STATISTICAL PREDICTION OF THE EXTINCTION TIME OF EXTINCT MAMMALIAN SPECIES IN ANATOLIA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ELÇİN EKŞİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ARCHAEOMETRY

FEBRUARY 2023

Approval of the thesis:

STATISTICAL PREDICTION OF THE EXTINCTION TIME OF EXTINCT MAMMALIAN SPECIES IN ANATOLIA

submitted by **ELÇİN EKŞİ** in partial fulfillment of the requirements for the degree of **Master of Science in Archeometry, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _________________

Prof. Dr. Gülay Ertaş
Head of the Department, **Archaeometry** _________________

Prof. Dr. Zeynep Kalaylıoğlu
Supervisor, **Statistics, METU** _________________

Examining Committee Members:

Prof. Dr. Mehmet Somel
Biological Sciences, METU _________________

Prof. Dr. Zeynep Kalaylıoğlu
Statistics, METU _________________

Assoc. Prof. Dr. Derya Silibolatlaz Baykara
Institute For Migration, Gaziantep University _________________

Date: 22.02.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Elçin Ekşi

Signature :

ABSTRACT

STATISTICAL PREDICTION OF THE EXTINCTION TIME OF EXTINCT MAMMALIAN SPECIES IN ANATOLIA

Ekşi, Elçin
Master of Science, Archeometry
Supervisor : Prof. Dr. Zeynep Kalaylıoğlu

February 2023, 95 pages

Extinction and origination of species are fundamental to the process of evolution, a dynamic force that shapes the history of life. These processes are studied with different motivations in both paleontology and ecology. In some cases, obtaining sufficient population data for the studied species is impossible. For example, fossils of an extinct species or data from an endangered and rare species offer limited analysis. In such cases, it is possible to predict the extinction time of a species by analyzing the ordered time point data in which the species was detected in nature or stratigraphic sections using statistical methods. In this study, statistical estimation was carried out to reveal the extinction date of the wild mammalian species that lived and went extinct in Anatolia ten thousand years ago. Archaeological records of non-domesticated animals obtained from literature and databases were brought together. The ages of the archaeological remains were used for statistical analysis. Species records can be evaluated as a Poisson process and modeled according to various distributions or independently of the distribution to estimate when they disappeared. In this study, we investigate the basic features of the compiled archeological species records using a histogram, PPCC test, and Q-Q graph and whether the data were uniformly distributed. The study resulted in a list of MAMMALIAN species that became extinct in the last ten thousand years and calculated the date on which the species went extinct in the past with a 95% confidence limit. In addition, the difference between extinction date estimation methods according to the characteristics of the data has been revealed.

Keywords: Extinction, Extinct Species, Extinction Time Estimation, Optimal Linear Estimation, Confidence Intervals For Temporal Range

ÖZ

ANADOLU'DA NESLİ TÜKENMİŞ MEMELİ TÜRLERİNİN YOK OLMA ZAMANLARININ İSTATİSTİK TAHMİNİ

Ekşi, Elçin
Yüksek Lisans, Arkeometri
Tez Yöneticisi: Prof. Dr. Zeynep Kalaylıoğlu

Şubat 2023, 95 sayfa

Türlerin ortaya çıkışı ve yok oluşu, evrim sürecinin temel unsurlarıdır ve canlılık tarihini şekillendiren dinamik güçlerdir. Bu süreçler hem paleontoloji hem de ekoloji alanında farklı motivasyonlarla incelenir. Kimi durumlarda incelenen türlere ait yeterli populasyon verisinin elde edilmesi mümkün değildir. Örneğin yok olmuş bir türün fosilleri ya da yok olmak üzere olan ve nadir rastlanan bir türe ait veriler kısıtlı analiz imkanı sunar. Böyle durumlarda bu canlıların doğada ya da stratigrafik kesitlerde tespit edildiği sıralı zaman noktası verileri istatistiksel olarak incelenerek canlının yok olma tarihine dair kestirimde bulunmak mümkündür.

Bu çalışmada on bin yıl öncesinden günümüze kadar Anadolu'da yaşamış ve yok olmuş yabani memeli hayvan türleri ve ne zaman yok olduklarını ortaya çıkarmak üzere istatistiksel kestirim gerçekleştirilmiştir. Literatür taraması ve veri tabanlarından elde edilen evcilleştirilmemiş hayvanlara ait arkeolojik kayıtlar bir araya getirilmiştir. Arkeolojik kalıntıların yaşları istatistiksel analiz için kullanılmıştır. Canlılara ait kayıtlar, Poisson süreci olarak değerlendirilip çeşitli dağılımlara göre ya da dağılımdan bağımsız olarak modellendiğinde, hangi tarihte yok olduğuna dair kestirimde bulunmak mümkündür.

Bu çalışmada derlenen arkeolojik tür kayıtlarının öncelikle histogram grafik, PPCC testi ve Q-Q grafiği ile temel özellikleri ve verilerin uniform dağılıp dağılmadığı araştırılmıştır.

Çalışma sonucunda son on bin yıl içinde nesli tükenen memeli türlerinin listesi çıkarılmış ve yok olmuş türlerin geçmişte hangi tarihte neslinin tükendiği %95 güven sınırıyla hesaplanmıştır. Bunun yanı sıra verilerin özelliklerine göre yok oluş tarihi kestirimi yöntemleri arasındaki fark ortaya konmuştur.

Anahtar Kelimeler: Yok Oluş, Nesli Tükenmiş Türler, Yok Oluş Zamanı Kestirimi, Optimal Doğrusal Kestirim, Zaman Aralığı İçin Güven Aralıkları

To Çapçik, Rosa, Venüs, Turgut, Tedi, Mars and Çirkin.

ACKNOWLEDGMENTS

I wish to express my heartfelt gratitude to my advisor, Prof. Dr. Zeynep Kalaylıoğlu, for her unwavering support and guidance throughout this thesis. Her invaluable insights and expertise have greatly expanded my horizons and shaped me into a better researcher.

As I wrote this thesis, I am also thankful to Irmak Güneş Yüceil, who provided support and helped me overcome problems related to my school. The encouraging comments of Elşen Aydın on mathematical approaches and the school were also very valuable to me. I am deeply grateful for their assistance.

Furthermore, I extend my gratitude to İlkem Gürgör and Jennifer Cress for their invaluable contributions to this research. Their generosity in sharing their data has been instrumental in the success of this project.

Lastly, I would like to thank my family and friends for their constant encouragement and support. Their unwavering belief in me has been a driving force in achieving my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The extinction of species is a part of evolution, together with speciation. The extinction idea emerged with an exploration of fossils and advances in geology. Geologist Georges Cuvier is accepted as the founder of extinction studies with his studies on extinct mammoth bones. He theorized extinction as a catastrophic event that occurs suddenly. Charles Darwin's view on extinction differed from that of Cuvier, who believed in the sudden disappearance of species. Darwin believed that extinction occurred gradually and that less fit species were more likely to become extinct (Reznick, 2009). The incremental extinction approach is mainly accepted as background extinction today.

Contrary to popular belief, it is not easy to detect a species' extinction status directly in ecology and geology. In ecology, it is often not possible to systematically observe the last individuals of a species, with exceptions. However, estimating whether or how close species are to extinction is essential to clarify their international conservation status. In geology, the situation is more complex. Fossil data were obtained from stratigraphic sections. Fossilization is a rare event, as evidenced by the fact that the vast majority of living organisms throughout history have not been preserved in the fossil record. Fossilization and taphonomic processes are the most critical points find these species' fossils. Therefore, it does not indicate that the first fossil record was the first appearance of the species or that the last fossil record was the last individual before extinction. Statistical approaches have been developed to estimate species' extinction dates in ecology and geology to overcome these problems.

Research interest in estimating species extinction time emerged in the second half of the 20th century. In their seminal study published in 1980, Alvarez et al. proposed that the terminal Cretaceous mass extinction was caused by a bolide impact event, providing evidence for this hypothesis through geological and paleontological investigations. The Alvarez hypothesis stimulated criticism of catastrophic extinction events and extinction patterns in fossil records. Signor and Lipps pointed to sampling bias and the gaps in the fossil records and showed that the last fossil record is not the latest member of any taxa (Signor, 1982). After the faulty assumptions of Alveraz's about extinction after bolide impact, attention increased to estimating extinction time in paleontology. After that, statistical methods were developed to estimate the extinction date of taxa according to their location in the stratigraphic section.

Estimating the extinction date of species is an important issue not only in paleontology but also in ecology. Estimation studies based on the last sight records of species are carried out to improve the conservation plans of especially endangered species with reduced population size (Robson, 1964).

Statistical analyses of extinct and endangered species data are often limited by insufficient sample sizes and incomplete datas, which can make it difficult to draw robust conclusions about the factors driving population declines and extinctions. Therefore, new methods were developed to estimate the extinction time of species with inadequate data. The extinction of species is affected by stochastic and deterministic processes, just like other evolutionary events. Population sizes, population structure (metapopulations), and growth rate determine which stochastic or deterministic processes will be more dominant in extinction. Small populations are generally accepted as more susceptible to stochastic processes and extinction. Larger populations with more data are more predictable than small populations. Therefore, stochastic, and deterministic approaches are developed to analyze populations according to the type of population (Shaffer, 1981). More data on populations are required to apply deterministic techniques. For example, population growth, death, and dispersal rates are needed to do viable population analysis. However, obtaining these data for every population in nature or fossil records is impossible. Mainly extinct, small, or declining populations are more challenging to observe population parameters. While there is an abundance of fossil, sedaDNA or new sources of data available for extinct species, they are often limited in their ability to provide detailed information on population dynamics. Therefore, extinction time estimation studies are generally modeled as a stochastic process of ordered animal record time data. Many studies model the occurrence of species as discrete random events over time. In general, extinction models are considered to follow a Poisson process. The Poisson process is widely used for modeling a system's entry time/occurrence time with the speed $\lambda(t)$ (Boakes, 2015).

This study aims to estimate the extinction dates of ancient Anatolian mammal species using statistical methods which have not been previously applied to this specific group of mammal species in the archaeological context. Understanding the extinction of Anatolian mammals during the Holocene epoch will make new contributions to the archaeological literature and give new perspectives to increase conservation efforts today.

1.1 Aim of Study

This study aimed to estimate the extinction dates of animal species whose remains were found in archaeological settlements in Anatolia over the past 10 thousand years, using appropriate statistical methods. First, animal remains data were obtained from archaeology literature, and databases were arranged as occurrence records of species data. Specimens defined at the species level were selected to use in this study. Secondly, the data is filtered, organized, and prepared for statistical analyses. Then, the most appropriate methods for estimating extinction dates were determined, and these methods were used to calculate the extinction dates of selected species. As a result, the first extinct mammal species listed for the last 10 thousand years have been found, and when these species have gone extinct is estimated. During the literature search, it was found that no studies have been conducted in Turkey using statistical approaches to estimate extinction times for either extinct or extant animals, including studies conducted on Turkish material. In addition, excluding paleontological studies and historical museum studies, similar studies using the extinction time estimation methods with archaeological animal remains are not found in the literature.

The main research question is, “When gone extinct selected species in Anatolia over the last 10 thousand years?”. The extinct animal list was determined by cross-checking archaeological and current ecological data to address that question. The chosen statistical methods were applied to animal records to estimate their extinction dates.

The thesis is organized as follows. Chapter 2 is an introduction to extinction estimation studies in different disciplines. This chapter summarizes the structure of data and methods used in various fields. Chapter 3 includes how data is handled, explored, and analyzed in this thesis. Details of methods are described. This chapter also contains results. Chapter 4 is the assessment of the results and overall conclusion. This chapter includes the extinct species list and the estimated extinction times of species. Chapter 5 is discussion and what are the future needs of the study.

