

IDK

Peter D. Smits^{1,*}

1. University of Chicago, Chicago, Illinois 60637.

* Corresponding author; e-mail: psmits@uchicago.edu.

Manuscript elements:

Keywords:

Manuscript type: Article

Prepared using the suggested L^AT_EX template for *Am. Nat.*

1 Materials and Methods

1.1 Taxon occurrence information

1.2 Supertree inference

1.3 Model specification

Define y as a $N \times T$ matrix of the observed 0/1 occurrence of species i at time t , and z as a $N \times T$ matrix of the “true” 0/1 occurrence of species i at time t . The observation presence of a taxon $y_{i,t}$ is modeled as following a Bernoulli distribution defined

$$y_{i,t} \sim \text{Bernoulli}(\rho_t z_{i,t}) \quad (1)$$

where ρ is a length T vector of the probability that a species is observed at time t given that it is present at time t ($\rho_t = \Pr(y_{i,t} = 1 | z_{i,t} = 1)$).

Observation probability ρ_t is modeled as Bernoulli distributed variable with probability equal to the inverse logit transformed length T vector ρ' (Eq. 2). The elements of ρ' are modeled as exchangeable draws from a Normal distribution with location ρ'' and scale $\sigma_{\rho'}$; these are given weakly informative Normal and half-Cauchy priors, respectively.

$$\begin{aligned} \rho_t &\sim \text{Bernoulli}(\text{logit}^{-1}(\rho'_t)) \\ \rho'_t &\sim \mathcal{N}(\rho'', \sigma_{\rho'}) \\ \rho'' &\sim \mathcal{N}(0, 1) \\ \sigma_{\rho'} &\sim \text{C}^+(1) \end{aligned} \quad (2)$$

The “true” presence/absence states of taxa z are modeled as a logistic regression (Eq. 3) where the probability of being present ($z = 1$) is a function of individual-level and group-level effects. The intercept of this regression α is a length T vector which varies by time t (Eq. 4). Individual-level covariates are a $N \times D$ matrix x whose regression coefficients are the length D column vector β .

18 These regression coefficients are given independent weakly informative Normally distributed priors.

$$\begin{aligned} z_{i,t} &\sim \text{Bernoulli}(\text{logit}^{-1}(\alpha_t + x_i\beta)) \\ \beta &\sim \mathcal{N}(0, 1) \end{aligned} \tag{3}$$

The value of α_t is itself defined as a regression (Eq. 4). The intercept of the this group-level regression is defined as μ . The effect of plant regime p is modeled as the length P vector ϕ whose elements are draws from a Normal distribution with a location of 0 and scale of σ_ϕ . The group-level covariates corresponding to global climate values of interest is defined as the $T \times U$ matrix u where γ is a length U column vector of regression coefficients. Finally σ_μ is positive real value. μ , the elements of γ , σ_μ , and σ_ϕ are all given weakly informative independent priors (Eq. 4).

$$\begin{aligned} \alpha_t &\sim \mathcal{N}(\mu + \phi_{p[t]} + u_t\gamma, \sigma_\mu) \\ \phi_p &\sim \mathcal{N}(0, \sigma_\phi) \\ \mu &\sim \mathcal{N}(0, 5) \\ \gamma &\sim \mathcal{N}(0, 1) \\ \sigma_\mu &\sim \text{C}^+(1) \\ \sigma_\phi &\sim \text{C}^+(1) \end{aligned} \tag{4}$$

Acknowledgements

26 I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for helpful discussion and advice. Miller for the epicontinental versus open-ocean assignments. This entire study would not have been possible without the Herculean effort of the many contributors to the Paleobiology Database. In particular, I would like to thank J. Alroy and M. Uhen for curating most of the mammal occurrences recorded in the PBDB. This is Paleobiology Database publication XXX.