

Lizard Perch Heights on Experimental Islands

Nicholas C. Herrmann

Project Aim

The aim of this project was to build and explore the limitations of a statistical model for analyzing the results of a field experiment. The experiment investigates how removal of a congeneric competitor affects the perch height of the lizard, *Anolis carolinensis*. Herein I describe a Bayesian, hierarchical statistical model that will be used to analyze this experiment. Using data simulations, I explore how changes in sample size and variance affect the model's performance.

Data Collection

This past summer, I measured the perch height of 24-29 individual *A. carolinensis* on six small islands in southeastern Florida. I recorded the sex of each lizard to account for potential differences in perching behavior between males and females.¹ A congeneric competitor, *A. sagrei*, is currently abundant on all six islands. Previous work suggests that *A. carolinensis* perches lower in the absence of *A. sagrei*, which is better adapted for life near the ground.² This upcoming spring, I will remove *A. sagrei* from three islands and leave it on the remaining three as controls. I will subsequently re-measure *A. carolinensis* perch height on each island. Thus the full data set for this experiment will encompass population-level perch height measurements (I did not uniquely tag individual lizards) from six islands across two years, with the first year serving as a pre-removal baseline.

Statistical Methods

Perch height, the response variable in the model, is treated as continuous and drawn from a truncated normal distribution with a lower bound of zero (*A. carolinensis* does not burrow belowground). Predictors include treatment (removal or control), sex (male or female), year (Year 0: pre-removal and Year 1: post-removal), and several interactions (see below) as fixed effects. To account for the

¹ Losos, J. B. *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles*. (Univ of California Press, 2009).

² Stuart, Y. E. *et al.* Rapid evolution of a native species following invasion by a congener. *Science* **346**, 463–466 (2014).

hierarchical structure of the data (lizards nested within islands), island identity is included as a random effect with varying intercepts. The full description of the model is

$$y \sim N(\hat{y}_i, \sigma_y^2) [0, \infty) \quad (1)$$

$$\hat{y}_i = \alpha_{j[i]} + \beta_{\text{treatment}} + \beta_{\text{sex}} + \beta_{\text{year}} + \beta_{\text{treatment} \times \text{year}} + \beta_{\text{treatment} \times \text{sex} \times \text{year}} \quad (2)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \quad (3)$$

where y is perch height, \hat{y}_i is the predicted perch height for lizard observation i , σ_y is the residual error around \hat{y}_i (within-island variation), and $\alpha_{j[i]}$ is the intercept for all lizard observations from island j , drawn from a normal distribution with mean μ_α and standard deviation σ_α . The values of predictor variables for individual lizard observations (e.g. *treatment_i*, *sex_i*) are omitted from line 2 above to avoid clutter. The *treatment* \times *year* interaction is intended to detect whether removal islands demonstrate a shift in perch height following competitor removal, relative to control islands. The *treatment* \times *sex* \times *year* interaction is intended to detect whether males and females on removal islands differ in their perch height shifts following competitor removal. The model is coded in the programming language Stan.^{3,4,5}

Using the model, I conducted six runs (Sims 1-6) of data simulation, model fitting, and posterior predictive checks in R to assess model performance.^{6,7} First, I generated “fake” data (hereafter “input” data) using a pre-determined set of parameter values (Table 1). I used the same fixed effects (β values) for each Sim. These values are based on the effects of competitors and sex on *A. carolinensis* perch height as measured previously in southeastern Florida.² I changed other parameter values ($i, j, \mu_\alpha, \sigma_\alpha$, and σ_y) to evaluate the performance of the model with different sample sizes and variances. Input data were either entirely simulated (Sims 1-4) or partly simulated using real, pre-removal data collected this past summer (Sims 5-6).

³ Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **20**, 1-37 (2016).

⁴ www.mc-stan.org

⁵ Supplemental file “perchheight_truncnormal_v2.stan”

⁶ R Development Core Team, 2017 version 3.4.1

⁷ Supplemental file “Herrmann_FinalProject_OEB201.R”

Second, I fit the model to input data with uninformed priors using the rstan package.⁸ I ran four chains simultaneously, with 500 warm-up iterations followed by 500 sampling iterations, resulting in 2000 posterior samples for each parameter. The model converged cleanly in all Sims except Sim 4, during which there were 77 divergent transitions. Parameter estimates for Sim 4 are still reported for comparison with other Sims, yet they should be interpreted with caution.

