A Sampling Strategy Designed to Maximize the Efficiency of Data Collection of Food Web Relationships

Department of Biology, University of Florida

Kristen E. Sauby and Mary C. Christman

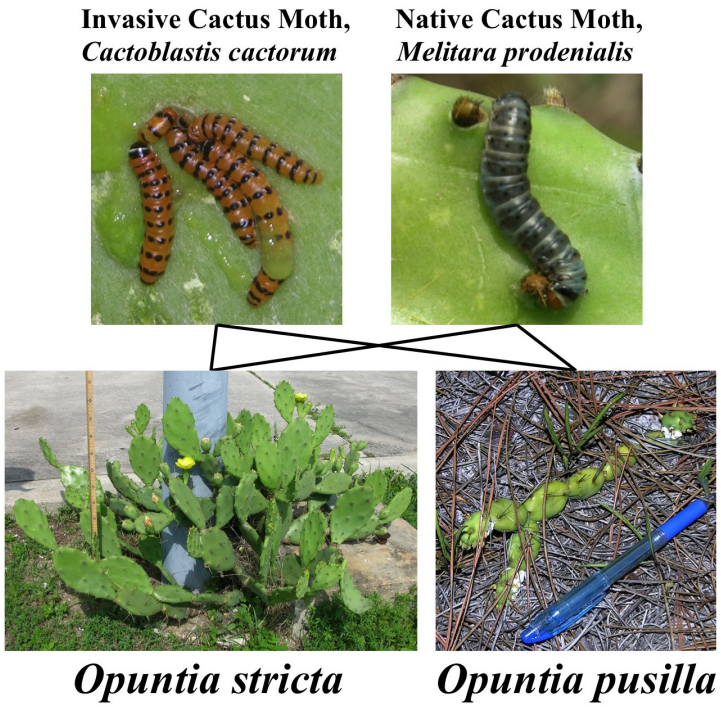
29 June 2016

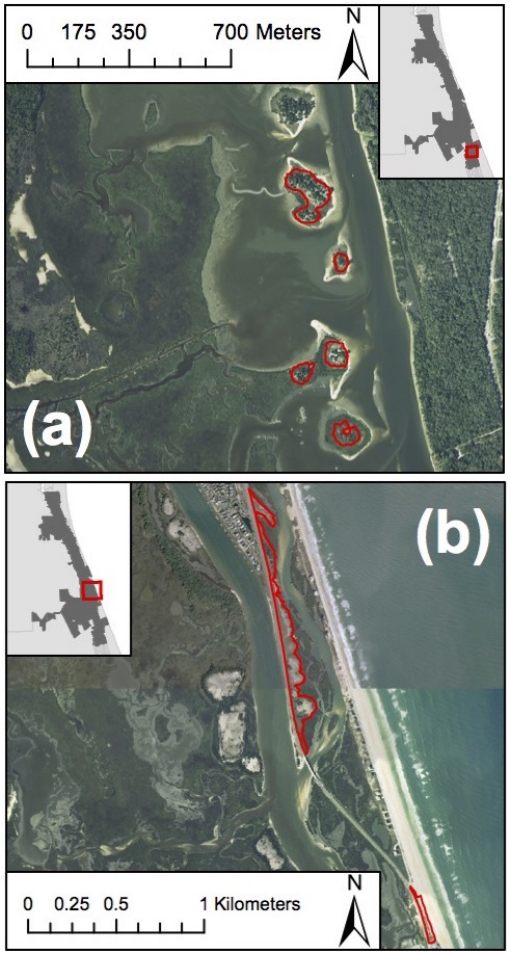
Introduction

- design efficient data collection methods
  - **minimize** costs of time/money
  - **maximize** information

Introduction

- **Variables:** occupancy, abundance
- Set up 1346 1-m^2 plots at 7 habitat patches from May-July 2012 in Northeastern Florida



