a Bayesian framework
- easily incorporates sampling variation
- the **Gibbs sampler** is a particularly simple algorithm (implemented by the **BUGS** program)



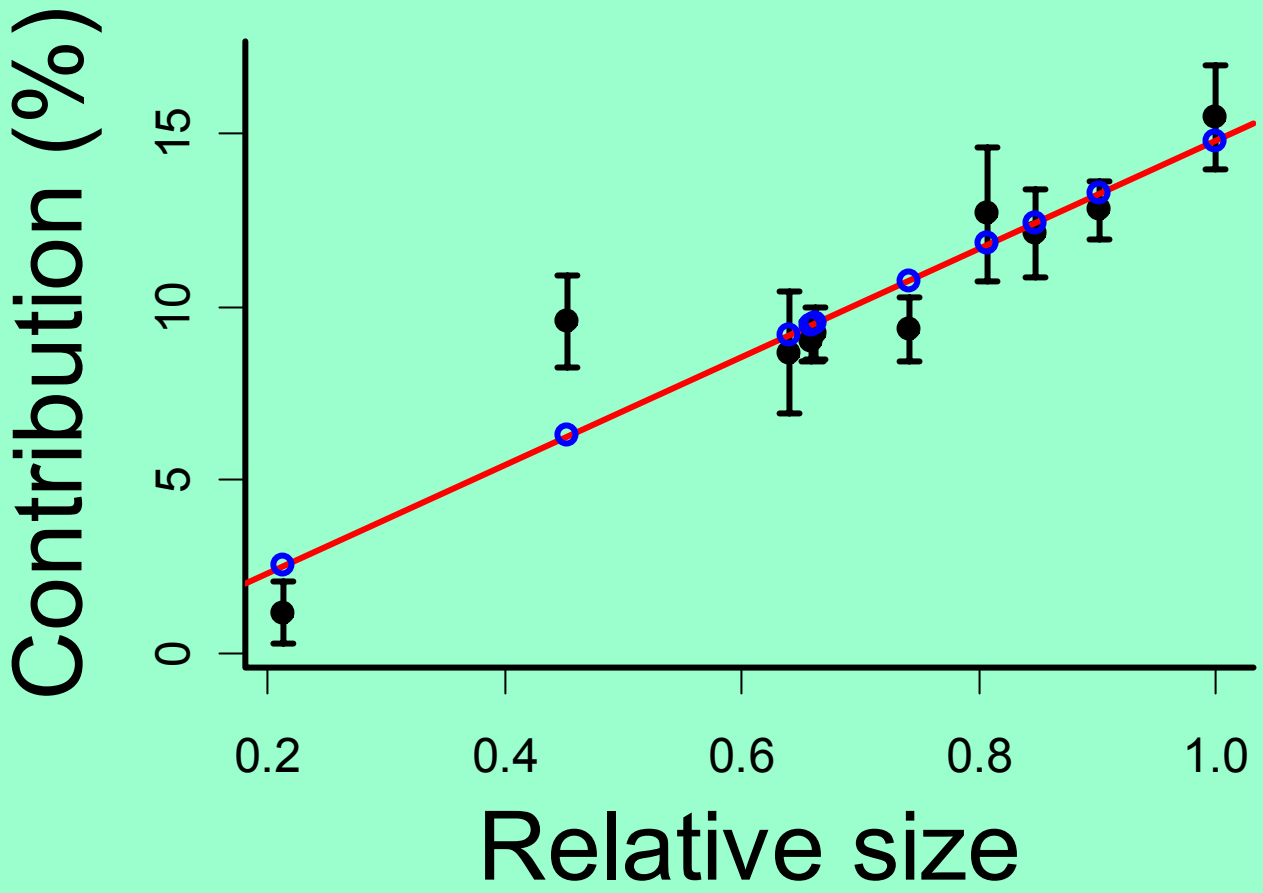
MCMC results (green turtles)



Bayesian analysis: much ado …

- Difference between mean and mode estimates
- Prior probabilities
  - less important than they're cracked up to be
- Technical capabilities
- Fixed vs. random effects