Third, I evaluated the model's performance by plotting 80% and 95% credible intervals for each parameter estimate and visually comparing these intervals to the true parameter values used to generate the input data. I also conducted a posterior predictive check (PPC) on each model fit using a single value (d) calculated as

$$d = (\mu_{\text{Year1Removal}} - \mu_{\text{Year0Removal}}) - (\mu_{\text{Year1Control}} - \mu_{\text{Year0Control}}) \quad (4)$$

where μ is the mean perch height across all lizard observations for a given island treatment in a given year. After each model fit, I used the posterior samples to generate 2000 new data sets equivalent in size to original input data. I then calculated d for each new data set and visually compared this distribution to the value of d for the original input data. The value d summarizes the extent to which removal islands and control islands differ in their between-year perch height shifts. When d is negative, lizards on removal islands show a larger decrease in perch height between years than lizards on control islands. In line with my expectation that *A. carolinensis* will perch lower following competitor removal, input data for each Sim had a negative value of d (driven by negative $\beta_{\text{treatment} \times \text{year}}$ and $\beta_{\text{treatment} \times \text{sex} \times \text{year}}$). The extent to which each model fit captured this signal can be inferred by the distribution of d generated from posterior samples. Specifically, if the model performed “well”, the distribution of d should be entirely negative with small variance around the value of d from the input data.

Results and Interpretation

Sim 1, which assumed large sample size and small variance at both hierarchical levels of the data, captured the true parameter values within narrow 80% CIs (Figure 1, top). A PPC showed that data

⁸ version 2.16.2

generated from posterior samples were highly consistent with input data for the value d (Figure 2, top). In other words, predictions from the model fit reconstituted a larger drop in perch height on removal islands than on control islands with high accuracy.

Sim 2, with a smaller sample size and slightly larger variances, still captured true parameter values within 80% CIs (results not pictured). Values of d from the PPC were all negative but had greater variance than in Sim 1, as would be expected given the larger variances in the input data.

Sim 3 was intended to evaluate how well the model would cope with a large treatment effect ($\beta_{\text{treatment}}$) at a sample size that closely matches reality (six islands with 26 observations per island per year). I can avoid introducing a large treatment effect into the experiment by assigning removal and control islands such that the groups have similar mean perch heights at baseline. Nonetheless, it will be impossible to eliminate the treatment effect entirely, and ideally the model will still capture true effects in the parameters of interest (β_{sex} , $\beta_{\text{treatment} \times \text{year}}$, and $\beta_{\text{treatment} \times \text{sex} \times \text{year}}$) even when a treatment effect exists. Indeed, the model performed relatively well in estimating these parameters of interest. The 95% CI around each estimate contained the true parameter value and did not overlap with zero (Figure 1, middle). The high uncertainty in the estimates of other parameters (e.g. $\beta_{\text{treatment}}$, μ_{α} , σ_{α}) is not concerning in and of itself, yet the PPC resulted in a tail of values for d suggesting little or no difference between perch height shifts on removal vs. control islands (Figure 2, middle). Still, the model fit overwhelmingly reconstituted the appropriate sign for d , with only 0.2% of values distributed above zero.

In Sim 4 I removed the treatment effect and substantially increased the variance, particularly the residual, within-island variation (σ_y). Even with incomplete convergence, the model captured true parameter values within 80% CIs, but these intervals were substantially wider than in Sim 3 and crossed zero for $\beta_{\text{treatment} \times \text{year}}$ and $\beta_{\text{treatment} \times \text{sex} \times \text{year}}$ (Figure 1, bottom). Furthermore, about 9% of PPC values were positive (Figure 2, bottom). It is unsurprising that uncertainty in parameter estimates increases with smaller sample size and larger variance. Unfortunately, the sample size and variances in Sim 4 closely match reality.

