Journal: Journal of Archaeological Science

Modelling caprine age-at-death profiles using the Gamma distribution

Adrian Timpson[1], Rosalind E. Gillis[2,4], Katie Manning[3], Mark G. Thomas[1,5]

1. University College London, Gower St, London WC1E 6BT
2. Institut für Ur- und Frühgeschichte, Christian-Albrechts-Universität, Johanna-Mestorf-Straße 2-6, D - 24098 Kiel
3. King's College London, Strand, London WC2R 2LS
4. CNRS – Muséum National d'Histoire Naturelle – Sorbonne Universités, Archéozoologie, Archéobotanique: Sociétés, Pratiques et Environnement, (UMR 7209), CP56, 55 rue Buffon, F-75005 PARIS, France
5. UCL Genetics Institute, University College London, Gower St, London WC1E 6BT, United Kingdom

Abstract

Age-at-death profiles constructed from archaeozoological data have been used for decades to infer the goals of prehistoric herd management strategies. Several 'ideal' profiles have been proposed as models for the optimal kill-off profiles that represent specific husbandry strategies, such as maximising milk or meat yields, which can then be compared to archaeological profiles. We evaluate the goodness of fit of ten caprine archaeological age-at-death profiles to five published idealised profiles, whilst properly accounting for sampling error and data where the age classes of observations are uncertain. We statistically reject all tested idealised profiles as plausible models to explain the data, and instead propose that a Gamma distribution provides a simpler and better general model to represent possible herd management strategies. Furthermore, we show that archaeological profiles can be summarised well using Gamma parameters, which allow multiple datasets (and models) to be easily compared and graphically represented together with minimal information loss, thus allowing clearer inferences to be drawn. Finally, we calculate likelihood distributions of the Gamma parameters, which provide confidence intervals that fully account for the uncertainties from small sample sizes and uncertain age classes. We have developed an R package 'GammaModel' to enable users to apply these tools to any age-at-death count data.

Highlights

- Existing 'ideal' kill-off profiles are implausible models of husbandry practices during the early Neolithic
- The Gamma distribution provides a simpler model that fits data better than the current nine-age class models.
- Summarising kill-off profiles with the Gamma distribution minimises compression and allows clearer graphical comparisons.
- We calculate kill-off profile model likelihoods whilst accounting for both sampling error, and uncertain age classes.
- R package 'GammaModel' developed to provide tools to perform these analyses on age-at-death count data.

Keywords

GammaModel; Gamma distribution; caprines; age-at-death profiles; MCMC; model comparison; Neolithic

Introduction

Since their initial domestication, caprines (Sheep, *Ovis aries;* Goat, *Capra hircus)* have formed an important component of present-day and prehistoric subsistence practices (Arbuckle, et al., 2014, Helmer, et al., 2007). Direct morphological evidence for caprine domestication has been identified at Pre-Pottery Neolithic (PPN) sites in the northern Levant and Zagros regions, dating to the 9th millennium BC (Peters, et al., 1999, Zeder, 1999). Farming arrived in Europe via two routes, a continental route associated with cattle herding, and a maritime route associated strongly with caprine husbandry in southern Europe (Çilingiroğlu, 2009, Coward, et al., 2008, Guilaine, 2001, Perlès, 2005). Large-scale analysis of age-at-death data and organic residues has demonstrated caprines were managed for dairy husbandry, particularly those belonging to Impressa/Cardial ware (ICW) sites (Debono Spiteri, et al., 2016). The role of caprines and their products (meat, milk, hair) during the Neolithic is often overlooked due to the symbolic and economic value of cattle, particularly for Central and Northern European cultures such as the Linearbandkeramik (LBK) (Gillis, et al., 2017, Manning, et al., 2013). In these regions caprines had a secondary role and have been proposed to have been primarily a source of meat (Marciniak, 2013).

Mortality profiles based on age-at-death data determined from dental development, replacement, and wear stages are commonly used by archaeozoologists to infer the goals of prehistoric subsistence strategies (Helmer, et al., 2005, Higham, 1969, Payne, 1973). The recovery of different age classes is governed by the mortality profile of the herd, taphonomic and preservation conditions, and excavation and sampling protocols. Payne (1973) and Redding (1981) proposed model profiles that represent the maximisation of specific product yields or herd sustainability. However the relevance of these models for prehistoric practices has been challenged (Halstead, 1998) since Payne's profiles were derived from observations of modern Turkish market herders and Redding's models are based on data from modern Middle Eastern caprines. Nevertheless, these mortality models have been widely used and cited to identify the targets of husbandry strategies from archaeological data (e.g. Helmer, et al., 2007). Here we test if these idealised profiles are plausible models to explain ten caprine archaeological profiles from early European Neolithic sites. Specifically, we aim to establish the degree to which such widely used kill-off profiles are appropriate models of prehistoric herding strategies.

Idealised age at death profiles

Payne (1973) theorised three ideal profiles, one each targeting meat, milk and wool production. These were presented graphically as survivorship curves, from which it is trivial to calculate the probability of death in each age class. Unfortunately, Payne did not provide a numerical summary, but our careful measurements of the original plots are summarised in Table 1:

Class	A	B	C	D	E	F	G	H	I
Age	0 to 2 months	2 to 6 months	6 to 12 months	1 to 2 years	2 to 3 years	3 to 4 years	4 to 6 years	6 to 8 years	8 + years
Meat	0.15	0.10	0.05	0.20	0.20	0.05	0.05	0.10	0.10
Milk	0.52	0.05	0.03	0.04	0.07	0.05	0.04	0.10	0.10
Wool	0.15	0.10	0.05	0.05	0.07	0.06	0.06	0.26	0.20

Table 1: summary of Payne's 1973 Figs 1:3. Caprine probability of death, in each age class.

Redding (1981) proposed two further ideal profiles targeting maximum energy offtake, and maximum herd security. Redding provided numeric probabilities in age classes, the durations of which differ slightly from Payne's (Classes A and B are aggregated), as summarised in Table 2.

class	AB	C	D	E	F	G	H	I
age	0 to 6 months	6 to 12 months	1 to 2 years	2 to 3 years	3 to 4 years	4 to 6 years	6 to 8 years	8 + years
Energy	0.096	0.128	0.30	0.226	0.011	0.078	0.098	0.063
Security	0.096	0.259	0.265	0.13	0.011	0.078	0.098	0.063

Table 2: summary of Redding's 1981 Table X2 and X3. Caprine probability of death, in each age class.