CHAPTER 2

LITERATURE REVIEW

Estimating the extinction time of species is a relatively new field and a subject that different disciplines. The extinction of species was first discovered by finding fossils in geology and then became one of the remarkable subjects in ecology. Extinction date estimation studies are of great importance for understanding natural history in geology, and examining past dynamics, besides using existing ecological conservation plans for species. Although the motivations of both fields are different, they also have similar aspects. Furthermore, other archeological materials have been analyzed with extinction date estimation methods to understand when cultures emerged and disappeared.

The main feature of estimating extinction date studies involved analyzing the statistical occurrence of a species' last individuals in the study area. The data in estimating extinction dates in geology and ecology have similar and different features. Their main similarity is that both data are ordered point data on a timeline. Because of that, many researchers handle these occurrence records of species data as time-series data. Moreover, fossil records have larger intervals than ecological records, for example, million years or thousands of years. Extinction date estimation methods are sensitive for every time scale. Their main differences are sourcing from their temporal position and their recording methods. While fossils aging with various methods are used in geology, the sight records of individuals in nature are used in ecology studies. Sources of species occurrence data shape characteristic features of data. For example, systematical field surveys give more information about species, both presence, and absence, in ecological studies; however, fossil records have limited information about species occurrence. Besides, record data certainties are different for these two primary data sources. Some ecological records can be uncertain. For example, anecdotal sight records from local people can be misidentified; therefore, these sight records are accepted as uncertain records. Fossil records can be weighted according to their "recovery potential" according to environmental factors similar effects on extinction time estimation models like "uncertain" records but not identical. Apart from these, finding animal remains at the archaeological excavation site also depends on the people who occupied these sites.

The frequency of occurrence records on the timeline is another crucial point for extinction date estimation models. The occurrence frequency of data is commonly examined in three groups, based on their distribution over time: constant, decreasing, and increasing. Understanding the occurrence frequency is essential for selecting appropriate extinction date estimation models that are consistent with these distributions. Generally, decreasing populations have decreasing occurrence frequency. Nevertheless, excavation efforts to find animal remains (or fossils) or sampling efforts in the field also impact the occurrence frequency trend. Rivadeneira et al. (2009) have described sampling probabilities in 5 different patterns that combine population dynamics and sampling efforts to generate simulated occurrence record data. These five patterns are uniform, down, down-up, up, and up-down. Uniform means sampling probability is constant over time, which is the most unrealistic pattern. Down and the up-down of sampling probability patterns are the most realistic ones. Up and down-up probability patterns can occur with intensive sampling effort. The sampling effort is similar for ecology, geology, and archaeology. But the sampling techniques are very different for ecology, geology, and archeology. Therefore, the sampling effort is not the same. Archaeology and geology are more similar to each other than ecology.

Extinction date estimation methods in archaeological studies are still very new, and the methods used in geology and ecology studies are used. Archaeological data is more familiar with geological data because archaeological data is obtained from excavations. Like the fossils of extinct species found in geological studies, some specimens obtained during archaeological excavations belong to extinct species. While presence information can be obtained from the data, it is impossible to get absence information. Various factors, taphonomy, affect the frequency of animal remains in the region and the excavation area. Since archaeological excavations are carried out in human settlements, the diversity and density of animal species found here are related to the behavior of the people who lived here and the past animal populations in the environment. The density of hunting of this species by humans, whether it came to human settlements accidentally, and the population density of the species in that region are the leading ones.

This chapter's first three sections are concerned with emerging extinction date estimation studies in different fields, paleontology, ecology, and archaeology. The last section concerns statistical methods used in extinction date estimation. In the last section, statistical methods are discussed in two groups frequentist and Bayesian approaches in extinction estimation time studies.

2.1 Extinction Date Estimation Studies in Paleontological Literature

Because the last fossil record cannot point to the extinction date of species, statistical approaches are needed to estimate possible proximate extinction dates. Fossils' occurrence range generally is estimated as a confidence interval at each endpoint of fossil records, the first occurrence, and the last occurrence of species (Patzkowsky, 2012). It can be applied to local stratigraphic fossils or global occurrence data.

Fossil taxa handle biostratigraphic data in paleontology and estimate temporal occurrence ranges of species with fossil records on biostratigraphic sections. Fossil records are highly biased in analytical paleontology (Smith, 1994). Therefore, understanding the biased sources is essential for each fossil record. Sedimentation rate, preservation potential of fossils in sediment, and sampling intensity are primary variation sources for fossil distribution over time (Smith, 1994).

Each statistical method to estimate extinction date has assumptions about fossil records and their distributions. Assumptions about data have limited the feasibility of these methods for every fossil record data. Radiometric dating errors, sampling rates, fossil preservation, and taphonomy are the main limiting issues about fossil records. Some methods accept these problems as uncertainty about the record and have additional equations to consider these uncertainties, which are handled as the recovery potential of fossils (Bradshaw, 2012).

Alvarez et al. (1980) published an early example of inferring the extinction time of species in paleontology. They assumed species went extinct instantaneously with the bolide impact associated with the Chicxulub crater, which is thought to have caused the Cretaceous–Paleogene mass extinction. Their study helped to emerge new questions and thoughts on more reliable estimation methods of extinction dates. Signor and Lipps pointed out gaps in fossil records of that period and brought a different approach. They suggested that extinction had begun before the bolide impact. Based on that suggestion, they have conceptualized that the last detected fossil record cannot be the last individual of that species. After this approach, many mathematical and statistical techniques were developed with different assumptions according to record type, period, observation effort, and many other factors.

Analysis of stratigraphic gaps between fossil recordings was firstly done by C.R.C. Paul (1982). Paul used intervals between fossil recordings to calculate the possibility of fossil range endpoints for the first time. Then Springer and Lilje modeled occurrence records as a broken stick model and they assumed gaps between records have exponential distribution, in 1987. Then Strauss and Sadler pointed to Springer and Lilje's faulty assumption about the distribution of occurrence records that models gaps as a broken stick model, and they developed confidence intervals for the endpoints method.

David Strauss and Peter M. Sadler's method (1989) is the earlier statistically valid example in the paleontological literature to estimate extinction time from fossil remains in the stratigraphic section. They developed a method that calculates confidence intervals for true endpoints (origination point and the extinction point of the species) of the fossil remains in a stratigraphic section. This method focuses on a total gap, occurrence record numbers, and the distribution of intervals between fossil occurrence records.

Strauss and Sadler accept that the Poisson probability distribution is valid for the distribution of fossil records. This fundamental method accepts that the probability of occurrence does not change over time. In other words, the model assumes that the fossil data is uniformly distributed (Strauss, 1989). This method mainly focused on the gap length between records, a record number, and the last record time.

Marshall (1997) has developed a generalized model to calculate intervals on the position of the fossils on a stratigraphic section with no strict assumptions about the probability of the occurrence records. This method has a "recovery potential" function which is the possibility of finding a fossil in different environments.

Holland implemented the multivariate ordination method to Marshall's, which considers environmental parameters in the stratigraphic section. Holland applies detrended correspondence analysis (DCA), a multivariate palaeoecological method, and calculates the probability of fossil occurrence with environmental gradients according to species' ecological needs (Holland, 2003).

Bradshaw et al. point to variation in the fossil data and modify McInerny et al.'s ecological method (in Chapter 2.2) to manage the effects of variation and uncertainties in the record data set. They apply Gaussian resampling to data to handle uncertainties, like dating error or sampling effort and weigh the samples depending on the most recent fossil record. Bradshaw et al. described this method as "The full Gaussian-resampled, inverse-weighted McInerny et al. method" (GRIWM)(Bradshaw, 2012).

Schueth et al. developed another method, the probable datum model, to model the origination and extinction of nannoplankton fossil species which are microscopic marine fossils. This method uses abundance data with preservation and the counting process. This method was developed for nannoplankton fossils and can be used for any marine microfossil taxa (Schueth, 2014).

Wang et al. developed a Bayesian method to estimate extinction time. This method uses the recovery potential of fossils as a Beta probability density function. This method is also flexible to work with or without fossil recovery data.

2.2 Extinction Date Estimation Studies in Ecological Literature

Systematic field survey records, DNA, fecal or hair samples from the field, a sound record of species, photographs, or coincidental sight records are the primary source of sight record data for ecological studies. Sight records can be gathered with different methods. All observation records, certain or uncertain sights, of a species under extinction threat are critical for ecological studies. For this reason, if a species extinction is not verified yet and there is a possibility of finding hidden populations of this species, unconfirmed, anecdotal, or suspicious observations are also considered in the evaluations and verified observation records. Statistical analysis of these inaccurate data is problematic to define in equations for the extinction date estimation. The first extinction prediction models are based on only verified certain sight records. Later new methods weighted uncertain data to estimate the extinction date. Some models only require certain or uncertain presence data, but others also require unsuccessful observation data. For example, Burgman (1995) proposed the Runs Test, which assesses the likelihood of observing zero sightings which is absence data over a period of time.

The method published by Robson and Whitlock (1964) is the most basic and earlier approach to estimating extinction time in ecology. This method only calculates the interval between the last two records as more time to change to sight this species.

Andrew Solow developed frequentist and Bayesian ecological methods (Solow, 1993a). This method accepted the probability of occurrence data has the same probability for each observation, which means the population is constant. This model also has restricted assumption about sampling effort and accept it is constant over time. In another study the same year, Solow developed models for populations whose number of individuals decreased during extinction. This model assumes extinction as a non-stationary Poisson process (Solow 1993b).

Burgman et al. developed another model from Solow's equation to infer an extinction threat from museum collections data which assumes the sight record rate is constant and sighting is a stationary Poison process (Burgman, 1995). This method is known as a Runs Test (McCarthy, 1998). The absence of records and the multiple records account for the extinction calculation's probability in Burgman's method.

McCarthy studied museum record data and presented the new version of Solow's method as a partial Solow equation that extends calculations with the collection effort of samples. However, this method is not well enough for the poorly recorded species.

The model developed by Solow and Robert in 2003 is based on the truncation point inference model developed by Robson and Whitlock (1964). Beyond that, it is a non-parametric method for modeling situations where the first sight date and sighting rate are unknown (Solow, 2003). This model is useful for a few sight records.

In the same year, Robert and Solow used the sight data of Dodo birds as an example and developed a new model for inferring extinction dates. This non-parametric model has an extreme order statistics perspective and only uses the upper tail of the distribution. This new method is based on the Optimal Linear Estimator (OLE), also known as Cooke's Estimator (Cooke, 1980). The advantage of this method is that it requires very few assumptions. This model only uses the last recent records (k) of species and assumes that this data has Weibull extreme value distribution (Robert and Solow, 2003). However, it is unclear what k should be here. If k is extensively small, it will not give the expected result due to the small sample size. However, if the k is remarkably large, this time, it conflicts with the asymptotic argument.

McInerny (2006) developed another approach to estimate extinction date in ecological literature that has assumed occurrence records have a uniform probability.

Solow developed a new Bayesian model in 2011 that considers uncertain sights. This model assumes that uncertain sights occur after certain sights on the timeline (Solow, 2011). Later, new Bayesian techniques emerged, and certain and uncertain sights were handled together.

2.3 Extinction Date Estimation Studies in Archaeological Literature

Estimation of species, or phenomena, extinction time in archaeology is a relatively new research area. Recent examples from archaeology mainly focus on prehistoric times. Archaeological data has more similar to geological information than ecological data. But museum data, frequently used in ecological studies, have some common points with archaeological data. Animals found at excavation sites have some unique properties. These wild animal remains are generally hunted and moved to human settlements. A small number of wild animals accidentally went to settlements. Therefore, their occurrence probability is also affected by the choices of an occupant of the settlement. That means the animals in human settlements did not have proper samplings like ecological studies, or just random environmental events, like geology, caused the accumulation of animal remains in human settlements. This situation is essential to consider choosing an extinction date estimation method for archaeological animal occurrence records.

