Estimating the Ontogenetic Age and Sex Composition of Faunal Assemblages with Bayesian Multilevel Mixture Models

Jesse Langdon Wolfhagen[1,2] ⬤

Abstract

Understanding the ontogenetic age and sex composition of zooarchaeological assemblages can reveal details about past human hunting and herding strategies as well as past animal morphology and behavior. As such, the accuracy of our estimates underlies our ability to ascertain details about site formation and gain insights into how people interacted with different animals in the past. Unfortunately, our estimates typically rely on only a small number of bones, limiting our ability to fruitfully use these estimates to make meaningful comparisons to theoretical expectations or even between multiple assemblages. This paper describes a method to use zooarchaeological remains with standard biometric measurements to estimate the ontogenetic age and sex composition of the assemblage, focused on immature, adult-sized female, and adult-sized male specimens. The model uses a Bayesian framework to ensure that the parameter estimates are biologically meaningful. Simulated assemblages show that the model can accurately estimate the biometry and composition of zooarchaeological assemblages. Two archaeological case studies also show how the model can be applied to produce tangible insights. The first, focused on sheep from Neolithic Pınarbaşı B, highlights the model's ability to elucidate site formation and function. The second, focused on cattle remains from four assemblages from 7th-6th millennium BCE northwestern Anatolia, showcases how to use the mixture modeling results to compare assemblages to one another and to specific hypotheses. This modeling framework provides a new avenue for investigating long-term trajectories in animal biometry alongside contextual analyses of past human choices in butchery and consumption.

Keywords Zooarchaeology · Biometry · Logarithmic size index (LSI) · Domestication · Bayesian statistics

✉ Jesse Langdon Wolfhagen
    jwolfhag@purdue.edu; wolfhagen@shh.mpg.de

1   Department of Anthropology, Purdue University, West Lafayette, IN, USA

2   Department of Archaeology, Max Planck Institute for Geoanthropology, Jena, Germany

🖄 Springer

Introduction

Different hunting and herding strategies target specific classes of animals among a herd that is determined by the animal's ontogenetic age and sex (Dahl & Hjort, 1976; Stiner, 1990). In addition to human-driven goals, sex differences in habitat use, diet quality, and reproductive capabilities among ungulate prey species contribute to the susceptibility and desirability of males and females at different ages to human exploitation (Corti & Shackleton, 2002; Post et al., 2001; Ruckstuhl, 2007; Ruckstuhl & Neuhaus, 2002; Saïd et al., 2011). These factors impact the formation of bone assemblages by affecting the probabilities that bones from different classes of animals (*e.g.*, immature, adult female, or adult male animals) are deposited before being mediated by other taphonomic processes (Lyman, 2008). The ontogenetic age and sex composition of zooarchaeological assemblages can, therefore, reflect anthropologically-relevant aspects of past hunting strategies—like seasonal site use and scale of exploitation (Speth, 2013)—or general management goals of past herding strategies (*e.g.*, Payne, 1973; Redding, 1984).

Reconstructing the ontogenetic age and sex composition of a zooarchaeological assemblage can enrich our understanding of past human-animal interactions by complementing mortality profiles and inter-assemblage comparisons. However, the disaggregated nature of faunal assemblages complicates efforts to conclusively identify the ontogenetic age and sex of a specimen. Because articulated remains are rare, zooarchaeologists typically cannot relate elements that are morphologically distinct between the sexes (*e.g.*, the pelvis) to other elements that can provide information about the animal's age-at-death (*e.g*., limb bones or mandibles). We can, though, take advantage of the general pattern of sexual dimorphism among ungulate taxa by using size differences in limb bones to distinguish between males and females.

Morphometric Sex Determination in Zooarchaeology

Some biometric methods to determine the sex of an animal bone are multivariate—using combinations of measurements with bivariate plots or discriminant functions to predict the sex of archaeological specimens based on distributions of known-sex specimens (*e.g.*, Munro et al., 2011; Speth, 1983; Widga, 2006). These methods typically combine dimensions from different planes of an element (*e.g.*, the breadth and depth of a distal articular end) to produce patterns that can be separated by a "cut point" between males and females, either visually in the case of bivariate plots or algorithmically in the case of discriminant functions. The analytical requirement that multiple dimensions of a bone be preserved in measurable condition, even on the same end of an element, may make it difficult to apply these methods to more heavily processed assemblages. Furthermore, specimens from animals that died before reaching adult body size may be misclassified as females, particularly for dimensions affected by post-fusion growth (Popkin et al., 2012).

Other sex determination methods are univariate—they use a single measurement from a specimen and typically use size index methods to associate those

measurements from different elements together (*e.g.*, Weinstock, 2006; Zeder & Lemoine, 2020). This approach allows general descriptions of the sex ratio in an assemblage that can be used to identify changes in these sex ratios or overall biometry over time (*e.g.*, Arbuckle & Atici, 2013; Grigson, 1989). Zeder and Lemoine (2020) go further by using inter-quartile ranges of log size index (LSI) values from their reference population to create 'cut-off' values between immature, female, and male specimens to calculate specific ontogenetic age and sex ratios for elements and assemblages.

Regardless of whether the method uses multivariate or univariate data, these sex determination methods tend to have the same weaknesses. Practically, these methods rely on direct comparisons with reference populations (typically, but not always, modern populations of known sex). Thus, the analysis implies that the biometry of the archaeological population is the same as the reference population. However, this implication is an untenable one in most cases, as animal biometry typically varies spatially and temporally due to population-level intra-taxonomic genetic differences caused by adaptation to local climates and ecologies (*e.g.*, Davis, 1982; Hill et al., 2008; Koch, 1986; Lebenzon & Munro, 2022; Wright & Viner-Daniels, 2015). Biometric variation in wild and domesticated taxa has also been attributed to anthropogenic pressures as a result of herding decisions or hunting pressure (*e.g.*, Arbuckle & Kassebaum, 2021; Grau-Sologestoa & Albarella, 2019; Manning et al., 2015; Trentacoste et al., 2021), though harvest pressure has also been attributed to biometric changes in wild taxa (*e.g.*, Wolverton, 2008; Munro et al., 2022. These environmental and anthropogenic pressures may affect males and females differently (*e.g.*, Tchernov & Horwitz, 1991; Zohary et al., 1998); pressures that reduce sexual dimorphism could interfere with analyses, as more specimens may be indeterminate or misclassified. Biometric variation between populations can complicate efforts to estimate changes in the demographic (ontogenetic age and sex) composition of assemblages over time. Furthermore, efforts to control for ontogenetic age (*e.g.*, removing unfused specimens or those from early-fusing elements) distorts the relationship between the analyzed specimens and the rest of the assemblage, decreasing our ability to make reliable inferences about the entire assemblage (Zeder & Hesse, 2000).

Philosophically, sex determinations made by these methods tend to be absolutist; specimens are identified as male or female (or immature) or are marked as indeterminate. As in taxonomic identifications, the use of absolutist determinations masks any underlying uncertainty in the determination (Wolfhagen & Price, 2017). Removing indeterminate specimens from consideration artificially reduces sample sizes and inflates reported accuracy rates. This produces a false sense of confidence in the sex determination results, especially when those results are then used to characterize the entire assemblage. More critically, any nuances or caveats in the sex determinations of an assemblage are lost when the results are used in synthetic analyses at larger spatial and temporal scales. What is necessary is a way to estimate the ontogenetic age and sex composition of a faunal assemblage that preserves the uncertainty inherent in the process.

Mixture Modeling in Zooarchaeology

Mixture modeling provides just such a method, producing probabilistic sex identifications rather than absolute ones by describing an assemblage of faunal measurements as a mixture of specimens from different animal groups (generally termed "mixture components") like male and female specimens—described by parameters for the proportion of the overall assemblage (π), average size (μ), and size variability (σ) of each animal group. A mixture model allows researchers to not only describe the overall composition of the assemblage but to also estimate the probabilities that a specific specimen belongs to a particular animal group (Dong, 1997; Monchot & Léchelle, 2002). Additionally, mixture modeling does not rely on a reference population, allowing biometric variation between populations and even changes in the extent of sexual dimorphism (*e.g.*, Helmer et al., 2005). These features allow mixture models the flexibility to track both biometric and demographic variation across assemblages over time and space (*e.g.*, Arbuckle et al., 2016; Arbuckle & Kassebaum, 2021).

Conceptually, a mixture model can be thought of as a "latent state" or "missing data" problem; we know that measured specimens come from particular animal groups, but that information has been lost (Marin et al., 2005). If we knew every specimen's group identity, then the calculation of the mixture model parameters (mixture proportion, average body size, and size variability for each animal group) would be trivial. In archaeological contexts, however, we cannot directly observe those group identities; we must, therefore, use probabilities of group membership and calculate group-specific parameters from those resulting probabilities (Monchot & Léchelle, 2002). Figure 1 describes a schematic example of a mixture model; Fig. 1A shows the distribution of LSI values from a reference population of 31 adult pig (*Sus domesticus*) tibia distal breadths (Tibia Bd: von den Driesch, 1976) described in Zeder and Lemoine (2020), with specimens colored by their known identity (females = blue, males = red). Figure 1B shows the results of fitting a two-component (females and males) mixture model to the data using standard approaches (*e.g.*, Arbuckle & Kassebaum, 2021; Monchot & Léchelle, 2002), ignoring those true identities.

The mixture model describes the assemblage as a mixture of the two "mixture components" (males and females); each component is described with three parameters: a proportion (π), an average size (μ), and a standard deviation (σ). Taken together, these parameters determine a specimen's probability of being in one of the groups, as shown in Fig. 1B; a specimen with an LSI_e value of 0.06 has a 31% probability of being female based on the model. This results in 31 sets of probabilities, one for each specimen; Fig. 1C shows four plausible simulated assemblages that result from the mixture model; every specimen's membership probability is used to simulate a "true" identity. Importantly, by leaving the mixture model results as specimen-specific probabilities of being female or male, mixture model results retain the uncertainty of the sex determination process; that is, a specimen with a 51% probability of being female is not treated as equivalent to one with a 95% probability of being female.
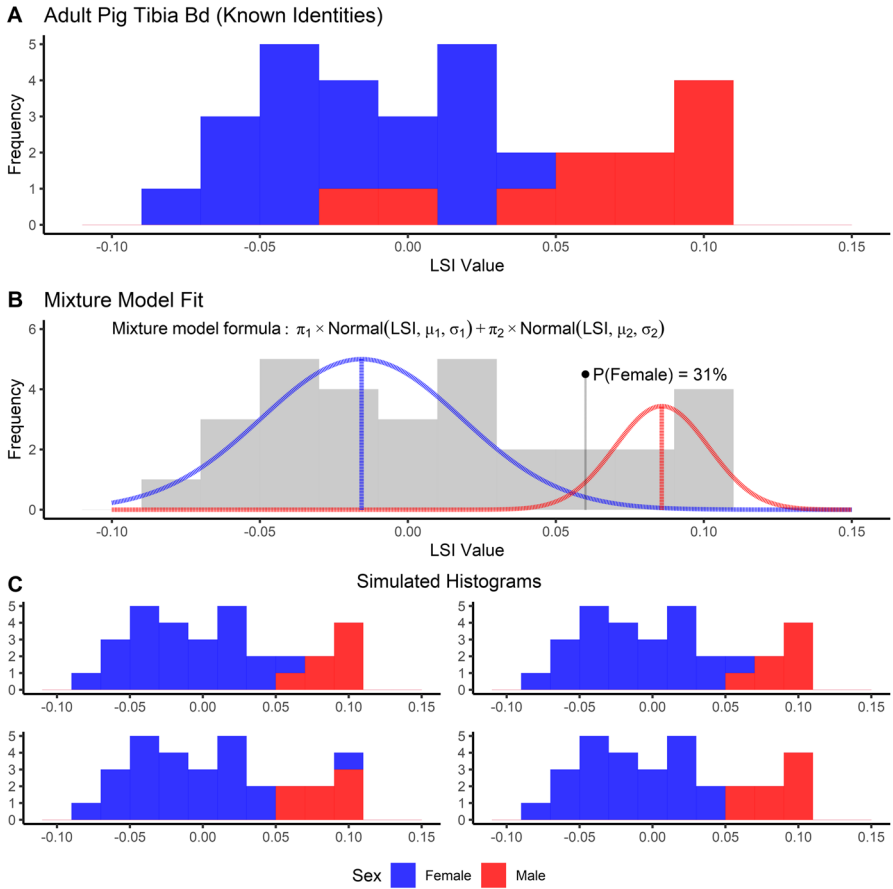
Fig. 1 Walkthrough of the mixture modeling procedure using pig Tibia Bd measurements from Zeder and Lemoine (2020). **A** The distribution of Tibia Bd LSI$_e$ values (standard value: 33.5 mm; Hongo & Meadow, 2000) of adult specimens, shaded by known sex (females = blue, males = red). **B** The result of a standard (maximum-likelihood estimation) mixture model analysis on the LSI$_e$ values, ignoring sex. Vertical lines show the estimated means for females and males; curves show the relative probability densities for the two distributions. The point shows the probability that a specimen with a particular LSI$_e$ value (0.06) would be considered female by the model results. **C** Four simulations using the membership probabilities of each specimen based on the mixture model results

The previous example showcases the benefits of mixture modeling as a flexible probabilistic sex determination method. First, the model does not require comparison to a reference population to estimate differences between female and male specimens. Second, the model produces parametric estimates of body size and size variability that can be used for inter-site comparisons, rather than just determinations for the included specimens. Finally, the model produces probabilistic estimates for every specimen, rather than leaving some specimens indeterminate and obscuring variation in the confidence of the sex assignments. These theoretical and practical advantages of mixture modeling and its potential for zooarchaeology have been

apparent since its introduction to the field (*e.g.*, Dong, 1997; Monchot et al., 2005; Monchot & Léchelle, 2002), though its application has been piecemeal over the past two decades despite the existence of free scientific software that can perform the analysis (*e.g.*, PAST: Hammer, 2022; R packages "mixtools" and "mclust": Benaglia et al., 2009; Scrucca et al., 2016).

The reasons for the patchy application of mixture modeling in zooarchaeology are less straightforward. High-profile early case studies of mixture modeling report size variability parameters (standard deviation σ) that vary widely and include very small values for some groups (*e.g.*, De Cupere et al., 2005; Monchot et al., 2005; Vigne, 2011). These results suggest that the very flexibility, that is, a great strength of mixture modeling, is actually identifying "groups" that are not necessarily consistent with biological expectations (*e.g.*, that the results are "overfitted" to the observed data). Such extreme differences in the standard deviation of different groups can result in counterintuitive implications; specimens may be considered more likely to come from the broad distribution (the one with the larger σ parameter) than the narrow distribution even when the value is more extreme than the narrow distribution's mean (*e.g.*, is larger than a larger mean or smaller than a smaller mean). Returning to the mixture model example in Fig. 1 can explain this issue more clearly. Table 1 shows the mixture model parameters for the two components; the standard deviation (σ) for females is more than twice the standard deviation for males ($\sigma_1 = 0.034$, $\sigma_2 = 0.016$). As such, higher numbers beyond the observed range will be considered likely females; an LSI_e value of 0.173 (Tibia Bd value: 39.83 mm) is more likely to be a female than a male using the mixture model's results (probability of being female: 51%).

Published mixture model examples show this issue, as well. De Cupere et al. (2005), Table 2) report three groups of chicken carpometacarpus lengths from bones with medullary bone, providing the full set of mixture model parameters (group 1: proportion $\pi = 0.285$, mean $\mu = 33.337$, standard deviation $\sigma = 0.3$; group 2: $\pi = 0.608$, $\mu = 35.416$, $\sigma = 0.433$; group 3: $\pi = 0.107$, $\mu = 37.866$, $\sigma = 0.094$). According to De Cupere et al. (2005), fig. 3), there is one carpometacarpus with a medullary bone whose greatest length is roughly 41.5 mm. Counterintuitively, the analysis would suggest that this specimen is most likely to be a member of group 2; it even determines that the specimen is more likely to be a member of group 1 than group 3. Vigne (2011), Table 3) reports mixture modeling results of LSI_{10} values from cattle recovered from Neolithic Shillourokambos, Cyprus, to estimate females and males, using PAST (Hammer, 2022). The reported values for the Recentes phase (female $\pi = 0.75$, $\mu = 0.120$, $\sigma = 0.042$; male $\pi = 0.25$, $\mu = 0.163$, $\sigma = 0.007$) produce

Table 1 Mixture model parameter estimates for the pig distal tibia Bd example

Group	Mixture component π	Average value μ	Size variability σ
Female	0.76	−0.016	0.034
Male	0.24	0.086	0.016

Estimates calculated using maximum-likelihood approaches

Table 2 Definitions of key terms used in this paper

Term	Definition
Element portion	A complete or partial skeletal element defined by the zooarchaeologist is used as the foundation of the multilevel model (*e.g*., distal humerus). The model produces parameter estimates for all defined element portions, so element portions must be non-overlapping. Analogous to the skeletal part type in Breslawski (2023).
Dimension	Specific type of observed measurement (*e.g*., humerus distal breadth) on a specimen. Dimension definitions typically follow von den Driesch (1976).
Measured assemblage	Assemblage of measured specimens from a defined number of element portions of a specific taxon.
Modeled assemblage	Assemblage of specimens from a defined number of element portions of a specific taxon. Includes measured and non-measured specimens, though all element portions must have some number of measured specimens. Measurability is assumed to be effectively random (*i.e*., unrelated to whether the specimen came from an immature, female, or male individual).
Full assemblage	Assemblage of specimens from a defined number of element portions of a specific taxon. Includes element portions that do not have any observed measurements. Measurability is assumed to be effectively random (*i.e*., unrelated to whether a specimen came from an immature, female, or male individual).

counterintuitive results; a specimen with an LSI value of 0.176—within the range of LSI values from this phase (Vigne, 2011: fig. 2)—would be considered more likely to be female than male. These issues extend to more recent publications. Arbuckle et al. (2016), fig. 5) report sex-specific LSI_{10} average sizes for cattle in the eastern fertile crescent during the early-mid Holocene; because they report their LSI_{10} data in a supplement, it can be shown that the smallest measurement from Ganj Dareh ($LSI_{10} = -0.044$, modeled female mean $= -0.019$, modeled male mean $= 0.024$) is considered more likely to be male than female due to the extreme differences in standard deviations.

These examples highlight the difficulties researchers face when interpreting the results of mixture analyses of zooarchaeological data. While mixture modeling provides the flexibility to model data from a pre-specified or unknown number of groups, there is no guarantee that the identified "groups" are biologically meaningful. Analysts may identify inconsistent results from mixture analyses and exclude the analysis from reports, leaving only mixture analyses that appear to have interpretable results (the "file drawer problem": Rosenthal, 1979). As these examples show, however, mixture analyses applied to more abstract quantities, like LSI values, or interpreted in light of less easily interpreted biological groups, like breeds, can have counterintuitive implications. These examples are not meant to highlight the errors; on the contrary, the fact that the authors report their full model results and/or data mean that such errors could be identified, highlighting the importance of open scientific reporting and publishing (Marwick, 2017; Ram & Marwick, 2018).

