Supplementary text for "Placing probabilities on taxon true absence: applications to the hominin genus *Paranthropus*"

Andrew Du, Eric Friedlander, John Rowan, Zeresenay Alemseged

Contents

1 Detailed overview of the mixture model **1**

1.1 Motivation 1

1.2 Derivation 2

1.3 Complete and expected likelihood functions 4

1.4 Expectation-maximization algorithm 7

2 Simulations to double-check the model **8**

3 Assessing model assumptions **13**

3.1 Assumption #1: independence of site data 13

3.2 Assumption #2: independence of specimen data 16

References **20**

1 Detailed overview of the mixture model

1.1 Motivation

While the methodology presented in this work is broadly applicable to any taxon, regarding the problem of inferring its true absence at each of a collection of sites, we present the model using the example of the hominin genus *Paranthropus* from eastern Africa, as was done in the main text. Inferring whether *Paranthropus* was truly absent at a site is difficult because observed absence is consistent with two mutually exclusive outcomes: (A) *Paranthropus* never occupied the site (i.e., "true absence"), or (B) it did occupy the site but has not been sampled yet. We are ultimately interested in estimating the probability of each of these two possibilities, conditional on not finding a *Paranthropus* specimen after sampling n specimens at a given site. To accomplish this, we propose a mixture model wherein each of the two outcomes is modeled with a different probability distribution called a *mixture component*. As we will see, this model, along with Bayes' rule, will allow us to compute the probability that a given site was generated from possibility (A) or (B), conditional on sampling n mammalian specimens without finding *Paranthropus*, and thus estimate the likelihood that *Paranthropus* was truly absent from a given site or has not been sampled yet.

1.2 Derivation

We now present the proposed model. In order to instill intuition, we outline the model in a generative fashion (i.e., how data would be generated from the model). Let ψ be the probability that a given site belongs to component (B) (i.e., *Paranthropus* is present). In particular, if one were to select a site uniformly at random (with no additional information about the recovered specimens), ψ would represent the probability that *Paranthropus* was present. Intuitively, one can think of ψ as the expected proportion of all sites that contains *Paranthropus*, regardless of whether it has been observed yet. Thus, there is a single ψ that needs to be specified for the model, which applies across all sites. It follows that $1 - \psi$ is the probability that a site belongs to component (A) (i.e., *Paranthropus* was truly absent from the site).

For site i, let Z_i be a latent (unobserved) variable that is equal to 1 if *Paranthropus* was present and 0 otherwise. It follows that $P(Z_i = 1) = \psi$ and $P(Z_i = 0) = 1 - \psi$. In the language of statistics, we say that Z_i is a $Bernoulli(\psi)$ random variable and write:

$$Z_i \sim Bernoulli(\psi). \tag{1}$$

For each value of Z_i (i.e., 0 or 1), we must specify the distribution on the number of *Paranthropus* specimens that will be sampled if n_i total mammalian specimens are collected. In general, if we were to know the sampling probability of *Paranthropus*, denoted p_i, then the number of sampled *Paranthropus* specimens can be modeled using a $Binomial(n_i, p_i)$ distribution. Using an example, a $Binomial(n, p)$ random variable represents the number of heads one may observe if flipping a weighted coin n times, given that the probability of obtaining heads from a single flip is p. In our setting, if we know that p_i is the probability that a randomly chosen specimen from site i will be *Paranthropus* (i.e., its sampling probability), then the number of *Paranthropus* specimens out of n_i mammalian specimens has a $Binomial(n_i, p_i)$ distribution.

We now use the value of Z_i to generate a value for p_i. Recall that if $Z_i = 0$, then *Paranthropus* never occupied the ith site, so the probability that a sampled mammalian specimen is *Paranthropus* is zero (i.e., $p_i = 0$). This is the first mixture component in our model, and we say that there is a *point mass* at 0 because all of the mass of the binomial distribution is concentrated at 0 (i.e., if $p_i = 0$, then the number of sampled *Paranthropus* specimens must also equal 0). If $Z_i = 1$, then *Paranthropus* did occupy the site. However, the sampling probability of *Paranthropus* varies from site to site. Therefore, we model p_i using the standard reflected beta distribution (Wang et al., 2016), which has the probability density function $(1 - \lambda)(1 - p_i)^{-\lambda}$ if $\lambda \leq 0$. There is a corresponding expression for the case $\lambda > 0$, but we restrict ourselves to the case when $\lambda \leq 0$ for reasons that will soon become clear. The standard reflected beta distribution is simply a reparametrization of the standard beta distribution with parameters $\alpha = 1$ and $\beta = 1-\lambda$, i.e., $Beta(1, 1-\lambda)$. The parameter λ is the same across all sites, and it determines the shape of the distribution, which in our case ($\lambda \leq 0$) can only be monotonically decreasing. This is appropriate here because such a distribution has the majority of its probability density concentrated towards lower values of p_i, reflecting the general ecological pattern that a species is rare, and thus harder to sample, at most sites within its geographic range (Brown et al., 1995; Murray et al., 1999). As with ψ, we estimate λ from the data. Figure 1 shows how different values of λ affect the shape of the distribution.

Combining scenario (A), where *Paranthropus* is truly absent at the ith site and $Z_i = 0$, with scenario (B), where *Paranthropus* is present at the ith site and $Z_i = 1$:

$$P(p_i | Z_i = 0) = \begin{cases} 1 & \text{if } p_i = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2a}$$

$$f(p_i | Z_i = 1) = (1 - \lambda)(1 - p_i)^{-\lambda}, \quad \text{if } \lambda \leq 0, \quad 0 \leq p_i \leq 1. \tag{2b}$$

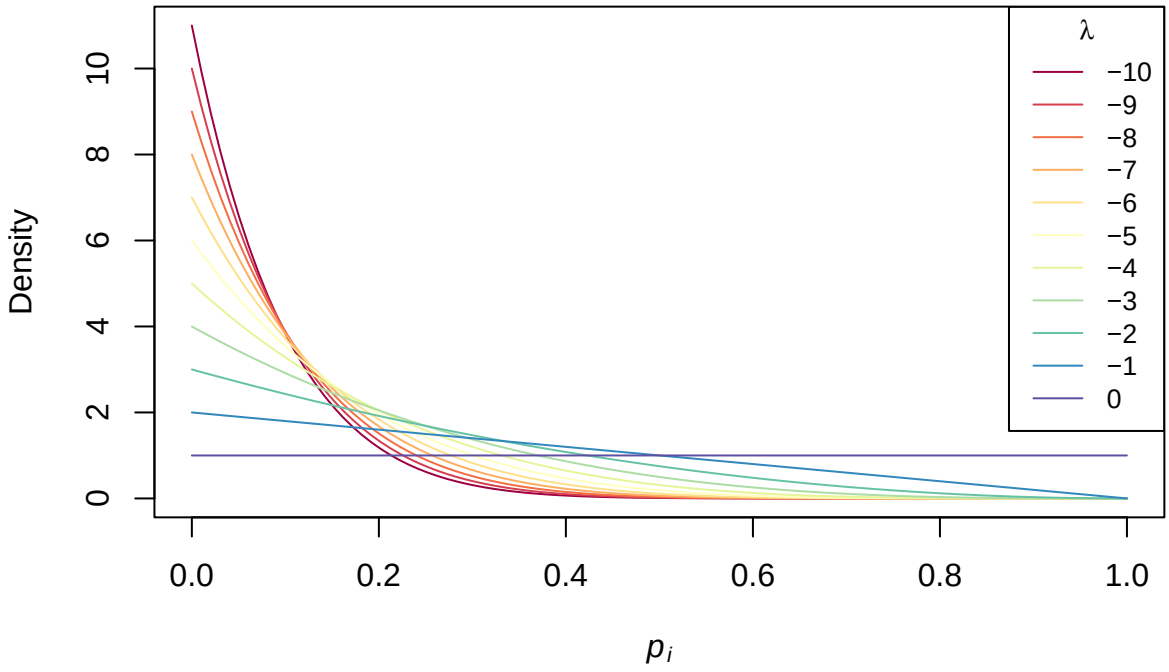

Figure 1: Shape of the standard reflected beta distribution under different values of the parameter, λ.