Du et al. estimate origination and extinction confidence intervals with Strauss and Sadler's method for an extinct hominin species *Australopithecus anamensis–afarensis* lineage (Du et al. 2020). The distribution of fossils in the timeline is accepted as a Poisson process, and they estimate the position of this species in the evolution of hominins. Du et al. first tested whether the temporal distribution of the fossil records adhered to the uniform distribution assumption or not. Then they applied Strauss and Sadler's method to assess extinction dates.

Key et al. used extinction estimation methods to reconstruct the temporal boundaries of archaeological phenomena with archaeological artifacts. Key et al. used Roberts and Solow's OLE method to infer the end of the Acheulean culture. Their study used data from the last ten archaeological settlements where stone tools were produced in the Acheulean culture. In this study, Key stated that the sporadic nature of the archaeological artifacts made them compatible with the OLE model assumptions (Key, 2021).

Briefly, Key et al. explained the relationship between the three assumptions in the OLE method and the archaeological artifacts. Firstly, the OLE method assumes that the temporal boundaries of the investigated phenomenon continue after the last seen date. This assumption can be safely accepted for the archaeological record. Secondly, OLE assumes that all observations are independent. Archaeologically, this assumption means that different people made these artifacts. For animal remains, this assumption implies that animal remains belong to other animal individuals.

Furthermore, thirdly, OLE assumes that the effort to observe the investigated phenomenon does not decrease to zero even if it changes. Key points out that archaeological search efforts may differ between different sediments. Nevertheless, ongoing archaeological studies and excavation efforts are generally consistent with this assumption.

2.4 Statistical Methods in Extinction time Estimation Studies

Observing all individuals in a population is impossible. Because of that, deciding on a species' extinction time only by observing occurrence records is biased. Many methods have been developed to detect situations of populations decrease or increase. These methods generally require demographic data and systematic observation. However, observing the population demographics of some species, and for fossil species, is unfeasible. Therefore, numerous statistical models were developed to estimate the extinction date for species.

The previous chapter briefly mention the historical development of statistical methods by study field. In this Chapter, the statistical approaches of these methods are discussed in more detail. Parametric and non-parametric frequentist methods and several Bayesian methods have been developed to estimate extinction time. Frequentist methods are also often mentioned as probabilistic methods in the literature. Frequentist methods generally focused on calculating confidence intervals for extinction time. Bayesian methods are used in inferring extinction time as well. Bayesian methods give both an estimated interval for extinction time and a probability for extinction time being in a certain interval.

2.4.1 Frequentist Methods

Frequentist or probabilistic methods are developed to calculate confidence intervals for extinction dates based on occurrence records data. Occurrence records data is ordered time data, which detects species records on a timeline. These methods assume that extinction occurred after the last occurrence record and the sampling probability of occurrence records is uniform.

Both parametric and non-parametric models are developed according to the features of different data sets. In non-parametric models, the beginning and end of the sampling time are unclear. These methods focused on only given data as a part of the whole distribution of occurrence records.

Robson and Whitlock (1964), the first example of the extinction time estimation models in ecology, published a non-parametric method to estimate the truncation point of distribution of the independent samples for biological systems. The last two records are used in this earliest method to estimate the extinction period. This simple method is;

$$T_E = t_n + (t_n - t_{n-1}) \quad (1)$$

Where T_E is a time of extinction, t_n is the last record date, and t_{n-1} is the record date before the last record date.

This method could give some insights into the species' extinction status, but it is not fully functional. This method only adds time intervals between the last two records to the last occurrence records. The performance of this method in estimating extinction times has not met to expectations.

One of the most fundamental methods from paleontological literature is Strauss and Sadler's method (1989). This method is developed to estimate the extinction of the fossil taxon in a stratigraphic section. Strauss and Sadler argue that the Poisson probability distribution is valid for the distribution of fossil records. The probability of occurrence records does not change over the timeline for his method. In other words, it assumes that occurrence data is uniformly distributed (Strauss, 1989). Strauss and Sadler focused on gaps between occurrence records and calculated confidence intervals for extinction according to mean gap length. After these methods, many methods handle occurrence records distribution as a Poisson process.

Andrew Solow developed two models (frequentist and Bayesian approach) by considering the sight data of the Caribbean Monk Seal (Solow, 1993a). The Bayesian method is not widely used. Frequentist methods are commonly used in ecological studies. This method is developed for ecological studies, and data is collected from field surveys with an observation period of (0, T). The primary assumption about sight data is that the probability of sighting an organism is the same for each field observation. This model also makes a restrictive assumption about sampling effort by assuming it to be constant over time. It means that the sampling study in the field is carried out at a constant frequency, and each sampling is carried out with the same methods. Then ordered records ($t=(t_1, t_2, t_3, t_4, \ldots, t_n)$) taken at this field survey were assumed to follow a Poisson Process. The null hypothesis is that extinction has not occurred.

$$H_0: T_E \leq T \ , \ H_1: T_E < T$$

Where T_E is the extinction date and T is an upper end point of the observation.

In this model, sight data are accepted as a stationary Poisson process. In other words, in this model, records are at a constant speed as $\lambda(t) = \lambda$ and decrease to zero when extinction occurs. T_n is the last record of the n records. Therefore, T_n has the same distribution under H_0. Solow mentions that it follows the distribution theory of the sample maximum. That means the probability is conditional on a T and T_E. The P-value for the observed value is calculated with this equation;

$$P\ (T_n \leq t_n | T_E = T\ ) = (t_n/T)^n \qquad (2)$$

T is the number of years after the first sight record, n is the number of sights, and t_n is the years between the last sighting and the first sighting. In other words t_n refers to the time duration between the first and last occurrence records.

For desired significance level α power of the test given by;

$$P(T_n \leq \alpha^{1/n}\ T | T_E < T) = \alpha(T/T_E)^n \qquad (3)$$

In another study the same year, Solow developed models for populations whose number of individuals decreased during extinction. This model assumes extinction as a non-stationary Poisson process. Therefore sighting speed is not constant and it is accepted as $\lambda(t) = \exp\{(a + bt)\}$ (Solow 1993b).

Burgman et al. (1995) developed another model to estimate an extinction threat from museum collections data which assumes the occurrence record rate is constant and the occurrence of the records is a stationary Poison process (Burgman, 1995). Burgman's method approaches discrete periods of the observation timeline, allowing to consider of multiple occurrence records for estimation calculations. This method is known as a Runs Test (McCarthy, 1998). Burgman's method uses the absence of records and multiple records to calculate the probability of extinction.

McCarthy's method (1998) the partial Solow equation, extends calculations with the collection effort of samples. McCarthy accepted that sampling is a non-homogeneous Poisson process and uses a discrete-time equation. However, this method may not be reliable for species with poorly recorded occurrence data.

Marshall (1997) has developed another model with no strict assumptions about $\lambda(t)$ in paleontology literature. This method has a "recovery potential" function. This recovery potential function is specialized for fossil taxa in Marshall's method, and it needs the total probability of occurrence of the species. Therefore, applying this method to all occurrence records data is not possible.

The model developed by Solow and Robert (2003) is based on the truncation point inference model developed by Robson and Whitlock (1964). The occurrence record rate is unknown for this non-parametric method (Solow, 2003). The only restriction of this model is that records are accepted as an independent. This model is useful for low-number records.

In the same year, Robert and Solow developed a new model an extreme order statistics perspective for estimating extinction date. This non-parametric model only uses the timeline's upper tail of the occurrence distribution. This new method is based on the Optimal Linear Estimator (OLE), also known as Cooke's Estimator. This model assumes that data has Weibull extreme value distribution and use only the most recent records (Robert and Solow, 2003). It is questionable how many occurrence records will be optimal for the OLE method. If the most recent occurrence number, *k*, is too large, it conflicts with the asymptotic argument of the method. If *k* is overly small ($k > 5$), accuracy decreases (Boakes, 2015). This method is one of the least restricted methods. The only strict rule is that observations and records should be independent, and sampling effort never falls to zero. This method is highly used in both ecology and paleontology. Besides, it has been preferred for archaeological studies recently.

McInerny (2006)'s method assumed that occurrence records were uniformly distributed. McInerny aimed to develop a method that can quickly apply to conservation plans of newly discovered and poorly recorded species with variation in their occurrence records. To ensure that this method reduces the influence of the observation time (McInerny, 2006).

2.4.2 Bayesian Methods

Bayesian methods calculate the probability that a species is extant based on the given occurrence records. Generally, many Bayesian approaches need a prior distribution to estimate the posterior distribution. However, a prior of extinction is unknown in many cases. Many Bayesian methods for extinction time estimation studies developed different approaches to overcome it.

Strauss and Sadler (1989) define an equation to calculate posterior distribution to estimate extinction in their study as an earlier example of the Bayesian method. However, this Bayesian approach is not used because their prior assumption is only limited to the stratigraphic section and is also unrealistic (Alroy 2014).

Solow published a Bayesian approach and the frequentist method in his work in 1993 (Solow 1993a). Later Solow (2011) and his colleagues developed a way to use this method to use uncertain sight data in their extinction date estimation studies (Solow et al. 2011). In this method, it is assumed that uncertain sights follow the stationary Poisson process like certain sights. However, in this model, uncertain sights come after certain sights on the timeline. Later, Solow and Beet (2014) modified this method and introduced a new model in which certain and uncertain sight records could coexist in the same period. The same study also includes a second Bayesian method, which assumes that uncertain and certain records may be of different quality.

Later, Thompson (2013) and Lee (2014) developed new methods that added uncertain sights to the model with the Bayesian approaches. It is difficult to define the prior possibility of extinction or the prior possibility of extension in these Bayesian models. However, these models are usually treated as 0.5 for prior possibility. It is generally possible to test the hypothesis with the Bayes factor in these developed models.

The Adaptive Beta method, developed by Wang et al., is another Bayesian method that models the recovery potential of fossils. This method uses recovery potential as a Beta probability density function. The advantage of this method is that if the recovery potential is unknown, this method still works (2016).

In the Bayesian updating model, Thompson and his colleagues (2019) divided the surveys after the last sighting date into two groups as unsuccessful and successful and calculated the cumulative Bayes Factor over the years according to the probability of occurrence in the surveys after the last sighting and deduced when the species might have disappeared. This study estimates extinction according to active and passive surveys and whether the species is observed (successful or unsuccessful) (Thompson, 2019).

Kodikara et al. (2021) developed a Bayesian hierarchical approach to estimate extinction time for certain and uncertain sight records for ecological records. This method focused on posterior probabilities rather than the Bayes factor. They have developed two models to analyze two different data record types. One is for certain data only, and the second is for uncertain and certain sight data. This method uses the likelihood of sighting data as prior knowledge of the Bayesian approach in the model. This method assumes that the sighting rate is constant.

CHAPTER 3

METHODOLOGY

Extinction date estimation studies can be carried out with different data in different fields. Previous chapters are the summary of approaches to how species ordered point data was analyzed in the literature.

In this chapter, our methodology is explained to estimate the extinction times of extinct Anatolian mammals. Our methodology comprises of two parts. Section 3.1 is the methodology for data collection and organization. Section 3.2 is statistical methods to analyze the data collected.

3.1 Methodology For Data Collection

This thesis is aimed to reveal the wild mammal species that lived and disappeared in the last 10 thousand years in Anatolia and to estimate when they have gone extinct. For this, archeological records were collected and compared with the current Anatolian wild mammal list. Archaeological records of wild animal species that lived in the last 10 thousand years were compiled from literature and databases. In addition, the current Anatolian mammal fauna was found through a literature review. By comparing these two datasets, extinct mammal species were revealed. Of the species that were later found to be extinct, those with sufficient data for statistical analysis were selected for extinction date estimation studies. Records of the selected species were arranged and made ready for analysis.

First, data sources, collection and organization methods, and identification of extinct species were given in Sections 3.1.1., 3.1.2, and 3.1.3. Then data exploration using data description methods in section 3.1.4.