Helmer and Vigne (2007) (H&V) introduced two additional profiles for meat and milk (defined as type B to distinguish from Payne's meat and milk profiles) which were observed kill profiles from two archaeological sites. These profiles were proposed as representations of mixed economies (Milk B) and where tender meat is the focus (Meat B). As such, these represented 'typical data' that might be expected, rather than new models.

Marom and Bar-Oz (2009) (M&B-O) evaluated Payne and Redding's models, but summarised the H&V type A models as distinct from Payne's models, and included the H&V type B profiles as additional models. Unfortunately, M&B-O's numerical summary bears little resemblance to the original source data. For example, H&V's meat B histogram (2007 Fig 5 b) has a height of 7% in class A (or approximately 1% of the total area), whilst M&B-O reports this as 14% (from 100% to 86% between 0 and 2 months). Payne's wool profile (1973 Fig 3) shows a 5% loss in class C from 75% at 6 months to 70% at 12 months, whilst M&B-O reports this as a 10% loss from 75% to 65%. M&B-O reports Redding's data with a percentage reduction in class A, but no reduction in class B. In fact, Redding did not provide this precision and combined classes A and B. Therefore, when summarising the numeric values that describe the five published models, we discard M&B-O's published summary in favour of our own summary of Payne and Redding's source information (Tables 1 and 2, respectively).

Archaeological data

We used combined sheep and goat remains from ten Neolithic sites from Central Europe, the Northern European Plain and the northwestern Mediterranean. Polgár-Piócási-dűlő (PPI) and Polgár-Ferenci-hát (PFE) come from Hungary and belong to the Alföld Linear Pottery (ALP) dated 5650-4800 BCE. Polgár-Csaszhalom-dulo (PCS) belongs to the Hungarian Late Neolithic and dated 4840-4560 cal. BCE. Mold (MOLD), Dillingen-Steinheim (WIK) and Těšetice-Kyjovice (TES) belong the typical LBK culture of central Europe, dated 5500 to 4900 BCE. The remaining sites, Trasano (TRA1/TRA2), Font Juvénal (FON1) and La Draga (LAD) are from the northwestern Mediterranean. These belong collectively to the Early Neolithic ICW cultures, in which ceramics are decorated with impressed designs often using the *cardial* shell. Trasano (Impressa, Italy) dated 6400 to 5320 BCE and Font Juvénal (Cardial, France) and La Draga (Cardial, Spain) both dated 5600 to 4800 BCE. All the sites are open-air settlements apart from Font Juvénal, which is a rock shelter site and assumed to be a spring/summer birthing station. The age-at-death was determined from teeth using eruption, replacement (Helmer, et al., 2007, Payne, 1987) and dental wear stages (Grant, 1982) for the fourth lower deciduous pre-molar (D4), first molar (M1), second molar (M2) and third molar (M3). Crown height, distance anterior to posterior (DAP) and distance transversal (DT) (Ducos, 1968) for the molars were also recorded and used to age dental remains. Occlusal eruption stage, wear stage and crown height index provide a less precise estimate of age-at-death given the inter-species and inter-individual variation in rates of eruption, wear and growth. Broken roots and long wear stages cause greater uncertainty of the age-at-death, resulting in a tooth being assigned to a multiple-age class. Data from these sites are summarised in Table 3.

Code	ABCD	A	AB	B	BC	C	CD	D	BCD	DEF	DEFG	EF	G	EFG	HI	GHI	EFGHI	DEFGHI
FON1	2	5	2	0	2	6	4	1	0	0	0	2	1	6	0	0	2	2
TRA1	0	1	0	1	3	10	1	7	0	0	0	12	10	1	0	0	1	0
MOLD	1	0	0	0	0	6	5	8	0	2	0	12	6	0	3	9	3	0
TES	2	0	0	3	3	15	2	7	1	0	0	18	2	0	4	0	0	0
TRA2	0	0	0	1	1	9	12	6	0	7	0	5	0	18	0	5	6	0
PPI	0	0	0	3	2	7	0	4	0	0	0	29	25	0	0	0	3	0
WIK	0	0	0	3	0	8	1	11	0	1	0	32	18	0	1	0	0	0
PFE	0	0	0	12	2	26	0	6	0	0	0	11	18	4	0	3	0	0
PCS	0	0	0	0	0	41	0	20	0	0	0	20	6	0	10	0	6	0
LAD	1	0	1	2	1	8	5	33	0	26	2	28	14	1	0	0	0	0

Table 3: summary of archaeological data showing teeth counts in age Payne's age classes A to I. Note the variation in uncertainty, such that some observations can only be assigned to a broader age range comprising several age classes.

Previous statistical approaches

The comparision of archaeological mortality profiles and idealised models has generally been based on a simple visual examination or on the comparison of the dominant age classes (Helmer, 1991). Occasionally statistical approaches have been employed to evaluate the (dis)similarity between archaeological profiles and /or model profiles. Greenfield and Arnold (2015) employed Mann-Whitney U-tests after transforming counts into proportions in each age class. However, this is problematic since these derived proportions become compositional, therefore age classes cannot be tested separately. Marom and Bar-Oz (2009) presented ideal profiles in the form of survivorship curves (percentage of a theoretical cohort still alive) and used Kolmogorov-Smirnov (KS) tests for continuous variables to evaluate if these curves are distinguishable. However, this is problematic for several reasons. Firstly, M&B-O correctly note that the archaeological counts in each age class are discrete variables, but argue that by presenting as survivorship curves the data becomes continuous. Clearly this is not the case, the survivorship curves are merely the cumulative counts in each age class, scaled such that the total sample size equals 100%. Secondly, M&B-O's conclusion that '*Many theoretical survivorship curves are not statistically different"* suggests a muddle between data and a model. A KS test can evaluate if two datasets are significantly different, or if a dataset can be reasonably explained by a model, and its power to detect this is largely a matter of the sample size of the data. The theoretical survivorship curves are models, not data, and therefore do not have sample sizes. As such, a KS test is not applicable since there are no data to test. Instead M&B-O used the model percentages to act a sample size of 100. Finally, M&B-O proposed aggregating data into fewer broader age classes, but this results in a loss of information content, and therefore cannot logically offer any improvements.