Equation 2a can also be expressed more conveniently using an *indicator function*: $P(p_i|Z_i = 0) = I(p_i = 0)$, where $I(p_i = 0)$ equals 1 if $p_i = 0$ and 0 otherwise.

Finally, with p_i defined for each site, we can now model the number of *Paranthropus* specimens recovered at site i, denoted X_i, as:

$$X_i \sim Binomial(n_i, p_i) \tag{3}$$

where n_i is the total number of mammalian specimens recovered at site i and is given by the observed data.

Since we have now specified the entire generative model, we can use this model to compute the probability of recovering, for any i, X_i specimens of *Paranthropus* from the ith site. First, focus on the case of $Z_i = 0$. Recall that if $Z_i = 0$, then $p_i = 0$, and our sampling probability becomes a point mass at 0. Therefore, the number of sampled *Paranthropus* specimens must be zero (i.e., $X_i = 0$), so:

$$P(X_i = x|Z_i = 0) = I(X_i = 0). \tag{4}$$

Alternatively, if $Z_i = 1$, we must account for both the randomness in selecting p_i and the randomness in sampling X_i given p_i, making the sampling probability hierarchical. Recalling that p_i is distributed according to $Beta(1, 1 - \lambda)$, we have:

$$P(X_i = x|Z_i = 1) \sim Binomial(n_i, p_i), \text{ where } p_i \sim Beta(1, 1 - \lambda). \tag{5}$$

In Bayesian statistics, this commonly used distribution is referred to as the beta-binomial distribution and has the following probability mass function:

$$P(X_i = x|Z_i = 1) = f_B(x_i; n_i, \lambda) = \binom{n_i}{x_i} \frac{B(x_i + 1, n_i - x_i + 1 - \lambda)}{B(1, 1 - \lambda)} \tag{6}$$

where B denotes the beta function.

Combining Equations 1, 4, and 6 by adding over the two possible values of the latent variable Z_i, the law of total probability implies that:

$$\begin{aligned} P(X_i = x) &= P(X_i = x|Z_i = 0)P(Z_i = 0) + P(X_i = x|Z_i = 1)P(Z_i = 1) \\ &= I(x = 0)(1 - \psi) + f_B(x; n_i, \lambda)\psi. \end{aligned} \tag{7}$$

Equation 7 is the probability of recovering x number of *Paranthropus* specimens for the ith site if the sample size is n_i.

The general logic of the mixture model and how the parameters, variables, and data are related to each other are illustrated in Figure 2.

1.3 Complete and expected likelihood functions

In the previous section, we discussed how the data are generated by our model, but ultimately, we are interested in going in the opposite direction: that is, given the data, $X_1, X_2, ...X_{52}$ (where X_1 is the number of *Paranthropus* specimens in the first site, X_2 is the number of *Paranthropus* specimens in the second site, and we have a total of 52 sites), what are the parameters (ψ, λ)? Furthermore, we are interested in using these

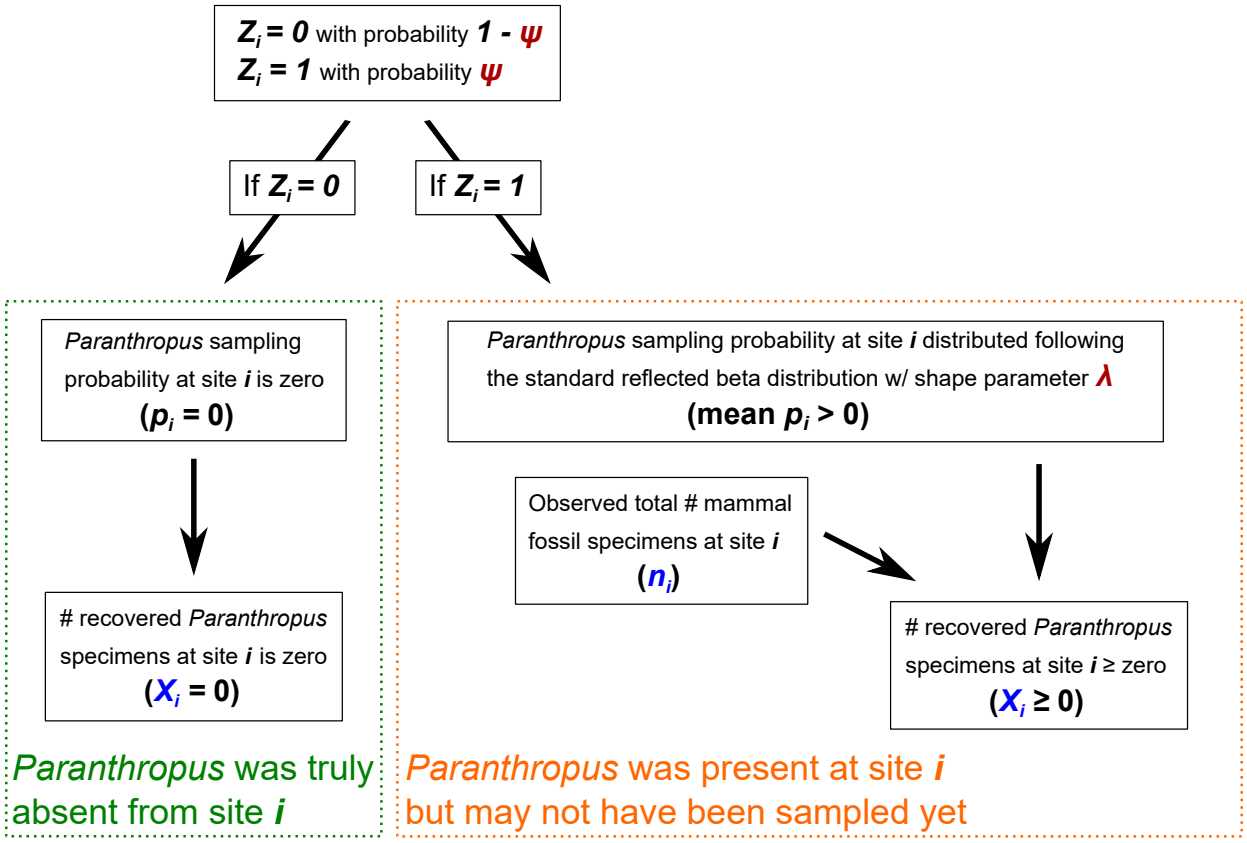

Figure 2: Flow chart illustrating the general logic of the model and how the parameters, variables, and data are related to each other. Estimated parameters are in red, while observed data are in blue.

parameter estimates to infer the latent variables $Z_1, Z_2, ..., Z_{52}$, using data from both within and outside the ith site. In other words, we would like to assess the probability that a given site has *Paranthropus* present or not, conditional on observations from the site and parameter estimates in our model (which are informed by data from all sites). In order to estimate the parameters (ψ, λ), we use a method called maximum likelihood estimation. Intuitively, this method selects the parameter values that will make the data most probable, with larger probabilities indicating more likely parameter estimates (Wang, 2010). Using the estimated parameter values, we can estimate the probability that $Z_i = 1$ for the ith site, which is the probability that *Paranthropus* was present at the site (the complement of this probability is the probability that *Paranthropus* was absent from the ith site). We note that the case that is most interesting is when no *Paranthropus* specimens have yet been observed at the given site. To do all this, we operate in the framework of likelihood and define a *likelihood function*.