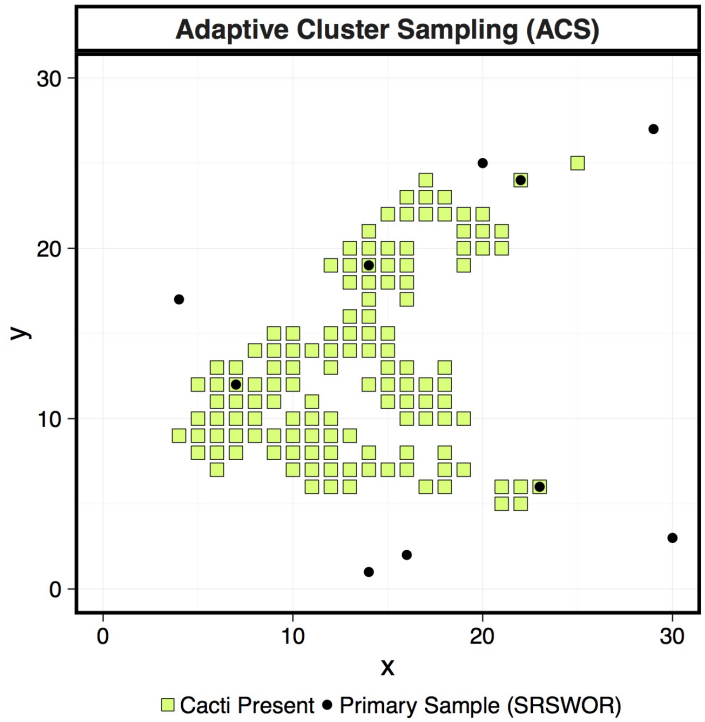


Adaptive Cluster Sampling (ACS)

Proposed in Thompson, S. (1990). Adaptive cluster sampling. Journal of the American Statistical Association, 85(412), 1050–1059.

Design:

- A primary sample of n_1 units is selected from N units, based on some probability sampling scheme

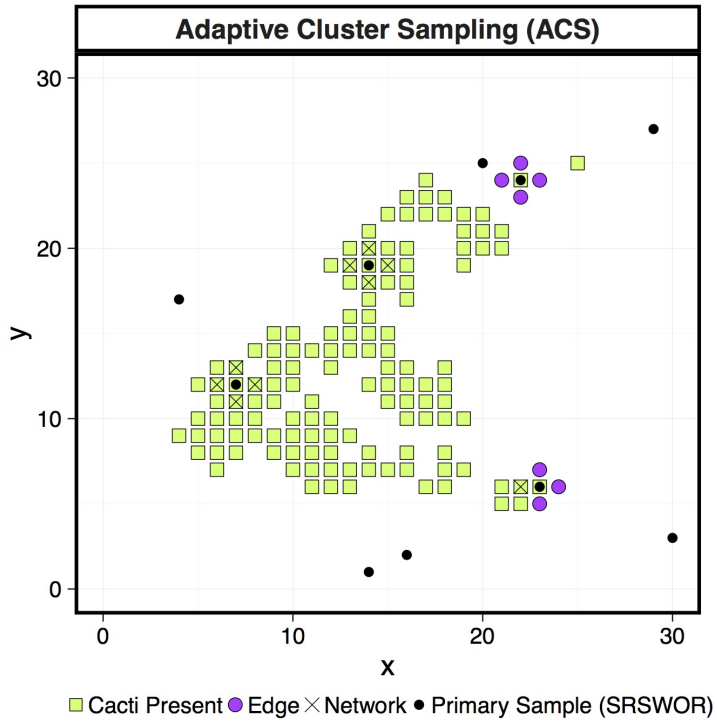


Adaptive Cluster Sampling (ACS)

Design:

- **Adaptive**: A secondary sample is selected if primary units satisfy a condition (e.g., cacti are present)

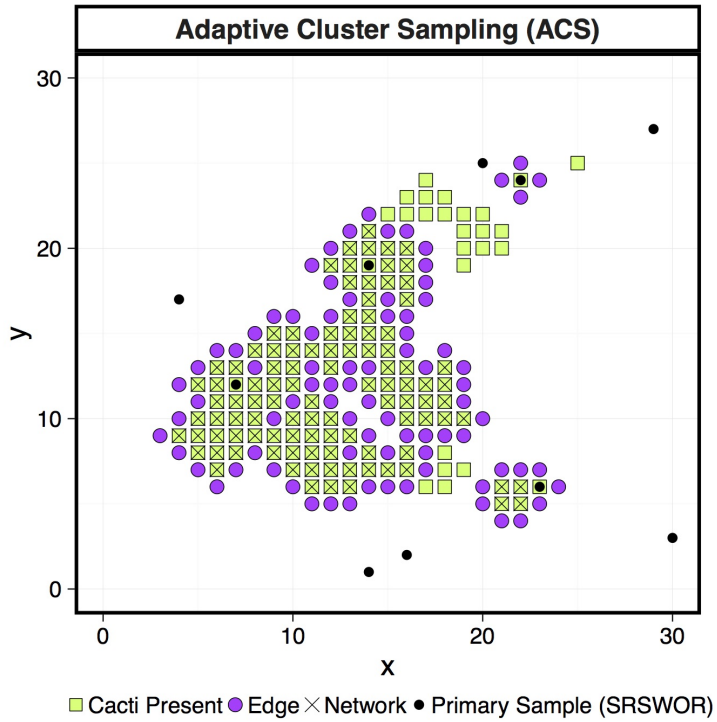
1. Adjacent units sampled in the four cardinal directions