Zooarchaeologists have a wealth of reference information that can inform them about the impacts of diet, sex, castration, and other factors on the size and variability of animal bones. These reference populations provide raw measurements from several taxa and generally include specimens of known age-at-death and sex, though

Table 3 Software packages used in the analytical script for this paper

Use	Package	Version	Citation
Data aggregation, analysis, and multi-core processing	boot	1.3.28	Canty & Ripley, 2021; Davison & Hinkley, 1997
	data.table	1.14.2	Dowle and Srinivasan, 2021
	readxl	1.4.1	Wickham and Bryan, 2022
	parallel	4.1.3	Microsoft Corporation & Weston, 2022
	doParallel	1.0.17	Microsoft Corporation & Weston, 2022
Creation of RMarkdown files	kableExtra	1.3.4	Zhu, 2021
	knitr	1.39	Xie, 2015, 2022
Bayesian analysis and summarization	cmdstanr	0.5.0	Gabry & Češnovar, 2022
	rstan	2.21.3	Stan Development Team, 2021
Mixture model analysis and standard animal measurements	mixtools	1.2.0	Benaglia et al., 2009
	zoolog	0.4.1	Pozo et al., 2022
Visualization	Cairo	1.5.15	Urbanek and Horner, 2022
	ggplot2	3.3.5	Wickham, 2016
	ggdist	3.2.0	Kay, 2022
	ggpubr	0.4.0	Kassambara, 2020
	ggrepel	0.9.1	Slowikowski, 2021
	rnaturalearth	0.1.0	South, 2017a
	rnaturalearthdata	0.1.0	South, 2017b
	sf	1.0.8	Pebesma, 2018

A   "Rare Extremes" θ Prior, Normal(0, 1.25)

Percentage Extreme (> 90% or < 10%): 8%

B   "Common Extremes" θ Prior, Normal(0, 7)

Percentage Extreme (> 90% or < 10%): 75%

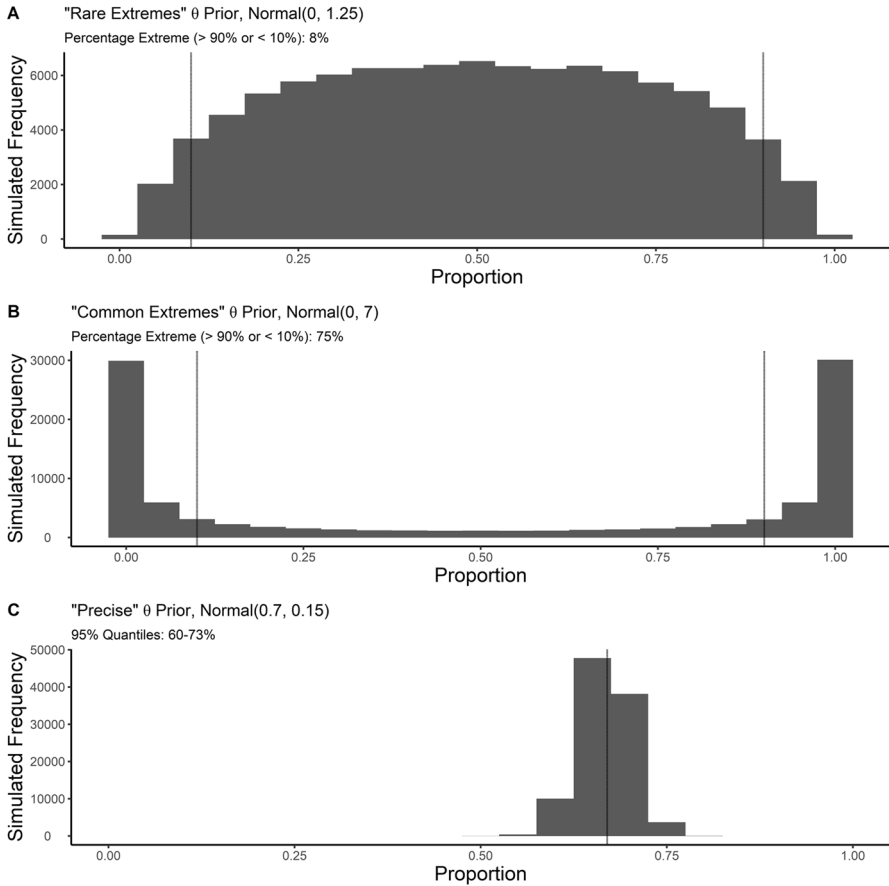**C**   "Precise" θ Prior, Normal(0.7, 0.15)

95% Quantiles: 60-73%

Fig. 2 Examples of prior distributions for demographic parameters (proportion of immature animals or adult sex ratio) with different expectations. **A** A distribution where extreme values are considered unlikely but otherwise most values are about equally as likely. **B** A distribution where extremely high or extremely low values are likely but indeterminate values are much less likely. **C** A distribution where it is extremely likely that values are centered around 67%. The last scenario would only be appropriate if there is sufficient prior knowledge about the context.

sometimes these include archaeological data of (relatively) complete individuals that can be assigned to sex (*e.g.*, sheep: Popkin et al., 2012; Davis, 1996, 2000; pigs: Zeder & Lemoine, 2020; Payne & Bull, 1988; aurochsen/cattle: Degerbøl, 1970; bison: Speth, 1983; Todd, 1983). These data can provide useful information that could be relevant for interpreting a mixture model analysis; ideally, one could take advantage of relevant information from reference populations while still maintaining some aspects of a mixture model's flexibility. Unfortunately, standard mixture modeling algorithms do not provide a straightforward way to ensure that the model parameters (μ and σ) for the groups accord with our understanding of these parameters from reference populations. Bayesian inference, however, does provide a way to do this very thing by using data from reference populations to create prior distributions

for mixture model parameters. Using prior distributions improves overall model performance because the analyst can use these sources of 'prior knowledge' to inform them about the data that they have on-hand (Otárola-Castillo et al., 2022).

This paper describes a Bayesian approach to the mixture model analysis of faunal measurements that addresses these weaknesses of mixture modeling as currently applied. The model uses informative priors derived from a "prior assemblage" of known age-at-death and sex individuals to constrain population parameter estimates to be biologically interpretable (Popkin et al., 2012). It also uses multilevel modeling to take advantage of partial pooling and address aggregation issues to directly estimate parameters for each measured dimension in the analysis (Gelman, 2006a; Wolfhagen, 2020). In addition to modeling females and males, the model includes a third group consisting of "immature" specimens that died before reaching adult body size. The model also emphasizes inference of the entire assemblage rather than just the measured specimens by incorporating observations of the sex ratio (from morphological data) and the proportion of immature specimens (from fusion data) to inform population parameters of the proportions of these different groups. The model is used on sixteen simulated assemblages derived from the (Popkin et al., 2012) Shetland sheep (*Ovis aries*) population to test its ability to accurately estimate the age and sex composition of assemblages. Two archaeological case studies then show the applicability of the model to archaeological assemblages for reconstructing the age and sex composition of assemblages and to highlight the importance of incorporating immature specimens into mixture modeling analyses.

A Bayesian Multilevel Mixture Model for Zooarchaeological Measurements

The Bayesian model developed for this paper improves on standard mixture modeling for zooarchaeological measurements in four distinct ways. First, it addresses complications caused by measurements from unfused specimens and post-fusion growth by modeling three groups within the mixture: immature animals, (adult-sized) females, and (adult-sized) males, each with distinct size parameters. Second, the multilevel structure allows the model to balance bias due to aggregation and overfitting from small sample sizes. Third, the Bayesian foundation of the model provides an avenue for synthesizing information about the ontogenetic age and sex composition of the assemblage from non-metrical data (*e.g*., fusion rates, sex ratios based on morphological data) to inform the results of the mixture model. Finally, researchers can create prior distributions for mixture model parameters from prior assemblages or other sources, ensuring that the mixture model results are biologically interpretable. This section outlines these benefits; specific details of the model are described in a Model Supplement and in the analytical code (available at the project's GitHub page: https://github.com/wolfhagenj/ZooarchMixMod).

Observed measurements from different dimensions (*e.g*., humerus distal breadth "Humerus Bd," radius proximal breadth "Radius Bp," abbreviations following von den Driesch 1976) are first converted to logarithmic size index (LSI) values using a natural logarithm base to take advantage of the normalization LSI standardization

provides (Meadow, 1999; Wolfhagen, 2020). The model uses a single LSI value per specimen, so any specimens with multiple observed dimensions are first summarized by estimating the mean LSI value from the observed dimensions; these specimen-level LSI values are the basis for the mixture model. These specimen-level LSI values can be clustered into different element portions—partial or complete elements that are the basic categorical unit of an analysis (*e.g.*, "distal humerus" or "first phalanx"; compare to "skeletal part type" in Breslawski 2023). Table 2 provides a glossary of the key terms used in this text.

Benefits of the Bayesian Multilevel Mixture Model

Body size is affected by both ontogenetic age and sex; animals killed before reaching adult body size pose a complication for most sex determination models, which exclusively focus on distinguishing between (adult-size) female and male animals (but see Zeder & Lemoine, 2020). Measurements from known age-at-death Shetland sheep show that specimens killed younger than 1 year of age are significantly smaller than those killed at older ages, regardless of fusion status and sex; after 1 year of age, size is no longer significantly impacted by age (Popkin et al., 2012). Thus, any measurement from an unfused epiphysis or from an element portion that does not fuse or exhibits significant post-fusion growth should be considered potentially immature and needs to be modeled with a three-member mixture model (group 1 = immature, group 2 = adult female, and group 3 = adult males). On the other hand, the model excludes the possibility that measurements from specimens that are conclusively not immature due to their fusion status could be from the immature group ($\pi_1 = 0$), effectively fitting a two-member mixture model (adult females and adult males).

Typically, biometric analyses aggregate LSI values from different element portions (*e.g.*, Arbuckle & Kassebaum, 2021; Sasson & Arter, 2020; Vigne, 2011); aggregation produces bias because it assumes that every element portion has the same parameter value (Wolfhagen, 2020). Multilevel modeling uses partial pooling to allow the cluster-specific parameters to vary between clusters while reducing overfitting caused by small sample sizes (Fernée & Trimmis, 2021; McElreath, 2020). In the case of this mixture model, element portions are the relevant clusters—the multilevel model produces a set of mixture model parameters for each element portion (a set of π, μ, and σ parameters for each of the three animal groups). These cluster-specific parameters are related to each other through "hyper-parameters" that describe the average value of the mixture model parameters and the variability of model parameters across element portions (Wolfhagen, 2020). This structure reduces overfitting caused by small sample sizes among some clusters while also avoiding the bias caused by aggregating all clusters together.

Of course, biometric data are not the only source of information on an assemblage's ontogenetic age and sex composition. Fusion rates of elements that fuse around the age that animals reach adult body size can provide relevant information on the proportion of immature specimens in the assemblage (*e.g.*, first and second phalanges in sheep: Popkin et al., 2012), just as sex ratios derived from

morphologically distinct adult elements provide information about the adult sex ratio in an assemblage (*e.g.*, fully fused pelvises: Stiner et al., 2022; horn cores: Twiss & Russell, 2009). These estimates of assemblage composition do not supersede those produced by a mixture model, but they are also not irrelevant to the composition from a mixture model. Unlike other sex determination methods, the multilevel structure of the Bayesian multilevel mixture model allows the analyst to inform their model results with relevant fusion and morphological sex data from the assemblage. These data do not determine the proportion of immature animals and the adult sex ratio of the mixture model, but they do help the model make more precise estimates of the ontogenetic age and sex composition of the assemblage than possible with the measurement data alone.

Relevant information from a prior distribution can inform an analyst about reasonable values for the model's hyper-parameters, which can be summarized as prior distributions. Creating informed prior distributions allows the model to ensure that the hyper-parameters have biologically interpretable results (*e.g.*, average size and size variability parameters for a measurement that align with reasonable expectations for a taxon). The multilevel structure of the model then ensures that mixture model parameters can vary between different element portions while still being informed by these hyper-parameters to maintain biological interpretability, even with small numbers of observations. Prior distributions draw explicit links between our sources of prior knowledge (*e.g.*, reference populations, ethnographic data, ecological data) and our archaeological data. Unlike absolutist models, we can define the prior distributions used in the Bayesian multilevel mixture model to be less specifically focused on the parameter values of the prior assemblage. Increasing the uncertainty of assemblage-derived prior distributions allows the mixture model to adjust to biometric differences between the prior assemblage and the assemblage being fit by the model. Care must still be taken to ensure that prior distribution definitions are at appropriate scales for the observations and not so broad as to include values that are known to be physically impossible (*e.g.*, Gabry et al., 2019).

Developing Prior Distributions from a Prior Assemblage

Prior distributions are central to Bayesian inference and describe one's prior beliefs in the potential values of a model parameter. Prior distributions can be likened to a "filter" from which parameter values are drawn to evaluate their fit with the data (Smith & Gelfand, 1992). Several approaches exist for deciding how to describe this prior belief, ranging from "objective" priors that provide equal weight to all possible values of a parameter to distinct distributions defined by a synthesis of previous or related research (Gelman, 2006b). Objective priors poorly reflect our intuition about phenomena we are modeling, waste computing effort by sampling parameter values that poorly fit the data, and can introduce errors into our analyses (Gabry et al., 2019); instead, "weakly informative priors" or "reference priors" use transformations of parameter values—like centering and scaling element portion-specific parameters—to describe variation in parameter values within reasonable values, with small deviations being more likely than large deviations (Gelman et al., 2008).

Informative priors are derived from relevant knowledge, be it the results of earlier studies on the same subject, the quantification of expert opinion, or parameter values for related subjects (McCarthy & Masters, 2005; Otárola-Castillo et al., 2022). Regardless of the distribution's source, it is important to evaluate how well the distribution reflects your prior knowledge about the system under study because the prior distributions influence the results of the analysis.

The mixture proportions summarize the composition of the assemblage and mediate the relative likelihoods of the different animal groups, adjusting a specimen's membership probabilities. Prior distributions for the mixture components reflect our prior beliefs about the relative proportions of immature, adult-sized females, and adult-sized males in the assemblage. Instead of estimating the prior belief for each of these three related categories, the model uses two prior distributions to estimate independent variables: the proportion of immature animals (π_1) and the adult sex ratio estimated through the relative proportion of adult females ($\frac{\pi_2}{\pi_2+\pi_3}$) (see Model Supplement for more details). The following examples show some of the flexibility researchers have when describing their prior belief about the proportion of immature animals or the adult sex ratio in an assemblage; the Model Supplement shows the mathematical details necessary to create relevant prior distributions for a model.

Figure 2 shows three examples of prior distributions for one of these mixture proportion concepts (proportion of immature or adult sex ratio) that reflect different expectations based on prior knowledge. Note that the Bayesian mixture model uses the observations of fusion rates and morphological sex ratios as observed data, so these prior distributions reflect knowledge prior to even those observations. Figure 2A shows a relatively broad, or uncertain, prior distribution where a researcher does not believe that the proportion is extreme (*i.e*., that the assemblage is either dominated by or bereft of immature animals or that the sex ratio is dominated by females or males) but has little opinion otherwise. Figure 2B shows a somewhat inverse situation, where the researcher is confident that the proportion (either the proportion of immature animals or the adult sex ratio) is at either one extreme or the other but is not sure which extreme it is. Figure 2C displays a scenario where the researcher is confident that the proportion is centered around 67% before looking at the faunal data, presumably based on prior research or other contextual information.

Prior distributions for parameters governing average body size (μ) and size variability (σ) are the keys to ensuring that the mixture model produces biologically feasible and interpretable results. These prior distributions are based on analysis of a "prior assemblage" created by sampling immature, adult female, and adult male/castrate Shetland sheep from the Popkin et al., (2012) population (150 specimens for each animal group; see Model Supplement for more details). Castrates were considered males for the purposes of the model, as proximal and distal bone width measurements as a whole did not vary significantly between intact males and castrates (Popkin et al., 2012: 1783–1784). Modeling the average body size and size variability of these animal groups using a multilevel model created the starting point for prior distributions that could be used in the model. To generalize the prior distributions so that they are applicable to a variety of zooarchaeological scenarios, the

results from the prior assemblage were given larger standard deviations to increase the uncertainty, which allows the model to better fit the data (see Fig. 3 and Model Supplement for more details). The prior distributions act to prevent the model from accepting parameter values that are implausibly large or small given our prior knowledge about size variability and size differences between animal groups.

Prior predictive checks, shown in the Model Supplement, show the implications of the prior distribution definitions used in the model. The results show that the chosen prior distributions do not exclude the possibilities of extreme values in mixture components and cover a wide range of potential size measurements. On the other hand, these model definitions are not so broad as to include much prior weight on biologically impossible values (*i.e.*, impossibly large measurements) that would
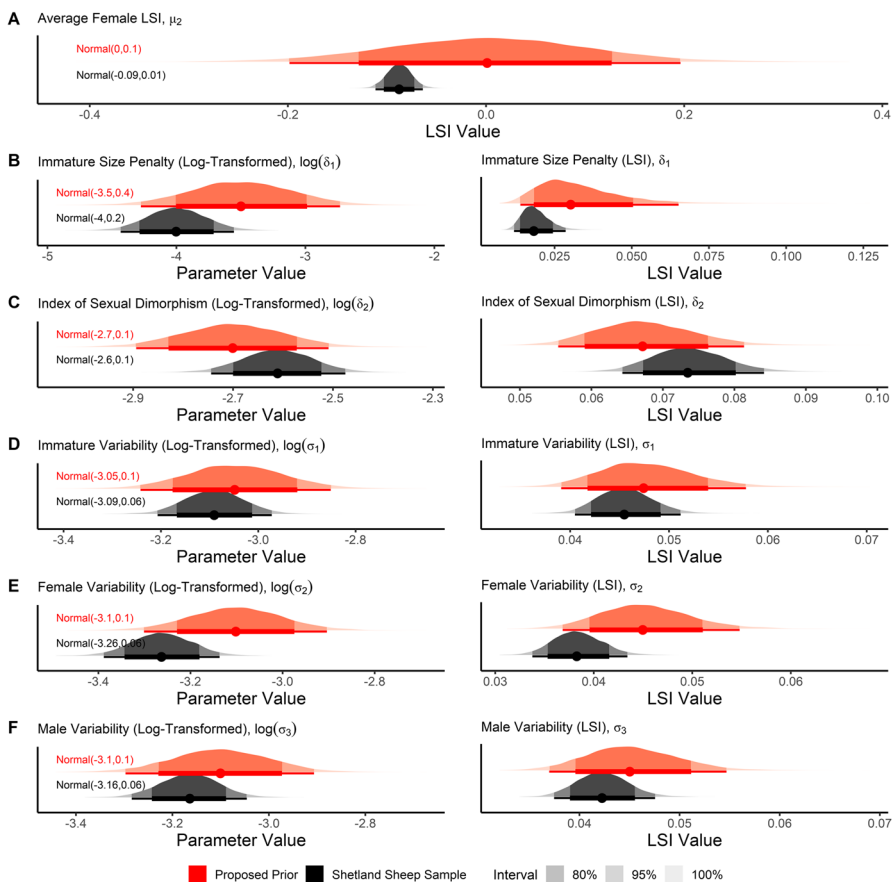
Fig. 3 Posterior distributions of model hyper-parameters from a sample of known-identity sheep specimens (red) and proposed prior distributions for mixture model applications (black). **A** Average female size (in LSI_e scale). **B** Immature size penalty (average size difference between females and immature animals). **C** Index of sexual dimorphism (average size difference between males and females). **D**: Size variability for immature animals. **E** Size variability for female animals. **F** Size variability for male animals. Proposed prior distributions provide a useful baseline in the absence of relevant biometric information regarding sexual dimorphism and size variability

slow down how quickly the model runs because it must evaluate the fit of extremely poorly fitting data. Such behavior could also produce biologically implausible results of actual model fits when only a small amount of data are available.