For now, assume that Z_i is known. This allows us to compute the *complete likelihood function*, where the word "complete" indicates that our latent variables, Z_i, are explicitly accounted for. Assuming the site data are independent from each other and adopting the convention that $0^0 = 1$, the complete likelihood function can be written as:

$$L(\psi, \lambda) = P(X, Z|\psi, \lambda) = \prod_{i=1}^{N} [I(x_i = 0)(1 - \psi)]^{1-Z_i} \times [f_B(x_i; n_i, \lambda)\psi]^{Z_i}. \tag{8}$$

where N is the total number of sites. It is worth spending some time dissecting Equation 8 and understanding how it works. If $Z_i = 0$ (i.e., *Paranthropus* is truly absent from the ith site), the second element in the product in Equation 8 equals 1 and is not considered in the likelihood function. Conversely, if $Z_i = 1$ (i.e., *Paranthropus* is present), the first element equals 1 and is not considered. Thus, different values of Z_i will "turn on and off" different mixture components, so the proper one is considered.

To make the likelihood function more mathematically tractable, we log-transform it to get the log-likelihood function:

$$\ell(\psi, \lambda) = log(P(X, Z|\psi, \lambda)) = \sum_{i=1}^{N} [(1 - Z_i)log(1 - \psi) + Z_i(log[f_B(x_i; n_i, \lambda)] + log[\psi])]. \tag{9}$$

We can only calculate the complete log-likelihood, however, if we know Z_i in Equation 9, and we do not. To get around this obstacle, we compute the expected value of $\ell(\psi, \lambda)$, conditional on the data. While this may sound daunting, it amounts to computing the posterior distribution of $Z_i = 1$, given the data, which can be easily accomplished using Bayes' rule. More precisely, this posterior distribution is composed entirely of $P(Z_i = 1|X_i)$, and one minus this probability gives $P(Z_i = 0|X_i)$. We then use $P(Z_i = 1|X_i)$ as a stand-in for Z_i in Equation 9 to calculate the expected log-likelihood. Denoting the posterior probability as τ_i, which is a function of parameters ψ and λ,

$$\tau_i(\psi, \lambda) = P(Z_i = 1|X_i, \psi, \lambda) = \frac{P(X_i|Z_i = 1)P(Z_i = 1)}{P(X_i)} \tag{10a}$$

$$= \frac{f_B(x_i; n_i, \lambda)\psi}{I(X_i = 0)(1 - \psi) + f_B(x_i; n_i, \lambda)\psi} \tag{10b}$$

where the first line (10a) follows from Bayes' rule, and the second line (10b) follows from Equations 1 and 6 (numerator) and 7 (denominator). Note that this is the complement of the posterior probability of interest in the main text (i.e., the probability of true *Paranthropus* absence at site i, given the data and parameters, or $1 - \tau_i(\psi, \lambda)$).

Now we can calculate the expected log-likelihood by substituting τ_i (Equation 10) in for Z_i in Equation 9. Calling the expected log-likelihood Q,

$$Q(\psi, \lambda) = \mathbb{E}[\ell(\psi, \lambda)|X] = \sum_{i=1}^{N}[(1 - \tau_i(\psi, \lambda))log(1 - \psi) + \tau_i(\psi, \lambda)(log[f_B(x_i; n_i, \lambda)] + log[\psi])]. \tag{11}$$

To estimate parameters ψ and λ, we want to maximize Q with respect to ψ and λ. For ψ, this can be accomplished analytically by taking the derivative of Q with respect to ψ and setting it equal to zero as follows:

$$\frac{\partial Q}{\partial \psi} = \sum_{i=1}^{N}[(1 - \tau_i(\psi, \lambda))\frac{1}{1 - \psi} + \tau_i(\psi, \lambda)\frac{1}{\psi}] = 0. \tag{12}$$

Solving Equation 12 then gives the following expression for the maximizing value of ψ:

$$\psi_{max} = \frac{1}{N}\sum_{i=1}^{N}\tau_i(\psi, \lambda) \tag{13}$$

which can be thought of as the average probability that $Z_i = 1$, given the data (cf. Equation 10). This should make sense, since this is how we defined ψ originally, which can also be interpreted as the proportion of sites in which *Paranthropus* was present.

Due to the complicated nature of f_B (Equation 6), it is intractable to solve for λ analytically. Instead, we will numerically solve for λ_{max} in Equation 11 via the L-BFGS-B optimization algorithm (Byrd et al., 1995).

1.4 Expectation-maximization algorithm

1.4.1 Description

We are now in a conundrum: if we knew the parameters ψ and λ, we could compute the posterior probability τ_i (Equation 10), which is what we are ultimately interested in. If we knew τ_i, we could compute the parameters ψ (Equation 13) and λ (numerical optimization of Equation 11). In fact, we can alternate performing each of these steps iteratively, which constitutes an algorithm known as expectation-maximization (EM). The EM algorithm is a standard numerical method used to compute the maximum likelihood estimates of parameters for models with unobserved latent variables (e.g., Z_i) (Searle et al., 2006). In the EM algorithm, we:

1) Take initial guesses of ψ and λ (call them $(\psi^{(0)}, \lambda^{(0)})$).

2) [E-Step] Calculate all τ_i's (Equation 10) using the most recent estimates of ψ and λ. We will call them $\psi^{(j)}$ and $\lambda^{(j)}$ to indicate that these are the jth values of ψ and λ.

3) Use the τ_i's to compute the expected log-likelihood, $Q(\psi^{(j)}, \lambda^{(j)})$ (Equation 11). Calculating Q is not really necessary, but it can serve as an indicator of how goodness of fit changes as the model improves.

4) [M-Step] Using the new τ_i's, estimate a new $\psi^{(j+1)}$ using Equation 13 and a new $\lambda^{(j+1)}$ by numerically optimizing Equation 11.

5) [Stopping Criteria] Evaluate the new parameter estimates compared to the old ones. If they have changed by less than some small prespecified amount, stop. Otherwise, go back to step 2.

One more efficiently expresses these steps by simply writing:

$$(\psi^{(j+1)}, \lambda^{(j+1)}) = \underset{(\psi,\lambda)}{\text{argmax}} Q(\psi, \lambda | \psi^{(j)}, \lambda^{(j)}) \tag{14}$$

Once we have estimated ψ and λ using the EM algorithm (Equation 14), we can compute τ_i for each site using Equation 10. This is the posterior probability that *Paranthropus* was present at site i, given the estimated parameters and data. We are mainly interested in the complement of this probability: 1 - τ_i (i.e., the posterior probability that *Paranthropus* was truly absent from site i, given the estimated parameters and data).

2 Simulations to double-check the model

For this section, we simulate *Paranthropus* abundances across sites, given a known ψ and λ. We then estimate ψ and λ with our model (Equation 14) and compare the estimates to their predetermined values. We use the observed number of sites and number of mammalian specimens across sites for the simulations, thereby only varying ψ and λ. We conduct the simulations as follows:

1) To determine at which sites *Paranthropus* was truly absent or present (i.e., Z_i), we take a random draw from a binomial distribution, where the number of trials is the number of sites, and the probability of success (i.e., *Paranthropus* is present) is the prespecified value of ψ. The output is a vector of 0s and 1s, where the length of the vector is the total number of sites, 0 denotes true *Paranthropus* absence, 1 denotes true *Paranthropus* presence, and the proportion of 1s, on average, equals ψ.

2) For those sites that have *Paranthropus* (i.e., $Z_i = 1$), we simulate *Paranthropus* abundances (i.e., X_i) as a random draw from the beta-binomial distribution, given a prespecified λ value (Equation 6) and the observed number of mammalian specimens at each site. We do this using the `rbetabinom` function in the `rmutil` package (Swihart & Lindsey, 2020).

3) Steps 1-2 are iterated 1000 times.

