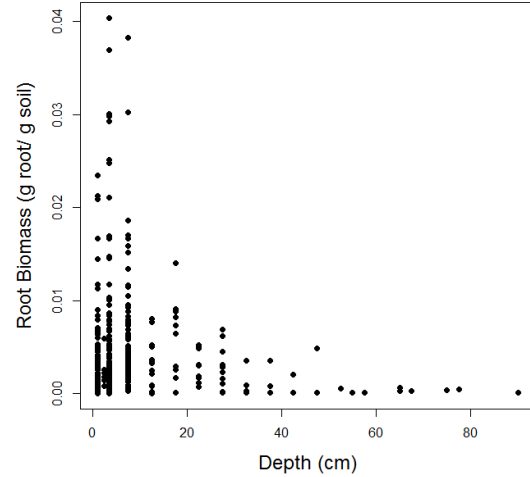


## Roots example

**The problem:** In deserts, root biomass is important for understanding plant water use and soil carbon dynamics. Roots are commonly sampled by taking soil cores at incremental depths. However, this method poses potential issues due to the small cross-sectional area that is sampled, thus increasing the likelihood of obtaining samples without roots, resulting in zero-inflated data. Models of zero inflated data often estimate the probability of detecting (or obtaining) a sample or observation. The use of a detection model can help estimate the probability that a soil core will not sample roots, thus improving estimates of root biomass at each layer in the soil.

**The data:** Root mass was collected from soil cores obtained from the Chihuahuan, Great Basin, Sonoran, and Mojave deserts. Soils were cored to a depth of 90-100 cm and divided into ~5 cm increments. The cores were taken under 10 different types of vegetation that included different shrubs, grasses, and bare ground. Root biomass was normalized by the mass of the soil core, yielding root density (g dw/g soil). Multiple plots are missing root density data because soil masses were not always obtained for each sample, thus introducing missing data. Data are plotted to the right.



**The model and assumptions:** We will simultaneously fit an exponential decay model to the root density data and a simple detection model that assumes that the probability of detecting roots is a function of the true or mean root density at a given depth. We will address the issue of missing data and compare parameter estimates based on different model versions (e.g., hierarchical vs. non-hierarchical parameter models). We will fit a model that assumes that root density decays exponentially with depth ( $D$ ), which in-turn implies that  $\log(\text{root density})$  varies linearly with depth. Let  $y_i$  denote the  $\log(\text{root density})$  for observation  $i$  ( $i = 1, 2, \dots, 1549$ ). We will implement the following model:

$$y_i \sim \text{Normal}(\mu_i, \sigma^2) \quad (\text{Eq 1. Likelihood})$$

The mean, or expected  $\log(\text{root density})$ , is given by a simple linear regression on depth ( $D$ ):

$$\mu_i = \alpha + \beta D_i \quad (\text{Eq 2. Mean model})$$

This initial model assumes that the intercept ( $\alpha$ ) and slope ( $\beta$ ) are the same for all deserts and vegetation types. You will later explore modifications to this assumption that allow these parameters to vary by desert or vegetation type.

Equations 1-2 describe the model for root biomass, but recall that we may have lots of zeros in our data. We will assume that a zero observation could either represent (1) a *real* zero such that there actually are not any roots at the depth increment represented by the sample or (2) a detection error such that the small sample had no roots because we failed to detect roots at that depth increment. We will account for this detection problem by assuming that the probability ( $p$ ) of *not getting* roots in a soil core sample is related to the “true” or predicted root density at that depth (e.g., we expect that if the predicted root density is really small, then the probability of not getting roots in the sample should be pretty high). Thus, let  $S_i = 0$  if roots were recorded

(detected) in sample or observation  $i$ , and let  $S_i = 1$  if roots were not detected in sample  $i$ . Thus, the detection model part is:

$$x_i \sim \text{Bernoulli}(p_i) \quad (\text{Eq 3. Likelihood})$$

We will model the detection error probability via a logistic regression on  $p$  is assumed to be a function of  $\mu$ :

$$\text{logit}(p_i) = a + b\mu_i \quad (\text{Eq 4. Logistic regression})$$

To complete this model, we need to specify priors; we will use fairly non-informative priors for each root node parameter:

$$\begin{aligned} \alpha, \beta, a, b &\sim \text{Normal}(0, 1000) \\ \tau &\sim \text{Gamma}(0.01, 0.01) \end{aligned} \quad (\text{Eq 5. Priors})$$