Adaptive Cluster Sampling (ACS)

Design:

1. For neighboring units whose values satisfy the condition, their neighbors are also surveyed
2. **Cluster Sampling**: Sampling of neighboring units continues until no additional units satisfy the condition



Adaptive Cluster Sampling (ACS)

Advantages

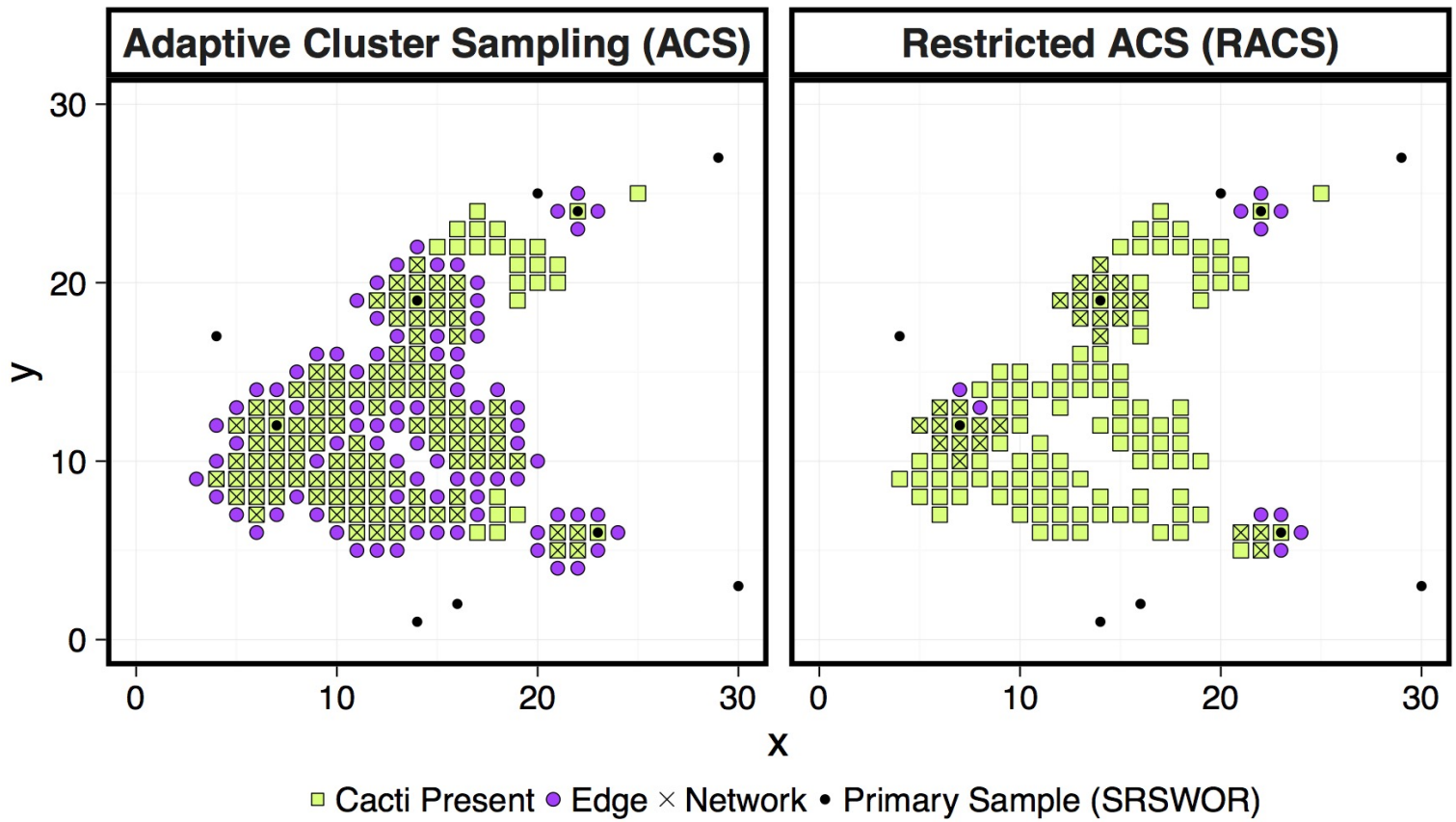
- Best for studies involving populations of rare and clustered individuals
- Compared to designs such as simple random sampling:
  - Increased sampling efficiency
  - Reduced variance