Chosen values for ψ are (0.1, 0.3, 0.5, 0.7, 0.9), and those for λ are (-200, -100, -50, -25, -10, -5, -3, -1). The λ values cover a range of standard reflected beta distribution shapes for true *Paranthropus* relative abundance across sites (i.e., p_i) (Figure 3), while encompassing the estimated value from our main analyses (i.e., -148). We explore all pairwise combinations of ψ and λ, and for each combination of values, we fit our model to the simulated data, iterated 1000 times. When estimating λ by numerically optimizing Equation 11, we cap the number of iterations in the optimization algorithm at 5000 due to the computationally intensive nature of these simulations (visual inspection of trace plots shows that the optimization algorithm typically converges on its parameter estimate well before 5000 iterations). We then calculate the mean and median estimated parameter values across all 1000 iterations for each pairwise combination of ψ and λ. We also calculated the median because our parameters are bounded (ψ is bounded by 0 and 1, and λ is bounded by $-\infty$ and 0), so their sampling distributions are likely to be skewed as the prespecified values approach these bounds. Calling the mean/median estimated parameter $\hat{\mu}$ and the true parameter μ, we measure relative bias as $(\hat{\mu} - \mu)/ \mid \mu \mid$. We use the absolute value of μ in the denominator to account for the fact that λ is negative, so a negative relative bias translates to $\hat{\lambda} < \lambda$ (e.g., -150 < -100).

All of this is operationalized with the `simulateData` function in our R script file. Note that we use R v.4.1.2 (R Core Team, 2021) for all our analyses.

When assessing relative bias using the mean parameter estimate over all 1000 iterations for each pairwise combination of ψ and λ, results show that $\hat{\psi}$ is fairly unbiased, though there is a slight positive bias when true ψ is 0.1 and when true λ becomes more negative (Figure 4). This is due to a large number of simulated zero

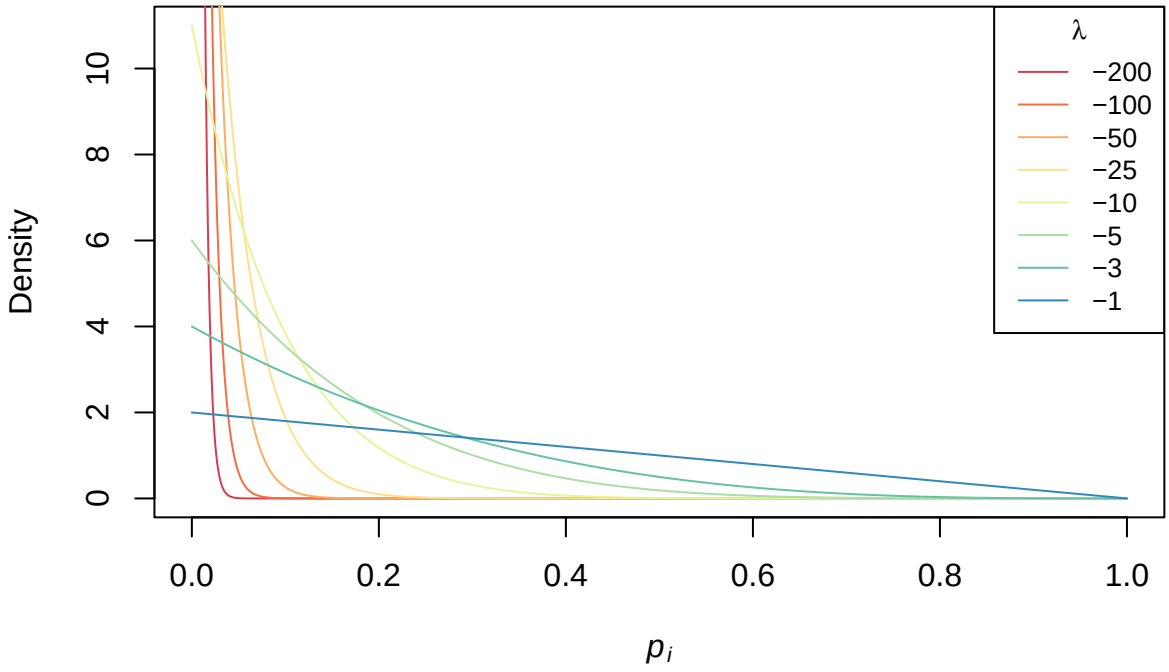

Figure 3: Shape of the standard reflected beta distribution for different values of λ in our simulations.

Paranthropus abundances across sites when true ψ is low (i.e., low proportion of sites with true *Paranthropus* presences) and true λ is very negative (i.e., probability density of *Paranthropus* relative abundance across sites is concentrated more towards zero; Figure 3). As a result, the model is interpreting some of the observed zero abundances as unsampled presences, hence the larger $\hat{\psi}$'s. Indeed, larger $\hat{\psi}$'s are associated with more negative $\hat{\lambda}$'s (Figure 5): λ's are only estimated for those sites where *Paranthropus* is inferred to be present, and very negative $\hat{\lambda}$'s are associated with more unsampled presences. Overall, however, the relative bias for $\hat{\psi}$ is small (range: -0.01 to 0.1).