This section describes one approach for defining prior distributions of mixture models that are relevant for a wide range of zooarchaeological cases, particularly if researchers do not have strong preconceptions about the relevant parameters from prior research. It is important to remember that a model's prior distributions are choices made by the researcher to fit particular research questions, regardless of whether the distributions are informed by advice on reference priors, prior assemblages, or mathematical summaries of existing research. Zooarchaeological assemblages resulting from catastrophic kills would be expected to have a different ontogenetic age and sex composition compared to assemblages derived from sustained hunting or herding take-off (*e.g.*, Lyman, 1987). Similarly, other research contexts may provide an analyst with different prior expectations about animal size variability and overall biometry. Other reference populations, particularly those from other taxa, could also be used to create prior assemblages and help determine the limits on biological feasibility. Researchers could and should adapt their prior distributions to best reflect their intuition about likely parameter values for their research context. Regardless of the prior distributions a researcher uses, it is crucial to formally describe the prior distributions that are used in Bayesian analysis to ensure replicability. Furthermore, researchers should examine the implications of the different candidate prior distributions while developing a Bayesian model to test a research question; prior distributions should be regularly tested even before models are fit to datasets (Gelman et al., 2020).

Extending the Multilevel Analysis to Multiple Sites

The multilevel structure of the model that allows parameters to vary across element portions can also be used to extend the modeling approach to examine multiple assemblages at once. Combining multiple assemblages into a single model allows researchers to investigate regional variation in herd management strategies or outline diachronic trends in body size that may relate to population turnover (*e.g.*, Arbuckle et al., 2016; Arbuckle & Atici, 2013). It also allows researchers to model diachronic changes over the course of a multi-period site's occupation, as each occupation layer can be defined as a separate assemblage (*e.g.*, Hongo et al., 2009; Wolfhagen et al., 2021). By including the assemblages in the same model, estimates share the same hyper-parameters, which improves the precision of these estimates and allows researchers to directly compare assemblage-specific parameters by using contrasts. Furthermore, adopting this structure provides the foundation for more sophisticated analyses that test specific hypotheses about variations in biometric or compositional parameters, such as spatiotemporal autocorrelation in body size.

An important consequence of extending the model to evaluate multiple sites at once is that the interpretation of the overall hyper-parameters that the researcher inputs into the model changes. Instead of describing the overall estimates for a specific assemblage, these hyper-parameters now describe a "grand mean" of the

parameter value for all the included assemblages. These overall summaries could be interpretively useful if, for instance, all the assemblages come from a discrete archaeological culture or region. In other scenarios, however, the interpretation of these overall hyperparameters may be less meaningful than comparisons of assemblage-specific estimates that still account for anatomical variation within each assemblage (see Model Supplement for more details).

Interpreting Model Results: Measured, Modeled, and Full Assemblages

The results of the Bayesian multilevel mixture model include specimen-specific membership probabilities (π_{Specimen}) based on the mixture model parameters. While these membership probabilities can be used to calculate "critical size limits" where the largest membership probability shifts from one group to another (*e.g.*, Monchot & Léchelle, 2002), they can also be used to simulate assemblages of known-group specimens to examine age/sex-stratified estimates of body part representation and sex-stratified fusion rates. Membership probabilities (π_{Specimen}) are used to simulate the specimen's identity by sampling from the probabilities using a multinomial distribution; in each posterior sample, a single simulated assemblage is created, resulting in a distribution of simulated assemblages with known age/sex assignments (Crema, 2012). The characteristics of these assemblages can then be used to summarize the overall assemblage or identify differences in composition based on element types, fusion states, sub-assemblage features, or other pertinent factors that a researcher is interested in examining in relation to the composition of the assemblage.

The usual goal of a mixture model analysis—like any sex determination analysis—is to estimate the composition of the entire (or modeled) faunal assemblage, rather than just the measured assemblage used by the analyst. Typical analyses elide these differences, smoothly translating the results of an analysis on a measured assemblage (*i.e.*, the sex ratio) to describe the entire assemblage. Sometimes, disparate results from different element portions require explanation, such as different butchery strategies for males and females (*e.g.*, Speth, 1983), but even in these cases, the results from measured specimens are used to describe the entire set of bones from the same element portion. This elision creates a bias by ignoring the existence of unmeasured specimens in the assemblage and presents an interpretive dilemma for researchers, whose only recourse if they are unwilling to make this elision is to discount the model results as unrepresentative.

We can avoid this bias by formalizing the relationship between the measured and modeled assemblages by stating that the measured assemblage is a sample of the modeled assemblage, wherein inclusion is governed by a specimen's measurability—the preservation of specific bony portions that allow for biometric measurement(s). If we assume that measurability is unrelated to a specimen's ontogenetic age or sex, then we can assume that the measured assemblage is a random sample of the modeled assemblage. Thus, an unmeasured specimen will have the same model parameters (mixture proportions π, average size μ, and size variability σ) as the measured specimens from the same element portion. Crucially, this

means that we can include unmeasured specimens in our simulated assemblages by using the relevant mixture proportions π (adjusted for the specimen's fusion data as necessary) as that specimen's membership probabilities (π_{Specimen}). Leveraging the multilevel structure of the model further, we can assume that the overall mixture model hyper-parameters for modeled element portions are equally valid for unmodeled element portions. The Bayesian multilevel mixture model estimates hyper-parameters that describe the average value (μ_{Element}) and expected variability (σ_{Element}) of mixture model parameters for element portions; these hyper-parameters can be used to estimate the relevant mixture model parameters of an unobserved element portion (Gelman et al., 2020; McElreath, 2020). The resulting parameters, then, could be used to estimate π_{Specimen} membership probabilities for the unmodeled (and unmeasured) specimens, as in the first extension, creating an estimate of the composition of the full assemblage.

At first blush, these extensions may seem like a departure from the concrete results of a mixture analysis into proxy-upon-proxy esoterica. However, by formalizing the relationship between what data are in the mixture analysis (the measured assemblage) and what data we are interested in describing (the modeled or full assemblage), these extensions are critical for creating a principled interpretation of an assemblage based on the analysis' results. Mixture analyses are based on the measurable sample of specimens from the modeled subset of all element portions; this does not mean that these results cannot produce useful information, but it does mean that we must contextualize those results by understanding how small the measured assemblage is in comparison with the modeled (or full) assemblage we are interested in describing. These extensions provide a way to do this—measured specimens will have much more certain membership probabilities than unmeasured or unmodeled specimens, owing to the information gained from their size. Thus, including unmeasured and unmodeled specimens will produce less "extreme" results (*e.g.*, a lower probability that a majority of the assemblage is from a single group). This will be especially clear when the measured assemblage is much smaller than the modeled or full assemblage.

Computational Details of the Bayesian Analysis

The Bayesian multilevel mixture model is written in Stan, version 2.30.0 (Stan Modeling Team, 2022). All analyses in this paper use R version 4.1.2 (2021-11-01), in RStudio 2022.12.0.353 (Elsbeth Geranium) (R Core Team, 2022; RStudio Team, 2022); Table 3 lists the packages, versions, and citations for the packages used in the analytical scripts. The model Stan code and analytical R code necessary to replicate and apply the analyses in this paper are freely available on GitHub page and Open Science Framework page. The files include a copy of the Shetland sheep data file from the supplemental files published in Popkin et al., (2012) and archaeological datasets for the case studies downloaded from OpenContext (Buitenhuis, 2013; Carruthers, 2006; Galik, 2013; Gourichon & Helmer, 2013). The analytical code includes two script files—a script for replication and one for application. The R Markdown file (ZooarchMixMod.Rmd) replicates the entire analytical workflow of

the paper, with a specific seed set to ensure the exact replicability of the submitted manuscript. Another set of scripts is included for the application of the model to other datasets, standardizing the analytical workflow for faunal datasets structured like the OpenContext faunal datasets used in these case studies (see the GitHub for more details at https://github.com/wolfhagenj/ZooarchMixMod). All scripts (R and Stan) are released under the MIT license, and figures are released as CC-BY to encourage reuse and reproducibility (Marwick, 2017; Marwick & Pilaar Birch, 2018).

Testing the Bayesian Multilevel Mixture Model

Two sets of tests are used to evaluate different aspects of the Bayesian multilevel mixture model. First, the accuracy of the model's ability to reconstruct the age and sex composition of assemblages is tested using simulated faunal assemblages of known age and sex from the Shetland sheep population. This test evaluates both the single-assemblage model and the multi-assemblage model. Second, two archaeological case studies showcase the applicability of the model to archaeological data and the added insights gained from adopting Bayesian multilevel mixture models. The simulated assemblage case study and the single assemblage archaeological case study use sheep (*Ovis aries*) measurements, with standard measurements coming from a female wild sheep (*Ovis orientalis* FMC 57951: Uerpmann & Uerpmann, 1994: Table 12). The multiple assemblage case studies use cattle (*Bos taurus*) measurements, with standard measurements coming from a wild female aurochs (*Bos primigenius* "Ullerslev": Degerbøl, 1970). Two dimensions of the standard cow (Scapula GLP: 89 mm; and Calcaneus GB: 46 mm) were not included in the "zoolog" output and were included manually, drawn from the referenced source.

Simulated Assemblages

A series of simulated assemblages of known age and sex composition are created from the Shetland sheep population by randomly drawing element portions (and all associated measurements) from the total assemblage without replacement. Table 4 describes the measured dimensions included in the simulation analyses from the 10 element portions. The first test, using a single-assemblage model, uses 150 element portions from the Shetland sheep population where every element portion has an equal probability of being selected. There is no guarantee, however, that the element portions have equal representation or even that all element portions are present in the simulated assemblage, which better approximates archaeological assemblages. The result of this first simulation produces an assemblage of 231 measurements from 125 individual animals. Using the same procedure, the second test creates 15 simulated assemblages that are analyzed in a single multi-assemblage model. Demographic observations for phalanx fusion rates and pelvis sex ratios were also simulated from the Shetland sheep population using the same underlying probabilities as the measurement assemblages. Table 5 describes the sample sizes of the measurement assemblages, including any manipulations to the measurement values. The

Table 4 Measurements included in the simulation analyses

Element portion	Dimension
Scapula	GLP
Humerus	Bd
Humerus	BT
Radius	Bp
Radius	Bd
Metacarpus	Bp
Metacarpus	Bd
Femur	Bd
Tibia	Bd
Astragalus	Bd
Metatarsus	Bp
Metatarsus	Bd

Dimension definitions follow von den Driesch (1976)

specific elemental composition and measurements of the assemblages, along with the simulated demographic observations, used in both simulations can be recovered from the replication script with the recorded random seed (see also Supplemental Tables S1-S3); using another random seed would provide a conceptual replication of new assemblages drawn from the same underlying populations.

Table 5 Group composition of the simulated measurement assemblages (element portions)

Assemblage	Demographics	Size	Female	Immature	Male	Total
Single assemblage	13% immature, 46% female, 40% male	1.00	80	23	47	**150**
Site 01	13% immature, 46% female, 40% male	1.00	11	2	17	**30**
Site 02	13% immature, 46% female, 40% male	1.00	3	3	4	**10**
Site 03	13% immature, 46% female, 40% male	1.20	13	4	13	**30**
Site 04	13% immature, 46% female, 40% male	0.80	11	3	16	**30**
Site 05	13% immature, 46% female, 40% male	1.20*	13	4	13	**30**
Site 06	20% immature, 70% female, 10% male	1.00	24	5	1	**30**
Site 07	20% immature, 70% female, 10% male	1.00	8	0	2	**10**
Site 08	20% immature, 70% female, 10% male	1.20	23	4	3	**30**
Site 09	20% immature, 70% female, 10% male	0.80	25	4	1	**30**
Site 10	20% immature, 70% female, 10% male	1.20*	17	11	2	**30**
Site 11	5% immature, 35% female, 60% male	1.00	10	2	18	**30**
Site 12	5% immature, 35% female, 60% male	1.00	3	0	7	**10**
Site 13	5% immature, 35% female, 60% male	1.20	17	1	12	**30**
Site 14	5% immature, 35% female, 60% male	0.80	11	3	16	**30**
Site 15	5% immature, 35% female, 60% male	1.20*	12	2	16	**30**

Demographics in the single assemblage and sites 01–05 reflect the original Shetland sheep composition

*Size increased for males only

Total counts are shown in bold

While the simulated assemblages are derived from the same Shetland sheep population that was sampled to create the 'prior assemblage' (see "Developing Prior Distributions from a Prior Assemblage"), there are several key justifications for this double use. First, the prior distributions used in the model differ from the results from the "prior assemblage" (Fig. 3); prior predictive checks of the single assemblage and multisite models show that the prior distributions are flexible enough to allow a wide range of potential assemblages (see Model Supplement). Logistically (Popkin et al., 2012), the population is the most complete, fully published assemblage of standard measurements, particularly including immature, adult females, and adult males; Davis (1996, 2000) describes similar sheep but does not include any immature specimens. Finally, the simulated assemblages vary in sample size, and some assemblages are manipulated to vary in average body size and expected composition from the original Shetland sheep population. These modifications are important to try to avoid issues of "prior mimicry," as seen in survivorship modeling (*e.g.*, Millard, 2006); this also stresses, however, the importance of developing additional sources of "prior assemblages" to test or develop relevant prior distributions.

Rather than trying to reconstruct the exact parameter values of the simulated assemblages, parametric accuracy is focused on relating the parameter distributions of the assemblage (the sample) to the respective values in the full Shetland sheep assemblage (the population from which the sample is derived), including any relevant demographic or size modifications. In this sense, the goal is not 100% accuracy; instead, the goal is being well-calibrated, wherein credible intervals about a sample parameter contain the true population values the specified percentage of the time (*e.g.*, 95% of a model's 95% credible intervals contain the true population values). If too few population values are contained in the interval statements, then the model has overfitted to the sample and the posterior distributions are too narrow; a researcher may falsely distinguish between two assemblages from the same underlying population (*i.e.*, a false positive). If too many population values are contained in the interval statements, then the model has to underfit to the sample, and the posterior distributions are too wide; a researcher may be unable to distinguish between two assemblages that derive from different underlying populations (*i.e.*, a false negative).

Compositional accuracy does not have the same structure as parametric accuracy because there is no underlying population value for composition; there is only the true number of immature, adult female, and adult male specimens in the measured and modeled assemblages. Again, though, it is important to understand accuracy in the context of overfitting and underfitting. Overfitted results, wherein credible intervals about the number of immature, adult female, and adult male specimens contain the true abundances at a lower rate than designed (*e.g.*, fewer than 95% of the 95% credible intervals), could lead to a researcher declaring an imbalance in the demographic composition of an element where one does not exist (or is even imbalanced in the opposite direction). Underfitted results, by contrast, would mean that a researcher is unable to identify an imbalance where one exists because the credible intervals are too wide. It is important, then, to use the simulations to understand the kinds of errors the model is prone to making so that researchers avoid overinterpretation.

Modeled assemblages were created for the single-assemblage simulation and the multi-assemblage simulation by assuming that measured specimens represent 20% of the overall assemblage and sampling more specimens from the Shetland sheep population to create the remaining 80% of the assemblage. For example, in the single-assemblage simulation with 150 measured specimens, this means sampling 600 more specimens from the Shetland sheep population to create a total modeled assemblage of 750 specimens. Specimens could not be repeatedly sampled, though multiple specimens could be from the same individual. As described in the "Interpreting Model Results: Measured, Modeled, and Full Assemblages" section, unmeasured specimens use the relevant π parameters for the element portion. For the multi-site simulation, this potentially includes element portions where there are no relevant measurements in the specific assemblage.

Because the "grand mean" parameters in the multisite simulation no longer represent the same thing as in the single assemblage model (see Model Supplement), the prior distributions must also be changed to reflect different expectations. Again, the goal of these prior distribution definitions is to prevent extreme overfitting so that parameter estimates are biologically feasible (Gelman et al., 2008). In general, the centers of the distributions stayed the same, but the uncertainty was increased to reflect the fact that there is less certainty about biologically feasible values for multiple populations, especially if there is size variation expected between the assemblages. These prior distributions are listed in the Model Supplement.

Archaeological Case Studies

The Bayesian multilevel mixture model is applied to two archaeological case studies to showcase its utility for both interpreting a single assemblage and examining multiple assemblages. In both case studies, the sheep and cattle measurements have been previously published on OpenContext, and the general zooarchaeological summaries of the assemblages have been published as well (Buitenhuis, 2008, 2013; Carruthers, 2005, 2006; Galik, 2013; Gerritsen & Özbal, 2019; Gourichon & Helmer, 2008, 2013). Again, LSI_e values are calculated using the same standard animal as the simulation analysis for the single assemblage analysis, the *Ovis orientalis* female standard animal (FMC 57951) from Uerpmann and Uerpmann (1994, Table 12) and the *Bos primigenius* female standard animal (Ullerslev: Degerbøl, 1970; Grigson, 1989), operationalized through "zoolog" functions (Pozo et al., 2022). Alongside metric data, the OpenContext faunal tables provide demographic data that can be used to observe relevant estimates of the age and sex composition of the assemblages. The goal of applying the mixture model to these assemblages, then, is to use the metric data to improve estimates of the age and sex composition of the assemblage, biometric estimates, and sex-specific fusion rates.

Single Assemblage: Biometric Analysis of Sheep from 7th Millennium BCE Central Anatolia (Pinarbaşı B)

The site of Pinarbaşı, located in the Konya Plain of central Turkey, consists of a series of rock shelters and open-air sites at the foothills of the Karadağ volcanic region and Lake Hotamiş and its associated wetlands (Baird et al., 2011; Kabukcu,

2017). This case study examines the Pınarbaşı B Late Neolithic occupation, which is dated to the second half of the 7th millennium BCE and includes a large number of domesticated sheep and goat remains (Baird et al., 2011; Carruthers, 2005). Carruthers (2005) analyzed fauna from the 1994 to 1995 excavations by Trevor Watkins (Watkins, 1996), interpreting the presence of fetal sheep remains and other juvenile remains in the assemblage as evidence for herders penning sheep on-site. The Neolithic assemblage was, thus, described as the result of seasonal occupation by sheep and goat herders during the lambing season and the fall, with culling in the spring possibly focused on young males (Carruthers, 2005). This analysis makes several claims that can be evaluated with the Bayesian multilevel mixture model: the dominance of immature remains and a female-dominated adult sex ratio.