3.1.1 Data Sources, Collection, and Identifying Extinct Species

The extinct wild mammalian species in Anatolia during the Holocene are the main object of this thesis. Statistical analyses were carried out on animal remains occurrence records obtained from archaeological excavations' publications and databases. This study collected data from the Master's Thesis of İlkem Gürgör (2016), the Doctoral Thesis of Jennifer Crees (2013), the Open Context database, and other current excavation reports and recent publications. The study focuses on mammalian species because they have relatively high occurrence records. The accuracy of the data obtained from the studies of both İlkem Gürgör and Jennifer Cress was checked from their references and then included in occurrence record data collection.

Gürgör evaluated the socioeconomic and socio-cultural characteristics of societies in Anatolia from the Paleolithic era to the Ottoman era in her thesis titled "Anatolian zooarchaeological remains from the Paleolithic Age to the present" based on archeozoological studies. In this work, she classified the archeozoological occurrence records from various publications by dividing them into periods. Her primary sources are books, journals, articles, reports of excavation results meetings, theses, and databases. She compiled domestic and wild species occurrence records from all taxa and grouped them in periods and locations. In Gürgör's study, radiocarbon dating was given for some records, and only the era names were given for others. References of these uncertain dates were checked from the literature.

Jennifer Crees also compiled data from Turkey in her doctoral thesis titled "Dynamics of large mammal range shifts and extinction: evidence from the Holocene record of Europe," published in 2013. These data include domestic or large wild mammal species and do not cover all mammalian taxa.

In addition, we conducted a literature review to identify taxa and occurrence records that were not included in the studies by Gürgör or Crees.

The current Anatolian fauna checklists were used for comparison with the archaeological data to identify extinct species. For this purpose, Nihat Turan's book "Mammals" (Turan, 1984), Şakir Önder Özkurt's book "Turkey Mammals" (Özkurt, 2020) based on the Tramem database, and "The Vertebrate Biodiversity of Turkey" compiled by Ahmet Karataş et al. book chapter (Karataş, 2021) was used. After comparing mammals' archaeological and modern occurrence record data, extinct wild mammal species of Anatolia to be used in the extinction date estimation were determined.

3.1.2 Organizing Occurrence Records of Extinct Species

After the extinct mammal species in Anatolia were determined by comparing the archaeological data and current faunal data, the identification number (Sample_ID) given to each archaeological occurrence record, species name, the name of the excavation/location, age of the record, and its reference saved as a table. All dates in this study are reported in Before Present (BP), using 2022 as the reference point. BP is not accepted as 1950 because of some historical occurrence records after 1950.

The following procedure was used to organize the occurrence record data:

1-) The records of extinct mammal species presented by Gürgör in her thesis were recorded in separate Species Tables for each species.

2-) The data in Crees' thesis and the data obtained from Gürgör's thesis were compared and processed into these Species Tables. Missing occurrence records were identified from Crees or Gürgör's data, added to the Species Tables, or ignored the repeated records.

3-) Inconsistent records were checked from their sources, verified ones were included in the tables, and unconfirmed ones were removed.

4-) New excavations and remains belonging to extinct taxa, which are not in the data of Gürgör and Crees, were searched by literature review and Open Context (https://opencontext.org) database search, and missing occurrence records were added to the previous tables. Representation of a Species Tables (species occurrence record table) in Figure 1.

5-) The midpoint of the date range is calculated using the method described by Crees and Key. (Crees, 2013, Key, 2021). The average of lower and upper time points is calculated as a midpoint. This midpoint was used in extinction estimation studies.

Species	Sample ID	Country	Site	Latitude	Longitude	Date Info	Dating Method	Calibrated	Lower_interval	Upper_interval	Reference	Comments	Data Source
Bison bonasu	bb_01	Turkey	Körtepe	37,86837	40,09244	LN	REL	6000-3000 BC	-6000	-3000	von den Driesch 1976; Uerpmann 19		JC
Bison bonasu	bb_02	Turkey	Norşun Tepe	38,68097	39,22640	LN	REL	6000-3000 BC	-6000	-3000	Boessneck & von den Driesch 1976;		JC
Bison bonasu	bb_03	Turkey	Norşun Tepe	38,68097	39,22640	IA	REL	1200-300 BC	-1200	-300	Boessneck & von den Driesch 1976;		JC
Bison bonasu	bb_04	Turkey	Tepecik	38,58333	39,46667	LN and	REL	6000-3000 BC	-6000	-3000	Boessneck & von den Driesch 1979;		JC
Bison bonasu	bb_05	Turkey	Büyüktepe Hö	40,25000	39,83000	Iron Age	REL	1000-550 BC	-1000	-550	Howell-Meurs 2001		JC
Bison bonasu	bb_06	Turkey	Sös Höyük	39,99380	41,52230	LC	RC	3500/3300-3000 ca	-3500	-3000	Piro and Crabtree, pers. comm. 2010		JC
Bison bonasu	bb_07	Turkey	Sös Höyük	39,99380	41,52230	EBI	RC	3000-2800 cal B.C.	-3000	-2800	Piro and Crabtree, pers. comm. 2010		JC
Bison bonasu	bb_08	Turkey	Sös Höyük	39,99380	41,52230	EBII-III	RC	2800-2200 cal B.C.	-2800	-2200	Piro and Crabtree, pers. comm. 2010		JC
Bison bonasu	bb_09	Turkey	Sös Höyük	39,99380	41,52230	MB	RC	2000-1500 cal B.C.	-2000	-1500	Piro and Crabtree, pers. comm. 2010		JC
Bison bonasu	bb_10	Turkey	Pınarbaşı	37,60000	32,90000	4550±7	RC	4550±70 BP	-2670	-2530	Carruthers 2004 http://www.liv.ac.uk/s		JC
Bison bonasu	bb_11	Turkey	Pınarbaşı	37,60000	32,90000	BAI 572	RC	BAI 5725±65 BP	-3840	-3710	Carruthers 2004 http://www.liv.ac.uk/s		JC
Bison bonasu	bb_12	Turkey	Yeni Kapı	41,00607	28,95194	669–77	RC	669–770 AD	669	770	Vedat Onar	https://link.spr	EE

Figure 1 The Species Tables, which organize collected occurrence records as a spreadsheet, provide an excerpt from the dataset of occurrence records.

3.1.3 Occurrence Record Data Preparation for Extinction Time Estimation Analysis

This study aims to estimate the extinction dates of the extinct wild mammal species that have gone extinct in the last ten thousand years in Anatolia. Records of archeological animal remains are expressed as “occurrence records” in this study.

Extinct species with the required number of occurrence record data for extinction estimation methods were selected from whole extinct wild mammal species. Moreover, occurrence records need to be independent; that means only one record is enough for each occurrence time point for archaeological studies. Therefore, if there is more than one occurrence record for each date, one is kept, and the others are removed.

The data come from different sources and each occurrence record has been dated by different dating methods. In some studies, only relative dating methods were applied to animal remains or archaeological section, and in some other studies, absolute dating methods such as radiocarbon were applied. Since the consistency of the dating methods in the samples to be used and the dating error in the data will affect the extinction date estimation studies, the data are divided into two groups, relative dating, and absolute dating, according to the dating methods of each occurrence record. *Panthera leo* and *Castor fibe*r species have some historical occurrence records from the literature. Historical occurrence records are added to absolute dating records. All analysis methods were applied to both groups, absolute and relative dated records, separately. Representation of one example for prepared tables in Figure 2. This table represents the derived values from previous table. (Figure 1). Duplicated records at the same time points were removed. Relative dated occurrence records were removed to another table. This table shows only absolute dated (RC, radiocarbon) occurrence records of *Bison bonasus* as an example.

In addition, *Bos primigenus* species has 18 relative dating occurrence records and 31 absolute dating occurrence records. The optimum sample size is essential for the OLE method. Therefore, smaller samples are generated from both relative dated and absolute dated *Bos primigenus* records. Out of the 31 and 18 records on the timeline, only the 15 most recent or latest ones were selected. Smaller samples covering 15 points were created and analyzed separately.

No	Sample_ID	Lower	Upper	Mean	Before Present	Dating Error	Dating Method	Interval between next record
1	bb_11	-3840	-3710	-3775	-5797	130	RC	525
2	bb_06	-3500	-3000	-3250	-5272	500	RC	350
3	bb_07	-3000	-2800	-2900	-4922	200	RC	300
4	bb_10	-2670	-2530	-2600	-4622	140	RC	100
5	bb_08	-2800	-2200	-2500	-4522	600	RC	750
6	bb_09	-2000	-1500	-1750	-3772	500	RC	2470
7	bb_12	669	770	720	-1303	101	RC	

Figure 2 Prepared occurrence records for extinction analyses, provide an excerpt from the dataset of occurrence records.

3.1.4 Data Exploration Methods

Data in extinction date estimation studies can be defined as ordered time points of species occurrence. Briefly, data is a sequence of occurrence dates of the species on a timeline. These data are very restricted in ecological and paleontological studies, and the sample is generally incomplete. These ordered record data do not have population dynamics and biological information, which restricts the study. That makes it challenging to analyze the data with commonly used methods. For example, in the Dodo sample data used by Robert and Solow in their 2003 study, the last ten occurrence records between 1598 and 1662 were used. The limited number of “ordered time data” also makes it challenging to figure out the distribution of the data.

Saltre et al. (2015) underlined that many extinction time estimation models are shaped according to the data-sampling approach to clarify the distributional prerequisites. In other words, models have assumptions according to the occurrence record sampling methods and sampling intensity. However, the actual sampling frequency of the occurrence data cannot be known. Therefore firstly, understanding and examining the characteristics of the data is essential. For this, they looked at five variables. These variables are the number of records, mean and standard error of the intervals between occurrence records, and mean and standard error of the dating errors of occurrence records. Besides, they used the coefficient of variations (CV). To understand the variation between species occurrence records datasets, they computed the coefficient of variation (CV) for the standard deviation of error and interval lengths.

The timeline from the present to the past is drawn in Figure 3. The representation of data and characteristic features were explained visually. Three sample absolute dated occurrence records (n_a) of a species and other variables are represented on the timeline. n_a, n_{a+1}, and n_{a+2} indicate the archaeological occurrence record dates of a species at three different times. θ_2 means the actual extinction time. ε denotes the dating error interval of an occurrence record. i represents the time interval between two consecutive occurrence records. This thesis used these features to calculate variables and understand data.

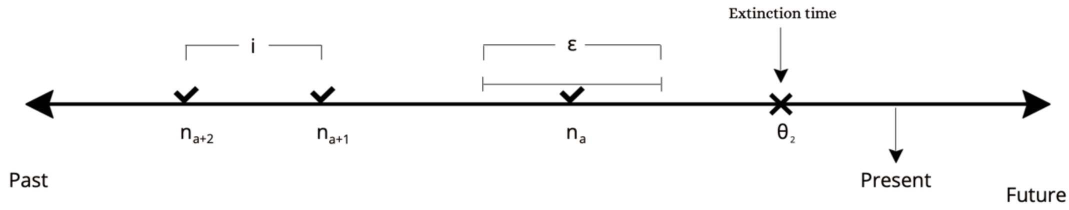

Figure 3 Visual representation of 3 absolute dated occurrence records.

Saltre et al. defined five variables for each species' occurrence records. Table 1 provides a summary of explanations of five characteristic variables (n, î, $σ_î$, ε, $σ_ε$) and other variables used to calculate those five characteristic variables used to explore data. A detailed explanation is below the table.

Table 1 Explanations of symbols, five characteristic variables, and other variables used to calculate five characteristic variables.