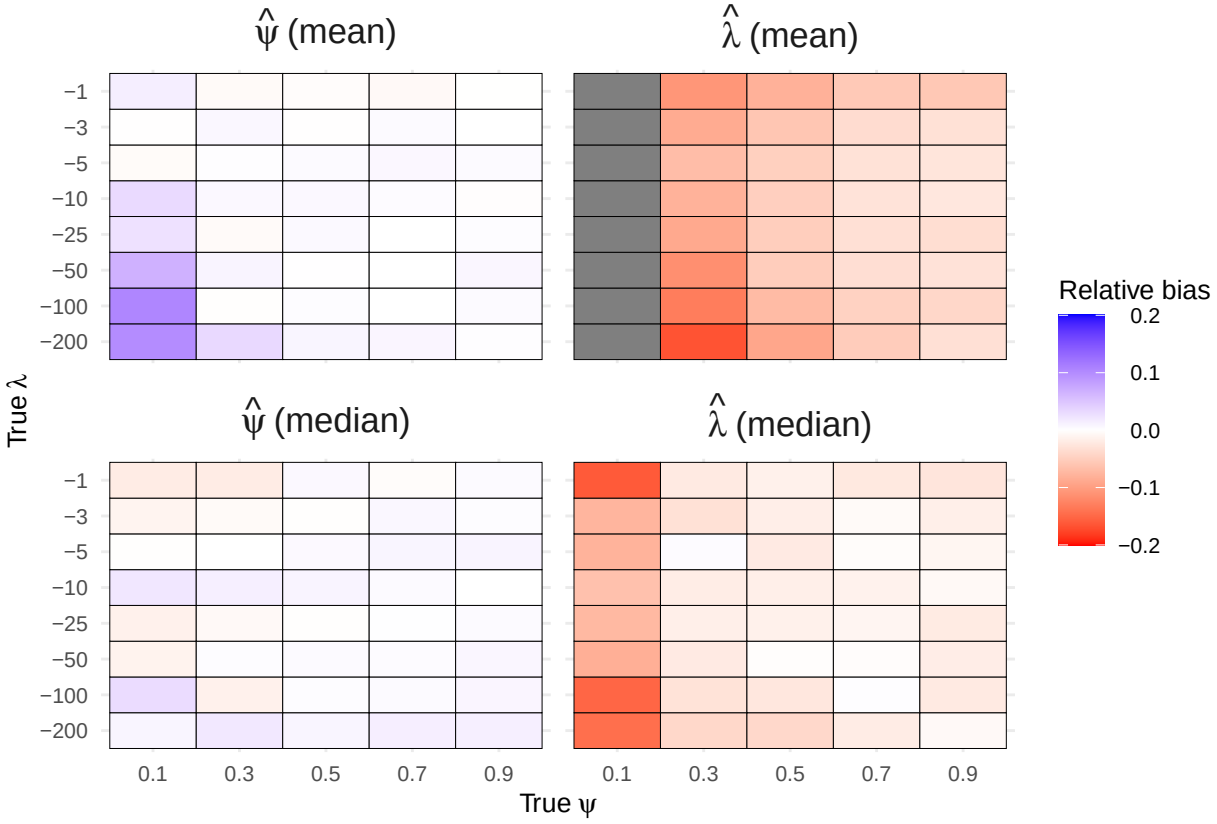

Figure 4: Relative bias of $\hat{\psi}$ (left) and $\hat{\lambda}$ (right) as inferred from simulations (1000 iterations for each pairwise combination of ψ and λ). Relative bias was calculated by substracting the true parameter value from either the mean (top) or the median (bottom) estimated value from each set of iterations and dividing the difference by the absolute value of the true parameter (see Section 2). Red shading indicates that the mean/median estimated parameter is less than the true value (i.e., $\hat{\psi}$ is too small, and the magnitude of $\hat{\lambda}$ is too large). Blue shading indicates the opposite. The gray column for $\hat{\lambda}$ when using the mean and when $\psi = 0.1$ indicates relative biases that are extremely negative due to estimating λ on those iterations where simulated *Paranthropus* abundances are zero across all sites (see Figure 6).

The relative bias of $\hat{\lambda}$, when calculated using the mean of all 1000 iterations, becomes more negative as true ψ decreases (Figure 4). This is again due to the increased number of simulated *Paranthropus* zero abundances across sites, which the model falsely interprets as unsampled presences (see previous paragraph). The gray column in Figure 4 when true $\psi = 0.1$ denotes $\hat{\lambda}$'s that are extremely negative (i.e., -395.8461744). This occurs when simulated *Paranthropus* abundances are zero across *all* sites, as might happen when true ψ is small and true λ is very negative. Zero abundances across sites means there is no information from which to distinguish true absences and presences, resulting in unstable, very negative $\hat{\lambda}$'s. Figure 6 shows a negative linear relationship between true λ and the number of unstable $\hat{\lambda}$'s (out of 1000 iterations for each unique true λ), which is to be expected if more negative true λ's lead to a higher probability of all sites having

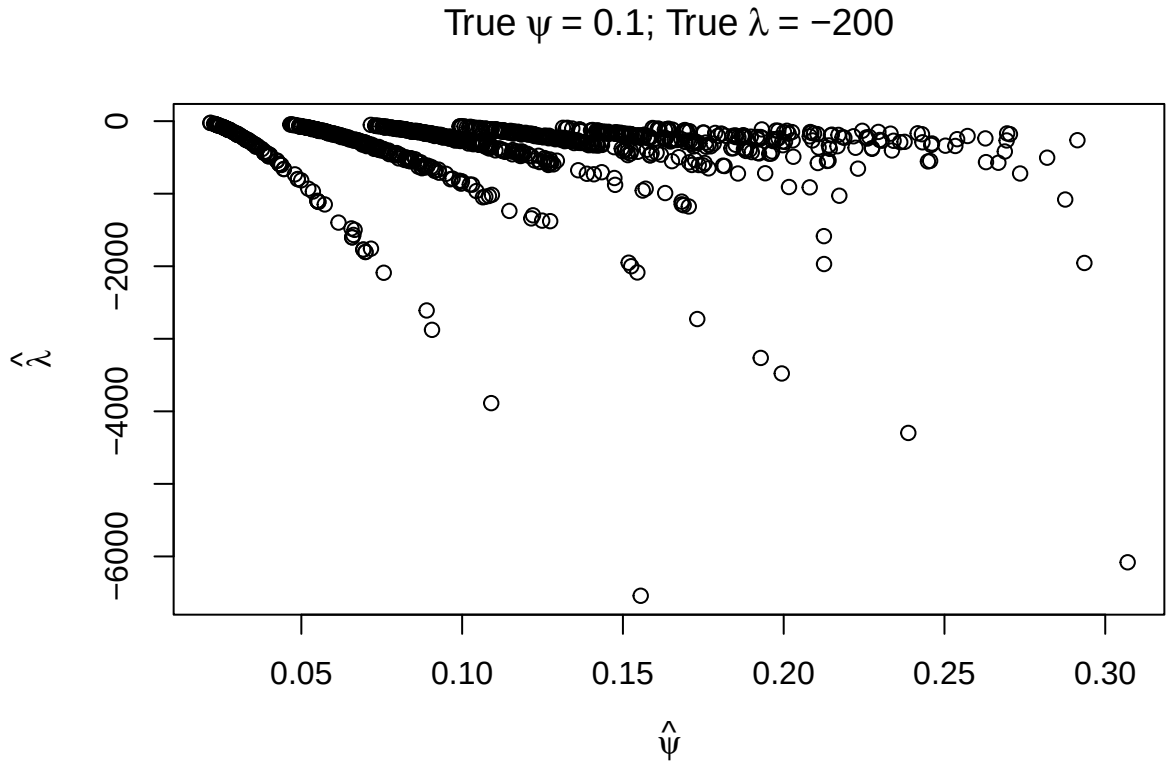

Figure 5: Scatterplot showing the negative relationship between $\hat{\lambda}$ as a function of $\hat{\psi}$ for simulations when true $\psi = 0.1$ and true $\lambda = -200$. Those iterations where $\hat{\lambda}$ are unstable are excluded (see Section 2) .

simulated *Paranthropus* abundances that are zero.

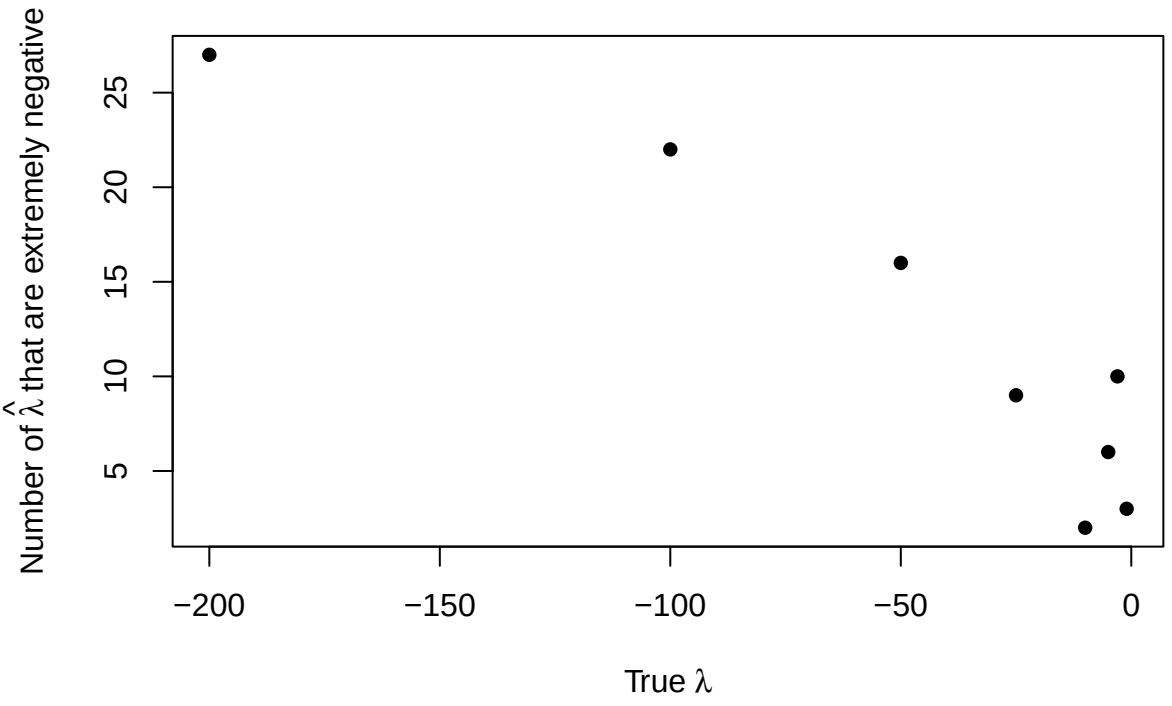

Figure 6: Scatterplot showing the negative linear relationship between the number of unstable, extremely negative $\hat{\lambda}$'s as a function of how negative true λ is. Unstable $\hat{\lambda}$'s occur only when simulated *Paranthropus* abundances are zero across all sites (see Section 2).