The Bayesian multilevel mixture model for the Late Neolithic Pınarbaşı B assemblage uses 44 sheep measurements from 44 specimens (see Table 6; Supplemental Table S4). In addition to these measurements, the observed proportion of immature animals from unfused first and second phalanges is 59/62 (95%), including specimens identified as sheep and sheep/goat. There are 0 observed sheep (or sheep/goat) pelvis bones with sex identifications; this is entered into the model by having an observed adult sex ratio for the assemblage of 0/0 (females/females + males). All data come from the Pınarbaşı faunal assemblage uploaded to OpenContext, focusing only on specimens in the Site B Neolithic contexts (Carruthers, 2006). The Pınarbaşı B sheep model uses the same prior distribution definitions for the model hyperparameters as the single assemblage simulation since both models, even though the sheep body sizes likely differ between the two populations, showcase the flexibility of the standard prior distribution definitions.

Multiple Assemblages: Biometric Analysis of Cattle from 7th to 6th Millennium BCE Northwest Anatolia (Barcın Höyük, Ilıpınar Höyük, Menteşe Höyük)

Understanding the development of Neolithic communities in northwestern Anatolia has long been of interest to researchers interested in studying the spread of agricultural lifeways from southwest Asia into Europe (*e.g.*, Çakırlar, 2013; Karul,

Table 6 Elemental composition of the Pınarbaşı B assemblage

Element portion	Dimension	*N*
Astragalus	Bd	10
Calcaneus	GB	9
Humerus (distal)	Bd	1
Metacarpal (distal)	Bd	2
Metatarsal (distal)	Bd	4
Metatarsal (proximal)	Bp	2
First phalanx	Bp	9
Radius (distal)	Bd	2
Tibia (distal)	Bd	3
Tibia (proximal)	Bp	2

Dimensions definitions follow von den Driesch (1976)

2019; Özdoğan, 2011, 2019). Agricultural communities first appear in the Marmara region in the mid-seventh millennium BCE in sites like Barcın Höyük (Gerritsen & Özbal, 2019; Karul, 2019). The domestic animal economies of these Late Neolithic and Early Chalcolithic communities appear to be focused on cattle and caprine (sheep and goat) herding, rather than pig husbandry (Buitenhuis, 2008; Çakırlar, 2013; Gourichon & Helmer, 2008). Milk residues on pottery recovered from these sites suggest these communities regularly consumed milk, potentially orienting herd management strategies of sheep, goats, and particularly cattle to specialize in milk production (Evershed et al., 2008; Thissen et al., 2010).

Four archaeological components from three sites are used in this case study, located near Lake İznik and on the Yenişehir Plain in the Bursa province of Turkey (Fig. 4). The Neolithic layers from Barcın Höyük (Phase VI) are the earliest of these assemblages, with occupation roughly from 6500 to 6000 cal BCE; excavations revealed a subsistence economy focused on cereal agriculture and the herding of cattle, sheep, and goat (Galik, 2013; Gerritsen & Özbal, 2019). Menteşe Höyük is located approximately 5 km west of Barcın Höyük on the Yenişehir Plain; the three Neolithic layers at the site date from 5800 to 5600 cal BCE (Gourichon & Helmer, 2013; Roodenberg et al., 2003). Previous faunal analysis of the Neolithic assemblage identified animal economies that shifted from predominantly cattle to sheep herding over the course of the occupation (Gourichon & Helmer, 2008). Ilıpınar Höyük is located near Lake İznik, separated from the Yenişehir Plain by a mountain ridge (Roodenberg, 2012a). The Neolithic/Early Chalcolithic occupation of the site spanned 6200–5400 cal BCE (Buitenhuis, 2013); the assemblage is divided into two sub-assemblages (Neolithic Ilıpınar = Phases X-VII, 6000–5700 cal BCE; Chalcolithic Ilıpınar = Phases VI-V, 5600–5400 cal BCE), marked by the introduction of mudbrick architecture and expanded storage (Roodenberg, 2012a, 2012b). Sheep and goat are common in the earlier assemblages of the site, with cattle becoming predominate in later phases of the site (Buitenhuis, 2008; Roodenberg, 2012a). Notably for this biometric analysis, Buitenhuis (2008) notes that cattle body sizes are stable throughout the site's occupation.
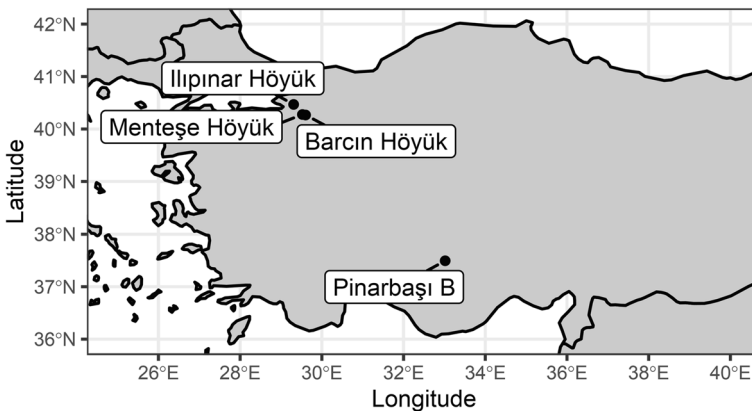

Fig. 4 Map of archaeological sites included in this analysis

The northwest Anatolian cattle bone assemblages consist of 614 measured specimens spread unevenly across the four components (Barcın Höyük $N = 67$, Menteşe Höyük $N = 45$, Neolithic Ilıpnar $N = 249$, Chalcolithic Ilıpnar $N = 253$; Supplemental Table S5). All measured *Bos* remains were included in the analysis, rather than separating out those identified as aurochs (*Bos primigenius*, N = 3) or identified only as *Bos* spp. ($N = 134$) in the Ilıpınar Höyük dataset; all specimens were only labeled as "*Bos*" in the Menteşe Höyük dataset. Table 7 shows the composition of the four northwest Anatolian measurement assemblages. Demographic observations of the proportion of immature animals and the adult sex ratio for each assemblage describe these parameters. For the four northwest Anatolian assemblages, estimates of the assemblage-level proportion of immature specimens based on the fusion rates of proximal and middle phalanges for cattle specimens are 28/87 (32%) for Barcın Höyük, 28/184 (15%) for Neolithic Ilıpınar, 8/25 (32%) for Menteşe Höyük, and 9/89 (10%) for Chalcolithic Ilıpınar. The observed adult sex ratios (females/females + males) based on cattle pelvis morphology are 3/4 (75%) for Barcın Höyük, 0/0 for Neolithic Ilıpınar, 0/0 for Menteşe Höyük, and 3/5 (60%) for Chalcolithic Ilıpınar. As in the Pinarbaşı B example, observations of 0/0 impart no information onto the prior distribution of the adult sex ratio. All demographic and measurement data come from the OpenContext datasets (Buitenhuis, 2013; Galik, 2013; Gourichon & Helmer, 2013); the associated RMarkdown file ("ZooarchMixMod.Rmd") in the GitHub project includes the steps for data processing and analysis.

Previous syntheses of the Late Neolithic and Early Chalcolithic animal economies in northwest Anatolia provide several prior inferences about the age and sex structure of cattle bone assemblages that can be evaluated with the results of the Bayesian multilevel mixture model. First, the general cultural continuity of the assemblages

Table 7 Elemental composition of the northwest Anatolia cattle measurement assemblages

Element portion	Barcın Höyük	Neolithic Ilıpnar	Chalcolithic Ilıpnar	Menteşe Höyük
Astragalus	5	19	14	15
Calcaneus	4	13	13	1
Femur (distal)	0	5	3	0
Femur (proximal)	0	4	4	0
Humerus	0	17	39	2
Metacarpal (distal)	2	6	28	0
Metacarpal (proximal)	4	10	10	1
First phalanx	16	76	35	8
Second phalanx	21	49	29	14
Radius (distal)	1	8	22	0
Radius (proximal)	2	10	35	0
Scapula	0	9	14	0
Tibia (distal)	4	9	21	0
Tibia (proximal)	0	2	4	0
Total	**67**	**262**	**296**	**46**

Total counts are shown in bold

suggests that the biometry and composition of cattle bone assemblages may be similar at the sites, having been produced by similar processes (Çakırlar, 2013; Özdoğan, 2019); Buitenhuis (2008, 312) explicitly states that there is no size change among cattle bones across the Ilıpınar assemblage. Second, the widespread evidence of milk consumption from pottery residue analyses from these sites and others in the region (Evershed et al., 2008; Thissen et al., 2010) has led some researchers to argue that cattle were managed for milk production (Gourichon & Helmer, 2008; Roodenberg, 2012a). Gourichon and Helmer (2008, 440) argue that the cattle tooth eruption and wear data at Menteşe indicate exploitation focused on milk consumption; one consequence of this pattern should be female-dominated adult sex ratios, including higher fusion rates for later-fusing elements among females than males (Zeder & Hesse, 2000). The multilevel modeling results can be used to evaluate the feasibility of these inferences by examining posterior distributions of relevant parameters and simulations of sex-specific fusion rates.

Because this application is a multisite model and deals with a different taxon than the original simulations, the prior distributions for the model hyper-parameters are redefined to reflect different expectations of biological feasibility. While the multisite simulation provides useful prior distribution definitions for most of the parameters, two other parameters (average body size of females μ_2 and index of sexual dimorphism $\log(\delta_2)$) should be changed because of different expectations modeling cattle rather than sheep. The change in the prior distribution definition of μ_2 reflects the fact that the standard measurements for cattle come from an aurochs female (Degerbøl, 1970), which is expected to be larger than the domestic cattle females in the assemblages. Cattle are expected to be more sexually dimorphic than sheep, which is reflected in increasing the average expected value of $\log(\delta_2)$, resulting in an expectation of 0.14 LSI_e units between males and females on average. This is slightly lower than the index of sexual dimorphism seen in the Degerbøl (1970) aurochs specimens (Grigson, 1989: Fig. 2, which uses LSI_{10}; the equivalent size difference is 0.06 on the LSI_{10} scale), though domestic cattle may be expected to be less sexually dimorphic than their wild counterparts (*e.g*., Tchernov & Horwitz, 1991); these prior distribution definitions are listed in the Model Supplement.

Results

Bayesian models that use the Monte Carlo methods, like the ones used here, rely on convergence diagnostics to ensure that the results (posterior distributions) have converged to the target distribution—that the results are not unduly affected by the random starting position of the analysis. To do this, analysts run multiple independent chains of the model—starting from different initial values—then evaluate how similar the chains are to one another using different diagnostic criteria (*e.g*., Gelman & Rubin, 1992). Supplemental Tables S6-S11 show the posterior estimates of the (overall and site-specific) model hyper-parameters and diagnostic criteria (R-hat and effective sample size). These results are consistent with the model successfully converging, as R-hat values are ≤ 1.01 and effective sample sizes are greater than ×100 the number of chains (Vehtari et al., 2021). Trace plots show the value of a

parameter at each posterior sample, with each chain overlain on top of each other; a converged model should have no directionality across the length of the chain and the independent chains should be indistinguishable from one another. Trace plots of each model's overall hyper-parameters are shown in Supplemental Figures S1-S4. None of the parameters show extreme deviations between chains, supporting the assertion that the model's posterior distributions properly describe the data and prior beliefs.

Simulated Assemblages: Testing Model Accuracy

Bayesian models work by updating prior information with new data to produce posterior distributions of parameters of interest (Otárola-Castillo et al., 2022). Thus, the difference between a model parameter's prior and posterior distribution shows the amount that the model "learns" from the data. If the data do not provide relevant information on a parameter's potential values, then the posterior distribution will resemble the prior distribution. Figure 5 compares the prior and posterior distributions of the main model hyper-parameters for the single assemblage simulation (for prior-posterior comparisons of the other models, see Supplemental Figures S5-S7). The results show that the data provides much more information about the likely values of the two demographic parameters (the proportion of immature animals, π_1, and the adult sex ratio, $\frac{\pi_2}{\pi_2 + \pi_3}$) and the average 594 body size for females (μ_2). This is largely to be expected, as the prior distribution definitions were weakly informative priors (Gelman et al., 2008), but it also shows how these choices did not appear to severely influence the resulting posterior distributions.

The prior distribution definitions for the size offsets (δ_1 and δ_2) and the size variability estimates (σ_1, σ_2, and σ_3) have a lot more overlap between the prior distributions and their respective posterior distributions. This overlap stresses the importance of using a Bayesian framework, particularly one relying on informative prior distributions, to produce meaningful parameter estimates from zooarchaeological data. But it also highlights the interpretive weight given to the reference population. However, the overlap is not necessarily a drawback of the model, as again, the prior distribution definitions were designed as informative priors, specifically to ensure that the resulting parameter estimates would be biologically feasible. Furthermore, the simulated population also has the same underlying biological population (the Shetland sheep population) that was used to develop the prior distributions, so it is possible that this overlap reflects that fact.

The parametric and compositional accuracy of both simulation tests are summarized in Table 8. The single assemblage model is well-conditioned when examining parametric accuracy, though the multisite model overfits in this respect; this is driven by poor performance on size variability (σ) parameters—the model estimates average body size (μ) parameters well. The multisite model also has a tendency to underfit when examining site-specific compositional accuracy. In both models, though, the compositional accuracy improves (in the sense of no longer underfitting) by using the modeled assemblages rather than the measured assemblages. This
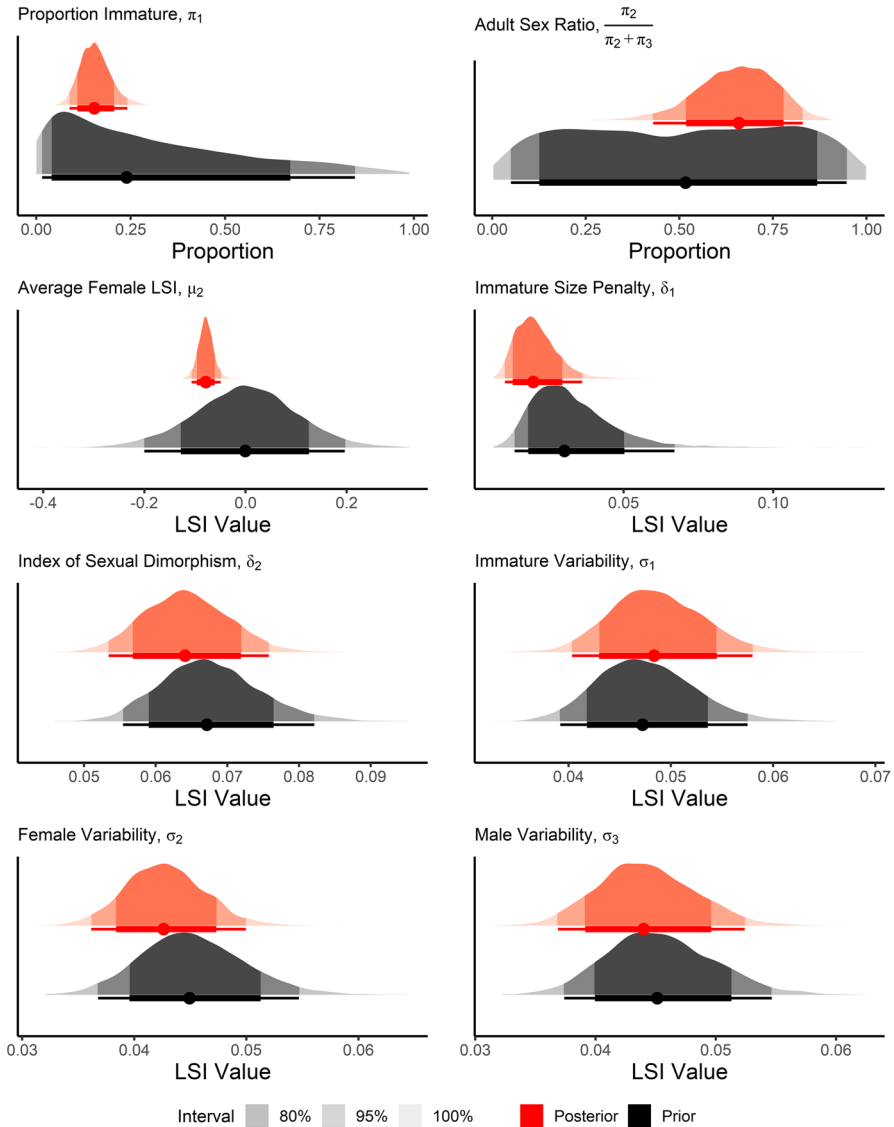
Fig. 5 Comparison of prior and posterior distributions for mixture model hyper-parameters of the simulated single assemblage. The model hyper-parameters serve as assemblage-wide estimates accounting for size and composition variation across element portions.

makes intuitive sense, as the measured assemblage is itself theoretically a sample from the modeled assemblage (based on the assumption that "measurability" is random). In these simulations, of course, this theory is held to be explicitly true, though the relationship between the measured and modeled assemblage is generally held to be true implicitly in zooarchaeology and can be explicitly tested (see "Interpreting Model Results: Measured, Modeled, and Full Assemblages").

Table 8 Parametric accuracy rates and expected accuracy ranges for the simulation models compared to the population values from which the samples were drawn

Quantity	Assemblage	Accuracy	Expected accuracy	Interpretation
Mixture model parameters (80%)	Single assemblage	9/9 (100%)	5–9	
	Multisite (overall)	79/135 (59%)	99–117	Overfit
	Multisite (average size)	41/45 (91%)	31–41	
Mixture model parameters (95%)	Single assemblage	9/9 (100%)	7–9	
	Multisite (overall)	114/135 (84%)	123–133	Overfit
	Multisite (average size)	45/45 (100%)	40–45	
Measurement assemblage composition (80%)	Single assemblage	24/24 (100%)	15–23	Underfit
	Multisite (overall)	21/24 (88%)	15–23	
	Multisite (by site)	331/339 (98%)	256–285	Underfit
Measurement assemblage composition (95%)	Single assemblage	24/24 (100%)	20–24	
	Multisite (overall)	24/24 (100%)	20–24	
	Multisite (by site)	338/339 (>99%)	314–329	Underfit
Modeled assemblage composition (80%)	Single assemblage	18/24 (75%)	15–23	
	Multisite (overall)	17/24 (71%)	15–23	
	Multisite (by site)	302/360 (84%)	273–303	
	Multisite (new elements)	18/21 (86%)	13–20	
Modeled assemblage composition (95%)	Single assemblage	21/24 (88%)	20–24	
	Multisite (overall)	22/24 (92%)	20–24	
	Multisite (by site)	349/360 (97%)	334–350	
	Multisite (new elements)	20/21 (95%)	18–21	

Expected accuracy ranges are based on the number of parameters being evaluated. If the modeled accuracy rate is above the range, then the model has to underfit the data. If the modeled accuracy rate is below the range, then the model has overfitted to the data

Figures 6 and 7 show the posterior distributions of the site-level parameters for the single-assemblage and multisite simulations. Figures 6A and 7A show the posterior distributions for the mixing proportions (π parameters), Figures 6B and 7B show the posterior distributions for the average body size (μ parameters), and Figures 6C and 7C show the posterior distributions for the size variability (σ parameters) for immature (black), adult female (blue), and adult male (red) specimens in the assemblage. Vertical bars summarize the 80 and 95% credible intervals for the parameter, while the solid horizontal lines denote the "true" parameter value for the Shetland sheep population from which the assemblage was sampled, including any deviations made for the multisite simulation (see Table 5). The model excels at estimating the average body sizes of the animal groups, even when those body sizes are manipulated (Figs. 6 and 7B). The multisite model's estimates of the overall proportions tend to be conservative (Fig. 7A), that is, underestimating the proportions for animal groups with extremely low "true" proportions (*e.g.*, adult males in sites 6–10, immature animals in sites 11–15) while overestimating the proportions for animal groups with extremely high "true" proportions (*e.g.*, adult females in sites 6–10). This is likely due to the multilevel modeling structure "shrinking" estimates towards a global mean, though it may also reflect overfitting from small sample sizes. This is particularly the case when examining variability parameters (σ: Fig. 7C), which the multilevel model tends to underestimate.