Gerbault, et al. (2016) and Gillis, et al. (2017) focused on estimating the uncertainty in the death probabilities in each age class resulting from small sample sizes by resampling the data from a Dirichlet distribution then using correspondence analysis to compress the information into two dimensions. Whilst the Dirichlet distribution provides a valid means of assessing uncertainty in population frequency estimates due to sampling error, and correspondence analysis provides a useful graphical representation, this remains a descriptive approach and does not provide an objective statistical test of any specific hypothesis. Nor does this approach account for multi-class assignments.

Calculating likelihoods exactly

Payne (1973) first proposed the discrete age classes A to I which represent a practical compromise between the resolution at which it is possible to confidently assign ages to archaeological teeth, and the resolution at which the data can then inform on competing hypotheses. Since a particular kill-off strategy model is defined by the chance of death in each of these age classes, and provided archaeological count data is available in the same age classes, the likelihood of the model (the probability of the data given the model) can be calculated exactly using the multinomial distribution. This can be thought of as a 9-sided die where the chance of death in each age class (the model parameters) are the chance of the die landing on each face, and the total observed counts are the number of rolls.

This relationship between model parameters and observed counts requires some simplifying assumptions. Firstly, we assume no differential taphonomy between age classes, which is generally justified by the use of tooth data, rather than bone fusion data. Secondly, we define 'slaughter' as any cause of death that could be represented in our archaeological data (for example natural deaths of new-borns would likely be included in the archaeological record whilst the predation and removal of a lamb would not). This is partly because we cannot determine a cause of death for our teeth, and partly because other causes of death are influenced by husbandry practices; for example providing shelter to avoid new-born deaths from predation or poor weather. Finally, we assume that the number of teeth in each age class is a good proxy for the number of individuals slaughtered.

Multi-class age assignments

A tooth may not fall neatly into one or another age class. A well-preserved tooth will contain more information, whilst another may have more uncertainty. As exemplified by our dataset, the archaeologist may be only able to assign it to a broader aggregate of multiple classes. For example, the observation of one tooth belonging to either A or B [AB = 1]. A persistent problem in previous studies has been the coercion of these counts into Payne's age classes by either removing teeth that cannot be assigned to a single class (losing information), or by 'spreading' counts across the individual classes, such that [A=0.5 and B=0.5] e.g. (Gerbault, et al., 2016) or sometimes weighted by the age class width; e.g. Payne 1973). Clearly these approaches are not equivalent, and the latter generates false precision in the data, resulting in incorrect probabilities downstream. The importance of handling multi-class assignments properly is not restricted to calculating likelihoods; representing the data graphically becomes challenging as even the humble histogram requires counts that fall into only one class. In fact, the example above is equivalent to [(A=1) or (B=1)]. Therefore, to correctly calculate the probability of the observed data where some counts are assigned to more than one age class, we must calculate the probabilities under every possible arrangement using the multinomial distribution, and these individual probabilities must then be summed.

Goodness of fit: Calculating p-values under each model

An archaeological profile may look very different to a model profile, even where the observed data could easily be generated under that model. Similarly, the observed data may be a fair outcome of several models, or none. This equifinality is a particular problem for small sample sizes, where there are many possible ideal profiles that could generate the same outcome and is confounded further by multi-class assignments.

We approach this by asking the question 'what is the probability of getting the observed data *or more extreme* (i.e. any outcome that is less likely than the observed data), from the proposed model'? This is the definition of a p-value when testing a null model (Pearson, 1900). Thus, we make the *a priori* assumption that the data could have been generated under any of the ideal models, and if the p-value is suitably low (e.g. below 0.05) we can then reject that model. We achieve this using a Chi-squared test (Pearson, 1900), and account for uncertain data assigned to multiple age classes by calculating a p-value for each arrangement of the observed data, then calculating the average p-value weighted by the frequency of each arrangement. Calculating the frequencies of each arrangement first requires clarity of what an uncertain assignment means (e.g. AB=1). We assume that the archaeologist's belief is that A or B are equally likely (0 to 2 months, or 2 to 6 months), based on the view that the age classes were devised to reflect the limits at which morphological differences in teeth can be distinguished. This contrasts with the view that the archaeologist is determining the tooth as having uniform probability distribution between 0 and 6 months, which would then require the relative probabilities to be adjusted for the different time width of these age classes.

In practice, an exact approach is computationally expensive, so we approximate this by sampling one million possible arrangements of the observed data under the above assumptions, from which we calculate an average p-value. Table 4 reports p-values from the goodness of fit tests, which evaluate the probability of each archaeological profile (or more extreme) being generated under each model profile. P-values are extremely low in all tests, allowing us to reject all five proposed ideal models for every archaeological assemblage considered.

		Payne			Redding	
Site	n	Meat	Milk	Wool	Energy	Security
FON1	35	1.4E-04	7.7E-11	5.0E-07	1.6E-04	1.6E-03
TRA1	47	4.6E-13	4.4E-30	9.5E-18	8.9E-07	1.9E-06
MOLD	55	4.0E-06	1.9E-22	1.0E-12	1.5E-04	3.2E-05
TES	57	3.5E-14	1.0E-37	3.8E-22	4.8E-07	4.7E-06
TRA2	70	5.0E-10	5.3E-38	2.8E-24	1.5E-06	3.9E-07
PPI	73	2.1E-34	2.2E-56	1.5E-36	7.9E-25	2.9E-30
WIK	75	9.7E-20	2.1E-46	8.9E-32	6.9E-16	1.0E-22
PFE	82	2.0E-41	5.9E-78	1.9E-41	8.3E-17	4.4E-11
PCS	103	4.6E-57	3.5E-124	1.1E-70	2.5E-17	4.2E-07
LAD	122	5.9E-19	8.2E-92	1.3E-69	2.6E-16	5.1E-25

Table 4: summary of Goodness of fit tests, showing none of these models are plausible explanations for the observed data (n= tooth counts).