Symbol	***Explanations of symbol***
n_a	Number of records with absolute dating
n_r	Number of records with relative dating
n	Total number of records (na + nr)
$\hat{i}$	Average of interval lengths
$σ_{\hat{i}}$	Standard error of interval lengths
$CV_{\hat{i}}$	Coefficient of variation of interval lengths
h	The whole interval between upper and lower dates
t_l	The first occurrence date
t_u	The last occurrence date
$ε$	Average of dating error
$σ_ε$	Standard error of dating error
$CV_ε$	Coefficient of variation of dating error

n_a Number of records with absolute dating

n_a represents the total number of records dated by absolute dating methods such as the radiocarbon method. Also, historical records are included in n_a.

n_r Number of records with relative dating

n_r represents the number of records dated by relative dating methods.

n Total number of records (n_a + n_r)

n is the sum of n_a and n_r, that is, the number of all records.

$$n = n_a + n_r$$

î Average of interval lengths

î is the average of successive gaps between all occurrence points of a species. For example, if we have 4 occurrence records from 4 different time points, n_1, n_2, n_3, and n_4, firstly occurrence records are ordered from oldest to newest. Then calculate the average of intervals between them.

$$\hat{\imath} = \frac{(n_1 - n_2) + (n_2 - n_3) + (n_3 - n_4)}{n - 1}$$

$\sigma_{\hat{\imath}}$ standard error of interval lengths

$\sigma_{\hat{\imath}}$ is the standard error of interval lengths between records.

$CV_{\hat{\imath}}$ Coefficient of variation of interval lengths

The coefficient of variation of interval lengths is calculated by dividing the standard deviation (STD) of interval lengths by î and multiple by 100. It is percentile. These relative measurements give us a variability of the mean of interval lengths between occurrence records.

$$\boldsymbol{CV_i} = (\text{STD}\hat{\imath}/\hat{\imath}) * 100$$

t_l The first occurrence date

t_l is the value that gives the oldest record of a sample.

t_u The last occurrence date

t_u is the date of the closest, youngest, record to the present.

h The whole interval between upper and lower dates

h indicates the time interval in which all the presence records were found. In other words whole time frame of occurrence records. It is found by subtracting t_u from t_l.

$$h = t_l - t_u$$

ε Average of dating error

All dating methods, absolute or relative, date the records of their occurrence records with a dating error. ε is the average of the dating errors in absolute or relative dated samples of the same species.

σ_ε standard error of dating error

σ_ε is the standard error of the ε.

CV_ε Coefficient of variation of error

The coefficient of variation of the error is divided by the standard deviation of the dating error by an average of the dating error and multiplied by 100. CV_ε gives variability of the dating error of the samples.

$$\boldsymbol{CV_\varepsilon} = (\mathrm{STD}\varepsilon/\varepsilon) * 100$$

Not all these five variables, n, î, $\sigma_{î}$, ε, σ_ε equally affect every extinction time estimation model. For example, the variance of the intervals between the occurrence records is essential for the methods that assume data is uniformly distributed, such as the Strauss and Sadler method. However, these methods do not consider the uncertainty/dating error, the dating error mean, and error variance variables of the occurrence records.

Descriptive statistics of occurrence data were examined according to the data set characteristics defined by Saltre et al. (2015). To reveal the general features of the Anatolian occurrence records, the five characteristics were calculated on Excel Spreadsheets for each species, with both absolute and relative dating. The number of archaeological occurrences for each species, the dating error, and intervals between the occurrence dates were examined separately. To explore data sets, the intervals between occurrence records, mean and the standard error of these intervals, mean and the standard error of dating error were calculated as characteristics of the occurrence record data set. To understand interval variance between species, the coefficient of variation is calculated (Table 4, Table 5, and Table 6).

3.2 Extinction Time Estimation Methods

Extinction date estimation methods have different approaches, from ecology to geology. Some of these methods are also used for archaeological records. In this thesis, two frequently preferred methods, Strauss and Sadler's method and Roberts and Solow's OLE method, were chosen. Strauss and Sadler's method is fundamental in geology to estimate the endpoints (origination and extinction) of fossil horizons, which is also applied to occurrence records. Roberts and Solow's OLE method is widely used in ecology to paleontology. It is also recently preferred in archaeological studies, providing flexibility with few prerequisites. Each method has advantages and disadvantages, according to the data.

The first reasons for choosing this widely used robust method are about variables and the approach of the method. The method of Strauss and Sadler does not use dating error, or uncertainties, in calculations as a variable. This feature provides an advantage for the data used in this thesis. Also, Saltre et al. analyze this method with various datasets, which are uniform or not. And they determine that the most crucial variable is the number of occurrence records. The performance of the methods is not directly affected by the different lengths of the intervals between datasets; rather, what matters is the distribution of these interval lengths. This is also important for this thesis because datasets with different characteristics were analyzed together.

The second method is the OLE method of Roberts and Solow, which are also widely used and preferred because of the minimal number of requirements of the method. The data used in this thesis does not come from a single geological stratigraphic section or obtained by field studies on living organisms in nature. That means data has many variations since it is a collection of many different data sets. This is critical to determine the distribution of the data and the amount of dating error. The non-parametric OLE method with minimum assumptions was appropriate for analyzing these highly variable data gathered from different sources. Apart from this, the fact that it is suitable for working with heterogeneous data is a relevant situation for analyzing the data in this thesis. In addition, the detailed information Key et al. provided on using the OLE method in archeology has been explained in detail in the previous chapter (2021).

3.2.1 Strauss and Sadler Method

The method calculates point estimation and confidence intervals for the origination and extinction dates. The assumptions and details of the method are given below.

Strauss and Sadler's method has strict assumptions about data: 1. Fossil occurrences and sampling intensity are assumed to be randomly distributed between two endpoints, 2. Moreover, intervals (or gaps) between fossils must be independent, 3. There is no correlation between intervals. This means fossils distribute uniformly along the timeline.

Strauss and Sadler indicate that randomness is referred to not for fossil records themselves but for the processes that form the fossil record. In this type of data, there are two variables of interest: 1. fossil records (occurrences), 2. the time gap between the fossil records. The definitions of these variables are given below:

1. Fossil Record: Fossil record means fossil deposit in sedimentary rock. Fossil records sequences are not always imprints of correct sequences of events. The fossilization process and environmental factors alter the fossil record in the long run.

Also:

First Fossil Record: This is the first occurrence of this species' fossil in sediments. The first fossil record is not the actual first organism of this species, but it is a reference point to estimate the origination time of this species.

Last Fossil Record: The last fossil record is this species' last appearance on the whole stratigraphic section. The last fossil record itself does not directly indicate extinction time. It is essential to estimate extinction time.

2. Time gap between fossil records: Time gaps between fossil records indicate how long each record comes after the previous record.

For statistical analysis of record data, one must determine the distributions used for these variables in (1) and (2) above. For (1), assuming that fossil records are independent of each other, the Poisson process is a valid distribution to use. For (2), gaps between fossil records can be assumed to follow Exponential distribution with the parameter lambda (λ), where lambda denotes the inverse of the average gap length. Its probability density function is given below (4).

$$f_x(x) = \lambda e^{-\lambda x}, x > 0 \qquad (4)$$

This method considers the fossil records observed between the origination point of a taxon (θ_1) and the extinction point of a taxon (θ_2) as a randomly selected sample from a uniform distribution. In this method, origination and extinction time (θ_1 and θ_2) are unknown parameters to be estimated using the data. Also, hypothesis testing on them is of interest.

Estimation of θ_1 and θ_2

Let "n" denote the number of uniformly distributed occurrence records, Y denotes the first fossil record, Z is the last fossil record, and Y and Z each have uniform distribution with endpoints θ_1 and θ_2. Upper-case letters denote random variables, and lower-case letters correspond to their realized values.

One of the powerful methods of estimation is Maximum Likelihood Estimation. Maximum likelihood estimators of θ_1 and θ_2 are z and y, respectively. However, $Y > \theta_1$ and $Z < \theta_2$ with probability 1. Therefore, Y and Z are biased estimators. To obtain unbiased estimators (E denoting expected value), we use that Y and Z are uniform on (θ_1, θ_2).

$$E(Y) = \frac{\theta_2 + n\theta_1}{n+1} \quad (5)$$

$$E(Z) = \frac{\theta_1 + n\theta_2}{n+1} \quad (6)$$

And rearrangement of equations,

$$E[(nY - Z)/(n-1)] = \theta_1 \quad (13, 7)$$

$$E[(nZ - Y)/(n-1)] = \theta_2 \quad (14, 8)$$

Thus, this led to the following unbiased estimators:

$$\widetilde{\theta_1} = (\mathrm{ny} - \mathrm{z})/(\mathrm{n} - 1) \quad (9)$$

$$\widetilde{\theta_2} = (\mathrm{nz} - \mathrm{y})/(\mathrm{n} - 1) \quad (10)$$

The variance of $\widetilde{\theta_1}$ and $\widetilde{\theta_2}$are,

$$\frac{1}{(n-1)^2} = [n^2Var(Y) + Var(Z) - 2nCov(Y,Z)] \quad (11)$$

Integration of joint density and marginal density functions to this equation leads to the following sampling variances,

$$Var(\widetilde{\theta_1}) = (\theta_2 - \theta_1)^2 \frac{n}{(n-1)(n+1)(n+2)} \quad (12)$$

$$Var(\widetilde{\theta_2}) = (\theta_1 - \theta_2)^2 \frac{n}{(n-1)(n+1)(n+2)} \quad (13)$$

To use (18) and (19) in practice, $(\theta_2 - \theta_1)$ is needed. This can be substituted with its estimate (14),

$$(z - y)(n + 1)/(n - 1) \quad (14)$$

To estimate the extinction time of the chosen taxa, Equation (10) was used to determine the point estimation of extinction time (θ_2). This is the best estimator in the sense that its variance is minimum. It is a Uniformly Minimum Variance Unbiased Estimator for the extinction time.

Confidence Intervals for θ_1 and θ_2

$\widetilde{\theta_1}$ and $\widetilde{\theta_2}$ are not distributed normally, even if n is large. Therefore Equations (9), (10), (12), and (13) cannot be used to calculate confidence intervals.

As suggested by Strauss and Sadler, the confidence interval should have the following form (Strauss, 1989). For a suitably chosen α (0,1),

For origination: $y - \alpha(z - y) < \theta_1 < y$ (15)

For extinction: $z < \theta_2 < z + \alpha(z - y)$ (16)

To calculate desired α by probability, Strauss and Sadler suggested that;

$$p_1 = 1 - (1 + \alpha)^{-(n-1)} \quad (17)$$

To find α Equation (17) can be inverted,

$$\alpha = (1 - p_1)^{-1/(n-1)} - 1 \quad (18)$$

Where p_1 is related to the level of confidence, and it can be suitably chosen with respect to the sample size. For instance, p_1 can be 0.95, 0.97, 0.99 if the sample size is sufficiently large and smaller otherwise. Another approach is to choose p_1 to attain a certain α which is related to the rate of false positives.

Spreadsheets were used to calculate point estimation of extinction and confidence intervals. The trial and error approach for α was applied with Wolfram Alpha for each calculation of trials.

3.2.1.1 Distribution of interval lengths between occurrence records

Strauss and Sadler's method depends on the interval length between occurrence records and the time between the last and the first occurrence. This method assumes occurrence records between the first and last appearance are distributed uniformly. Therefore, examining the uniformity of a sample, i.e., testing the distribution of the sample for uniformity, is essential to apply the Strauss and Sadler method appropriately.

The limited number of occurrences makes it challenging to figure out the distribution of these records. However, the distribution of data is essential for the accuracy of the Strauss and Sadler method. For this reason, different approaches are applied to explore and analyze the occurrence records distributions.

The Interval length between occurrence records described in the previous section as an “i” represents the time interval between two consecutive occurrence records in Figure 3. "i" was calculated for each species occurrence record, and distributions of these interval lengths are examined.

Du et al. visually examine the intervals between their samples using a histogram and calculate the skewness and kurtosis of interval length data because the importance of whole time length and occurrence record numbers highly affect Strauss and Sadler's methods (2020). Vogel et al. use the probability plot correlation coefficient test to understand distribution (2008).