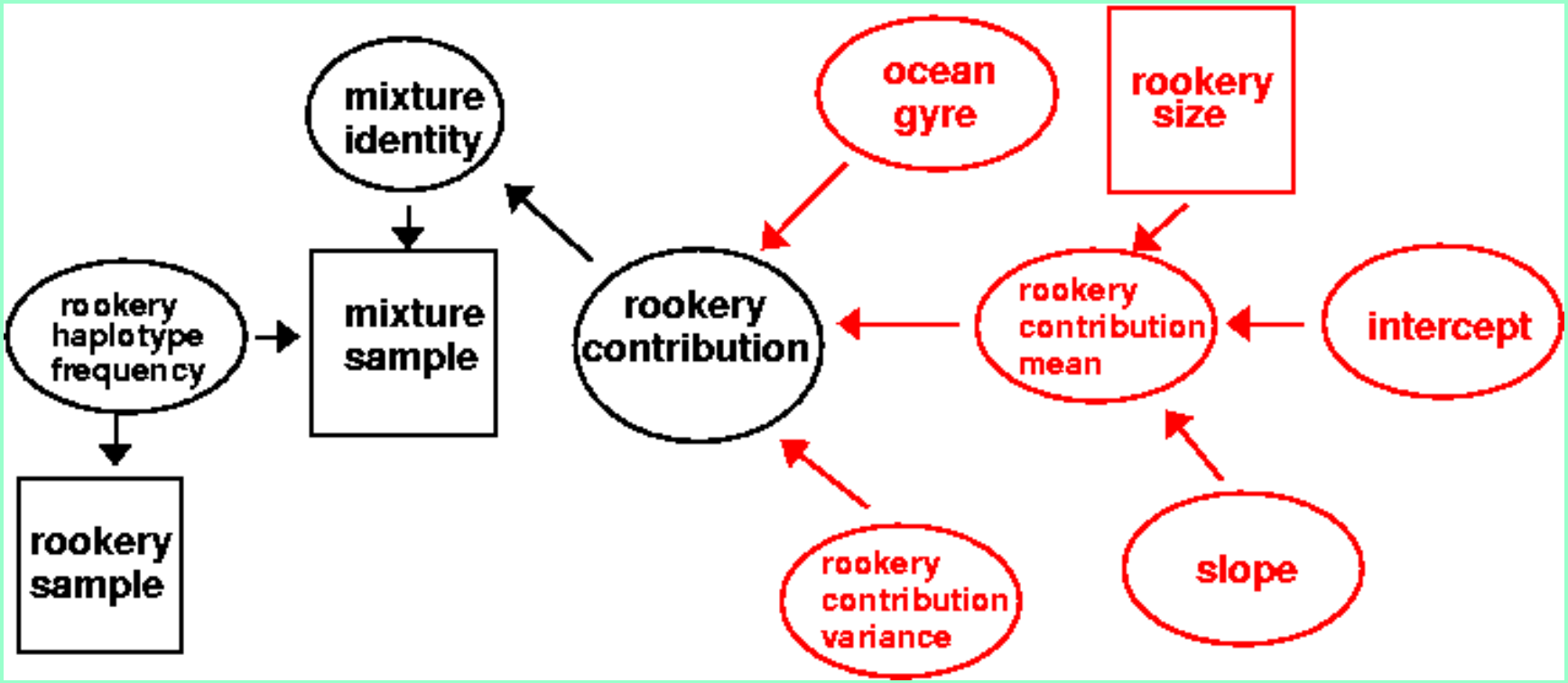
Hierarchical models vs. fixed regression