The Gamma distribution as a model for idealised profiles

The failure of all five tested idealised profiles to provide a statistically plausible model for any of the archaeological data sets suggests the need to re-evaluate the appropriateness of these models. This failure may be partly due to poor specification of the slaughter probabilities in each age class. However, each model is described using eight parameters, (the ninth age-class probability is not free as they must all sum to 1), which suggests they may also suffer from unjustified model complexity. Occam's razor requires us to first consider the simplest models with the fewest parameters, and to increase model complexity only where this can be justified by providing a substantial improvement in their fit to data. The Gamma distribution uses just two parameters μ and k, representing the mean slaughter age and the shape of the dispersion around this mean, respectively. The general properties and shape of the Gamma distribution fit well with the expectation of age-at-death under a single simple theoretical subsistence strategy – as a unimodal probability distribution it naturally represents the distribution of slaughter ages with a single peak (the most common kill age), and has a domain greater than zero (an age below which no kills can occur).

Using Gamma parameters to summarise age-at-death data

Archaeological age-at-death profiles are usually represented using histograms. This can work well for a single profile provided each count falls into a single age class, but becomes problematic for multi-class assignments. Furthermore, the comparison of several profiles cannot be achieved on the same axes, since each profile requires its own histogram. Any attempt to represent several profiles on a single two-dimensional plot requires some information compression. Gerbault, et al. (2016) approached this by using Correspondence Analysis, which illustrates the two dimensions with the most variation whilst the remaining dimensions (which together typically comprise a substantial fraction of variation) are lost.

A common strategy when representing complex or large data (either graphically or numerically) is to compress the information content into a few summary statistics. The objective is to communicate the largest amount of information as simply as possible. For example, box and whisker plots might be used to represent a large univariate dataset using just five statistics (minimum, maximum, median, first and third quartile), or the data might be compressed further and described using just the mean and variance.

We build on this approach by fitting a Gamma distribution to our age-at-death data using Markov Chain Monte Carlo (MCMC) (see MCMC section details) which provides the full joint parameter likelihood distribution. This has the advantage of representing the entire age-at-death profile using Gamma parameters μ and k, and the clear graphical representation and comparison of datasets in two-dimensions, with less information loss. Furthermore, since the Gamma distribution is a continuous probability distribution, it provides the flexibility to summarise and compare datasets with different age classes, such as modern ethnographic data, where the age at death may be known with an accuracy of just a few days.

We illustrate this in Fig 1, showing best fit Gamma distributions as a continuous function of age, for each archaeological dataset (using joint Maximum Likelihood Estimates (MLE) of μ and k), and in Fig 2 showing the 90% confidence interval (CI) of these Gamma parameters, representing the uncertainty due to small sample sizes and multi-class assignments.

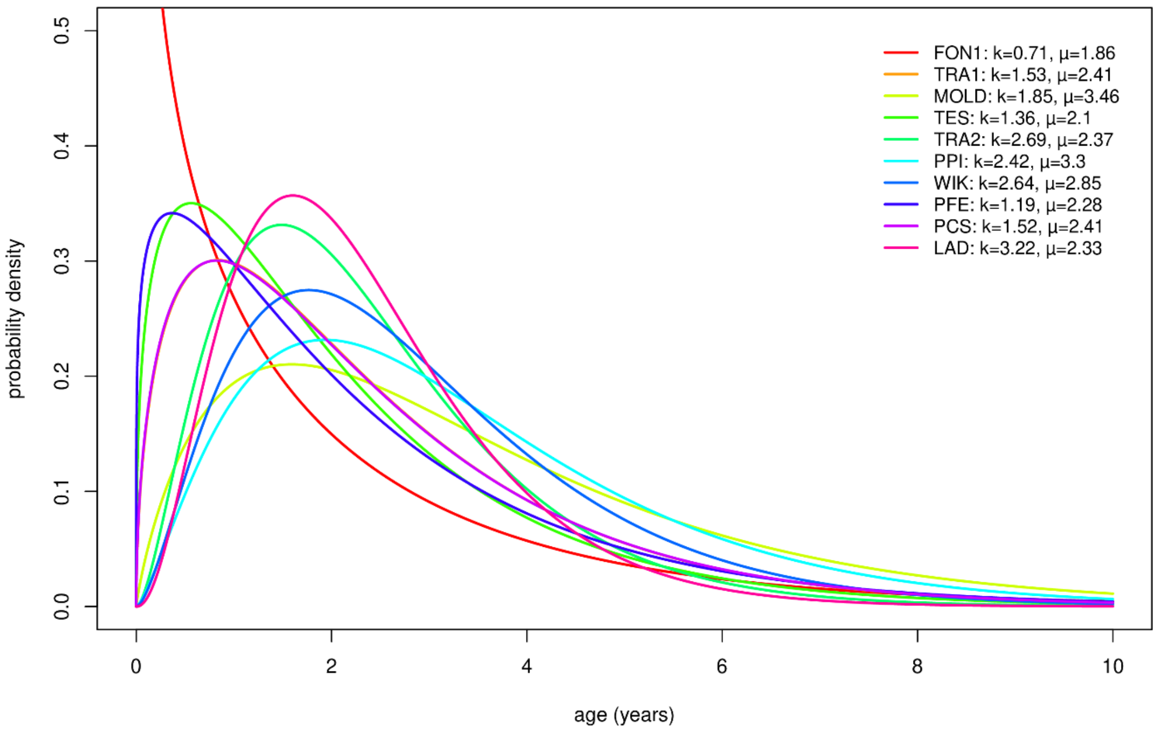

Fig 1: Best fit Gamma distributions as continuous functions of age, for each archaeological dataset, based on Maximum Likelihood Estimates of the Gamma parameters (k and μ). Note, the distribution of TRA1 is obscured behind the PCS distribution, since their MLE are almost identical.