This study examined skewness, kurtosis, and histograms of interval lengths between occurrence records were drawn. The distribution of the occurrence data was tested for uniformity with the probability plot correlation coefficient (PPCC) test as used by Vogel et al. (2008).

3.2.1.1.1 Skewness, Kurtosis, and Histogram

Skewness is a measure of the asymmetry of the distribution. These measurements help us to understand whether the distribution of interval lengths is close to being symmetrical. This measurement gives information about distribution because uniform distribution is symmetrical.

Kurtosis is a measurement related to the tails of a distribution. Uniform distribution is platykurtic distribution. And kurtosis measurement helps to understand the tails of interval length distributions.

A histogram is a visual examination tool for the frequencies and distribution of data. Histograms are visually valuable for understanding whether interval length data seem uniform.

Skewness, kurtosis, and histogram are calculated on R. Histograms were created with the *ggplot2* package, and all histograms are on the same scale and same bin. Skewness and kurtosis were calculated with the *datawizard* package on R. "Type 2" skewness method is chosen from this package. If the skewness value is around zero, that means distribution is normal, negative values indicate left-skewed distribution, and positive values are right-skewed distribution. Values below 1 indicate half-normal distribution, and values around 2 indicate exponential distribution (Makowski, 2021). Kurtosis around zero mean "mesokurtic", positive values mean "leptokurtic", and negative values mean "platykurtic" tails (Makowski, 2021).

3.2.1.1.2 Q-Q Plots and Probability Plot Correlation Coefficient Test

Wang et al. visually test the uniform distribution assumption with probability plots (2009). Probability plots are visual representations of how to fit the observed ordered occurrence record to the expected occurrence times of records. If expected times are the same as observed times, in other words, observed times represent on the y-axis and expected on the x-axis and y=x, this data is uniformly distributed. Also, this method gives a confidence limit for the uniform distribution. Although they stated that it is more important to determine the distribution of data points within the confidence interval rather than the result of the uniformity test, it is essential to demonstrate whether the data are uniformly distributed when evaluating the results of the selected extinction time estimation models. Many extinction time estimation models assume occurrence record data is distributed uniformly. However, uniformly distributed record data is idealized scenarios, and real-world data do not fit properly with uniform distribution (Wang, 2009). Moreover, they emphasize that exploring deviation from a uniform distribution is more critical than determining whether data is uniform (Wang, 2009).

Filliben described the probability plot correlation coefficient (PPCC) test in 1975 to test the normality of data. PPCC test combines both graphical and quantitative measurement. To test the uniformity of the ordered data, PPCC measures the linearity of a plot of the ordered data values against the expected values. The PPCC test to analyze the goodness of fit of probability distributions for occurrence records was applied by Vogel et al. (2008).

For example, if we have ten ordered values from i to m, $x_i \leq x_{i+1} \leq \cdot \cdot \cdot \leq x_m$, the expected values will be $p_i = i/(m + 1)$.

And test statistic of PPCC is r calculated with equation 34.

$$r = \frac{\sum_{i=1}^{m}\left(x_{(i)} - \bar{x}\right)(p_i - \bar{p_i})}{\sqrt{\sum_{i=1}^{m}\left(x_{(i)} - \bar{x}\right)^2 \sum_{i=1}^{m}(p_i - \bar{p_i})^2}} \tag{34}$$

and

$$\bar{x} = 1/m \sum_{i=1}^{m} x_{(i)} \tag{35}$$

$$\bar{p_i} = \frac{1}{m \sum_{i=1}^{m} \frac{i}{m+1}} = \frac{1}{2} \tag{36}$$

PPCC test is the best alternative to understand the distribution of occurrence records. Because the data set has low occurrence records, this method gives visual plots and a quantitative measure to understand the distribution.

The probability plot correlation coefficient (PPCC) test is conducted with 10,000 Monte-Carlo replications to calculate r-values and p-values. Then, calculated r values, p-values, and sample size (n) are interpreted as the null hypothesis is rejected or not rejected.

PPCC test hypotheses are;

H_0: Distribution is uniform.

H_1: Distribution is not uniform.

p-value < 0.05 means H_0 is rejected, and the distribution of the sample is not uniform. If p-value>0.05 H_0 is not rejected.

The R *ppcc package* (Pohlert 2020) was used for calculating the r values of the PPCC test and its p-value. *ppcc package* can perform the PPCC test for several distributions, normal, lognormal, Weibull, Gumbel, Rayleigh, uniform, and exponential, that are extended from Filliben's method by other authors. Data is tested for only uniform distribution with the "qunif" argument and 10,000 Monte-Carlo replications.

3.2.2 Roberts and Solow Method

Roberts L David and Andrew R Solow developed this non-parametric method to estimate the extinction time of Dodo. They use Cooke's Linear Estimator technique developed for estimating a range of variations of upper or lower bounds of an independent random variable with order statistics (Cooke, 1980). Moreover, this method was recently applied to archaeological records (Key, 2021).

This method assumes that the population of a species decreases before extinction and the frequency of occurrence records decreases and makes an extinction estimation only based on the distribution of the last records. Regardless of the distribution of all occurrence records, the distribution of the last records is considered to follow the Weibull distribution. In this approach, there is no presumption of sampling effort or distribution of occurrence records. This method only assumes that the sampling effort does not fall to zero. That means collecting data on the field by observations or excavations continues. There are no assumptions about sampling design or sampling intensity.

Key et al. describe the equations of this method for archaeological records. The following description is from Key et al. (2021). This method assumes that the joint distribution of the most recent occurrence records follows Weibull extreme value distribution. The method works as a sum of occurrence records weighted according to Weibull distribution.

Firstly, order each occurrence record (*T*) according to its age; k is the known oldest record: $T_1 > T_2 > ... > T_k$

Each record is weighted with a vector ($a = (a_1, a_2, ..., a_k)$), then these weighted records are summed to find the estimated extinction time ($\hat{\theta}$).

$$\hat{\theta} = \sum_{i=1}^{k} a_i T_i \tag{9}$$

$$a = (e^t \wedge^{-1} e)^{-1} \wedge^{-1} e \tag{10}$$

e, vector of k1's, and $k \times k$ matrix, $\wedge$ are needed to calculate a.

$\wedge$ is the symmetric matrix with a typical element

$$\lambda\_ij = \Gamma(2\hat{v} + i)\Gamma(\hat{v} + j)/\Gamma(\hat{v} + i)\Gamma(j)\ , j \leq i \tag{11}$$

and Γ is the standard gamma function.

T shape parameter of the joint Weibull distribution *($\hat{v}$)* of the *k* youngest occurrence records to calculate $\wedge$.

$$\hat{v} = \frac{1}{k-1}\sum_{i-1}^{k-2} log\frac{T_1 - T_k}{T_1 - T_{i+1}} \tag{12}$$

An approximate one-sided upper bound of a $(1 - \alpha)$ confidence interval for θ is*:*

$$S_U = \frac{T_1 - c(\alpha)T_k}{1 - c(\alpha)} \tag{13}$$

Where

$$c(\alpha) = (k/(-log\alpha))\hat{}(-\hat{v}) \tag{14}$$

This method is flexible for various data sets with different characteristics, and there is no assumption about data distribution. Moreover, this method has been implemented in archeological records recently. Therefore, this method is chosen as the second method of estimating the extinction time of species in this study.

To estimate extinction dates, the OLE function of the R software package *sExtinct* (Clements 2013) is used in this study.

CHAPTER 4

STATISTICAL DATA ANALYSIS

In this chapter results of the statistical analysis are presented. The chapter has two main sections. The first one is the extinct taxa list and data characteristics of occurrence records of extinct species. The second section is the results of the extinction estimation analysis of these species.

The first section represents extinct species in Anatolia in the past ten thousand years. 16 large mammalian species are found to have gone extinct. Then these species are filtered, and the ones which are not suitable for extinction estimation analyses are excluded. After filtering, only eight species remained eligible for inclusion in the extinction estimation analyses. The last part of this section presents the results of the data characteristics of these 8 species. The main aim of these parts is to analyze the general feature of data.

The second section is comprised of the results of the statistical estimation of extinction times. Estimated extinction time of both absolute and relative dated samples with two different analysis methods are presented. Strauss and Sadler method's extinction time estimation results are interpreted with distributional test results.

4.1 Extinct wild mammals and selected taxa for extinction time estimation analysis

In this section, the results of the comparison of the archeological records and the current Anatolian wild mammal list selected extinct taxa for estimation extinction time analysis and their data characteristics are given.

This list is the mammalian species found to be extinct in Anatolia, and their taxonomic positions are mentioned in section 4.1.1. Then, the occurrence records of these species, which were found to be extinct, were examined, and the species containing the appropriate number of occurrence records for extinction time estimation were selected for analysis in section 4.1.2. Finally, data characteristics of occurrence records of selected species for extinction date prediction studies are examined as relative dated and absolute dated samples in section 4.1.3.

4.1.1 Extinct wild species during the Holocene period in Anatolia

This study compared species records obtained from archeozoological data with modern Anatolian fauna checklists. The study was limited to mammal taxa, Artiodactyla, Perissodactyla, Lagomorpha, Rodentia (Castoridae), Proboscidea, and Carnivora, because of the lack of archeozoological data and the difficulty of identification. It was determined that 16 species were extinct in 6 orders. These species are listed in Table 2. Figure 4 shows percentages of species that disappeared. More than 50% of the extinct species are part of the ungulate group, which includes several taxa that have been critical for domestication and economic purposes.

Some species in the table have identification difficulties in systematic taxonomy. The taxonomic positions and taxonomic names of *Gazella subgutturosa* and some Ovis species have been changed by molecular studies recently. In addition, identifying some species from archaeological bone remains is difficult. *Dama dama mesopotamica*, Equus species, Ovis/Capra species are some of these species.

Table 2 Extinct mammal species list in Anatolia during the Holocene.

No	***Order***	***Family***	***Species***	***Common Order name***	***Common Family name***	***Common name***
1	Artiodactyla	Cervidae	*Dama dama mesopotamica*	Even-toed ungulate	Cervid	Fallow Deer
2	Artiodactyla	Bovidae	*Ovis ammon*	Even-toed ungulate	Ovis	Wild Sheep
3	Artiodactyla	Bovidae	*Bos primigenius*	Even-toed ungulate	Bovidae	Wild Ox
4	Artiodactyla	Bovidae	*Bison bonasus*	Even-toed ungulate	Bovidae	European Bison
5	Artiodactyla	Bovidae	*Gazella subgutturosa*	Even-toed ungulate	Bovidae	Goitered gazelle
6	Perissodactyla	Equidae	*Equus hemionus*	Odd-toed ungulate	Equidae	Onager
7	Perissodactyla	Equidae	*Equus africanus*	Odd-toed ungulate	Equidae	African Wild Ass
8	Perissodactyla	Equidae	*Equus hydruntinus*	Odd-toed ungulate	Equidae	European Wild Ass
9	Perissodactyla	Equidae	Equus ferus	Odd-toed ungulate	Equidae	Wild horse
10	Lagomorpha	Leporidae	*Lepus capensis*	Lagomorpha	Leporidae	Cape Hare
11	Proboscidea	Elaphantidae	*Elaphas maximus*	Proboscidea	Elephantidae	Asian Elephant
12	Carnivora	Canidae	*Vulpes corsac*	Carnivor	Canidae	Fox
13	Carnivora	Felidae	*Panthera leo*	Carnivor	Felidae	Lion
14	Carnivora	Felidae	*Acinonyx jubatus venaticus*	Carnivor	Felidae	Cheetah
15	Carnivora	Felidae	*Panthera tigris virgata*	Carnivor	Felidae	Tiger
16	Rodentia	Castoridae	*Castor fiber*	Rodentia	Castoridae	Beaver