The input data for Sim 5 was combination of real pre-removal baseline data and simulated post-removal data. Rather than assign treatments randomly, I assigned the removal treatment to the three islands on which I actually intend to remove the competitor. The model captured true parameter values within 80% CIs for β_{sex} , $\beta_{\text{treatment} \times \text{year}}$, and $\beta_{\text{treatment} \times \text{sex} \times \text{year}}$, but these intervals were relatively wide and crossed zero for the latter two parameters (Figure 3, top left). Additionally, the model estimated a large, positive β_{year} even though post-removal data were simulated with no year effect, except for its role in interaction terms. Estimated island intercepts were nonsensical, as perch height must be greater than or equal to zero (Figure 3, bottom left). Interestingly, this was not an issue in Sim 4 (Figure 3, bottom right) despite its similarities to Sim 5 in true parameter values (Table 1). Compared to Sim 4, fewer PPC values for Sim 5 were positive (about 7%; Figure 3, top right), although the overall variance in d was larger.

I have no reason to expect that the large within-island variance in perch height measured this past summer was anomalous. Regardless, I conducted Sim 6 identically to Sim 5 except with substantially reduced σ_y . The PPC was marginally better (about 5% positive values of d ; results not pictured), but a visual comparison of the raw input data for Sims 5 and 6 suggests that the reduced σ_y in Sim 6 does a poor job of maintaining between-year consistency on control islands and over-constricts the distribution of perch heights on removal islands (Figure 4). Therefore, Sim 6 may grossly misrepresent reality despite being partly based on real, pre-removal perch height data.

Using my best current assumptions for effect sizes and variances, along with a realistic sample size for this experiment, the model described herein does not completely fail, yet it does not inspire great confidence in its ability to detect true effects related to my hypothesis. To improve its performance, I will explore allowing different variances for different levels of treatment, which should permit the distribution of post-removal perch heights on removal islands to contract without distorting post-removal perch heights on control islands. I will also test whether using weakly informative priors for the hyperparameters that determine island intercepts (μ_α and σ_α) improve the model's ability to capture true signals in fixed effects.

Table 1. Description of input data and the parameters used to generate input data for Sims 1-6.

	Sim 1	Sim 2	Sim 3	Sim 4	Sim 5	Sim 6
Data Type	fake	fake	fake	fake	real and fake	real and fake
Balanced sampling?	Y	Y	Y	Y	N	N
Observations per island	50	26*	26	26	26‡	26‡
Number of islands (<i>j</i>)	16	6*	6	6	6	6
Total number of observations (<i>i</i>)	1600	312*	312	312	294*	294
$\beta_{\text{treatment}}$	0	0	-120*	0*	0	0
$\beta_{\text{treatment} \times \text{year}}$	-30	-30	-30	-30	-30	-30
β_{year}	0	0	0	0	0	0
β_{sex}	80	80	80	80	80	80
$\beta_{\text{treatment} \times \text{sex} \times \text{year}}$	-40	-40	-40	-40	-40	-40
μ_{α}	180	180	180	180	181.6*†	181.6†
σ_{α}	10	25*	25	40*	41.2*†	41.2†
σ_y	10	25*	25	100*	123.4*†	40*

* parameters that changed from the previous Sim

† calculated from real, pre-removal baseline data and used to simulate post-removal data

‡ for post-removal data only

Figure 1. Median (black dot), 80% CI (thick red bar), and 95% CI (thin black bar) for estimated fixed effect parameters and hyperparameters from Sims 1, 3 and 4. Yellow triangles point to the true parameter values used to generate input data. Values on the x-axis differ between panels.

