Evaluating Differential Animal Carcass Transport Decisions at Regional Scales using Bayesian Mixed-Effects Models

Ryan P. Breslawski

Department of Anthropology, Southern Methodist University, Dallas, Texas

1. Predicting element frequencies.

How can we infer relationships between archaeofaunal element frequencies and taphonomic or anthropological predictors of those frequencies? Archaeologists recover animal bones in frequencies that differ from their anatomical distribution in a living animal. This results from post-depositional taphonomic processes and differential transport decisions made by humans in the past. Zooarchaeologists typically assess the latter process using measures of dietary utility.

2. Existing approaches

rely on various measures of ordinal correlation, parametric correlation, or regression models, with bone counts converted to MAU values to standardize counts across anatomical frequencies. The importance of relationships is assessed with *p* values, effect size point-estimates (e.g., Spearman's ρ), and/or model parameter point-estimates (e.g., a regression coefficient). This has **numerous pitfalls:**

- *Point-estimates* can be noisy and unreliable, especially when MAU values are derived from low counts.
- MAU values reduce the sample of counted elements to a continuous rate variable defined by few observations. This obscures sample size differences between assemblages, making comparisons of element count-predictor relationships between assemblages unreliable.
- Significance threshold values have no zooarchaeological justification in these contexts, and "significance" is highly contingent on the number of element categories over which MAU values are calculated.

3. Proposed approach

: Bayesian generalized linear mixed-effects models (GLMMs; **Box 1**). These GLMMs have **numerous advantages:**

- Inferred model parameter values are represented by posterior probability distributions rather than point-estimates. These distributions are a function of the data and **skeptical prior distributions**. Posteriors are conservative and communicate uncertainty. Inferences are not limited to dichotomous decisions based on an arbitrary threshold.
- A Poisson likelihood allows these GLMMS to model element counts rather than MAU values. Variability in anatomical frequency is accommodated with a varying exposure term.
- Element counts from multiple assemblages can be modelled in a single GLMM, allowing for easy inter-assemblage comparisons. In a regional context, this also yields the "average" relationship for an assemblage and produces predictions about what other assemblages in the region should look like.
- These models can incorporate multiple predictors. Bone density and measures of dietary utility should be considered simultaneously, as we know that both factors have likely conditioned bone frequencies to some degree.
- Uncertainty in the values of taphonomic and dietary predictors can be incorporated into the model (**Box 2**). This is in contrast to using averaged sample values.

Box 1. Model formulas.

Likelihood:

$$element\ count_i \sim Poisson(\lambda_i)$$

Linear model:

$$log(\lambda_i) = \overbrace{\log(AF_i)}^{exposure\ for\ anatomical\ frequencies} + \overbrace{\hat{A}_i}^{inter.} + \overbrace{B_i^{dens}}^{bone\ density\ effect} + \overbrace{\sum_{j=1}^{J} B_i^j}^{sum\ of\ effects\ across\ J\ dietary\ util.\ predictors}$$

$$A_i = \overbrace{\alpha_{type[i]}}^{inter.for\ site\ type\ t} + \overbrace{\alpha_{site[i]}}^{offset\ for\ site\ s} + \overbrace{\hat{\alpha}_i}^{observation\ level\ overdispersion\ offset}$$

$$B_i^{dens} = (\overbrace{\beta_{type[i]}^{dens}}^{mean\ effect\ site\ type\ t} + \overbrace{\beta_{site[i]}^{dens}}^{offset\ for\ site\ s}) \times density_i$$

$$B_i^j = (\overbrace{\beta_{type[i]}^{j}}^{mean\ effect\ site\ type\ t} + \overbrace{\beta_{site[i]}^{j}}^{offset\ for\ site\ s}) \times utility_i^j +$$

$$\underbrace{(\overbrace{\beta_{type[i]}^{jxdens}}^{mean\ effest\ site\ type\ t} + \overbrace{\beta_{site[i]}^{jxdens}}^{offset\ for\ site\ s}) \times utility_i^j \times density_i}_{interaction\ term\ for\ utility\ x\ density}$$

Priors:

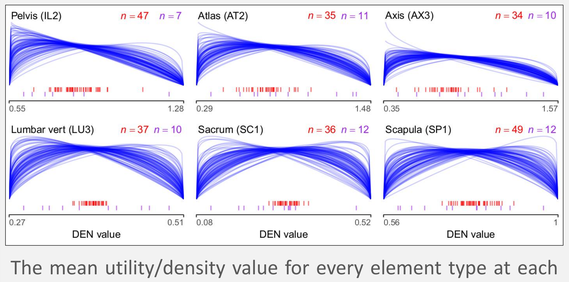
$$\alpha_{type}\ ,\ \beta_{type}^{density}\ ,\ \beta_{type}^{j}\ ,\ \beta_{type}^{jxdens} \sim N(0,1)$$

$$\alpha \sim N(0, \sigma_{site[s]}) \qquad \sigma_{site} \sim \Gamma(p1/p2, 1/p2) \qquad p1 \sim \exp(1) \qquad p2 \sim \exp(2)$$

$$\begin{bmatrix} \alpha_{site} \\ \beta_{site}^{dens} \\ \beta_{site}^{j=1}, \beta_{site}^{jxdens=1} \\ \vdots \\ \beta_{site}^{j=J}, \beta_{site}^{jxdens=J} \end{bmatrix}_{type[t]} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{cov. matrix composed of J x 2 +2 } \sigma^{[t]} \text{ terms and correlation matrix } R \right)$$

$$\sigma_1^{[t]}, \sigma_2^{[t]}, \ldots, \sigma_{J\times 2+2}^{[t]} \sim \exp(2) \qquad R \sim LKJcorr(4)$$

Box 2. Modelling uncertainty in predictor values.