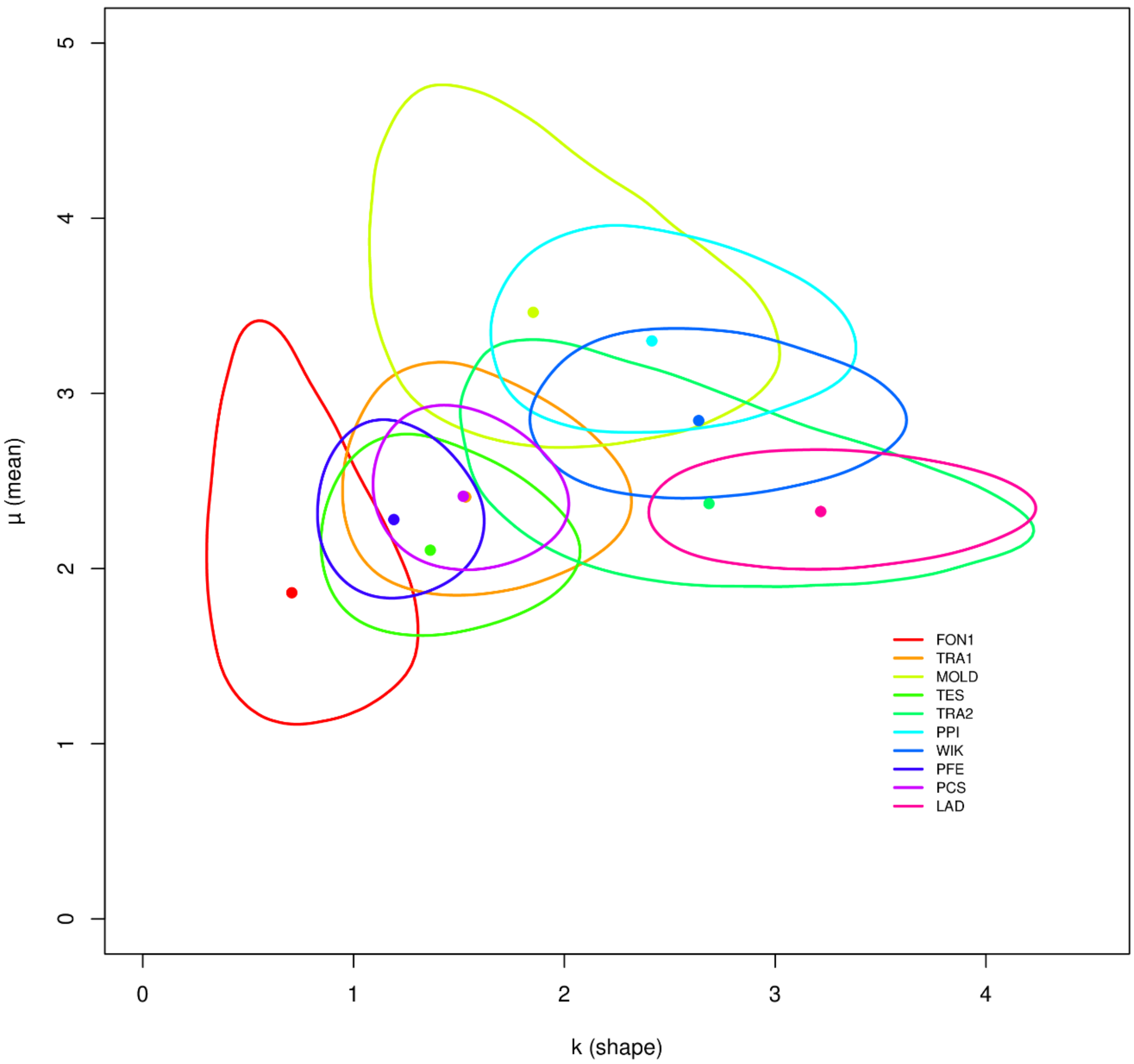

Fig 2: Summary of joint Gamma parameter distribution (*k* and μ) for each archaeological dataset, using 90% CI contours (lines) and MLE (dots) for each archaeological dataset. CI represents the uncertainty from sample sizes, and multi-class assignments. Note, although the MLE of TRA1 is obscured behind the PCS, the uncertainty contour lines of both sites clearly differ.

Fitting a Gamma distribution using MCMC

In order to estimate Gamma parameters for a dataset, we are interested in calculating the likelihood, which is the probability of the observed dataset given some proposed parameters of the Gamma distribution. To do this, our likelihood function first discretises the gamma probability density function (PDF) to match the Payne age classes (this is simply the PDF area within each class). This gives the probability of observing a tooth in each age class, given the Gamma parameters. Finally, the overall likelihood is calculated using these derived age class probabilities in a multinomial distribution. Data with uncertain assignments requires the likelihood to be calculated for every possible arrangement, before finally summing.

This approach of calculating likelihoods exactly allows us to find both the MLE and confidence intervals of Gamma parameters for a given archaeological profile, which we estimate using Markov Chain

Monte Carlo (MCMC) by implementing the Metropolis-Hastings algorithm (Hastings, 1970) (10 chains, each of 30,000 iterations, removing 2000 for burn-in, thinning to every 5th iteration).

Assessing the quality of the Gamma model using Information Criteria

Although the Gamma model provides the benefit of fewer parameters, its unimodal shape is more constrained than the current age class models. Whilst we have previously argued that the existing models of Payne and Redding cannot explain any of our datasets, it is plausible that a new age-class model with different probabilities might not be rejected.

Therefore, to compare the Gamma model with the age-class model without the constraint of Payne and Redding's specific probabilities, we use a model selection approach, using both the Akaike Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwarz, 1978). These provide estimators of the relative quality of both models, given our data, by balancing model likelihood against model complexity. For each dataset, we use a search algorithm to find the Maximum Likelihood Gamma parameters and the Maximum Likelihood age class probabilities, therefore estimating the MLE under both types of model. These are used to calculate the AIC and BIC of each model using the following formulas, where p is the number of parameters, and n is the sample size:

$$\mathrm{AIC} = 2p - 2\ln(\mathrm{MLE})$$

$$\mathrm{BIC} = \ln(n)p - 2\ln(\mathrm{MLE})$$

The results (Table 5) show that for most of the datasets the Gamma model provides the lower estimator, and therefore the better model (12/20 when aggregating all AIC and BIC comparisons, or 11/17 after excluding very similar IC scores). In all ten cases the maximum likelihood of the age class model is greater, which is to be expected since the age class model is more complex and therefore free to fit the data more closely. However, the model comparison shows that in most cases the amount of this improved fit cannot be justified by the amount of increased model complexity. This shows that the age class model often overfits, whilst the Gamma model provides a more justified and conservative model.

Nevertheless, in datasets PPI, PFE and PCS we observe enough structure in the data to warrant a model of greater complexity than the two parameter Gamma model can provide. This does not provide a counter argument in favour of the 8-parameter age class model, since it is likely that a slightly more complex model comprising just three parameters (for example a bimodal distribution) may provide a close enough fit to provide the lower IC estimator. This remains an area for further investigation, and the main objective of this model comparison is not to explore the entirety of model space to provide the best possible model, but merely to show that the current approach of using a 9-age class model is unjustifiably complex and overfitted.