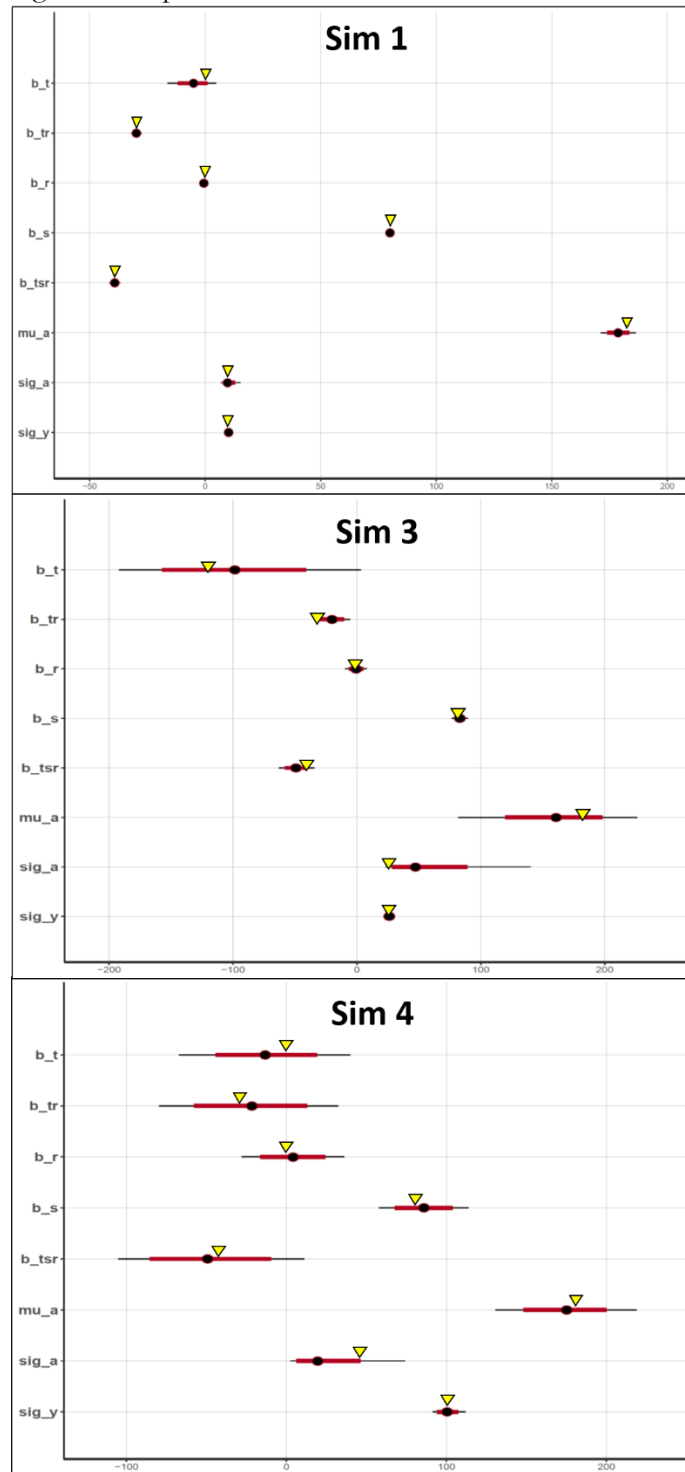


Figure 2. Histograms illustrating the posterior predictive checks on Sims 1, 3 and 4. “PHdiff” on the x-axis is synonymous with the value d (equation 4 in text). Vertical solid blue lines at $d=50$ represent the expected value of d in input data based on fixed effect sizes used to simulate the data. Vertical dashed red lines indicate the actual of d in input data, which differs from the expected value of d because of non-zero variances. Axis values differ between panels.

