

Extrapolation

Goal

Extrapolate to non-sampled spatial locations on the ranch with the learned regression for abundance. The following steps assume there is not any random effects (spatial structure, etc.) in the regression model.

This is just a posterior prediction and typically is done in the MCMC, but can be done outside.

Pseudocode

Currently, the abundance model is at the scale of the area of sampled sites (A_s) so in order to extrapolate to the area of unsampled sites (A_u), we need to calculate the ratio: $\rho = A_s/A_u$.

Input: MCMC samples of regression coefficients (beta0, beta1, beta2), covariates for unsampled sites (Herb_COH, Woody_SPLIT), ratio ρ

for each MCMC sample k :

for each unsampled site j :

- calculate scaled log-mean: $\log(mu^{(k)}) = \log(\rho) + beta0^{(k)} + beta1^{(k)}\text{Herb_COH}[j] + beta2^{(k)}\text{Woody_SPLIT}[j]$
- predict a $\sigma_{mu}[j]$ by drawing from the prior OR randomly sampling the σ_{mu} from the values for the sampled sites: $(\sigma_{mu}[1]^{(k)}, \dots, \sigma_{mu}[S]^{(k)})$ and scale by ρ to again match
- predict an abundance: $N[j] \sim \text{Normal}(\mu^{(k)}, \sigma_{mu}^{(k)})$

Random effect prediction

To be honest, writing up the prediction step for $\sigma_{\mu[j]}$ slowed me down. I read this blog on the Stan forum and it details three options of which I'd do one of the first two: <https://discourse.mc-stan.org/t/still-confused-about-how-sample-new-levels-works-in-posterior-predict/27255>