		Gamma model				Age class model			
Code	n	Log MLE	parameters	AIC	BIC	Log MLE	parameters	AIC	BIC
FON1	35	-3.00	2	9.99	13.10	-1.47	8	18.93	31.37
TRA1	47	-11.21	2	26.42	30.12	-5.13	8	26.26	41.07
MOLD	55	-4.43	2	12.87	16.88	-1.95	8	19.91	35.97
TES	57	-11.53	2	27.05	31.14	-5.47	8	26.95	43.29
TRA2	70	-2.24	2	8.49	12.98	-0.96	8	17.92	35.91
PPI	73	-21.10	2	46.19	50.77	-5.03	8	26.06	44.38

WIK	75	-12.97	2	29.94	34.58	-7.27	8	30.53	49.07
PFE	82	-30.79	2	65.58	70.39	-5.62	8	27.24	46.50
PCS	103	-36.03	2	76.06	81.33	-5.55	8	27.11	48.19
LAD	122	-4.63	2	13.26	18.86	-1.82	8	19.65	42.08

Table 5: Summary of information criteria tests. The Gamma model has a lower Information Criterion score (considering both AIC and BIC together) for the majority the 10 data sets (dark blue representing clear superiority, light blue only a marginal superiority), showing that in most cases, the Gamma model is a better explanation of the observed data.

Discussion

The mortality models proposed by Payne (1973) and Redding (1981) were derived from specialised market economies using modern improved breeds. Over the last few centuries intensive selective breeding has increased the capacity for animals to produce milk/wool and develop muscle/fat faster, which has been driven by the development of economic specialisation within market economies. As such it is unsurprising that these models can be statistically rejected as plausible explanations for archaeological data. Intensive slaughter strategies require developed social organisation to process animal products into storable products, or large-scale consumption, and furthermore require developed trade networks between economic specialists to ensure herder security (Davies, 2015, Sikana, et al., 1993). Early Neolithic herders would not have optimised for a single product, but would have managed their herds for a mix of products, including milk, meat and wool (Halstead, 1998, Helmer and Vigne, 2007) as indicated by the wide range of slaughter ages. The archaeological record for the early Neolithic in Europe indicates a diversified agro-pastoral economic system with supplementation from wild resources (Rowley-Conwy, et al., 2013); meaning labour input would have also been spread across a diverse range of economic activities. For this reason, the slaughter practices of modern subsistence agro-pastoralists keeping unimproved stock are likely to be more similar to those practiced in the Neolithic than the slaughter strategies of specialised market economy pastoral systems.

The Gamma parameters k and μ can be estimated for any archaeological profile, and represented either directly in two dimensions (as in Fig 2), or as a function of age (as in Fig 1). In either case, where the objective is to draw an inference about the true parameters of herd slaughter ages assuming our data is a random sample, confidence intervals around the joint parameter estimates can also be represented in order to reveal the uncertainty in these distributions, given small sample sizes and multi-class assignments (Figs 2 and 3).

Whether k and μ are used as summary statistics to describe archaeological data, or as model parameters to infer a slaughter strategy, their numerical values provide useful representations. Firstly, μ can be interpreted directly as the mean slaughter age of the herd. Secondly, the mode of each distribution can be calculated exactly as $(k-1)(\mu/k)$, *for* $k \geq 1$, which can be interpreted as the most likely slaughter age. However, it should be noted that although the mode is not mathematically defined for distributions where $k < 1$ (since the probability density = zero when age = 0), the probability density function asymptotically increases as the age approaches zero, so in terms of summarising real data this constraint becomes irrelevant, and such distributions can be described as having a peak kill age of ‘new-born’, which may be natural mortality given that this is highest in the first month (Mellor and Stafford, 2004).

Furthermore, we can estimate the 50% Highest Density Interval (HDI) of the Gamma distribution, which indicates the age range for the majority (nominally) of slaughters. For example, at FON1 we observe a peak death of newborns coupled with a relatively intensive slaughter strategy where the majority of animals were being killed across a narrow one year age range (0 to 1.09 yrs), as expected for a rock shelter site being used as a seasonal camp during the birthing season. The intensive slaughter may also be a reflection of milk exploitation; given that milk lipids have been recovered from ceramics and that caprine milk let-down is not sensitive to the removal of infants in comparison to cattle (Balasse, 2003). An element of natural mortality of young infants may also explain the Gamma distributions for TRA1, TES and PFE. In comparison, the peak slaughter was around 1.5 yrs at MOLD, TRA2 and LAD with a much broader range of kill ages (around 2 years).