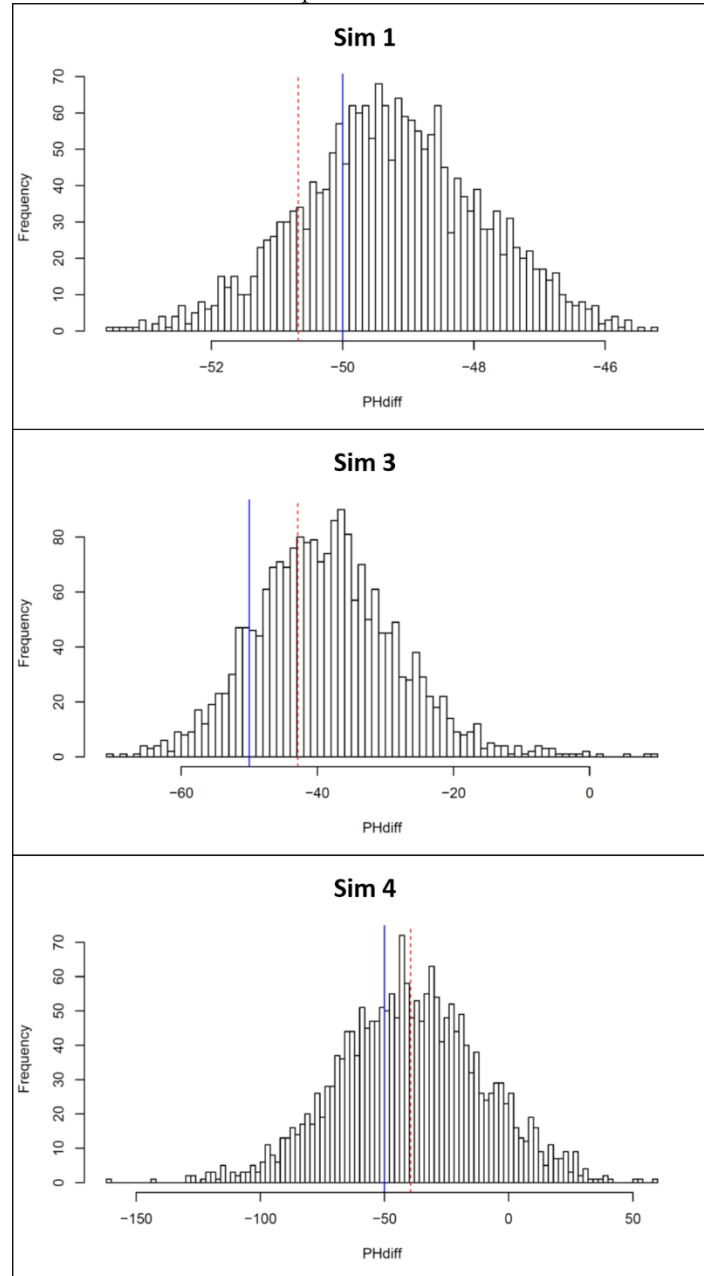


Figure 3. Top left: Parameter estimates and true parameter values for Sim 5. Symbols are as in Figure 1 except for blue crosses, which mark “true” parameter values calculated from real, pre-removal baseline data. Top right: Histograms illustrating the posterior predictive check on Sims 5. Axes and vertical lines are in Figure 2. Bottom row: Estimated island intercepts for Sim 5 (left) and Sim 4 (right). Axis values differ between panels.