Figure 8 shows posterior distributions of simulated group-specific compositions for both the single-assemblage and total composition of the multisite models (*i.e.*, all
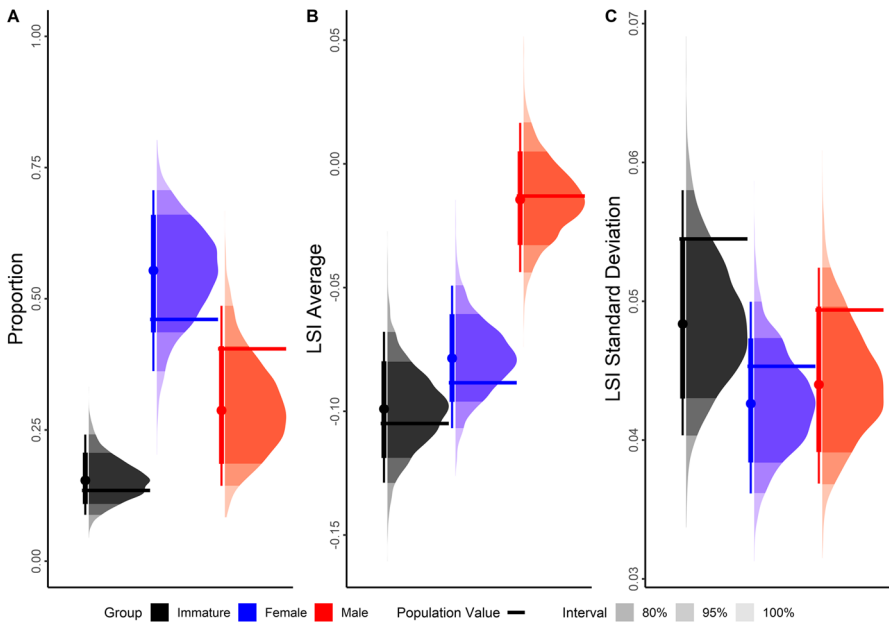
Fig. 6 Posterior distributions of the mixture model hyper-parameters (assemblage-level estimates) for the simulated single assemblage. **A** Mixture proportion (π). **B** Average LSI_e size (μ). **C** Size variability (σ). Horizontal lines denote the relevant parameter values from the Shetland sheep population from which the assemblage was sampled
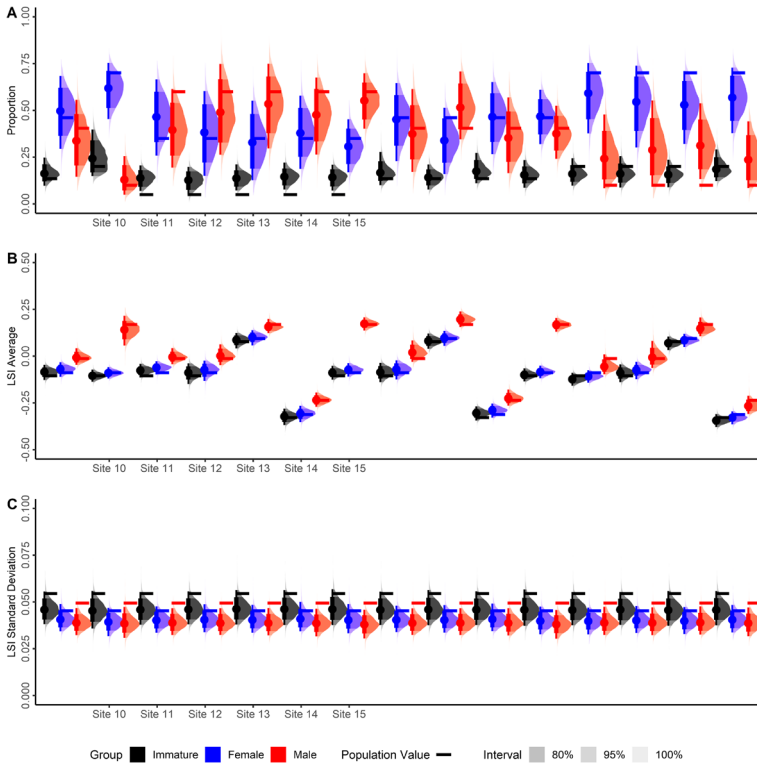
Fig. 7 Posterior distributions of assemblage-level mixture model parameters for the simulated multisite assemblages. **A** Mixture proportion (π). **B** Average LSI$_e$ size (μ). **C** Size variability (σ). Horizontal lines denote the relevant parameter values from the Shetland sheep population from which the assemblage was sampled, including modifications to the composition and body size for some assemblages (see Table 4)

sites combined) alongside true counts for each group. The underfitting performance of the multisite simulation, particularly at the level of individual sites, may be a consequence of low statistical power due to small sample sizes. The median number of specimens per element portion is 4 for the measured assemblage and 17 for the modeled assemblage. By contrast, the single-assemblage simulation has a median number of specimens per element portion of 19.5 for the measured assemblage and 92 for the modeled assemblage. More important than overall sample sizes, though, is the group-specific element counts, where the contrast between the median number of specimens in the measured and modeled assemblages from the multisite simulation (measurement assemblage: 1, modeled assemblage: 4) and the measured and modeled assemblages from the single-assemblage simulation (measurement assemblage: 5, modeled assemblage: 26) is starker. The larger sample sizes in the modeled assemblages also partially explain the reduced underfitting relative to the measured assemblages.

It is also noteworthy that the additional sampling to create the multisite simulation's modeled assemblage resulted in seven newly observed element portions at
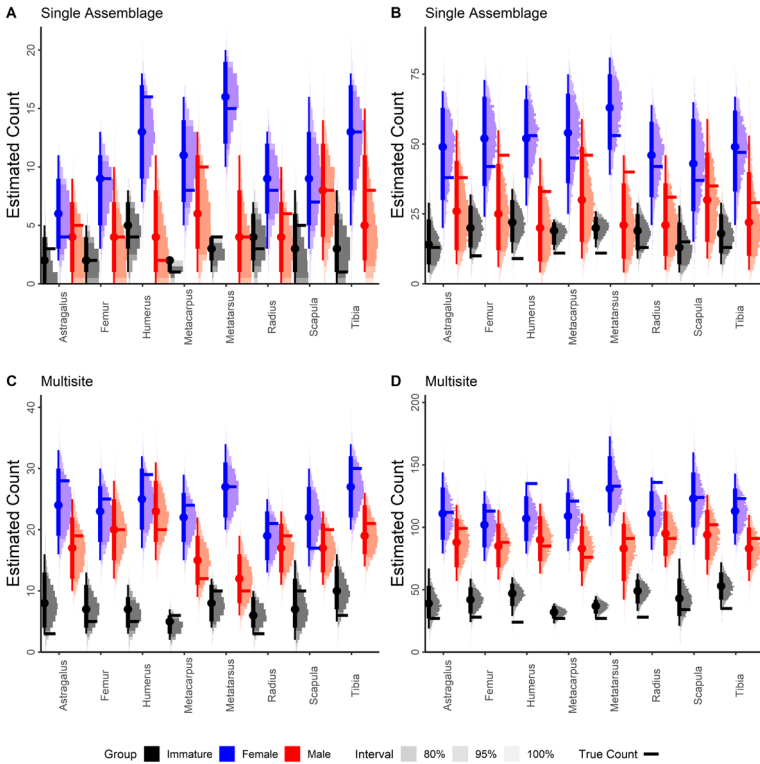
Fig. 8 Posterior distributions of simulated group-specific composition for the simulated assemblages. Top row: single assemblage model **A** measured assemblage and **B** modeled assemblage. Bottom row: multisite assemblage model (combined counts) **C** measured assemblages and **D** modeled assemblages. Horizontal lines denote the true group-specific composition of the relevant measured or modeled assemblage

certain sites. The model can estimate the composition of these element portions due to its multilevel structure, which estimates element-specific offsets and interaction terms ($v_{Element}$ and $v_{Interaction}$) for elements that are present in at least one site. The compositional accuracy of these newly observed element portions is well-calibrated (see Table 8), despite having no observed measured specimens from the element portion for those sites. That these element portions can be accurately modeled despite lacking observed measurements for the site lends support to the idea that researchers could extend the same multilevel model structure (element-specific offsets and interaction terms) to estimate the composition of unobserved (*i.e.*, completely unmeasured) element portions in an assemblage.

Pinarbaşı B Sheep: The Impact of Immature Specimens

The first three distributions in Fig. 9 show the posterior estimates of the assemblage-level proportion of immature, female, and male sheep at Pinarbaşı B. In general, the

Pinarbaşı B sheep assemblage is overwhelmingly composed of immature animals (posterior μ_{π_1} median = 87%; 95% posterior credible interval for μ_{π_1} = 77–94%), somewhat lower than the observed fusion rate of proximal and middle phalanges (59/62 = 95%). Even though the overall proportions of female and male animals in the assemblage are low relative to immature specimens, the Bayesian multi-level mixture model can produce a posterior distribution of the overall adult sex ratio that suggests that adult females are more common than adult males (median θ_{Female} = 64%; 95% posterior credible interval for θ_{Female} = 10–96%); the wide credible interval of this distribution reflects the small proportions of mature specimens in the assemblage overall. Despite this uncertainty, we can use this distribution to estimate the probability that adult females are more common than adult males (*i.e.*, that θ_{Female} is greater than 0.5); 68% of the posterior θ_{Female} samples are above 0.5, providing some confidence in the interpretation that the adult sex ratio is skewed towards females. Note that any threshold value could be chosen to test a hypothesis about the adult sex ratio—for instance, only 45% of the posterior θ_{Female} samples are above 0.67 (*i.e.*, a 2:1 female-to-male ratio), providing poor support that the mature portion of the assemblage is heavily skewed towards females.

The other distributions show element-specific composition estimates for the Pinarbaşı B assemblage. Most of the element-specific proportions of immature animals (π_1) broadly match the overall estimates, with posterior medians over 75%. However, these element-specific distributions also have long tails extending into lower π_1 values, conveying less certainty about element-specific π_1 estimates relative to the assemblage-wide estimate. This likely owes to small element-specific
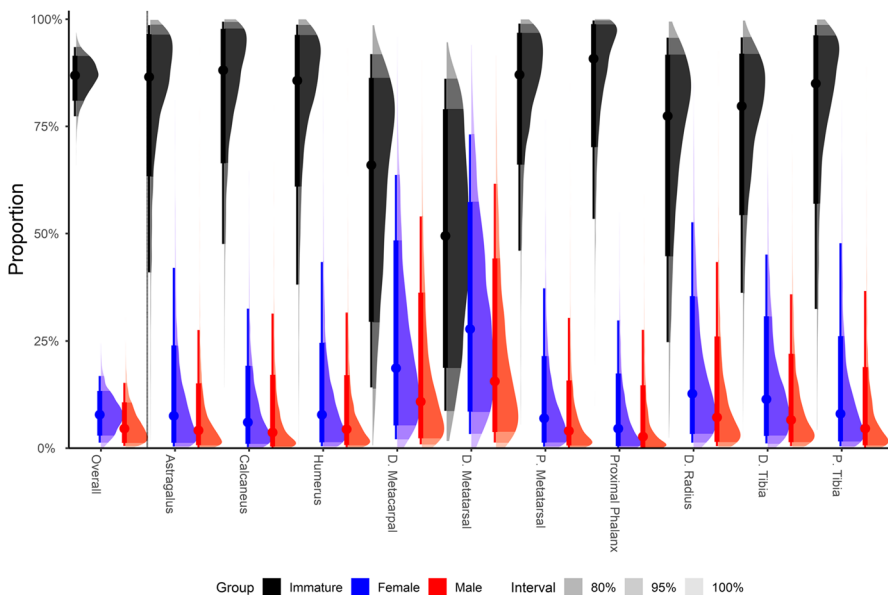
Fig. 9 Posterior distributions of composition estimates (π) for the Pinarbaşı B sheep assemblage. The first three distributions to the left show assemblage-level estimates, while the other distributions show element-specific composition estimates

sample sizes (the astragalus, calcaneus, and proximal phalanx have 9–10 specimens, all other element-specific samples sizes are 1–4, see Table 6) and to the presence of some element portions with lower modeled proportions of immature animals. Two element portions—the distal metacarpal and distal metatarsal–have posterior π_1 median values below 75%, though again, have long tails that extend in both directions (Distal metacarpal: π_1 posterior median = 66%, 95% posterior credible interval: 46–92%; distal metatarsal: π_1 posterior median = 49%, 95% posterior credible interval: 32–86%). Notably, all measured specimens from these two element portions have fused distal epiphyses, meaning that the model considers it impossible for the specimens to be immature.

Figure 10 shows the same comparison (overall and element-specific distributions) for the average size of female animals (μ_2) in the Pinarbaşı B assemblage. Average body sizes vary across elements, highlighting some allometric variation between Pinarbaşı B sheep and the standard sheep and the importance of using a multilevel model to account for allometric variation. The multilevel structure of the model provides a parameter ($\sigma_{\text{Element}}[3]$) that directly estimates this variation: the posterior mean of this distribution is 0.05, with an upper 95% quantile of 0.08 on the LSI_e scale. Combining all the measurements into a single LSI analysis would confound this element-level variation with differences in composition, muddling the ability to compare the biometry of Pinarbaşı B sheep with sheep from contemporary assemblages.
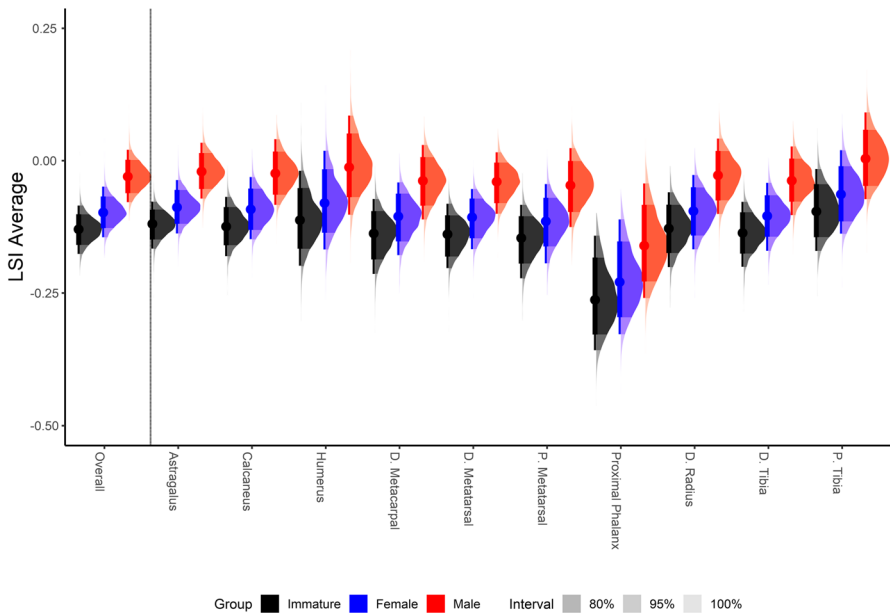
Fig. 10 Posterior distributions of average LSI value for female animals (μ_2) for the Pinarbaşı B sheep assemblage. The first three distributions to the left show assemblage-level estimates, while the other distributions show element-specific size estimates

Figure 11 shows the distribution of simulated compositions for immature, female, and male specimens in three Pınarbaşı B assemblages: the measured assemblage (*N* = 44), the assemblage of modeled element portions (*N* = 277), and the full sheep assemblage including five element portions that were not modeled due to lack of measurements (additional elements: proximal radius, ulna, proximal metacarpal, pelvis, and middle phalanx; total *N* = 428). Beyond visualizing the group-specific composition of the assemblage, the simulated compositions provide more insight into the assemblage's formation. Five element portions in the full assemblage (astragalus, calcaneus, proximal radius, proximal fused metacarpal bones 3 and 4, pelvis) contain no mature specimens in 18% of the simulations; by contrast, no element portion has more than 2% of the simulations without immature specimens (see Table 9). Additionally, element portions vary in the probability that they contain male
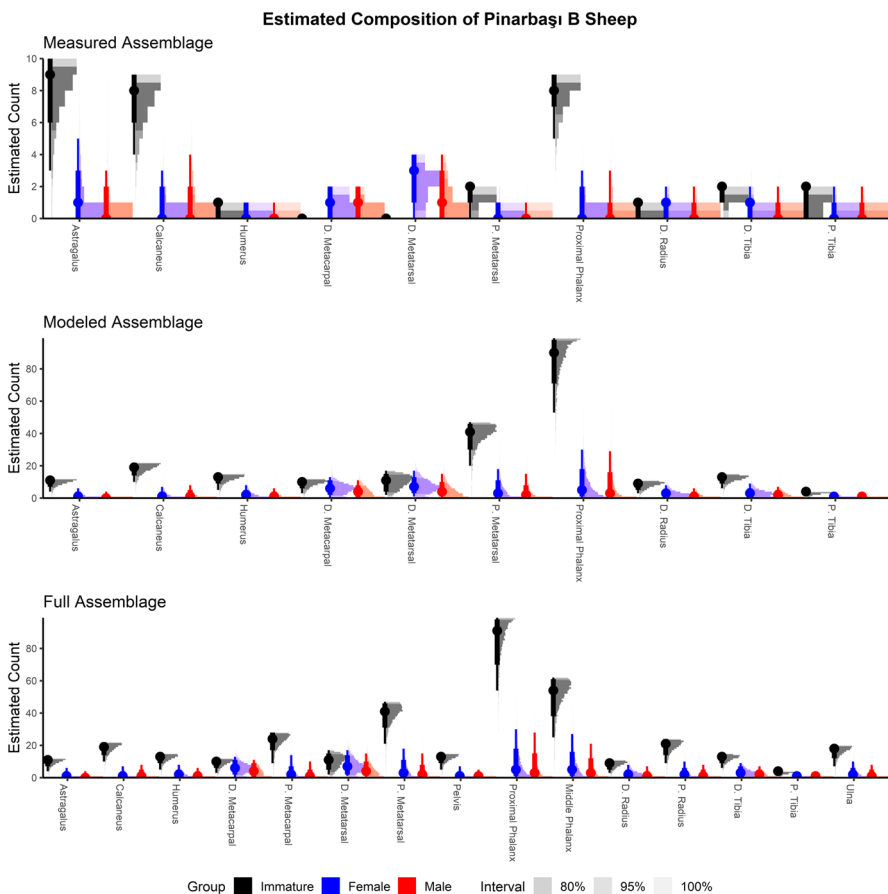
Fig. 11 Posterior distributions of simulated group-specific composition for the Pınarbaşı B sheep (top) measured, (middle) modeled, and (bottom) full assemblages. The full assemblage includes five additional element portions compared to the measured and modeled assemblages (proximal metacarpal, pelvis, middle phalanx, proximal radius, ulna)

specimens: distal metapodials (no males in 8% of simulations) and phalanges (no males in 16% of simulations) contain males much more frequently than other element portions (no males in 39% of simulations; Table 9). While this analysis focuses on appendicular elements, these results suggest that mature animals—particularly mature males—did not enter the assemblage as complete animals but potentially as raw materials for bone tools (*e.g.*, distal metapodials, distal tibia: Russell & Griffitts, 2013) or on skins (*e.g.*, proximal and middle phalanges). Immature animals, by contrast, appear to have entered the assemblage as complete animals; the element portions with the highest probability of missing immature specimens may be affected more by density-mediated attrition (*e.g.*, proximal tibia: Symmons, 2005) and identifiability of neonatal specimens (*e.g.*, distal metapodials; see "Discussion").