Disadvantage

- Final sample size is unknown at the start of sampling

Restricted ACS (RACS)

- same procedure as ACS, **except** that we restrict the number of neighboring units to q (in our example, $q = 12$)

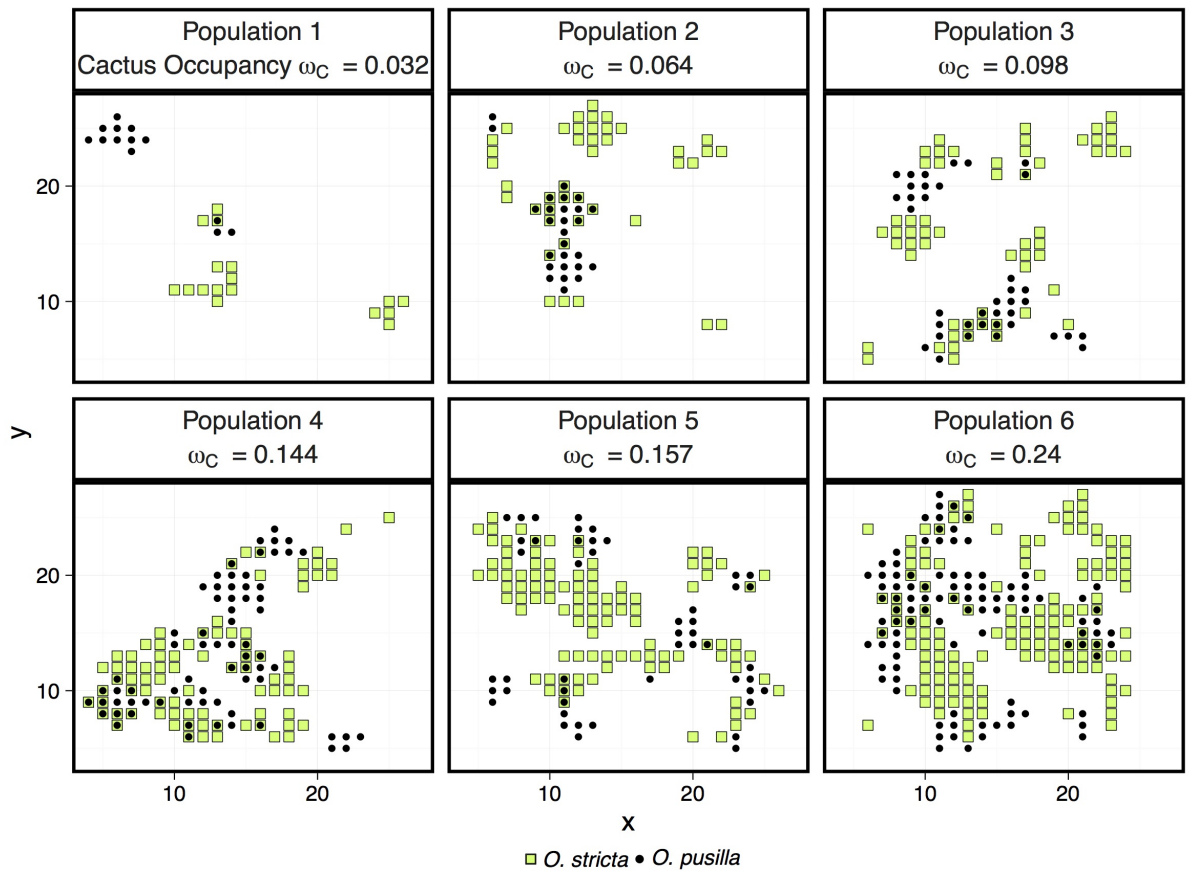


Estimators

- The unbiased (for ACS) Horvitz-Thompson estimator of the population mean of variable x is $\bar{x}_{HT} = \frac{1}{N}\sum_{i=1}^{v}(x_i J_i)/\pi_i$
  - N = population size, n_1 = primary sample size
  - v = the number of unique units in the sample
  - J_i is an indicator variable, equalling 1 if the ith unit in the sample satisfies the condition or was included in the primary sample and 0 otherwise
  - **inclusion probability:** $\pi_i = 1 - \binom{N-m_i}{n_1} \Big/ \binom{N}{n_1}$**, where** m_i **is the number of units in the network** ψ_i

Simulations

- **Objective**: Explore properties of the RACS design by bootstrapping populations created with a dataset of two cactus species and their associated insects.
- randomly sampled the survey data to generate 6 populations