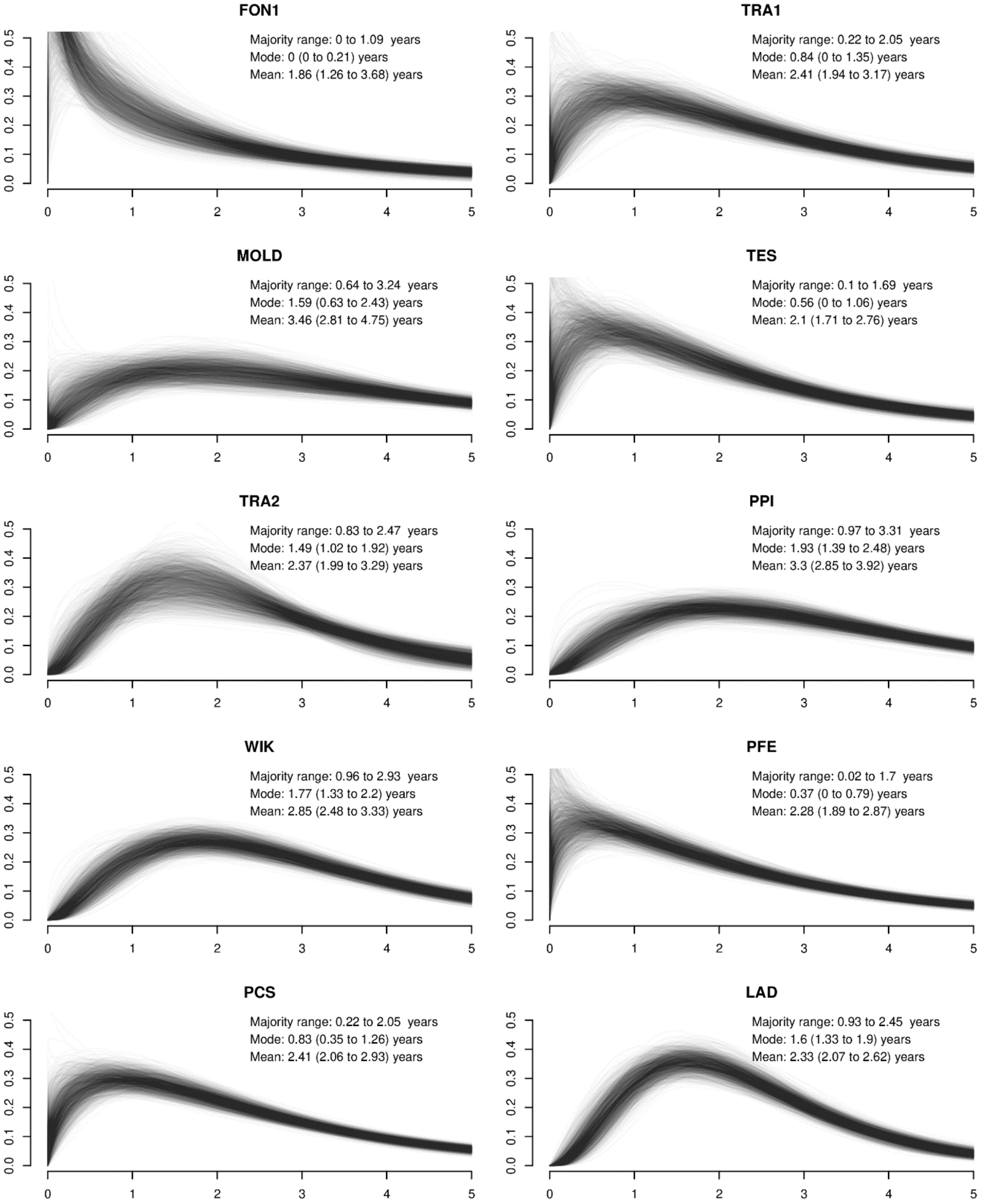

Fig 3: 2000 Gamma distributions drawn from the joint parameter distributions of shape (*k)* and mean (μ). The (nominal) majority slaughter range is the 50% HDI, using the MLE parameter values. The Slaughter mode is calculated exactly from the MLE parameters. 95% CI for both the Mode and Mean are calculated from the MCMC samples.

Conclusion

Because we know that the human past was complex, it is often assumed that more complex models of that past are more realistic. However, more complexity means more parameters, and so more ways (or degrees of freedom) for a model to differ from reality. This means that unless specified by secure information, more complex models are typically more wrong, not more realistic. We show that the Gamma distribution provides a better model for a theoretical slaughter strategy than the current age-class models of 9 probabilities, which are unjustifiably complex. Furthermore, we show how the Gamma distribution can be used to provide a more robust summary of archaeological age-at-death profile data using just two summary statistics, μ and *k,* with minimal information compression. This opens up the potential for in-depth exploration of animal husbandry practices over time and space by comparing the parameter distances between sites. Finally, we highlight the problems inherent in analysing and representing (both numerically and graphically) data with multi-class assignments, and show how the Gamma distribution permits the representation of data that accounts for uncertainty from both small sample sizes and multi-class assignments.

GammaModel R package

Tools to replicate the analyses performed in this paper are freely available as the R package GammaModel. (Timpson A, 2018) All source files are available from the Github repository https://github.com/UCL/GammaModel and users are advised to work through the vignette 'guide.pdf' which includes details of how to install, load and use the package. Help files can be accessed within the package, or from the manual.pdf. GammaModel utilises functions from R packages dplyr (Wickham H. et al. 2018), combinate (Chasalow S. 2002), DEoptimR (Conceicao L.T. et al. 2016), LaplacesDemon (Statisticat L. L. C. 2018) and stats (R Core Team 2018).

The GammaModel manual and vignette can be found at https://doi.org/10.1016/j.jas.2018.08.015

References

Akaike, H., 1974. A new look at the statistical model identification, IEEE transactions on automatic control 19, 716-723.

Arbuckle, B.S., Kansa, S.W., Kansa, E., Orton, D., Cakirlar, 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., Pollath, N., Pawlowska, K., Russell, N., Twiss, K., Wurtenberger, D., 2014. Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey, PLoS One 9, e99845.

Balasse, M., 2003. Keeping the young alive to stimulate milk production? Differences between cattle and small stock, Anthropozoologica, 3-10.

Chasalow S., Combinat: Combinatorics Utlilites. R Package Version 0.0-8, 2012 https://CRAN.R-project.org/package=combinatEduardo.

Çilingiroğlu, C., 2009. Of stamps, loom weights and spindle whorls: Contextual evidence on the function (s) of Neolithic stamps from Ulucak, Izmir, Turkey, Journal of Mediterranean Archaeology 22, 3-27.

Coward, F., Shennan, S., Colledge, S., Conolly, J., Collard, M., 2008. The spread of Neolithic plant economies from the Near East to northwest Europe: a phylogenetic analysis, Journal of Archaeological Science 35, 42-56.

Davies, M., 2015. Economic specialisation, resource variability, and the origins of intensive agriculture in Eastern Africa, Rural Landscapes: Society, Environment, History 2.

Debono Spiteri, C., Gillis, R.E., Roffet-Salque, M., Castells Navarro, L., Guilaine, J., Manen, C., Muntoni, I.M., Sana Segui, M., Urem-Kotsou, D., Whelton, H.L., Craig, O.E., Vigne, J.D., Evershed, R.P., 2016. Regional asynchronicity in dairy production and processing in early farming communities of the northern Mediterranean, Proc Natl Acad Sci U S A 113, 13594-13599.

Ducos, P., 1968. origine des animaux domestiques in Palestine.

Gerbault, P., Gillis, R., Vigne, J.D., Tresset, A., Brehard, S., Thomas, M.G., 2016. Statistically robust representation and comparison of mortality profiles in archaeozoology, Journal of Archaeological Science 71, 24-32.