The mean utility/density value for every element type at each component is modelled based on the distribution of known values, as recorded in experimental studies. These panels show **modelled density distributions** for six element types, including **known values**, and **median posterior values for each component** where those elements were recorded as present.

Example: Bison from 52 Early Holocene components in North America.

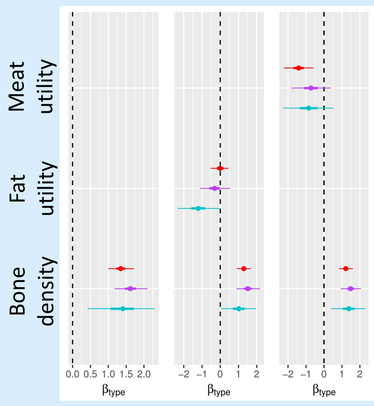
32 kill site components, **15 camp site components**, and **5 components of unknown type**. Components contain 5–10,846 elements, with a median value of 234.5. I specified three models with alternative predictors in the Stan language. Posterior parameter samples were generated via Hamiltonian Monte Carlo simulation:

Element counts = Density
Element counts = Density + Fat utility
Element counts = Density + Meat utility

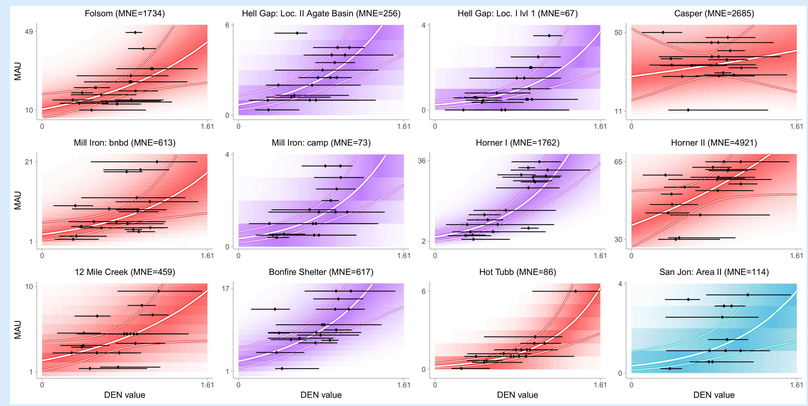
Photo by RP Breslawski

Photo from nps.gov

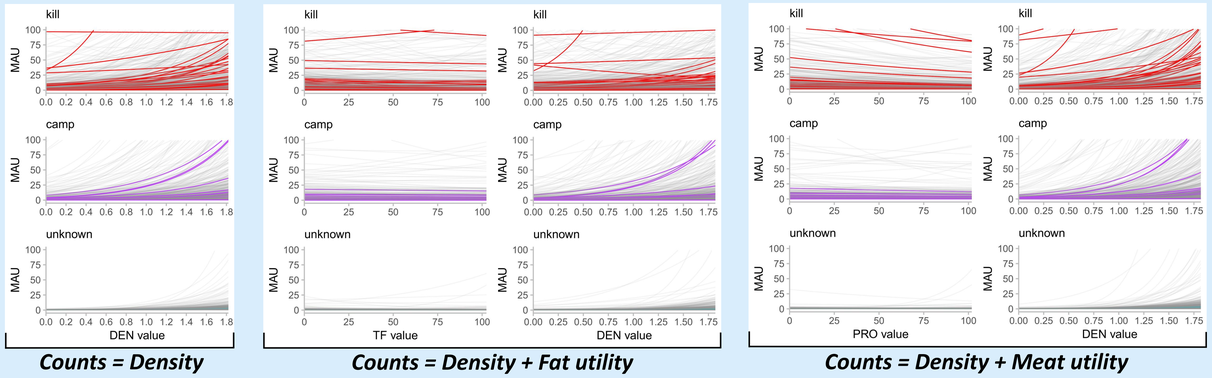
Density data: Kreutzer 1992 & Utility data: Emerson 1990



Marginal posterior parameter estimates for predictor effects in each model. Dots and segments show posterior median, 50% HPDI, and 95% HPDI values.



Relationships for 12 example components (Counts = Density model). Black dots and segments are posterior medians and 95% HPDIs for density values. White lines depict the posterior mean and 95% HPDIs for relationships between bone density and element counts. Color transparency corresponds to posterior predictive distributions for these relationships.



Counts = Density ***Counts = Density + Fat utility*** ***Counts = Density + Meat utility***

Lines in color are for observed components, grey lines are components simulated from posterior model parameters.

References and libraries

Emerson AM (1990) Archaeological Implications of Variability in the Economic Anatomy of *Bison bison*. PhD dissertation, Washington State University, Pullman. University Microfilms International, Ann Arbor.

Kreutzer LA (1992) The Taphonomy of the Mill Iron, Montana (24CT30) Bison Bone Bed. Unpublished PhD dissertation, University of Washington, Seattle.

Pederson TL (2017) patchwork: The Composer of ggplots. V 0.0.1. https://github.com/thomas85/patchwork

Stan Development Team (2018) Rstan: the R interface to Stan. V 2.18.2. http://mc-stan.org/

Venables WN, Ripley DB (2002) *Modern Applied Statistics with S*. 4th ed. Springer, New York.

Wickham H (2007) Reshaping data with the reshape package. *J Stat Softw* 21(12), 1-20. v 1.4.3

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. V 3.1.0

Acknowledgments

This work was completed on the ManeFrame II cluster, Center for Scientific Computation, Southern Methodist University.

csv data, Stan code, R code, and poster PDF

SMU