Hypotheses

Horvitz-Thompson estimators will be

1. **biased** for the variable on which the RACS criterion is based
   - *i.e.*, cactus presence
2. **unbiased** for variables that
   - are not used to generate the RACS criterion
   - are uncorrelated with the variable on which the RACS criterion is based
   - *i.e.*, insect presence, evidence of insect damage

Simulations

- We used cactus presence as our criterion to initiate cluster sampling
- We ran 5000 simulations of the RACS and ACS designs per
  - population
  - primary sample size (20, 30, 40, 75, and 100)
- For each simulation we estimated:
  - *O. stricta*, *O. pusilla*, and cactus occupancy
  - insect occupancy given cactus presence

Simulations

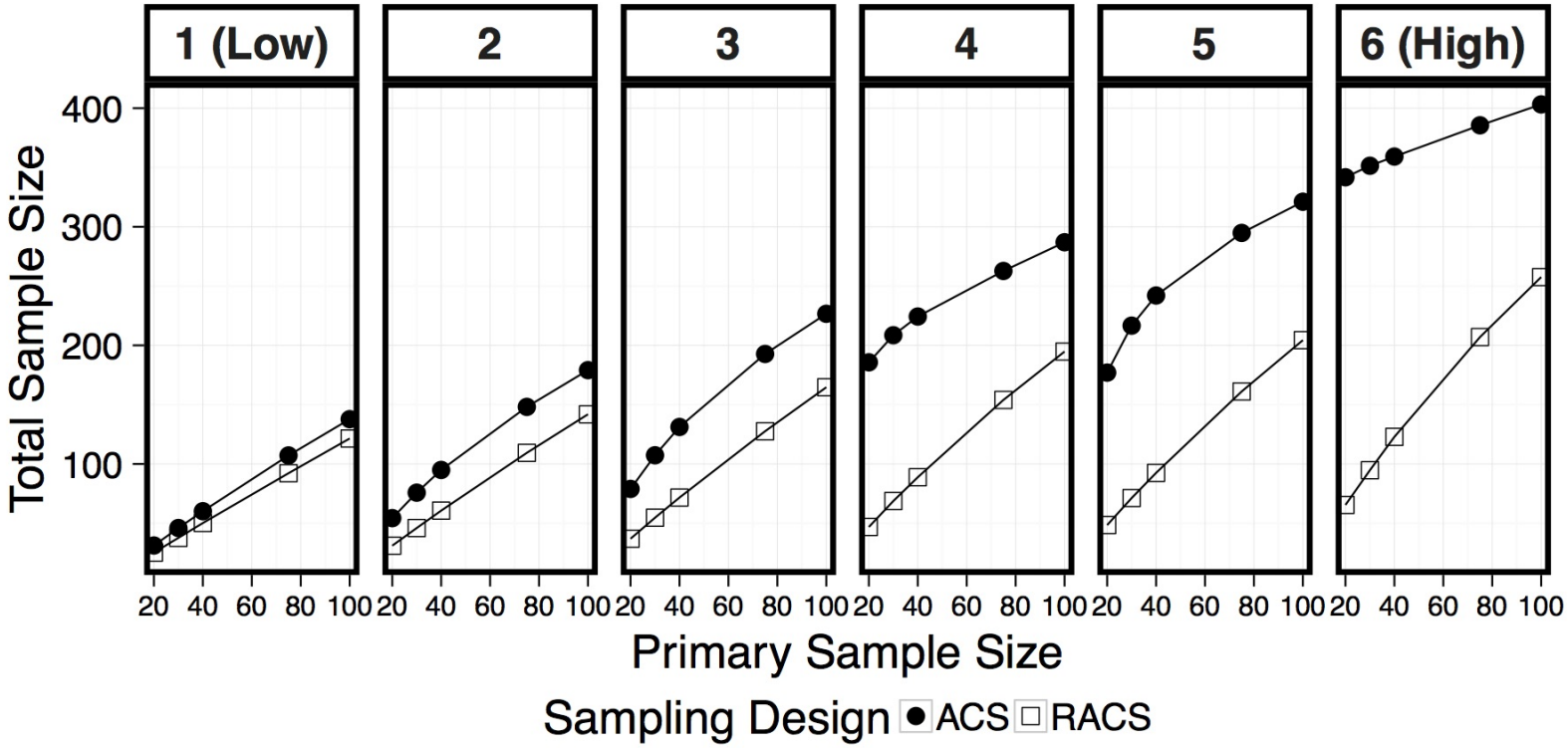
- relative bias of the mean:

$B_{\bar{x}_{HT,i}} = (\bar{x}_{HT,i} - \mu)/\mu$

- μ is the true population mean
- $\bar{x}_{HT,i}$ Horvitz-Thompson mean of variable x from the ith simulation
- mean percent relative bias for the S simulations

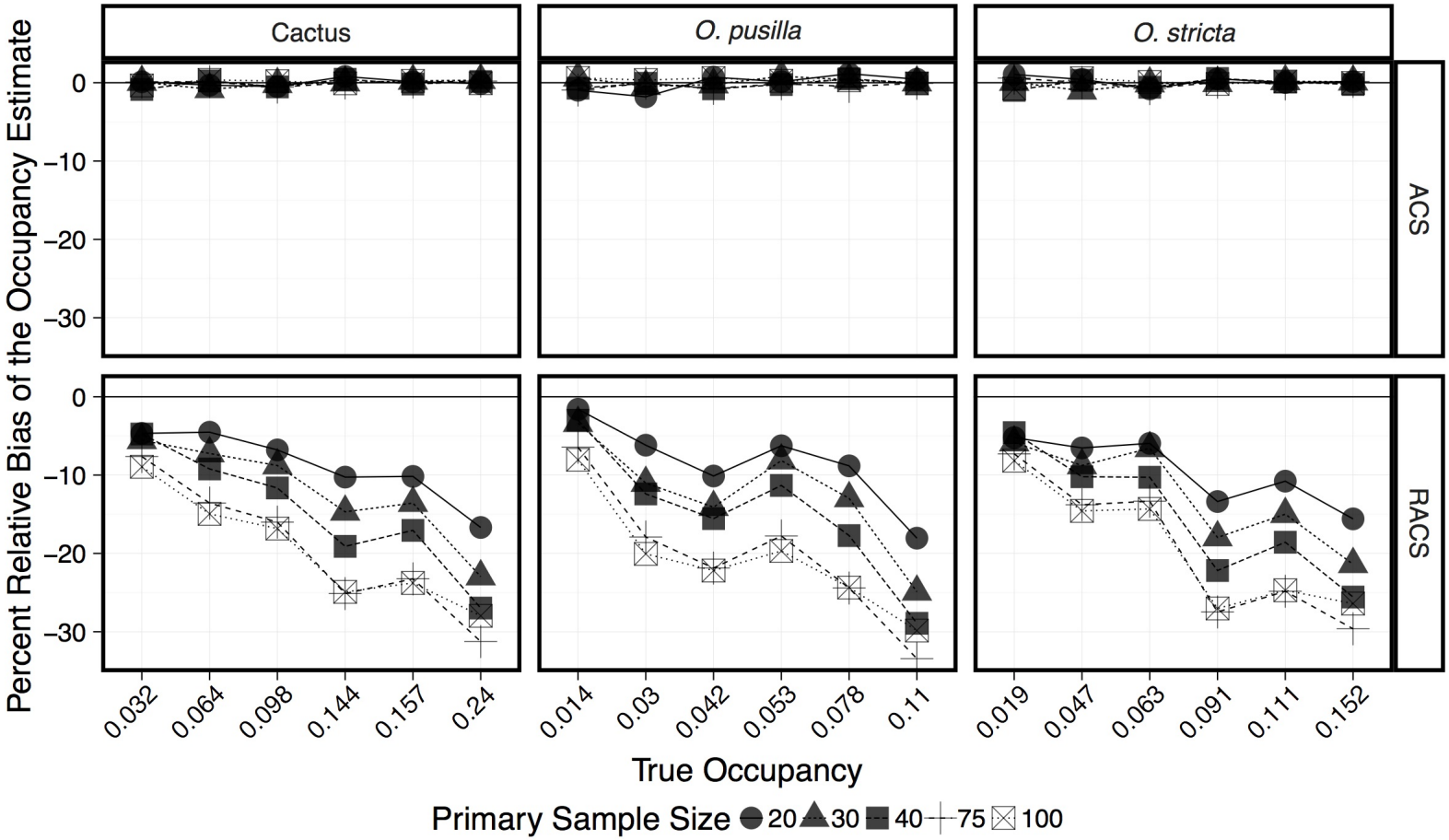
$\text{mean}(B_{\bar{x}_{HT}}) = 100 \times S^{-1} \sum_{i=1}^{S} B_{\bar{x}_{HT,i}}$