Gillis, R.E., Kovacikova, L., Brehard, S., Guthmann, E., Vostrovska, I., Nohalova, H., Arbogast, R.M., Domboroczki, L., Pechtl, J., Anders, A., Marciniak, A., Tresset, A., Vigne, J.D., 2017. The evolution of dual meat and milk cattle husbandry in Linearbandkeramik societies, Proc Biol Sci 284.

Grant, A., 1982. The use of tooth wear as a guide to the age of domestic ungulates., in: Wilson, R., Grigson, C., Payne, S. (Eds.), Ageing and sexing animal bones from archaeological sites., BAR International series, Oxford, pp. 91-108.

Greenfield, H.J., Arnold, E.R., 2015. 'Go (a) t milk?'New perspectives on the zooarchaeological evidence for the earliest intensification of dairying in south eastern Europe, World Archaeology 47, 792-818.

Guilaine, J., 2001. La diffusion de l'agriculture en Europe: une hypothèse arythmique, Zephyrus 53, 267-272.

Halstead, P., 1998. Mortality models and milking: problems of uniformitarianism, optimality and equifinality reconsidered, Anthropozoologica 27, 3-20.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57, 97-109.

Helmer, D., 1991. Les faunes Chasséenes du sud-est de la France, Essai de synthèse, Identité du Chasséen: Actes du Colloque International Mémoires du Musée de Préhistorie d'Ile-de-France, Nemours, pp. 343-349.

Helmer, D., Gourichon, L., Sidi Maamar, H., Vigne, J.D., 2005. L'élevage des caprinés néolithiques dans le sud-est de la France : saisonnalité des abattages, relations entre grottes-bergeries et sites de plein air, Anthropozoologica 40, 167-189.

Helmer, D., Gourichon, L., Vila, E., 2007. The development of the exploitation of products from Capra and Ovis (meat, milk and fleece) from the PPNB to the Early Bronze in the northern Near East (8700 to 2000 BC cal.), Anthropozoologica 41, 41-69.

Helmer, D., Vigne, J., 2007. Was milk a ''secondary product''in the Old World Neolithisation process, Its role in the domestication of cattle, sheep and goats. Anthropozoologica 42, 9-40.

Higham, C.F.W., 1969. Towards an Economic Prehistory of Europe, Current Anthropology 10, 139-150.

Manning, K., Stopp, B., Colledge, S., Downey, S.S., Connolly, J., Dobney, K., Shennan, S., 2013. Animal exploitation in the Early Neolithic of the Balkans and Central Europe, in: Colledge, S., Conolly, J., Dobney, K., Manning, K., Shennan, S. (Eds.), The origins and spread of domestic animals in Southwest Asia and Europe, Left coast press, Califonia, pp. 237-252.

Marciniak, A., 2013. Origin of stock-keeping and the spread of animal exploitation strategies in the early and middle Neolithic of the North European Plain, in: Colledge, S., Conolly, J., Dobney, K., Manning, K., Shennan, S. (Eds.), The origins and spread of domestic animals in southwest Asia and Europe, Left Coast Press, California, pp. 221-236.

Marom, N., Bar-Oz, G., 2009. Culling profiles: the indeterminacy of archaeozoological data to survivorship curve modelling of sheep and goat herd maintenance strategies, Journal of Archaeological Science 36, 1184-1187.

Mellor, D.J., Stafford, K.J., 2004. Animal welfare implications of neonatal mortality and morbidity in farm animals, Vet J 168, 118-133.

Payne, S., 1973. Kill-off patterns in sheep and goats: the mandibles from Aşvan Kale, Anatolian studies 23, 281-303.

Payne, S., 1987. Reference Codes for Wear States in the Mandibular Cheek Teeth of Sheep and Goats, Journal of Archaeological Science 14, 609-614.

Pearson, K., 1900. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50, 157-175.

Perlès, C., 2005. From the Near East to Greece: Let's reverse the focus. Cultural elements that didn't transfer, How did farming reach Europe, 275-290.

Peters, J., Helmer, D., Von Den Driesch, A., Saña Segui, M., 1999. Early animal husbandry in the northern Levant, Paléorient 25, 27-47.

R Core Team, R: a Language and Environment for Statistical Compution, 2018, R Foundation for Statistical Computing; Vienna, Austria https://www.R-project.org/.

Redding, R., 1981. Decision making in subsistence herding of sheep and goats in the Middle East. University of Michigan, Ann Arbor. Ph. D.

Rowley-Conwy, P., Gourichon, L., Helmer, D., Vigne, J., Colledge, S., Conolly, J., Dobney, K., Manning, K., Shennan, S., 2013. The origins and spread of domestic animals in southwest Asia and Europe.

Schwarz, G., 1978. Estimating the dimension of a model, The annals of statistics 6, 461-464.

Sikana, P.M., Kerven, C.K., Benkhe, R., 1993. From subsistence to specialised commodity production: commercialisation and pastoral dairying in Africa.

Zeder, M.A., 1999. Animal domestication in the Zagros: a review of past and current research, Paléorient 25, 11-25.
Scott Chasalow (2012). combinat: combinatorics utilities. R package version 0.0-8. https://CRAN.R-project.org/package=combinatEduardo

L. T. Conceicao (2016). DEoptimR: Differential Evolution Optimization in Pure R. R package version 1.0-8. https://CRAN.R-project.org/package=DEoptimR

Statisticat, LLC. (2018). LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com. R package version 16.1.1. https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software

Timpson, A (2018). GammaModel: Modelling age-at-death profiles using the Gamma distribution. R package version 1.0. https://github.com/UCL/GammaModel

Wickham H, Francois R, Henry L, Muller K, Dplyr: a Grammar of Data Manipulation. R Package Version 0.7.6, 2018 https://CRAN.R-project.org/package=dplyr.

Acknowledgements

This research was funded by the NeoMilk project (ERC advanced grant awarded to R.P. Evershed (FP7-IDEAS-ERC/324202)). We are grateful to Jean Guiliane, Isabelle Carrère, Maria Saña Seguí, Giovanna Radi, Joachim Pechtl, Henriette Obermaier, Alexandra Anders and Erich Pucher for allowing access to the archaeozoological material. We acknowledge the use of the UCL Legion High-Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work. We also thank Pascale Gerbault, Alan Outram, Marie Balasse, Anna Rudzinski and especially Jean-Denis Vigne and Yoan Diekmann for their useful discussions and advice, and to Mike Charlton and three anonymous reviewers for their constructive feedback.