These results strongly reinforce the interpretation that Pinarbaşı B was used by herders as a camp where sheep gave birth, with on-site culling largely reflecting either the first seasonal cull of animals before winter or animals that died naturally in their first year of life (Carruthers, 2005; Martín et al., 2015). This method reinforces previous analyses, which is not surprising; however, the mixture modeling results go beyond these earlier interpretations by creating group-specific biometric estimates and providing a probabilistic framework to estimate a specimen's membership into the three groups. The biometric estimates not only account for the presence of immature specimens in the assemblage but also for allometric variation across element portions. The probabilistic identifications allow analysts to simulate assemblage compositions, highlighting potential differences in the ways that bones from immature, adult female, and adult male entered the Pinarbaşı B assemblage.

Table 9 Percentage of simulated group-specific compositions of the full Pinarbaşı B sheep assemblage that exclude a different population category by element portion

Element portion	P (no immature)	P (no adults)	P (no females)	P (no males)
Metatarsal (distal)	1%	0%	2%	6%
Metacarpal (distal)	0%	0%	3%	10%
Proximal phalanx	0%	0%	7%	16%
Radius (distal)	0%	0%	7%	33%
Middle phalanx	0%	0%	8%	17%
Tibia (distal)	0%	0%	9%	22%
Metatarsal (proximal)	0%	7%	16%	27%
Ulna	0%	0%	16%	32%
Humerus	0%	0%	18%	36%
Metacarpal (proximal)	0%	12%	24%	36%
Radius (proximal)	0%	12%	26%	39%
Tibia (proximal)	2%	0%	30%	49%
Calcaneus	0%	18%	35%	41%
Pelvis	0%	22%	38%	50%
Astragalus	0%	28%	41%	64%

Unmodeled element portions included in the composition analysis are the proximal metacarpal, pelvis, middle phalanx, proximal radius, and ulna

7th–6th Millennium BCE Northwest Anatolian Cattle: Examining Differences Between Assemblages

Figure 12 shows the posterior distributions of average body sizes for female cattle ($\mu 2$) from the four analyzed assemblages. These distributions are produced from posterior samples; assemblage-specific estimates from a single posterior sample share the same relevant hyper-parameters (μ_{μ_2} and σ_{Site} [3]), meaning that they covary with one another to an extent. To compare these distributions, then, a contrast is necessary to account for this potential covariation. This is done by simply evaluating the difference between two parameters
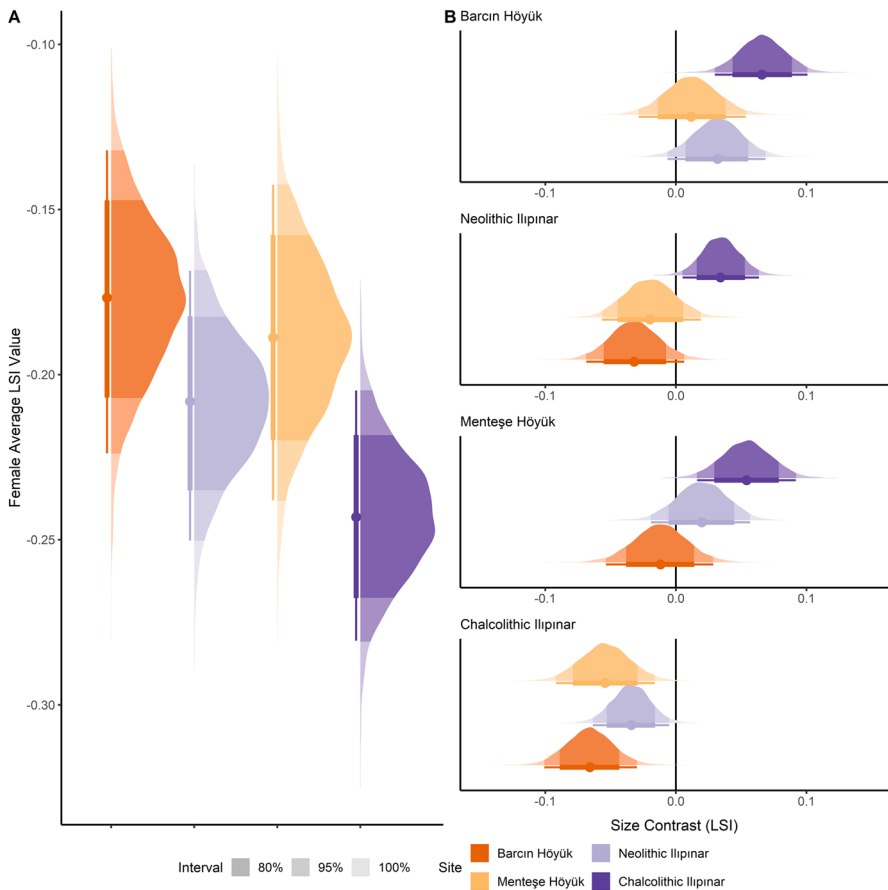
Fig. 12 Posterior distributions of site-specific average LSI value for female animals (μ_2) for the northwest Anatolian cattle assemblages. **A** The distributions in broadly temporal order from left to right; **B** site-specific contrasts for average female body size, indicating specific size differences between pairs of sites. The title describes the focal assemblage, while the color scheme for the distributions is consistent across both panels. If the contrast distribution is greater than zero, then the focal assemblage is larger than the displayed assemblage (*e.g.*, the top-most distribution in the top panel shows that Barcın Höyük is significantly larger than Chalcolithic Ilıpınar Höyük)

(*e.g.*, between the average female LSI$_e$ value μ_2 for Barcın Höyük and μ_2 for Neolithic Ilıpınar) in each posterior sample, shown in the right-hand panel of Fig. 12. These contrasts show that the female cattle from Chalcolithic Ilıpınar are likely smaller, on average, than female cattle from the other sites. These cattle measurements are 3–7% smaller, on average, than those from the other northwestern Anatolian sites relative to the standard animal measurements. Thus, the mixture modeling approach not only identifies a size difference that was unrecognized when using standard LSI analysis (*e.g.*, Buitenhuis, 2008), but also reveals a diachronic trend in cattle body size that can be studied to evaluate factors like climate or herding practices that affected animal body size over time (*e.g.*, Munro et al., 2022; Wright & Viner-Daniels, 2015). Accounting for differences in the elemental and demographic composition of different assemblages provides more accurate reconstructions of body size that allow us to critically examine patterns of animal size change in the past (*e.g.*, Arbuckle & Kassebaum, 2021; Manning et al., 2015).

In addition to identifying size differences among the assemblages, the modeling results also reveals variation in the age and sex composition of the four assemblages. Figure 13 shows the distributions of assemblage-level demographic variables—the proportion of immature animals and the adult sex ratio (the proportion of adults that are female)—for the four northwest Anatolian assemblages. The assemblages have broadly similar estimates for the adult sex ratio (right-hand panel of Fig. 13), with strong evidence that they contain more females than males. However, modeling shows significant variation in the proportion of immature animals in the assemblages (top panel of Fig. 13): the proportion of immature cattle in the Menteşe Höyük and Barcın Höyük assemblages is significantly higher than the proportions of immature cattle in either Ilıpınar Höyük assemblage. This difference between the sites could be investigated further to understand whether there are taphonomic factors that have selectively removed immature specimens from the Ilıpınar Höyük assemblage post-depositionally or whether herd management strategies differed among communities at these sites (*e.g.*, Gillis et al., 2014, 2015).

Simulating sex-specific fusion rates for late-fusing elements (proximal femur, distal femur, proximal humerus, distal radius, proximal tibia, proximal ulna: Grigson, 1982) from the full northwest Anatolian assemblages highlights the complexities of examining sex-specific fusion rates in zooarchaeological assemblages. In each assemblage, estimates of male fusion rates are extremely uncertain, owing to the small number of estimated males in each iteration and thus large potential shifts in the denominator for fusion rates (Fig. 14). This uncertainty makes it difficult to clearly establish whether fusion rates differed between males and females; regardless, in 68% of the posterior samples female fusion rates were higher than male fusion rates for Chalcolithic Ilıpınar. These results complicate regional syntheses that tie the presence of milk residues to milk-oriented cattle management (*e.g.*, Evershed et al., 2008); the ability to directly estimate sex-specific fusion rates allows researchers to test the validity of these exploitation models for past assemblages (*e.g.*, Arbuckle & Atici, 2013).
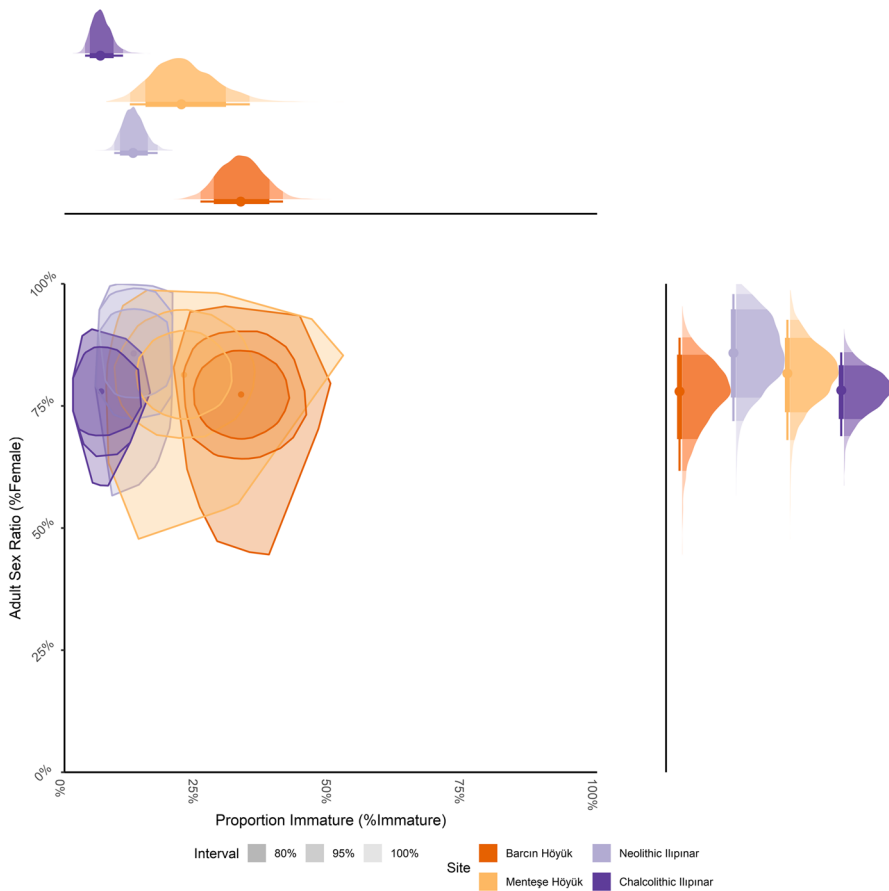
Fig. 13 Comparison of posterior distributions of site-specific demographic parameters (proportion of immature $\pi_{Immature}$ and adult sex ratio θ_{Female}) for the northwest Anatolian cattle assemblages. The top panel shows the marginal plot for the proportion of immature specimens, while the right-hand panel shows the marginal plot for the adult sex ratio (in the proportion of females among mature animals)

Discussion

The simulation analyses show that the Bayesian multilevel mixture model presented here can accurately reconstruct age- and sex-specific biometry of a faunal population represented in a measured assemblage, while also producing relatively accurate estimates of the "demographic" (ontogenetic age and sex) composition of the assemblage. The archaeological applications of the mixture model demonstrate how the model can highlight meaningful variation in the composition and relative size of specimens across element portions. The results can point to potential differences in how animals entered an assemblage, as between immature and mature sheep at Pınarbaşı B, or reveal variation in demographic proportions that could highlight taphonomic differences or variation in management strategies across the sites, as
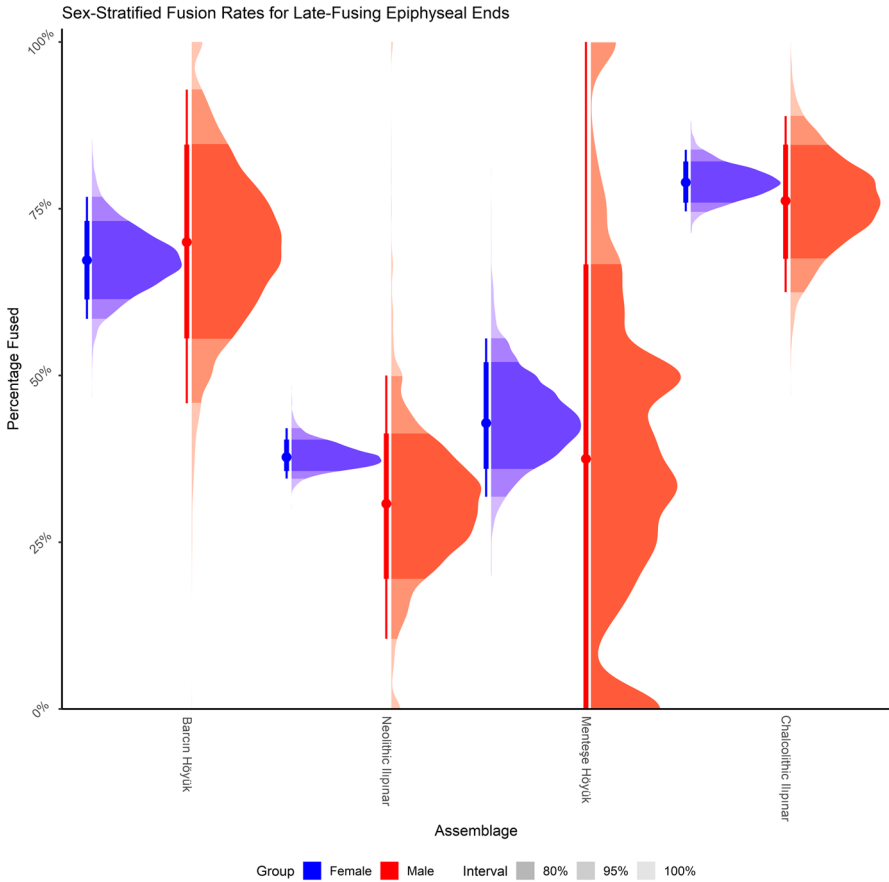
Fig. 14 Posterior distributions of simulated sex-specific fusion rates for late-fusing elements among northwest Anatolian cattle full assemblages. Included element portions are the distal tibia, distal metapodials, calcaneus, femur, proximal ulna, distal radius, proximal tibia, and proximal humerus

in the northwest Anatolian cattle. While broadly supporting the earlier analyses of these assemblages, applying mixture models to the measurement assemblages also opened new lines of inquiry based on exploring the drivers of variation in body size and assemblage composition.

The performance of the Bayesian multilevel mixture models relies on the prior distributions, which provide constraints against overfitting and ensure that the model produces biologically reasonable parameter estimates. The prior distribution definitions in this paper were derived largely from the measurements of a herd of known-age, known-sex population of Shetland sheep (Popkin et al., 2012), though for the multisite cattle model, some of the definitions were changed based on data on European aurochsen (Degerbøl, 1970). It is important to note that prior distribution definitions can be derived from many different sources—including quantification based on one's judgment (*e.g.*, Gelman et al., 2008; McCarthy & Masters, 2005). This

could allow researchers to use different kinds of demographic priors to adjust for their expectations about the ways that assemblages are created, for instance, if one may expect catastrophic profiles (*e.g*., Stiner, 1990). More important than the source of one's prior distribution definitions is investigating the expectations of those prior distribution definitions by performing prior predictive checks as in the Model Supplement (Gabry et al., 2019; Gelman, Vehtari, et al., 2020). Furthermore, emphasis should be paid to increasing the diversity of known-age, known-sex animal populations with individual measurement data (*e.g*., Lebenzon & Munro, 2022; Zeder & Lemoine, 2020), which could help develop prior distributions relevant to different taxa and to understand how variable different parameters, especially size variability (σ) parameters, are across populations.

One of the central tenets of the mixture model's extension to modeled assemblages is the idea that "measurability" (adequate preservation to maintain a biometric dimension) is unrelated to a specimen's status as immature, female, or male. Variation in the mixture proportions among elements, especially the proportion of immature specimens π_1 may highlight group-specific biases in the deposition of specimens but could also indicate issues with the assumption that "measurability" is random. The Pinarbaşı B sheep assemblage potentially demonstrates this issue, as the distal metapodials have much lower element-specific π_1 estimates than other element portions. While metapodial bones from mature sheep could have been selectively over-represented in the assemblage, it is also likely that distal metapodials from immature animals—particularly very young animals—are less likely to be measurable compared to adult animals. Because the distal breadth measurement requires both distal condyles to be present, distal metapodial specimens from neonatal or extremely juvenile individuals may be missed while those from other element portions (*e.g*., proximal metapodial, distal humerus) would still be theoretically measurable (Martín & García-González, 2015). The inclusion of condyle-specific measurements could address this issue, though it would require identifying whether the isolated condyle is medial or lateral (*e.g*., width of condyle: Payne, 1969).

The ability to create accurate simulated estimates of age and sex composition provides many opportunities for further analyses. For instance, a comparison of the composition of animals in different depositional contexts could support contextual taphonomic analyses (*e.g*., Meier, 2020). Access to certain kinds of animals could highlight systems of provisioning or status-related restrictions (Arbuckle 2012; Twiss, 2019: 73–97). Differences in the ontogenetic age and sex composition of different body parts could also highlight ritual behaviors reflected in the use of certain contexts or sites (*e.g*., Madgwick & Mulville, 2015). In a similar vein, tracking adult sex ratios could identify the use of castrates to take advantage of secondary products like wool or labor; the models used here would identify castrates as males, though alternative prior distributions and measured dimensions could distinguish these groups (*e.g*., Popkin et al., 2012: Fig. 7). On a more practical level, providing specimen-specific probabilities of being immature, female, or male can provide a useful baseline for sampling strategies focused on ancient DNA or stable isotopes, allowing researchers to explore potential sex differences in diets (*e.g*., Post et al., 2001) or more easily identify male specimens to isolate Y-chromosomal DNA to explore sex-specific selection (*e.g*., Daly et al., 2021; McGrory et al., 2012).