The *Normal* priors (above) are parameterized in terms of a (large) variance, and we give a *Gamma* prior to the precision,  $\tau = 1/\sigma^2$ , in the Likelihood in Eq 1. Note that  $\beta$  is the exponential decay parameters, and it describes how quickly root density decreases with depth; it might seem biologically reasonable to use a prior that restricts  $\beta < 0$  since  $\beta > 0$  would imply that root density *increases* with depth.

**Example data:** The original data were provided as in Table 1, with actual root density (see above plot). The RootDensity data were transformed into two variables, one giving  $Y = \log(\text{root density})$  for the samples with RootDensity  $> 0$ , and one for the detection variable,  $S$  (Table 2). You will work with the data shown in Table 2. (D indicates the midpoint depth of the sample.)

Table 1. Original data format.				Table 2. Re-formatted data.				
D	Desert	Veg	RootDensity	D	Desert	Veg	S	Y
2.5	1	5	0.002533	2.5	1	5	0	-5.9784
7.5	1	5	0.004501	7.5	1	5	0	-5.4035
12.5	1	5	0.002135	12.5	1	5	0	-6.1491
...				...				
57.5	1	5	0.000147	57.5	1	5	0	-8.8246
62.5	1	5	0	62.5	1	5	1	NA
67.5	1	5	0.000326	67.5	1	5	0	-8.0283
72.5	1	5	0	72.5	1	5	1	NA
77.5	1	5	0.000500	77.5	1	5	0	-7.6017
90	1	5	0.000118	90	1	5	0	-9.0489
...				...				
1	6	9	0.008425	1	6	9	0	-4.7766
3.5	6	9	0.003191	3.5	6	9	0	-5.7476
7.5	6	9	0.002637	7.5	6	9	0	-5.9382
15	6	9	NA	15	6	9	0	NA
25	6	9	NA	25	6	9	0	NA
35	6	9	0	35	6	9	1	NA
45	6	9	NA	45	6	9	0	NA
1	3	11	0	1	3	11	1	NA

### Activities:

1. Implement the above model in JAGS to obtain posterior statistics for the parameters of interest. Hints: You may notice poor mixing and convergence of your MCMC chains; you might consider “covariate centering” in both Eqs 2 and 4.
2. Does root density decrease exponentially with depth? Does the probability of “missing” (not detecting) roots increase as the predicted number of roots decrease?
3. How would you evaluate model fit?
4. Modify the above model to allow for the parameters  $\alpha$  and  $\beta$  in the root density model (Eq 2) to vary by desert, and provide a hierarchical model for these parameters, such as:

$$\alpha_d \sim \text{Normal}(\hat{\alpha}, \sigma_\alpha^2)$$

Specify relatively non-informative priors for the population-level mean (e.g.,  $\hat{\alpha}$ ) and precision (e.g.,  $\tau_\alpha$ ) similar to the priors in Eq 5. Implement / run this model. Does the root density depth profile vary among deserts?

5. Following a similar procedure and assumptions as in 4 above, now allow  $\alpha$  and  $\beta$  to vary by vegetation type. Implement / run this model. Does the root density depth profile vary among vegetation types?
6. You may try a non-hierarchical version of the models in 4 or 5 such that, for example,

$$\alpha_d \sim \text{Normal}(0, 1000)$$

How does the MCMC behavior and posterior results differ between the hierarchical model (4 or 5) and this non-hierarchical version?

### Additional reading:

A more detailed / complicated version of the above model was applied to minirhizotron data on root tip counts obtained from a long-term FACE experiment. For additional details, see:

Sonderregger, D., K. Ogle, R. D. Evans, R.S. Nowak, S. Ferguson (2013). Temporal dynamics of fine roots under long-term exposure to elevated CO<sub>2</sub> in the Mojave Desert. *New Phytologist* 198:127-138