When using the median parameter estimate, instead of the mean, across all 1000 iterations for each pairwise combination of ψ and λ, the relative bias in $\hat{\psi}$ is generally small across all parameter values (range: -0.02 to 0.03) (Figure 4). The slight positive relative bias in $\hat{\psi}$ using the mean, and when true $\psi = 0.1$ and true λ is very negative, disappears because the sampling distribution of $\hat{\psi}$ is right skewed when the true value is close to its lower bound (i.e, zero); such a skewed distribution "drags" the mean upwards, but the median is less affected.

Relative bias of $\hat{\lambda}$ when using the median across all 1000 iterations is only notable when true $\psi = 0.1$, and this bias is still small overall (largest relative bias is -0.16) (Figure 4). This pattern can again be attributed to the large number of simulated *Paranthropus* abundances that are zero across all sites, which the model is mistaking for unsampled presences. The unstable, extremely negative $\hat{\lambda}$'s disappear (gray column in Figure 4) because this happened rarely across all 1000 iterations (Figure 6), such that $\hat{\lambda}$ is not affected by these few outliers when the median is used.

In sum, the model produces relatively unbiased parameter estimates except when there are a lot of zero *Paranthropus* abundances in the data, as to be expected (i.e., it is difficult to produce robust parameter estimates when there are not a lot of *Paranthropus* abundances to work with). The result is that the model is mistaking true absences as unsampled presences, causing $\hat{\lambda}$ to be too negative (Figure 4). However, this bias is still small in determining the overall shape of the beta distribution. For example, the largest relative bias when using the median (-0.16) occurs when $\psi = 0.1$ and $\lambda = -1$. This means median $\hat{\lambda}$ is too low by 0.16 (i.e., median $\hat{\lambda}$ = -1.16). This does not change the estimated shape of the beta distribution in any

meaningful way (cf. Figure 3). Either way, our main analysis gives $\hat{\psi} = 0.54$, where $\hat{\lambda}$ is expected to have a negligible bias (Figure 4).

3 Assessing model assumptions

Our model makes two assumptions regarding data independence:

1) Site data are independent from each other.
2) Within sites, specimen data are independent from each other.

3.1 Assumption #1: independence of site data

3.1.1 Regarding ψ

Regarding assumption #1 in estimating ψ, this would be violated if true *Paranthropus* presence/absence is spatiotemporally autocorrelated across sites. That is, sites that are closer together in space and/or time are more likely to exhibit the same state of *Paranthropus* presence/absence (e.g., closer sites all have *Paranthropus* present). To assess this assumption, we first take observed *Paranthropus* presence/absence data across sites and assume that they reflect true presence/absence. We know that this is strictly not true as an observed absence can reflect an unsampled presence (after all, this is the general problem we are trying to address), but it can also be reasonably assumed that observed presence/absence is more reflective of true presence/absence, rather than complete noise. We then consider all pairwise comparisons of sites in terms of (1) whether their observed *Paranthropus* presence/absence states match (1 = match, 0 = otherwise) and (2) their temporal and spatial distances from each other. We finally use multiple logistic regression to model (1) as a function of (2) with an interaction term between temporal and spatial distance (after centering and scaling the independent variables) to see if sites that are closer together in time and/or space are more likely to exhibit the same observed *Paranthropus* presence/absence state.

Table 1 shows that the logistic regression coefficient estimates for temporal distance and the interaction term are small, while it is larger for spatial distance. We will not discuss the interaction term given its negligible magnitude (indeed, coefficient estimates are virtually identical when no interaction term is included). Firstly, the coefficient estimates are expected to be negative: an increase in spatial and/or temporal distance between sites should lead to differences in *Paranthropus* presence/absence states (i.e., *Paranthropus* is present at one site but absent at the other); recall that we coded different states as 0s, while matches are 1s. We can exponentiate the logistic regression coefficients to make them more interpretable using the language of odds. Odds are defined, in our case, as the probability of *Paranthropus* presence/absence states matching divided by the probability that they do not match. Therefore, an odds of 2 means it is twice as likely that *Paranthropus* presence/absence states will match rather than not. When exponentiated, the intercept is 1.16, meaning that the probability of *Paranthropus* states matching is only 16% more likely than them not matching when the independent variables are set to zero (zero values in our case indicate the average pairwise temporal and spatial distance in our dataset because we centered the variables prior to analysis). The exponentiated coefficients for "scale(temporal distance)" and "scale(spatial distance)" are 1.09 and 0.9, respectively. Thus, when holding spatial distance constant, a one standard deviation increase in temporal distance results in a 9% *increase* in the odds of *Paranthropus* presence/absence states matching, opposite the direction we would expect. Holding temporal distance constant, a one standard deviation increase in spatial distance leads to a 10% decrease in the odds of *Paranthropus* presence/absence states matching. This translates to a decrease in odds from 1.16 to 1.05 when the variable "scale(spatial distance)" increases from 0 to 1 (while setting the other variables to zero). However, because we centered and scaled the data prior to analysis, this increase is quite substantial (Figure 7). In sum, there appears to be no temporal autocorrelation in *Paranthropus* presence/absence states, but there might be some slight spatial autocorrelation.

Even if the assumption of non-independence is violated and there is some spatial autocorrelation, our likelihood function is what is called a composite-likelihood, and the parameter estimates are consistent (i.e. they

will converge to the true values as the sample size approaches infinity) (Lindsay, 1988). The impact of such autocorrelation would be to increase the variance of the parameter estimates and decrease their efficiency (i.e., the speed at which the estimates converge to their true values). However, based on the analysis above, it seems that the spatial autocorrelation is small and this effect will be minimal.

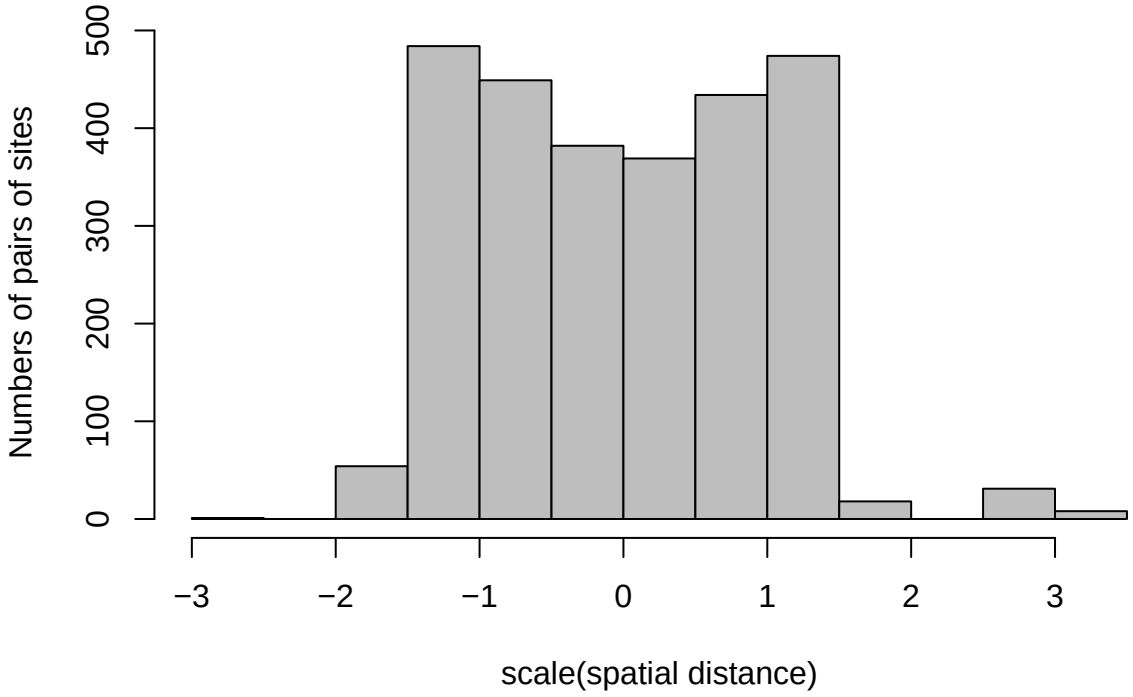

Figure 7: Histogram of the centered and scaled spatial distances between every pairwise combination of sites. Centering the data results in a mean of zero, while scaling transforms the data into standard deviation units.