The archaeological case studies highlight the importance of considering the presence of immature specimens and elemental variation in body size when summarizing the biometry and composition of an assemblage. Variation in the proportion of immature animals in the assemblage, as in the multisite case study for northwest Anatolia, may point to differences in culling strategies or even the seasonality of animal presence at the sites. Most of the Pinarbaşı B material derives from immature specimens, which could complicate inter-assemblage biometric analyses that do not use sex-specific size estimates (*e.g.*, Arbuckle et al., 2014; Helmer et al., 2005). Restricting measurements only to fused specimens removes useful information, particularly when fusion rates may differ between male and female animals (Zeder & Hesse, 2000); furthermore, it does not resolve the problem of immature animals in the measurement assemblage if early-fusing elements like the distal humerus are still included because of post-fusion growth (Popkin et al., 2012). The ability to create sex-specific biometric estimates is important to document large-scale spatial and temporal dynamics in animal body size (*e.g.*, Arbuckle et al., 2016; Arbuckle & Kassebaum, 2021; Wolfhagen et al., 2021).

Examining LSI_e (or LSI_{10}) values of Pinarbaşı B sheep without accounting for the impact of immature animals would mislead a researcher into believing that sheep were smaller, on average than sheep from contemporaneous sites (*e.g.*, late Çatalhöyük: Baird et al., 2011). While stable isotopic analyses have suggested that the Pinarbaşı B and Çatalhöyük sheep had similar diets (*e.g.*, Baird et al., 2011, 2018; Middleton, 2018), biometric analyses had not previously supported the idea that these assemblages derived from the same animal population (*e.g.*, Arbuckle et al., 2014: Fig. 4). Only eight measurements from the Pinarbaşı B sheep are excluded from being immature due to fusion status, four of which are distal metatarsals; the median LSI_e of these specimens is −0.12, compared to a median of −0.06 for the 362 Late Çatalhöyük sheep that must be mature based on fusion status (Wolfhagen et al., 2021). Even though the mixture model's estimate for the average LSI_e for female sheep (μ_2) at Pinarbaşı B is relatively uncertain (95% credible interval: −0.14 – −0.05 LSI_e), it still provides a useful reference point for comparison that supports the idea that these assemblages have similar biometrics (95% credible interval of μ_2 for late Çatalhöyük sheep: −0.11 – −0.06 LSI_e: Wolfhagen et al., 2021).

The increased ability to specify the ontogenetic age and sex composition of faunal assemblages with Bayesian multi-level mixture models also highlights the limitations of our current language used to describe and interpret these compositions. Results like the sex-specific fusion biases in northwest Anatolian cattle complicate straightforward expectations of clear sex-specific patterns in archaeological assemblages, which is perhaps unsurprising. Many discussions that examine changes in the composition of faunal assemblages to identify shifts in exploitation patterns use terms like "prime-dominated age structure" (*e.g.*, Stiner, 1990), "dominance of females," (*e.g.*, Peters et al., 2005 or "young male slaughter/kill-off" (*e.g.*, Arbuckle & Atici, 2013; Zeder & Hesse, 2000). These terms are deceptive in their utility—they describe some empirical pattern but are ordinal-scale; thus, it is up to the individual researcher to define the cut-off between a "dominant" and "non-dominant" assemblage (Wolverton et al., 2016). In the case of the adult sex ratio for Barcın Höyük, 98% of the posterior samples are above 60% (1.5 females:1 male), but only

68% of the posterior samples are over 75% (3 females:1 male). Meanwhile, 79% of the posterior samples for the adult sex ratio for Neolithic Ilıpınar are over 80% (4 females:1 male). Are both assemblages "dominated by females"? More formalized language in our hypotheses—or, rather, the adoption of statistical modeling frameworks (McElreath, 2020: 4–17)—is necessary to clarify what changes in assemblage-level estimates of biometry and composition mean for past human-animal interactions.

Conclusions

This paper describes a new method for estimating the biometry and ontogenetic age and sex composition of faunal assemblages based on standard measurement data using Bayesian multilevel mixture modeling. The model produces accurate estimates of sex-specific biometry, which can provide a more useful framework for inter-assemblage analysis (*e.g.*, Arbuckle et al., 2016; Helmer et al., 2005). Such a framework could better explore broad spatial and chronological patterns in animal biometry while accounting for differences in assemblage composition across the assemblages, ensuring reliable comparisons of animal body size in relation to other variables. These analyses could investigate the processes behind size fluctuation in animals, particularly in relation to changing human-animal interactions and adaptation to new lifeways and anthropogenic environments.

Furthermore, the estimates of the age and sex composition of the assemblage can be used to simulate assemblages of specimens with known group assignment (immature, female, and male). These simulations are the baseline for comparing differences in the composition of sub-assemblages. Using these simulations allows researchers to make testable statements about the composition of the assemblage and to directly test hypotheses about differences in the age and sex composition of animal bones from different parts of a site, different fusion groups, or other categories. The Bayesian structure of the model allows researchers the flexibility to create hypotheses that can be tested directly, rather than relying on null hypothesis testing for inference (Otárola-Castillo et al., 2022; Otárola-Castillo & Torquato, 2018). Thus, the mixture modeling framework described here provides a foundation for biometric and compositional analyses that operate at multiple scales and presents a new avenue for summarizing and comparing zooarchaeological assemblages.

Data Availability Data and analytical scripts are available at the associated GitHub project (https://github.com/wolfhagenj/ZooarchMixMod) and Open Science Framework page (https://osf.io/4h9w6/).

Declarations

Competing Interests The author declares no competing interests.

References

Arbuckle, B. S. (2012). Animals and inequality in Chalcolithic Central Anatolia. *Journal of Anthropological Archaeology, 31*(3), 302–313. https://doi.org/10.1016/j.jaa.2012.01.008

Arbuckle, B. S., & Atici, L. (2013). Initial diversity in sheep and goat management in Neolithic southwestern Asia. *Levant, 45*(2), 219–235. https://doi.org/10.1179/0075891413Z.00000000026

Arbuckle, B. S., & Kassebaum, T. (2021). Management and domestication of cattle (*Bos taurus*) in Neolithic southwest Asia. *Animal Frontiers, 11*(3), 10–19. https://doi.org/10.1093/af/vfab015

Arbuckle, B. S., Price, M. D., Hongo, H., & Öksüz, B. (2016). Documenting the initial appearance of domestic cattle in the eastern fertile crescent (northern Iraq and western Iran). *Journal of Archaeological Science, 72*, 1–9. https://doi.org/10.1016/j.jas.2016.05.008

Arbuckle, B. S., Whitcher Kansa, S., Kansa, E., Orton, D. C., Çakirlar, C., Gourichon, L., Atici, L., Galik, A., Marciniak, A., Mulville, J., Buitenhuis, H., Carruthers, D., De Cupere, B., Demirergi, A., Frame, S., Helmer, D., Martin, L., Peters, J., Pöllath, N., et al. (2014). Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS ONE, 9*(6), e99845. https://doi.org/10.1371/journal.pone.0099845

Baird, D., Carruthers, D., Fairbairn, A., & Pearson, J. A. (2011). Ritual in the landscape: Evidence from Pınarbaşı in the seventh-millennium Cal BC Konya Plain. *Antiquity, 85*(328), 380–394. https://doi.org/10.1017/S0003598X0006782X

Baird, D., Fairbairn, A., Jenkins, E., Martin, L., Middleton, C., Pearson, J. A., Asouti, E., Edwards, Y., Kabukcu, C., Mustafaoğlu, G., Russell, N., Bar-Yosef, O., Jacobsen, G., Wu, X., Baker, A., & Elliot, S. (2018). Agricultural origins on the Anatolian plateau. *Proceedings of the National Academy of Sciences, 115*(14), E3077–E3086. https://doi.org/10.1073/pnas.1800163115

Benaglia, T., Chauveau, D., Hunter, D., & Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software, 32*(6), 1–29 http://www.jstatsoft.org/v32/i06/

Breslawski, R. (2023). Minimum animal units and the standardized count problem. *Journal of Archaeological Method and Theory, 30*(1), 268–309. https://doi.org/10.1007/s10816-022-09563-9

Buitenhuis, H. (2008). Ilıpınar: The faunal remains from the Late Neolithic and Early Chalcolithic levels. In E. Vila, L. Gourichon, A. M. Choyke, & H. Buitenhuis (Eds.), *Archaeozoology of the Near East VIII (Tome I, pp. 299-322)* Maison de l'Orient et de la Méditerranée.

Buitenhuis, H. (2013). *Ilıpınar zooarchaeology*. Open. *Context.* Retrieved March 17, 2023 from http://opencontext.org/projects/D297CD29-50CA-4B2C-4A07-498ADF3AF487

Çakırlar, C. (2013). Rethinking Neolithic subsistence at the gateway to Europe wtih new archaeozoological evidence from Istanbul. In M. Groot, D. Lentjes, & J. Zeiler (Eds.), *Barely Surviving or More Than Enough?: The Environmental Archaeology of Subsistence, Specialisation and Surplus Food Production* (pp. 59–80). Sidestone Press.

Canty, A., & Ripley, B. (2021). *Boot: Bootstrap r (s-Plus) functions*. In *The Comprehensive R Archive Network* Retrieved March 17, 2023 from https://cran.r-project.org/web/packages/boot/index.html

Carruthers, D. (2005). Hunting and herding in Central Anatolian prehistory: The sites at Pınarbaşı. In H. Buitenhuis, A. M. Choyke, L. Martin, L. Bartosiewicz, & M. Mashkour (Eds.), *Archaeozoology of the Near East VI* (pp. 85–95). ARC.

Carruthers, D. (2006). *Pınarbaşı 1994: Animal nones*. Open Context. Retreived March 17, 2023 from https://opencontext.org/projects/1677643a-15d0-1b2c-3a35-37f04c765387

Microsoft Corporation, & Weston, S. (2022). *doParallel: Foreach parallel adaptor for the 'parallel' package*. The Comprehensive R Archive Network. Retrieved March 17, 2023 from https://CRAN.R-project.org/package=doParallel

Corti, P., & Shackleton, D. (2002). Relationship between predation-risk factors and sexual segregation in Dall's sheep (*Ovis dalli dalli*). *Canadian Journal of Zoology, 80*(12), 2108–2117. https://doi.org/10.1139/z02-207

Crema, E. (2012). Modelling temporal uncertainty in archaeological analysis. *Journal of Archaeological Method and Theory, 19*(3), 440–461. https://doi.org/10.1007/s10816-011-9122-3

Dahl, G., & Hjort, A. (1976). *Having herds: Pastoral herd growth and household economy*. University of Stokholm.

Daly, K. G., Mattiangeli, V., Hare, A. J., Davoudi, H., Fathi, H., Doost, S. B., Amiri, S., Khazaeli, R., Decruyenaere, D., Nokandeh, J., Richter, T., Darabi, H., Mortensen, P., Pantos, A., Yeomans, L., Bangsgaard, P., Mashkour, M., Zeder, M. A., & Bradley, D. G. (2021). Herded and hunted goat genomes from the dawn of domestication in the Zagros Mountains. *Proceedings of the National Academy of Sciences, 118*(25), e2100901118. https://doi.org/10.1073/pnas.2100901118

Davis, S. J. M. (1982). Climatic change and the advent of domestication: The succession of ruminant artiodactyls in the late Pleistocene-Holocene in the Israel Region. *Paléorient, 8*(2), 5–15.

Davis, S. J. M. (1996). Measurements of a group of adult female Shetland sheep skeletons from a single flock: A baseline for zooarchaeologists. *Journal of Archaeological Science, 23*, 593–612. https://doi.org/10.1006/jasc.1996.0056

Davis, S. J. M. (2000). The effect of castration and age on the development of the Shetland sheep skeleton and a metric comparison between bones of males, females and castrates. *Journal of Archaeological Science, 27*(5), 373–390. https://doi.org/10.1006/jasc.1999.0452

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge University Press.

De Cupere, B., Van Neer, W., Monchot, H., Rijmenants, E., Udrescu, M., & Waelkens, M. (2005). Ancient breeds of domestic fowl (*Gallus gallus* f. domestica) distinguished on the basis of traditional observations combined with mixture analysis. *Journal of Archaeological Science, 32*(11), 1587–1597. https://doi.org/10.1016/j.jas.2005.04.015

Degerbøl, M. (1970). The Urus (*Bos primigenius* Bojanus) and Neolithic domesticated cattle (*Bos taurus domesticus* Linné) in Denmark. *Det Kongelige Danske Videnskabernes Selskab Biologiske Skrifter, 17*(1), 5–177.

Dong, Z. (1997). Mixture analysis and its preliminary application in archaeology. *Journal of Archaeological Science, 24*, 141–161. https://doi.org/10.1006/jasc.1996.0100

Dowle, M., & Srinivasan, A. (2021). Data.table: Extension of 'data.frame'. *The Comprehensive R Archive Network*. Retrieved March 17, 2023 from https://CRAN.R-project.org/package=data.table

Evershed, R., Payne, S., Sherratt, A. G., Copley, M. S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Özdoğan, M., Özdoğan, A. E., Nieuwenhuyse, O., Akkermans, P. M. M. G., Bailey, D., Andeescu, R., Campbell, S., Farid, S., Hodder, I., Yalman, N., Özbaşaran, M., Bıçakcı, E., et al. (2008). Earliest date for milk use in the near east and southeastern Europe linked to cattle herding. *Nature, 455*(7212), 528–531. https://doi.org/10.1038/nature07180

Fernée, C., & Trimmis, K. (2021). Detecting variability: A study on the application of Bayesian multilevel modelling to archaeological data. Evidence from the Neolithic Adriatic and the Bronze Age Aegean. *Journal of Archaeological Science, 128*, 105346. https://doi.org/10.1016/j.jas.2021.105346

Gabry, J., & Češnovar, R. (2022). *Cmdstanr: R interface to 'CmdStan'*. Retrieved March 17, 2023 from https://mc-stan.org/cmdstanr/

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society, Series A, 182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Galik, A. (2013). *Barcın Höyük zooarchaeology*. Open Context. Retrieved March 17, 2023 from http://opencontext.org/projects/74749949-4FD4-4C3E-C830-5AA75703E08E

Gelman, A. (2006a). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics, 48*(3), 432–435. https://doi.org/10.1198/004017005000000661

Gelman, A. (2006b). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis, 1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2020). *Bayesian data analysis* (3rd ed.). CRC Press.

Gelman, A., Jakulin, A., Pittau, M., & Su, Y. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics, 2*(4), 1360–1383. https://doi.org/10.1214/08-aoas191

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–511. https://doi.org/10.1214/ss/1177011136

Gelman, A., Vehtari, A., Simpson, D., Margossian, C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P., & Modrák, M. (2020). Bayesian workflow. *arXiv*. Advance online publication. https://doi.org/10.48550/arXiv.2011.01808

Gerritsen, F., & Özbal, R. (2019). Barcin Höyük, a seventh millennium settlement in the eastern Marmara Region of Turkey. *Documenta Praehistorica, 46*, 58–67. https://doi.org/10.4312/dp.46.4

Gillis, R., Arbogast, R. M., Piningre, J. F., Debue, K., & Vigne, J.-D. (2015). Prediction models for age-at-death estimates for calves, using unfused epiphyses and diaphyses. *International Journal of Osteoarchaeology, 25*(6), 912–922. https://doi.org/10.1002/oa.2377

Gillis, R., Carrère, I., Saña Seguí, M., Radi, G., & Vigne, J.-D. (2014). Neonatal mortality, young calf slaughter and milk production during the early Neolithic of north western Mediterranean. *International Journal of Osteoarchaeology, 26*(2), 303–313. https://doi.org/10.1002/oa.2422

Gourichon, L., & Helmer, D. (2008). Etude de La Faune Néolithique de Menteşe. In J. Roodenberg & S. Roodenberg (Eds.), *Life and death in a prehistoric settlement in northwest Anatolia. The Ilipinar Excavations (Volume III, pp. 435-438)* Nederlands Instituut voor het Nabije Oosten.

Gourichon, L., & Helmer, D. (2013). *Faunal data from Neolithic Menteşe*. Open Context Retrieved March 17, 2023 from https://opencontext.org/projects/42ead4dbbaed-4a58-9b9b-7ec85266d2a9

Grau-Sologestoa, I., & Albarella, U. (2019). The 'long' sixteenth century: A key period of animal husbandry change in England. *Archaeological and Anthropological Sciences, 11*(6), 2781–2803. https://doi.org/10.1007/s12520-018-0723-6

Grigson, C. (1982). Sex and age determination of some bones and teeth of domestic cattle: A review of the literature. In B. Wilson, C. Grigson, & S. Payne (Eds.), *Ageing and Sexing Animal Bones from Archaeological Sites BAR British Series 109* (pp. 7–23). British Archaeological Reports.

Grigson, C. (1989). Size and sex: Evidence for the domestication of cattle in the Near East. In I. Hershkovitz (Ed.), *People and Culture in Change* (pp. 77–109). Archaeopress.

Hammer, Ø. (2022). *PAST: Paleontological statistics reference manual, Version 4.12b*. University of Oslo Natural History Museum Retrieved March 17, 2023 from https://www.nhm.uio.no/english/research/resources/past/

Helmer, D., Gourichon, L., Monchot, H., Peters, J., & Saña Segui, M. (2005). Identifying early domestic cattle from pre-pottery Neolithic sites on the middle Euphrates using sexual dimorphism. In J.-D., Vigne, J. Peters, & D. Helmer (Eds.) *First Steps of Animal Domestication: New Archaeozoological Approaches* (pp. 86–95). Oxbow Books.

Hill, M. E., Hill, M. G., & Widga, C. (2008). Late quaternary *bison* diminution on the Great Plains of North America: Evaluating the role of human hunting versus climate change. *Quaternary Science Reviews, 27*(17-18), 1752–1771. https://doi.org/10.1016/j.quascirev.2008.07.002

Hongo, H., & Meadow, R. (2000). Faunal remains from prepottery Neolithic levels at Çayönü, Southeastern Turkey: A preliminary report focusing on pigs (*Sus* sp.). In M. Mashkour, A. M. Choyke, H. Buitenhuis, & F. Poplin (Eds.), *Archaeozoology of the Near East IV A* (pp. 121–140). ARC.

Hongo, H., Pearson, J. A., Öksüz, B., & İlgezdi, G. (2009). The process of ungulate domestication at Çayönü, southeastern Turkey: A multidisciplinary approach focusing on *Bos* sp. and *Cervus elaphus*. *Anthropozoologica, 44*(1), 63–78.

Kabukcu, C. (2017). Woodland Vegetation History and Human Impacts in South-Central Anatolia 16,000-6500 cal BP: Anthracological results from five prehistoric sites in the Konya Plain. *Quaternary Science Reviews, 176*, 85–100. https://doi.org/10.1016/j.quascirev.2017.10.001

Karul, N. (2019). Early farmers in northwestern Anatolia in the seventh millennium. In A. Marciniak (Ed.), *Concluding the Neolithic: The Near East in the Second Half of the Seventh Millennium BC* (pp. 269–286). Lockwood Press.