Results - Final Sample Sizes



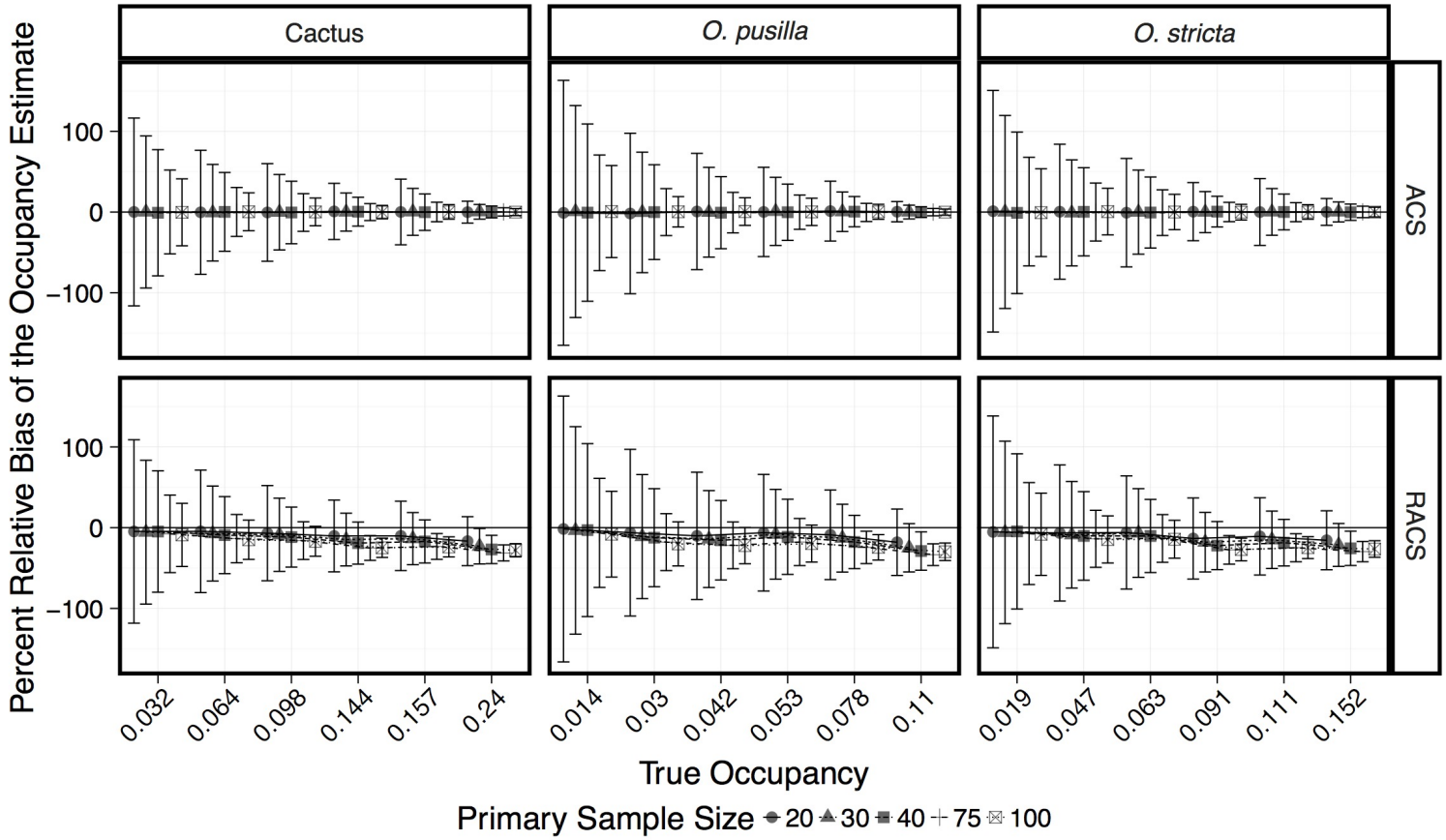
Results - Cactus Occupancy

- with RACS, we underestimate occupancy (negative percent relative bias)
- this underestimate increases with cactus occupancy