3.1.2 Regarding λ

Regarding assumption #1 in estimating λ, this would be violated if true *Paranthropus* relative abundance is spatiotemporally autocorrelated across sites. That is, sites that are closer together in space and/or time are more likely to have similar *Paranthropus* relative abundance values. As with *Paranthropus* presence/absence before, we take observed *Paranthropus* relative abundances across sites and assume that they reflect true relative abundances. This assumption is safer than the previous one concerning presence/absence data, given that observed relative abundances track their true counterparts but with a positive bias that increases as a function of taxon rarity and sampling incompleteness (Chao et al., 2015). We then consider all pairwise comparisons of sites in terms of their (1) differences in observed *Paranthropus* relative abundance and (2) temporal and spatial distances from each other. As with the the presence/absence logistic regression, we model (1) as a function of (2) with an interaction between the centered and scaled independent variables, but this time using multiple ordinary least square regression. Again, the goal is to assess whether sites that are closer together in time and/or space are more similar in their observed *Paranthropus* relative abundances.

Table 1 shows that all regression coefficients are small, and the multiple R^2 of the entire model is only 0.02. Considering all pairwise comparisons between sites, Figure 8 plots observed *Paranthropus* relative

abundance difference as a function of temporal (Figure 8A) or spatial (Figure 8B) distance. We expect the relationships to be positive, where sites that are further away temporally or spatially are expected to have more dissimilar relative abundance values. The observed relationships are in fact *negative* (Figure 8), opposite the expected direction. Therefore, it appears that observed *Paranthropus* relative abundances across sites are not spatiotemporally autocorrelated.

Table 1: Coefficient estimates for the multiple logistic (presence-absence) and multiple ordinary least squares (relative abundance) regression models. "scale" indicates that these independent variables were centered and scaled prior to fitting the models.

	Presence-absence logistic regression	Relative abundance OLS
Intercept	0.149	0.007
scale(temporal distance)	0.086	-0.001
scale(spatial distance)	-0.101	-0.001
Interaction term	0.076	0.000

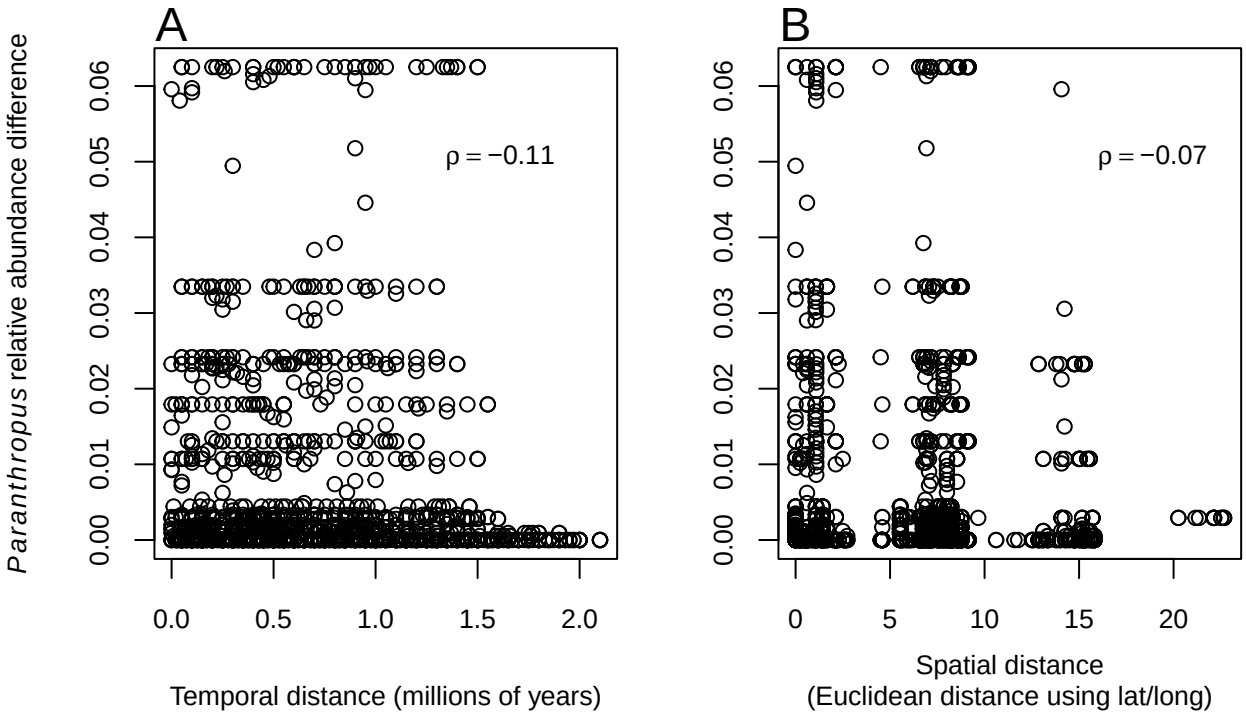

Figure 8: Scatter plots of observed *Paranthropus* relative abundance difference as a function of (A) temporal distance and (B) spatial distance for all pairwise comparisons of sites. ρ is the Spearman's rank correlation coefficient.

It should be noted that our models exhibit negligible multicollinearity: the Spearman's rank correlation between temporal and spatial distance between pairs of sites is only $\rho = 0.08$.

3.2 Assumption #2: independence of specimen data

Assumption #2 would be violated if multiple specimens belonged to the same individual (e.g., a partial skeleton), and this was the case for a large portion of specimens at each site. Non-independence would artificially inflate the number of independent data points for each site, which would bias posterior probabilities

of absence upwards. We assessed independence of specimen data for those sites where data on the number of specimens per individual are available (i.e., databases). Results show that the majority of individuals at each site are comprised of one specimen (Figure 9). In fact, for each site, the proportion of individuals made up of one specimen exceeds 0.8, except for Member G of the Shungura Formation (0.68) (Figure 10). Because all sites in our dataset are taphonomically similar in a general sense (e.g., eastern African large mammalian fossil assemblages that were preserved in fluvio-lacustrine settings and predominantly surface collected), we assume that the dominance of single-specimen individuals is a general pattern that can be extrapolated to those sites not represented in Figures 9 and 10. Thus, given the low proportion of individuals with non-independent specimen data at each site, we view Assumption #2 as not grossly violated, and the estimated posterior probabilities of absence are robust.