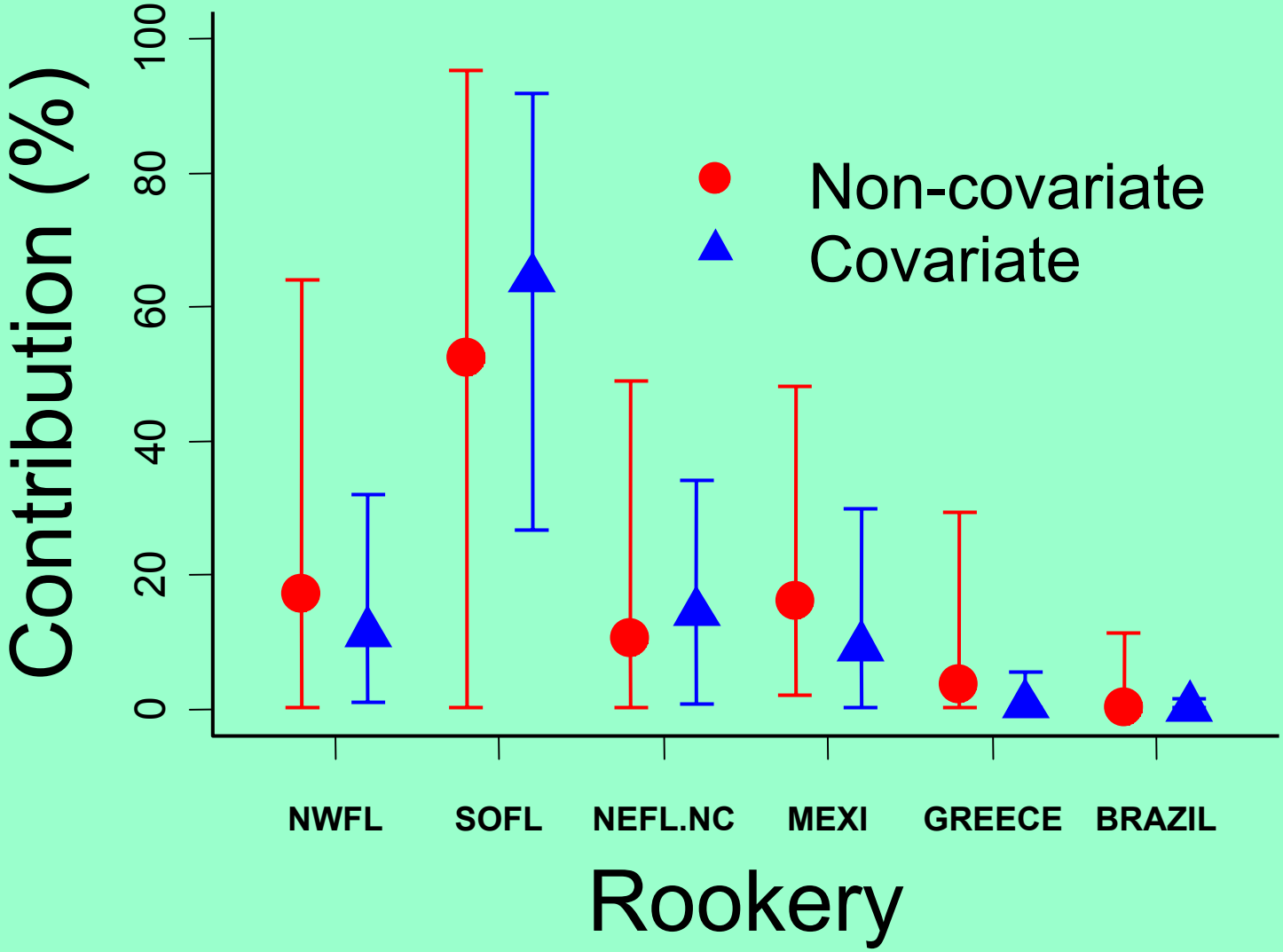
Ecological covariates

- Rookery size, distance … others?
- How?
  - Size-based prior probabilities?
  - Regression models?
  - **Hierarchical models**
    - "hyperparameters" describe distributions of model parameters (e.g. contribution $\propto$ size, with variation)
    - fairly objective — let data decide on relationships
    - more flexible than regression approaches

A hierarchical model using rookery size and gyre …



Covariate-based results (loggerheads)



Conclusions so far

- For Atlantic sea turtles, new methods refine conclusions rather than overturning them
  - conditional and unconditional maximum likelihood: mitigates sparse source sampling
  - MCMC/parametric bootstrapping: improves confidence limits
  - Bayesian and frequentist methods are two sides of the same coin: differences sometimes overstated
  - hierarchical models: increase statistical power by incorporating covariates; **test hypotheses**
- Hierarchical models will be most useful with **large but diffuse** data sets — many small sampling units

Current & future directions

- Controlling model complexity: the **deviance information criterion (DIC)**
- Developing and testing spatial hypotheses
- Multiple-mixed-stock problems
- Incorporating evolutionary dynamics??

Acknowledgements

- co-workers: Toshinori Okuyama, Karen Bjorndal, Alan Bolten
- graphics: Jennifer Piascik
- funding: NOAA/PFRP/JIMAR (of course)