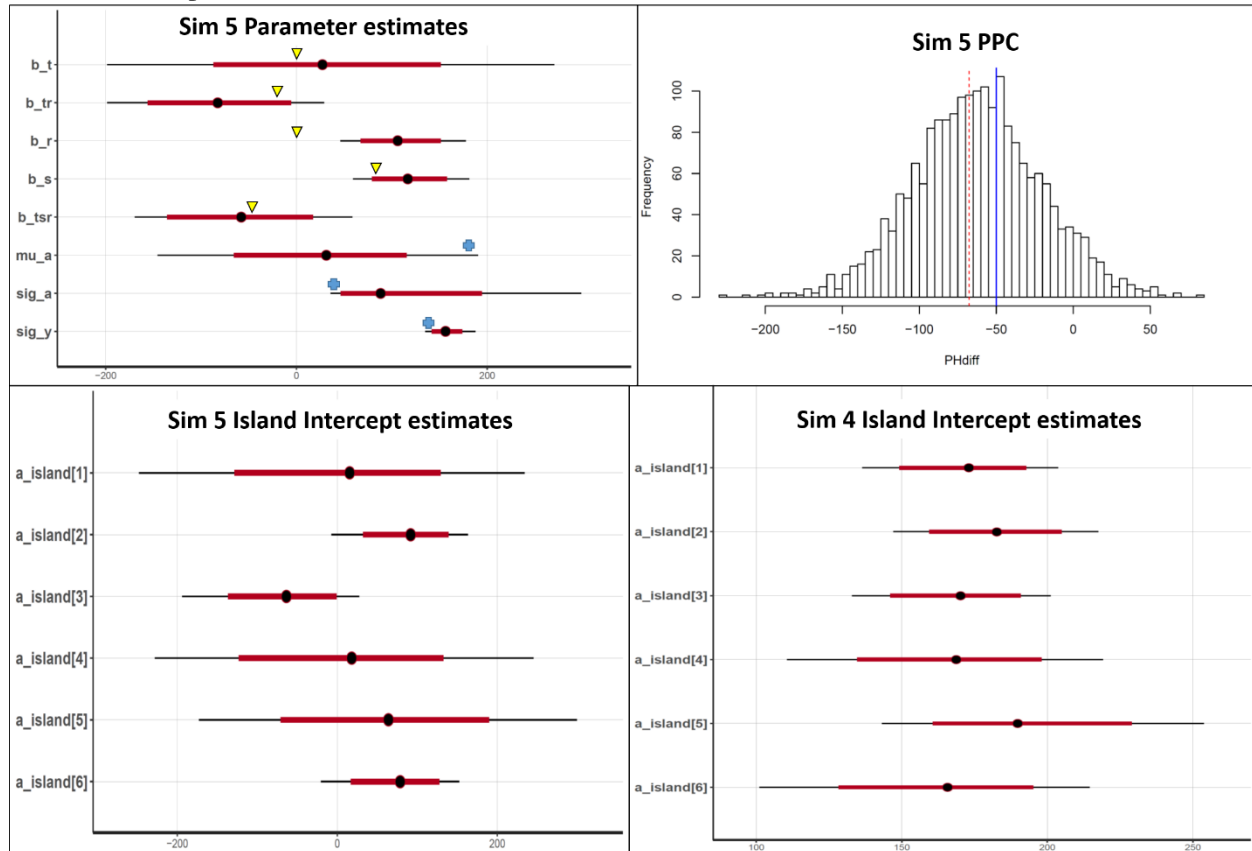
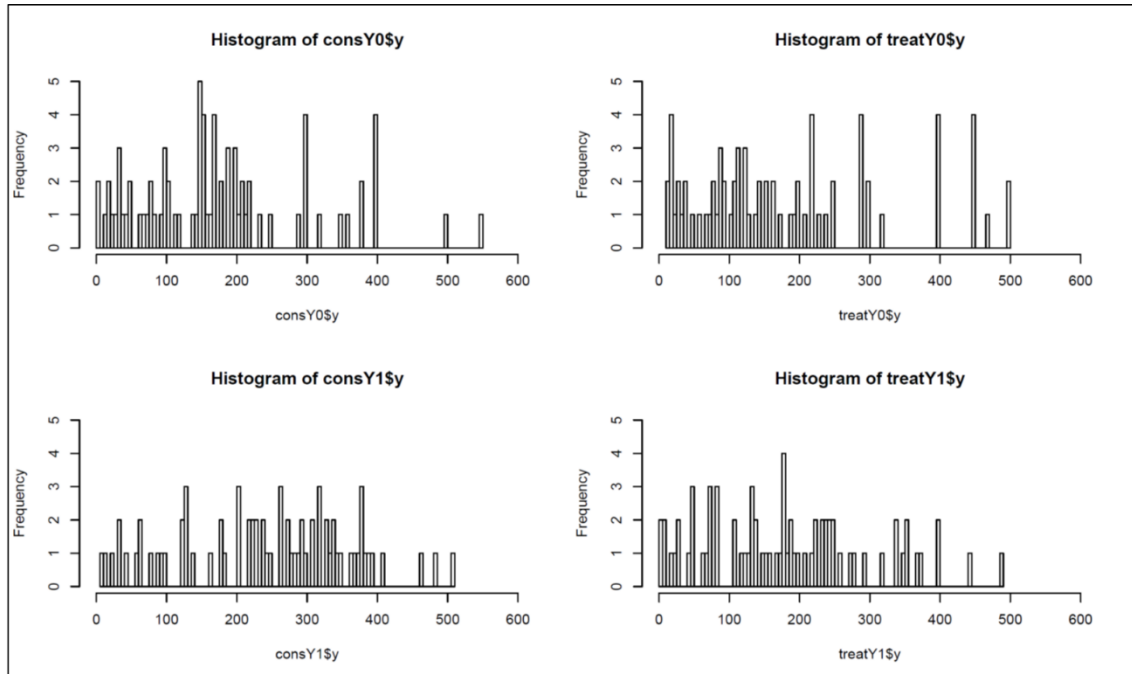


Figure 4. Histograms of raw input data for Sim 5 (top) and Sim 6 (bottom). Within each Sim, perch heights on control islands and removal islands are displayed in the left and right columns, respectively. Perch heights in Year 0: pre-removal and Year 1: post-removal are displayed in the top and bottom rows, respectively. Therefore, to compare between treatments within a given year, look left to right. To compare within a particular treatment across years, look up and down. Relative to Sim 5, perch heights on control islands in Sim 6 are poorly matched between years, suggesting that the within-island variation (σ_y) used to simulate post-removal data in Sim 6 was too small.

Sim 5



Sim 6