Kassambara, A. (2020). *ggpubr: 'ggplot2' based publication ready plots*. The Comprehensive R Archive Network. Retrieved March 17, 2023 from https://CRAN.R-project.org/package=ggpubr

Kay, M. (2022). *ggdist: Visualizations of distributions and uncertainty*. The Comprehensive R Archive Network. Retrieved March 17, 2023 from https://cran.r-project.org/web/packages/ggdist/index.html

Koch, P. (1986). Clinal geographic variation in mammals: Implications for the study of chronoclines. *Paleobiology, 12*(3), 269–281. https://doi.org/10.1017/S0094837300013774

Lebenzon, R., & Munro, N. D. (2022). Body size variation in a modern collection of mountain gazelle (*Gazella gazella*) skeletons. *Journal of Archaeological Science: Reports, 41*, 103285. https://doi.org/10.1016/j.jasrep.2021.103285

Lyman, L. (2008). *Quantitative paleozoology*. Cambridge University Press.

Lyman, R. L. (1987). On the analysis of vertebrate mortality profiles: sample size, mortality type, and hunting pressure. *American Antiquity, 52*(1), 125–142. https://doi.org/10.2307/281064

Madgwick, R., & Mulville, J. (2015). Reconstructing depositional histories through bone taphonomy: Extending the potential of faunal data. *Journal of Archaeological Science, 53*, 255–263. https://doi.org/10.1016/j.jas.2014.10.015

Manning, K., Timpson, A., Shennan, S., & Crema, E. (2015). Size reduction in early European domestic cattle relates to intensification of Neolithic herding strategies. *PLoS ONE, 10*(12), e0141873. https://doi.org/10.1371/journal.pone.0141873

Marin, J.-M., Mengersen, K., & Robert, C. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics, 25*, 459–509. https://doi.org/10.1016/S0169-7161(05)25016-2

Martín, P., & García-González, R. (2015). Identifying Sheep (*Ovis aries*) Fetal remains in archaeological contexts. *Journal of Archaeological Science, 64*, 77–87. https://doi.org/10.1016/j.jas.2015.10.003

Martín, P., García-González, R., Nadal, J., & Maria Vergès, J. (2015). Perinatal ovicaprine remains and evidence of shepherding activities in early Holocene enclosure caves: El Mirador (Sierra de Atapuerca, Spain). *Quaternary International, 414*, 316–329. https://doi.org/10.1016/j.quaint.2015.08.024

Marwick, B. (2017). Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. *Journal of Archaeological Method and Theory, 24*(2), 424–450. https://doi.org/10.1007/s10816-015-9272-9

Marwick, B., & Pilaar Birch, S. (2018). A standard for the scholarly citation of archaeological data as an incentive to data sharing. *Advances in Archaeological Practice, 6*(2), 125–143. https://doi.org/10.1017/aap.2018.3

McCarthy, M., & Masters, P. (2005). Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology, 42*(6), 1012–1019. https://doi.org/10.1111/j.1365-2664.2005.01101.x

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

McGrory, S., Svensson, E. M., Götherström, A., Mulville, J., Powell, A. J., Collins, M. J., & O'Connor, T. P. (2012). A novel method for integrated age and sex determination from archaeological cattle mandibles. *Journal of Archaeological Science, 39*(10), 3324–3330. https://doi.org/10.1016/j.jas.2012.05.021

Meadow, R. (1999). The use of size index scaling techniques for research on archaeozoological collections from the Middle East. In C. Becker, H. Manhart, J. Peters, & J. Schibler (Eds.), *Historia Animalium Ex Ossibus: Beiträge Zur Paläoanatomie, Archäologie, Ägyptologie, Ethnologie Und Geschichte Der Tiermedizin* (pp. 285–300). Verlag Marie.

Meier, J. (2020). The contextual taphonomy of middens at Neolithic Kfar HaHoresh. *Journal of Archaeological Science: Reports, 33*, 102531. https://doi.org/10.1016/j.jasrep.2020.102531

Middleton, C. (2018). The beginning of herding and animal management: The early development of caprine herding on the Konya Plain, Central Anatolia. *Anatolian Studies, 68*, 1–31. https://doi.org/10.1017/s0066154618000017

Millard, A. R. (2006). A Bayesian approach to ageing sheep/goats from toothwear. In D. Ruscillo (Ed.), *Recent Advances in Ageing and Sexing Animal Bones* (pp. 145–154). Oxbow Books.

Monchot, H., & Léchelle, J. (2002). Statistical nonparametric methods for the study of fossil populations. *Paleobiology, 28*(1), 55–69. https://doi.org/10.1666/0094-8373(2002)028<0055:SNMFTS>2.0.CO;2

Monchot, H., Mashkour, M., & Vigne, J.-D. (2005). Kernel smoothing and mixture analyses for the determination of the sex ratios at death, at the beginning of the domestication of ungulates. In J.-D. Vigne, J. Peters, & D. Helmer (Eds.), *First steps of animal domestication: New archaeozoological approaches* (pp. 55–60). Oxbow Books.

Munro, N. D., Bar-Oz, G., & Hill, A. C. (2011). An exploration of character traits and linear measurements for sexing mountain gazelle (*Gazella gazella*) skeletons. *Journal of Archaeological Science, 38*(6), 1253–1265. https://doi.org/10.1016/j.jas.2011.01.001

Munro, N. D., Roxanne, L., & Sapir-Hen, L. (2022). Revisiting late Pleistocene-early Holocene mountain gazelle (*Gazella gazella*) body size change in the southern Levant: A case for anthropogenic impact. *PLoS ONE, 17*(8), e0273024. https://doi.org/10.1371/journal.pone.0273024

Otárola-Castillo, E., & Torquato, M. G. (2018). Bayesian statistics in archaeology. *Annual Review of Anthropology, 47*(1), 435–453. https://doi.org/10.1146/annurev-anthro-102317-045834

Otárola-Castillo, E., Torquato, M. G., Wolfhagen, J., Hill, M. E., & Buck, C. E. (2022). Beyond chronology, using Bayesian inference to evaluate hypotheses in archaeology. *Advances in Archaeological Practice, 10*(4), 397–413. https://doi.org/10.1017/aap.2022.10

Özdoğan, M. (2011). Archaeological evidence on the westward expansion of farming communities from eastern Anatolia to the Aegean and the Balkans. *Current Anthropology, 52*(S4), S415–S430. https://doi.org/10.1086/658895

Özdoğan, M. (2019). Early farmers in northwestern Turkey: What is new? In A. Marciniak (Ed.), *Concluding the Neolithic: The Near East in the Second Half of the Seventh Millennium BC* (pp. 307–328). Lockwood Press.

Payne, S. (1969). A metrical distinction between sheep and goat metacarpals. In P. Ucko & G. Dimbleby (Eds.), *The Domestication and Exploitation of Plants and Animals* (pp. 295–305). Duckworth.

Payne, S. (1973). Kill-off patterns in sheep and goats: The mandibles from Aşvan Kale. *Anatolian Studies, 23*, 281–303. https://doi.org/10.2307/3642547

Payne, S., & Bull, G. (1988). Components of variation in measurements of pig bones and teeth, and the use of measurements to distinguish wild from domestic pigs. *Archaeozoologia, 2*(1), 27–66.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal, 10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Peters, J., von den Driesch, A., & Helmer, D. (2005). The Upper Euphrates-Tigris basin: Cradle of agro-pastoralism? In J.-D. Vigne, J. Peters, & D. Helmer (Eds.), *The First Steps of Animal Domestication* (pp. 96–124). Oxbow Books.

Popkin, P. R. W., Baker, P., Worley, F., Payne, S., & Hammon, A. (2012). The sheep project (1): Determining skeletal growth, timing of epiphyseal fusion and morphometric variation in unimproved Shetland sheep of known age, sex, castration status and nutrition. *Journal of Archaeological Science, 39*(6), 1775–1792. https://doi.org/10.1016/j.jas.2012.01.018

Post, D. M., Armbrust, T. S., Horne, E. A., & Goheen, J. R. (2001). Sexual segregation results in differences in content and quality of bison (*Bos bison*) diets. *Journal of Mammalogy, 82*(2), 407–413. https://doi.org/10.1644/1545-1542(2001)082<0407:SSRIDI>2.0.CO;2

Pozo, J. M., Trentacoste, A., Nieto-Espinet, A., Guimarães Chiarelli, S., & Valenzuela-Lamas, S. (2022). *Zoolog: Zooarchaeological analysis with log-ratios*. In *The Comprehensive R Archive Network* Retrieved March 17, 2023 from https://cran.r-project.org/web/packages/zoolog/index.html

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ram, K., & Marwick, B. (2018). Building toward a future where reproducible, open science is the norm. In J. Kitzes, D. Turek, & F. Deniz (Eds.), *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences* (pp. 69–78). University of California Press.

Redding, R. W. (1984). Theoretical determinants of a Herder's decisions: Modeling Variation in the sheep/goat ratio. In J. Clutton-Brock & C. Grigson (Eds.), *Animals and Archaeology: 3. Early Herders and their Flocks BAR International Series 202* (pp. 223–241). British Archaeological Reports.

Roodenberg, J. (2012a). Change in food production and its impact on an early 6th millennium community in northwest Anatolia. *The example of Ilıpınar. Praehistorische Zeitschrift, 87*(2), 223–235. https://doi.org/10.1515/pz-2012-0015

Roodenberg, J. (2012b). Ilipinar: A Neolithic settlement in the eastern Marmara Region. In G. McMahon & S. Steadman (Eds.), *The Oxford Handbook of Ancient Anatolia (10,000-323 BCE)* (pp. 950–967). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195376142.013.0044

Roodenberg, J., van As, A., Jacobs, L., & Wijnen, M. H. (2003). Early settlement in the plain of Yenişehir (NW Anatolia): The basal occupation layers at Menteşe. *Anatolica, 29*(1), 17–59. https://doi.org/10.2143/ANA.29.0.2015511

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

RStudio Team. (2022). *RStudio: Integrated development environment for R*. RStudio.

Ruckstuhl, K. E. (2007). Sexual segregation in vertebrates: Proximate and ultimate causes. *Integrative and Comparative Biology, 47*(2), 245–257. https://doi.org/10.1093/icb/icm030

Ruckstuhl, K. E., & Neuhaus, P. (2002). Sexual segregation in ungulates: A comparative test of three hypotheses. *Biological Reviews, 77*(1), 77–96. https://doi.org/10.1017/s1464793101005814

Russell, N., & Griffitts, J. (2013). Çatalhöyük worked bones: South and 4040 areas. In I. Hodder (Ed.), *Substantive Technologies at Çatalhöyük: Reports from the 2000-2008 Seasons* (pp. 277–306). Cotsen Institute of Archaeology Press.

Saïd, S., Tolon, V., Brandt, S., & Baubet, E. (2011). Sex effect on habitat selection in response to hunting disturbance: The study of wild boar. *European Journal of Wildlife Research, 58*(1), 107–115. https://doi.org/10.1007/s10344-011-0548-4

Sasson, A., & Arter, S. (2020). Earliest utilization of chicken in Upper California: The zooarchaeology of avian remains from the San Diego Royal Presidio. *American Antiquity, 85*(3), 516–534. https://doi.org/10.1017/aaq.2020.27

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021

Slowikowski, K. (2021). *ggrepel: Automatically position non-overlapping text labels with 'ggplot2'*. The Comprehensive R Archive Network. Retrieved May 18, 2023, from https://CRAN.R-project.org/package=ggrepel

Smith, A. F. M., & Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician, 46*(2), 84–88. https://doi.org/10.1080/00031305.1992.10475856

South, A. (2017a). *rnaturalearth: World map data from natural earth*. The Comprehensive R Archive Network. Retrieved March 17, 2023 from https://CRAN.R-project.org/package=rnaturalearth

South, A. (2017b). *rnaturalearthdata: World vector map data from natural earth used in 'rnaturalearth'*. The Comprehensive R Archive Network. Retrieved March 17, 2023 from https://CRAN.R-project.org/package=rnaturalearthdata

Speth, J. (1983). *Bison kills and bone counts: Decision making by ancient hunters*. University of Chicago Press.

Speth, J. (2013). Thoughts about hunting: Some things we know and some things we don't know. *Quaternary International, 297*, 176–185. https://doi.org/10.1016/j.quaint.2012.12.005

Stan Development Team. (2021). *RStan: The R interface to stan*. Retrieved March 17, 2023 from https://mc-stan.org/users/interfaces/rstan

Stan Development Team. (2022). *Stan modeling language user's guide and reference manual*. Retrieved March 17, 2023 from https://mc-stan.org/docs/stan-users-guide/index.html

Stiner, M. (1990). The use of mortality patterns in archaeological studies of hominid predatory adaptations. *Journal of Anthropological Archaeology, 9*, 305–351. https://doi.org/10.1016/0278-4165(90)90010-B

Stiner, M., Munro, N. D., Buitenhuis, H., Duru, G., & Özbaşaran, M. (2022). An endemic pathway to sheep and goat domestication at Asikli Hoyuk (Central Anatolia, Turkey). *Proceedings of the National Academy of Sciences, 119*(4), e2110930119. https://doi.org/10.1073/pnas.2110930119

Symmons, R. (2005). New density data for unfused and fused sheep bones, and a preliminary discussion on the modelling of taphonomic bias in archaeofaunal age profiles. *Journal of Archaeological Science, 32*(11), 1691–1698. https://doi.org/10.1016/j.jas.2005.05.011

Tchernov, E., & Horwitz, L. K. (1991). Body size diminution under domestication: Unconscious selection in primeval domesticates. *Journal of Anthropological Archaeology, 10*, 54–75. https://doi.org/10.1016/0278-4165(91)90021-O

Thissen, L., Özbal, H., Bıyık, A. T., Gerritsen, F., & Özbal, R. (2010). The land of milk? Approaching dietary preferences of Late Neolithic communities in NW Anatolia. *Leiden Journal of Pottery Studies, 26*, 157–172.

Todd, L. C. (1983). *The Horner site: Taphonomy of an early Holocene bison bonebed*. Ph.D. Dissertation. The University of New Mexico.

Trentacoste, A., Nieto-Espinet, A., Guimaraes, S., Wilkens, B., Petrucci, G., & Valenzuela-Lamas, S. (2021). New trajectories or accelerating change? Zooarchaeological evidence for Roman transformation of animal husbandry in northern Italy. *Archaeological and Anthropological Sciences, 13*(1), 25. https://doi.org/10.1007/s12520-020-01251-7

Twiss, K. (2019). *The archaeology of food*. Cambridge University Press.

Twiss, K., & Russell, N. (2009). Taking the bull by the horns: Ideology, masculinity, and cattle horns at Çatalhöyük (Turkey). *Paléorient, 35*(2), 19–32.

Uerpmann, M., & Uerpmann, H.-P. (1994). Animal bones. In F. Hojlund & H. Hellmuth (Eds.), *Qala'at Al-Bahrain* (Volume 1, pp. 417–454). Aarhus University Press.

Urbanek, S., & Horner, J. (2022). *Cairo: R graphics device using cairo graphics library for creating high-quality bitmap (PNG, JPE, TIFF), vector (PDF, SVG, PostScript) and DISPLAY (X11 and Win32) output*. The Comprehensive R Archive Network. Retrieved May 18, 2023, from https://CRAN.R-project.org/package=Cairo

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with Discussion). *Bayesian Analysis, 16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Vigne, J.-D. (2011). Le Mouton (*Ovis aries*). In J. Guilaine, F. Briois, & J.-D. Vigne (Eds.), *Shillourokambos: Un Établissement Néolithique Pré-Céramique à Chypre* (pp. 1021–1038). Editions Errance.

von den Driesch, A. (1976). *A guide to the measurement of animal bones from archaeological sites*. In *Peabody Museum Bulletin 1*. University.

Watkins, T. (1996). Excavations at Pınarbaşı: The early stages. In I. Hodder (Ed.), *On the Surface: Çatalhöyük 1993-1995* (pp. 47–57). McDonald Institute for Archaeological Research.

Weinstock, J. (2006). environment, body size and sexual dimorphism in late glacial reindeer. In D. Ruscillo (Ed.), *Recent Advances in Ageing and Sexing Animal Bones* (pp. 247–253). Oxbow Books.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham, H., & Bryan, J. (2022). *readxl: Read Excel Files*. The comprehensive R archive network. Retrieved May 18, 2023, from https://CRAN.R-project.org/package=readxl

Widga, C. (2006). Niche variability in late Holocene bison: A perspective from Big Bone Lick. *KY. Journal of Archaeological Science, 33*(9), 1237–1255. https://doi.org/10.1016/j.jas.2005.12.011

Wolfhagen, J. (2020). Re-examining the use of the LSI technique in zooarchaeology. *Journal of Archaeological Science, 123*, 105254. https://doi.org/10.1016/j.jas.2020.105254

Wolfhagen, J., & Price, M. D. (2017). A probabilistic model for distinguishing between sheep and goat postcranial remains. *Journal of Archaeological Science: Reports, 12*, 625–631. https://doi.org/10.1016/j.jasrep.2017.02.022

Wolfhagen, J., Twiss, K., Mulville, J., & Demirergi, A. (2021). Examining caprine management and cattle domestication through biometric analyses at Çatalhöyük East (north and south areas). In I. Hodder (Ed.), *Peopling the Landscape of Çatalhöyük: Reports from the 2009-2017 Seasons* (pp. 181–198). Cotsen Institute of Archaeology Press.

Wolverton, S. (2008). Harvest pressure and environmental carrying capacity: An ordinal-scale model of effects on ungulate prey. *American Antiquity, 73*(2), 179–199. https://doi.org/10.1017/S0002731600042244

Wolverton, S., Dombrosky, J., & Lyman, R. L. (2016). Practical significance: ordinal scale data and effect size in zooarchaeology. *International Journal of Osteoarchaeology, 26*(2), 255–265. https://doi.org/10.1002/oa.2416

Wright, E., & Viner-Daniels, S. (2015). Geographical variation in the size and shape of the European aurochs (*Bos primigenius*). *Journal of Archaeological Science, 54*, 8–22. https://doi.org/10.1016/j.jas.2014.11.021

Xie, Y. (2015). *Dynamic documents with R and knitr* (Second Edition). Chapman and Hall/CRC

Xie, Y. (2022). *knitr: A general-purpose package for dynamic report generation in r*. Retrieved May 18, 2023, from https://yihui.org/knitr/

Zeder, M., & Hesse, B. (2000). The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago. *Science, 287*(5461), 2254–2257 https://www.science.org/doi/10.1126/science.287.5461.2254

Zeder, M., & Lemoine, X. (2020). A method for constructing demographic profiles in Sus scrofa using logarithm size index scaling. *Journal of Archaeological Science, 116*, 105115. https://doi.org/10.1016/j.jas.2020.105115

Zhu, H. (2021). *kableExtra: Construct Complex Table with 'kable' and pipe syntax*. The Comprehensive R Archive Network. Retrieved May 18, 2023, from https://CRAN.R-project.org/package=kableExtra

Zohary, D., Tchernov, E., & Horwitz, L. K. (1998). The role of unconscious selection in the domestication of sheep and goats. *Journal of Zoology, 245*(2), 129–135. https://doi.org/10.1111/j.1469-7998.1998.tb00082.x