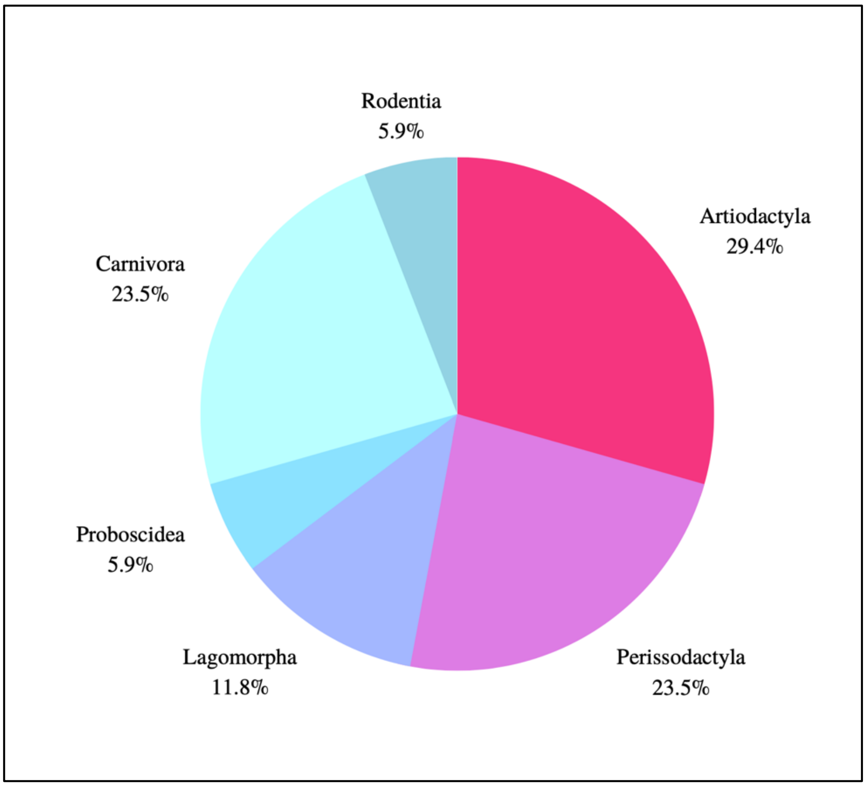

Figure 4 The figure represents the percentage of extinct species by order.

Four of the extinct species in Anatolia have disappeared globally. *Panthera tigris virgata* (Caspian tiger), *Bos primigenius* (Wild ox), *Equus hydruntinus* (European wild ass), and *Elaphas maximus asurus* (Syrian elephant) are globally extinct species. The populations of other extinct (or extirpated, locally extinct at Anatolia) species in Anatolia generally have some decreasing populations in different regions, or they face extinction threats currently. Conservation programs and re-releases are carried out in various countries for most at-risk species.

The last record dates of extinct species in Anatolian fauna vary between the beginning of the Holocene and to 20th century. *Panthera leo* (lion), *Castor fiber* (beaver), and *Panthera tigris virgata* (tiger) are the species whose existence was determined in the early and late periods of the 20th century. *Vulpes corsac* is a single record from the beginning of the Holocene, 11,710 years ago. However, the Caspian tiger *P. tigris virgata*, seen from the 1950s to 1970, has not been detected in the archaeological occurrence records.

4.1.2 Selected taxa for extinction time estimation studies

Species were selected for extinction estimation analysis in the studied date range in Anatolia. The criteria for selection of species included having sufficient occurrence record data and avoiding taxonomic identification problems such as those that result in uncertainties in the data, such as in the case of Ovis/Capra. Specifically, species with less than three occurrence records were excluded from the analysis.

It has been determined that eight species are suitable for extinction estimation studies. These species are *Bison bonasus*, *Bos primigenius*, *Dama dama mesopotamica*, *Equus hydruntinus*, *Equus ferus*, *Equus hemionus*, *Castor fiber*, and *Panthera leo*. Out of 166 occurrence records, 71 of them were dated with relative dating methods, and 95 of them were dated with absolute dating methods. The number of the absolute and relative dated occurrences of the species are given in Table 3.

Table 3 Selected extinct mammalian species with the number of occurrence records (Absolute Dating (n_a) and Relative Dating (n_r)).

Species	*Occurrence record of Extinct Species with Absolute Dating (n_a)*	*Occurrence record of Extinct Species with Relative Dating (n_r)*	*Total occurrence record of Extinct Species (n)*
Panthera leo	6	10	**16**
Castor fiber	15	9	**24**
Equus hemionus	15	12	**27**
Equus ferus	12	5	**16**
Equus hydruntinus	6	2	**8**
Dama d. mesopotamica	3	12	**15**
Bos primigenius	31	18	**50**
Bison bonasus	7	3	**10**
Total occurrence records	***95***	***71***	***166***

4.1.3 Data characteristics

To explore data sets, the intervals between occurrence records (i), mean (î) and standard error ($\sigma_{\hat{i}}$) of these intervals, mean (ε) and standard error (σ_{ε}) of dating error was calculated as characteristics of occurrence record data set. Also, the coefficient of variation (CV) of î and CV of ε were calculated. Besides the whole time length from first to last occurrence (h), the first occurrence (t_l), and the last occurrence record (t_u) are given with these data characteristics in Table 4 and Table 5 for absolute dated and relative dated samples.

4.1.3.1 Data characteristics for relative dated samples

The occurrence records of 8 species, dated using relative dating methods, were examined, with a total of 71 records included in the analysiss. Eight species have a different number of occurrence records between 2 to 18. *Equus ferus* and *Bison bonasus* have inadequate sample sizes for extinction time estimation, with 2 and 3 occurrence records respectively. *Equus ferus* also has an insufficient record number; however, it can still be analyzed with the OLE method. Table 4 provides data characteristics for each species of relative dated occurrence record data.

Averages of the interval length between records (î) vary between 400 to 1875 years. *Equus hydruntinus* has only 2 relative dating occurrence records; calculating the standard error (SE) of interval length is not applicable. *Equus ferus*, with a coefficient of variation of 51.9%, has the least variation in interval lengths between records. *E. ferus* followed by *Equus hemionus*, *Panthera leo*, *Dama dama mesopotamica*, and *Castor fiber*, respectively. *Bos primigenius* and *Bison bonasus* have high variation between interval lengths of records. Their CVs are 105.3% and 139.5%, respectively. *B. bonasus* only has 3 samples and 2 intervals; however, *B. primigenius* have the highest sample number, 18, and has 17 intervals.

Table 4 Data characteristics for relative dating records of 8 species (BP: Before Present, n_r represent the number of occurrence records, others are years.).

Species	n_r	$\hat{\imath}$	$\sigma_{\hat{\imath}}$	$CV_{\hat{\imath}}$ %	h	t_u (BP)	t_l (BP)	ε	σ_ε	CV_ε %
Panthera leo	10	722.22	208.10	86.44	6500	1172	7672	1490.00	501.54	106.44
Castor fiber	9	715.63	234.99	92.88	8522	2797	8522	944.44	366.61	116.45
Equus hemionus	12	675.00	148.60	73.01	7425	2997	10422	683.83	145.96	73.94
Equus ferus	5	937.50	243.56	51.96	3750	2772	6522	1300.00	496.99	85.49
Equus hydruntinus	2	400.00	NA	NA	400	7872	8272	1100.00	400.00	51.43
Dama mesopotamica	12	790.45	210.90	88.49	8695	772	9467	989.17	281.84	98.70
Bos primigenius	18	494.12	126.22	105.32	8400	622	9022	859.50	193.07	95.30
Bison bonasus	3	1875.00	1850.0	139.54	3750	2772	6522	1450.00	785.81	93.87

The average dating error(ε) for each species varies between 683.8 years to 1490 years. Dating error is calculated with each record's upper and lower dating points. The relative dating method is not as sensitive as absolute dating methods, such as radiocarbon dating. The relative dating method uses stratigraphic analysis and seriation techniques and has larger time periods for each record. For example, a *B. bonasus* sample dated to Late Neolithic dating as 8022 - 5022 years BP. It has 3000 years period between the start and end of the occurrence of this sample. However, the average error for *B. bonasus* species is 1450 years. Other species' mean errors are 1490 years for *P. leo*, 1300 years for *E. ferus*, 1100 years for E. hydruntinus, 989 years for *D. mesopotamica*, 944 years for *C. fiber*, 859.5 years for *B. primigenius*, 683.8 years for *E. hemiounus*.

Consequently, the average error can be close to the average intervals between occurrence records or longer than the intervals between records in some species and samples. This situation also needs to be evaluated when interpreting the estimation of extinction time analysis. The standard error between dating errors also varies between species. CVε gives variability of the dating error of the samples. The coefficient of variation in *Equus hydruntinus* is 51.4%. CVε of *E. hemiounus*, *E. ferus*, *D. mesopotamica*, *B. primigenious*, and *B. bonasus* vary between 98.7% and

73.93%. CVε of *P. leo* and *C. fiber* are 106.44% and 116.45%, respectively. *E. hydruntinus* has fewer dating error variations, while *E. hemiounus*, *E. ferus*, *D. mesopotamica*, *B. primigenious*, and *B. bonasus* have more, *P. leo* and *C. fiber* the most, respectively.

4.1.3.2 Data characteristics for absolute dated samples

A total of 95 absolute dated occurrence records of 8 different species were evaluated. *B. primigenius* has the maximum occurrence with 31 records. *D. mesopotamica* has the minimum number of records with only 3. *P. leo* and *E. hydruntinus* have 6 records, *B. bonasus* has 7, *E. ferus* has 11 records. *E. hemionus* and *C. fiber* have 15 occurrence records each. Data characteristics for absolute dated occurrence record data are provided in Table 5.

Table 5 Data characteristics for absolute dating records of 8 species. BP: Before Present, n_r represents the number of occurrence records, others are years.).

Species	n_a	î	σî	CVî %	h	tu (BP)	t_l (BP)	ε	$σ_ε$	$CV_ε$%
Panthera leo	6	1544.8	1899.76	122.98	7724	148	7872	94.67	231.89	244.95
Castor fiber	15	848.0	1011.16	119.23	11873	49	11922	161.00	382.33	238.06
Equus hemionus	15	564.0	641.23	113.64	7900	4022	11922	288.00	351.12	121.83
Equus ferus	11	491.0	187.62	38.22	5400	4622	10022	221.00	215.47	97.39
Equus hydruntinus	6	1033.0	500.96	48.50	5165	4622	9787	370.00	393.14	106.25
Dama mesopotamica	3	367.0	367.70	100.19	734	8897	9631	214.00	153.73	71.84
Bos primigenius	32	261.0	376.12	144.30	8080	3672	11752	288.00	205.67	71.45
Bison bonasus	7	749.0	870.90	116.26	4495	1303	5797	310.14	213.46	68.83

A total of 95 absolute dating occurrence records of 8 different species were evaluated. *B. primigenius* has the maximum occurrence records with 31 records. *D. mesopotamica* has the minimum number of records with only 3 records. *P. leo* and *E. hydruntinus* have 6 records, *B. bonasus* has 7, *E. ferus* has 11 records. *E. hemionus* and *C. fiber* have 15 occurrence records each.

The mean interval between records(î) varies between 261 years to 1544.8 years, *P. leo* and *B. primigenious*, respectively. CVs of interval length between records *E. ferus* and *E. hydruntinus* are 38.2% and 48.49%. Other species' CVs of interval length between records vary between 100.18% to 144.30%. Variation of interval length between records is considerably high, *E. ferus* and *E hydruntinus*. *B. primigenius* have the highest variation between interval lengths between records.

The mean dating error (ε) of species is relatively low for absolute dated samples than for relative dating samples. The lowest one is *P. leo*, 94.6 years. The *B. bonasus* is the highest one with 310.14 years. *B. primigenious* has a higher dating error average than its interval length average when comparing averages of the dating error and interval lengths between records. The highest CV of dating error is in *P. leo* with 244.9%. The lowest one is *B. bonasus* with 68.8%. *C. fiber* is 238%, *E. hemionus* is 121.8%, *E. hydruntinus* is 106.2%, *E. ferus* is 97.3%, *B. primigenius* is 71.4%, *D. mesopotamica* is 71.8%, and *B. bonasus* is 68.8%. CVs of dating error of *P. leo* and *C. fiber* are remarkably higher than others due to this species having historical records with 0 dating error. Dating error is essential for some extinction estimation models. Selected models for this thesis are not affected by dating errors.