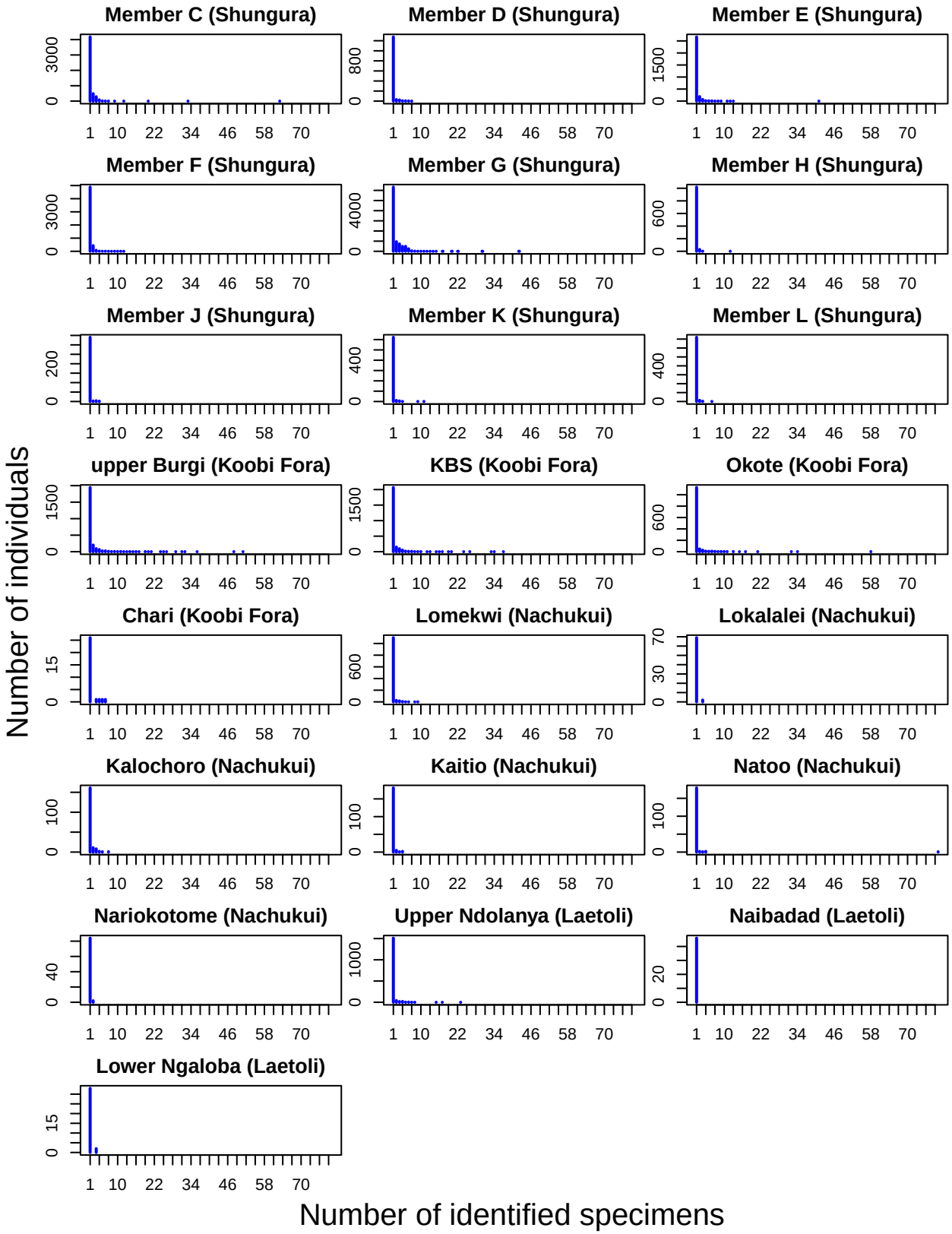

Figure 9: Frequency distributions depicting the number of individuals with a given number of specimens for each site, where those data are available (i.e., databases).

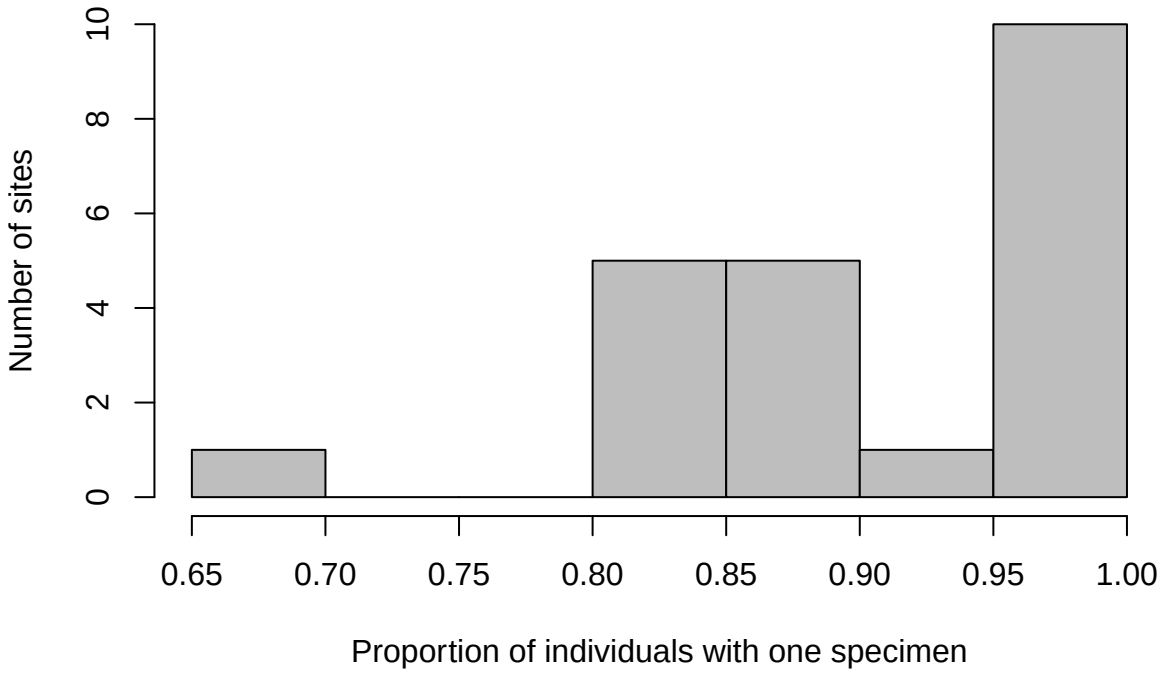

Figure 10: Histogram of the number of sites with a given proportion of individuals with one specimen. The proportion was calculated for each site as the number of individuals represented by one specimen divided by the total number of individuals.

References

Brown, J. H., Mehlman, D. W., & Stevens, G. C. (1995). Spatial variation in abundance. *Ecology*, *76*(7), 2028–2043. https://doi.org/10.2307/1941678

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208. https://doi.org/10.1137/0916069

Chao, A., Hsieh, T. C., Chazdon, R. L., Colwell, R. K., & Gotelli, N. J. (2015). Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology*, *96*(5), 1189–1201. https://doi.org/10.1890/14-0550.1

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*(1), 221–239.

Murray, B. R., Rice, B. L., Keith, D. A., Myerscough, P. J., Howell, J., Floyd, A. G., Mills, K., & Westoby, M. (1999). Species in the tail of rank-abundance curves. *Ecology*, *80*(6), 1806–1816. https://doi.org/10.1890/0012-9658(1999)080%5B1806:SITTOR%5D2.0.CO;2

R Core Team. (2021). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components* (Vol. 391). John Wiley & Sons.

Swihart, B., & Lindsey, J. (2020). *Rmutil: Utilities for nonlinear regression and repeated measurements models* [Manual].

Wang, S. C. (2010). Principles of statistical inference: Likelihood and the Bayesian paradigm. In J. Alroy & G. Hunt (Eds.), *Quantitative Methods in Paleobiology* (Vol. 16, pp. 1–18). Paleontological Society.

Wang, S. C., Everson, P. J., Zhou, H. J., Park, D., & Chudzicki, D. J. (2016). Adaptive credible intervals on stratigraphic ranges when recovery potential is unknown. *Paleobiology*, *42*(2), 240–256. https://doi.org/10.1017/pab.2015.37