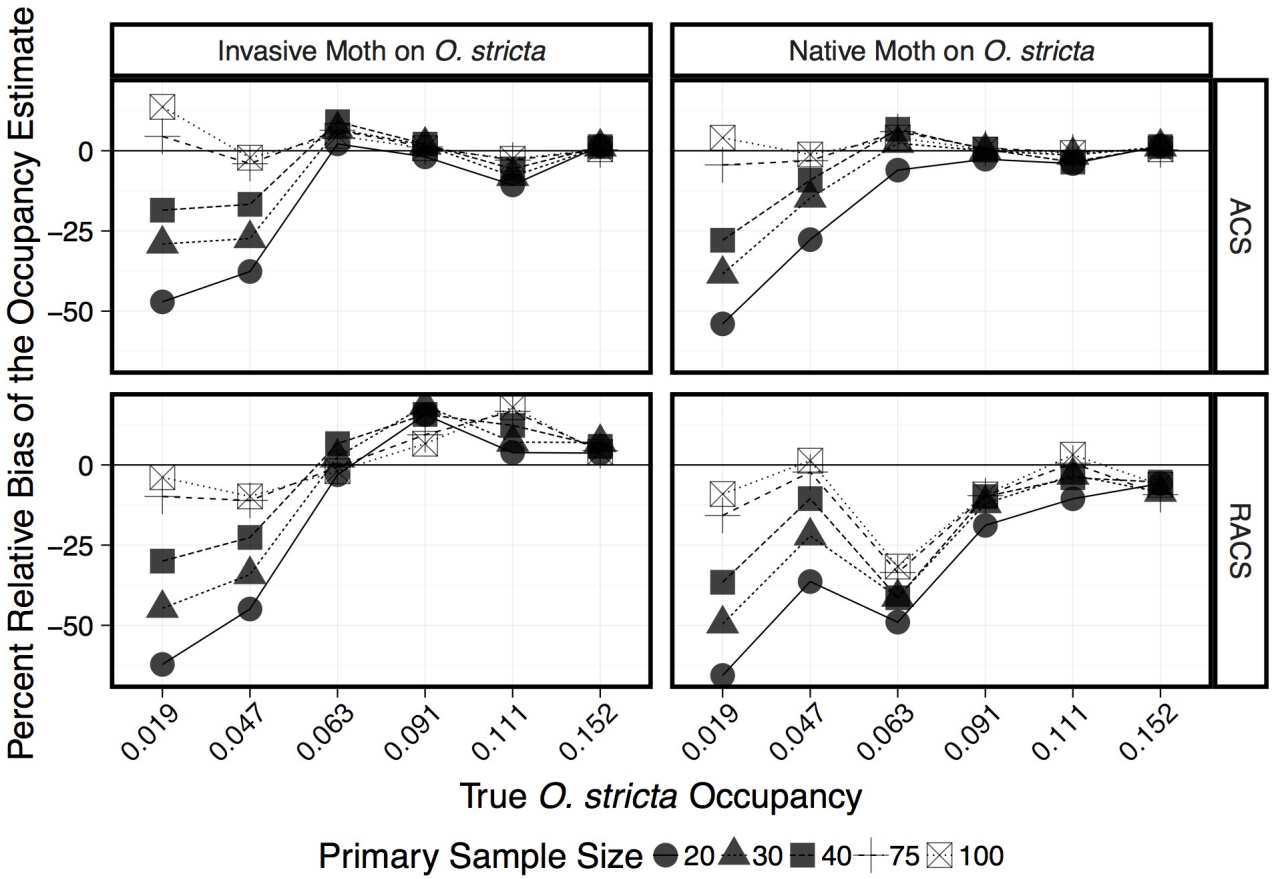
Results - Cactus Occupancy

- Variability decreases with increased primary sample size and occupancy



Moth Occupancy on *O. stricta*

- bias decreases with increasing primary sample size
- noise suggests that something other than just occupancy (e.g., spatial structure) is influencing the bias



Conclusions (So Far)

- Compared to ACS, RACS generates smaller final sample sizes, particularly for populations with relatively high occupancy
- confirmed that estimate of cactus occupancy is biased
  - may need to fix calculation so that bias does not increase with n_1
- with larger sample sizes, our estimates of moth occupancy is less biased

Bias Correction Techniques

- correct bias by correcting inclusion/intersection probabilities
- ways to do this:
  - analytical - come up with a formula based on average behavior
  - use nonparametric technique such as bootstrapping
    - keep track of how many times a unit is sampled we can estimate the probabilities

Additional Work

- we have also looked at bias for:
  - variance
  - insect damage occupancy
  - max. plant height
  - percentage of plot covered by cacti
- we have created a R package, "ACSampling," that we will release as a public Github repository

Future Work

- Estimate bias of the variance
- Run more simulations
- Test bias reduction techniques
- compare efficiency of RACS, ACS, and SRSWOR
  - primary sample size = 40 + RACS = 40 vs. SRSWOR = 80?

Acknowledgements

- R.D. Holt

Data Collection

- Guana Tolomato Matanzas National Estuarine Research Reserve
- Katie Petrinec
- Matt Welsh
- Polly Harding, undergrad. volunteer

Funding

- QSE3 IGERT (NSF Award # 0801544)
- Cactus and Succulent Society of America
- Garden Club of America/Center for Plant Conservation
- University of Florida Graduate Student Council