4.1.3.3 Data Characteristics of *Bos primigenious* with reduced sample size

Occurrence records number have an impact on the confidence intervals of OLE extinction estimation analysis. It is controversial in the literature how many samples should be used in the OLE analysis (Boakes, 2015). When working with data containing many samples, the upper bound of the confidence interval gives high results in OLE analysis (Rivadeneria, 2009). This is because the given data, regardless of the main distribution, is considered to follow the Weibull extreme values distribution.

Bos primigenius has 18 and 32 samples for relative and absolute dated samples. Reduced *B. primigenius* data sets were analyzed to investigate the sample size effect on extinction time estimation analysis. Relative and absolute dated data sets are reduced to the last 15 samples, and these data characteristics are separately calculated.

Table 6 Data characteristics for limited *Bos primigenius*.

	n_{Bos}	$\hat{\imath}$	$\sigma\hat{\imath}$	$CV\hat{\imath}$ %	h	tu (BP)	tl (BP)	ε	$\sigma\varepsilon$	$CV\varepsilon$
Absolute	15	358	504,34	140.93	5010	3672	8682	218	140.51	64.59
Relative	15	421.4	534.66	126.87	5900	622	6522	721.4	809.53	112.22

Both the Absolute and Relative dated samples have 15 occurrence records in Table 6. The average interval between occurrence records is 358 and 421.4 years, respectively. CVs of interval length between records is 140.9% for absolute and 126.8% for relative dated samples. The mean dating error of absolute dated sample is only 218 years, but it is 721.4 years for relative dating samples. Besides, the CV of dating error is also high in relative dated occurrence records.

4.2 Extinction Time Estimation Results

In this section, estimated extinction times of absolute dated and relative dated samples of each species are evaluated, where the estimates are obtained by the Strauss and Sadler method and Roberts and Solow's OLE method.

Each extinction estimation result is interpreted with the assumptions of each model.

4.2.1 Strauss and Sadler Method for Extinction Time Estimation

There are statistical assumptions to be satisfied by the data for the proper interpretation of the results of the Strauss and Sadler method. It assumes that the sample is uniformly distributed. In this, firstly, the distribution of occurrence records of the samples was examined by various methods. The extinction estimation results of the samples that did not meet the assumptions of the Straus and Sadler method were evaluated accordingly.

4.2.1.1 Test of Assumptions of Strauss and Sadler Method and Uniformity of Data

Strauss and Sadler's method is one of the most fundamental methods for paleontological studies to estimate the extinction date of fossils. This method assumes data is distributed uniformly. Some of the data in this study violates this assumption. According to Wang et al., finding uniform fossil data is unrealistic (2009). This situation can be easily accepted for archaeological data. Wang. et al. also point out that data cannot fit uniform distribution but analyzing data with probability plots makes sense to understand the extinction of the study object (2009). Estimation of extinction time results are interpreted above with results of histograms, skewness, kurtosis, q-q plot, and ppcc test of interval length between occurrence records shown below.

4.2.1.1.1 Histograms, skewness, and kurtosis of the interval between records

Distribution of interval lengths between occurrence records examined with skewness and kurtosis. Also, histograms are drawn for visual evaluation. If the skewness value is around zero, that means distribution is normal, and negative values indicate left-skewed distribution, and positive values are right-skewed distribution.

Histogram, skewness, and kurtosis of interval length between occurrence records of absolute dating data set are below. Figure 7 presents histogram of interval length between absolute dating occurrence records by species, *B. Bonassus*, *B. Primigenius, B. Primigenius-14, C. Fiber, D. d. messopotamica, E. ferus, E. hemiounus, E. hydruntinus,* and *P. leo* respectively.

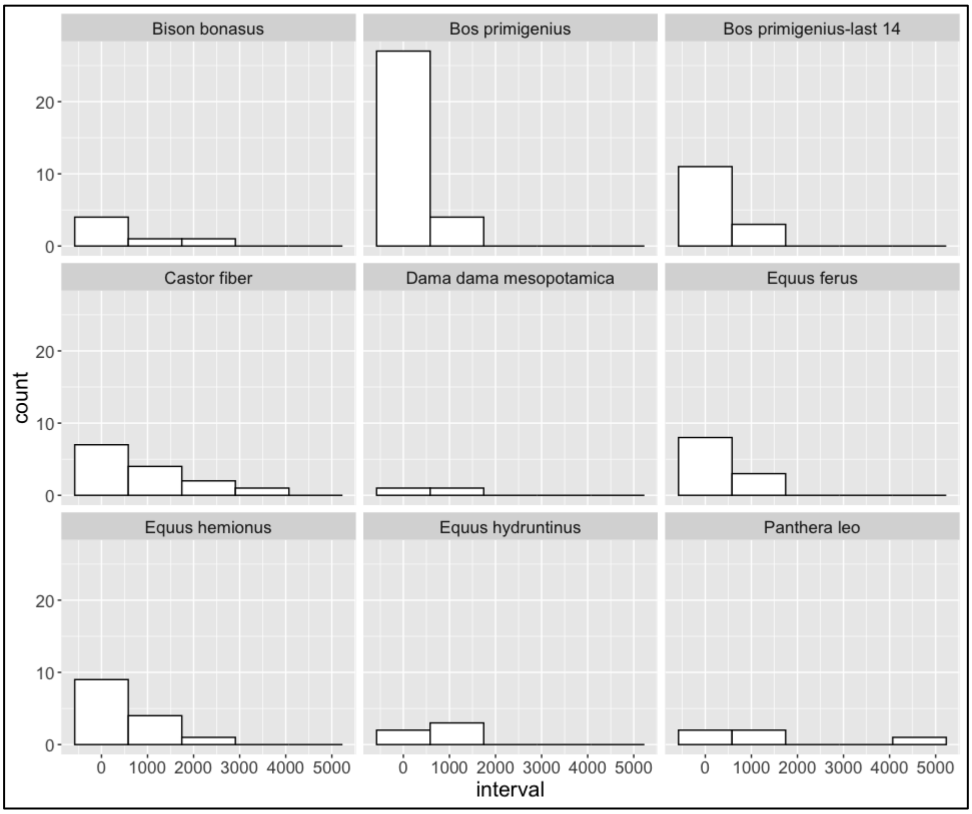

Figure 5 Histogram of interval length between absolute dating occurrence records by species.

Table 7 Skewness and kurtosis for interval length between absolute dating occurrence records by species.

Absolute	*Bison bonasus*	*Bos primigenius*	*Dama dama mesopotamica*	*Equus hydruntinus*	*Equus ferus*	*Equus hemionus*	*Castor fiber*	*Panthera leo*	*Bos primigenius-last 14*
Skewness	2.114	2.425	NA	-0.337	0.072	1.830	1.083	1.381	1.757
Kurtosis	4.710	5.586	NA	-2.745	-0.568	3.484	0.624	1.985	1.979
Intervals (i)	6	31	2	5	10	14	14	5	14

Interval lengths between absolute dating occurrence records are examined on a histogram, and skewness and kurtosis are calculated for each species. *B. primigenius* has 31 intervals, *E. hemionus* and *C. fiber* have 14, *E. ferus* has 10, *B. bonasus* has 6, *E. hydruntinus* and *P. leo* have 5, and *D. dama mesopotamica* has only 2 intervals between each record. The last 15 occurrence records and 14 intervals of *B.*

primigenious were also evaluated. *D. dama mesopotamica* has an insufficient number of intervals to calculate skewness and kurtosis.

The skewness of *E. ferus* (n=10) is close to 0.072. Kurtosis is negative and platykurtic. *E. ferus* seems skewed to the right on the histogram. *E. hydruntinus* (n=5) seems left skewed on the histogram with a negative value of -0.337 and platykurtic. Other species' skewness change between 2.425 to 1.083, and all of them are skewed on the histogram. Their kurtosis change between 0.6240 to 5.560.

Histogram, skewness, and kurtosis of interval length between occurrence records of relative dating data set are below. Figure 8 shows histogram of interval length between relative dating occurrence records by species, *B. bonasus, B. primigenius, B.primigenius*-14, *C. fiber, D. d. mesopotamica, E. ferus, E. hemiounus, E. hydruntinus,* and *P. leo* respectively.

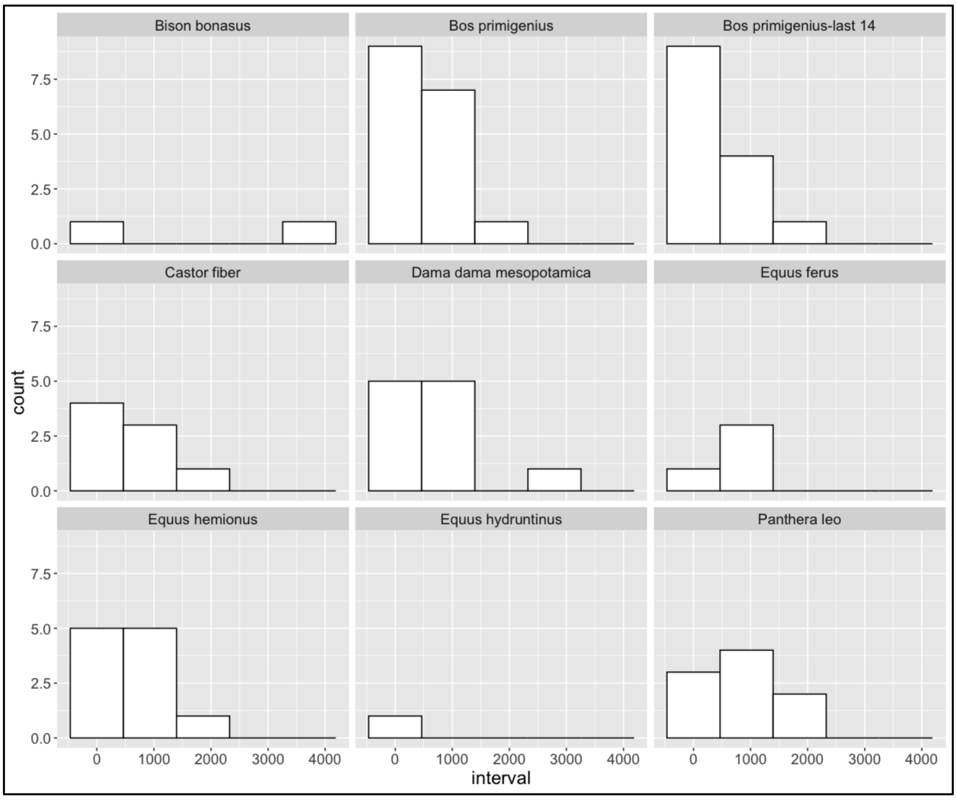

Figure 6 Histogram of interval length between relative dating occurrence records by species.

Table 8 Skewness and kurtosis for interval length between relative dating occurrence records by species.

	Bison bonasus	*Bos primigenius*	*Dama dama mesopotamica*	*Equus hydruntinus*	*Equus ferus*	*Equus hemionus*	*Castor fiber*	*Panthera leo*	*Bos primigenius-last 14*
Skewness	NA	1.548	1.529	NA	-0.223	1.049	0.877	1.039	2.107
Kurtosis	NA	2.544	2.889	NA	-4.726	1.559	-0.209	-0.29	4.754
Intervals(i)	3	18	12	2	5	12	9	10	14

Histograms are examined frequencies of the interval length between relative dating occurrence records, and skewness and kurtosis are calculated for each species to understand the shape of distributions. *B. primigenius* has 18 intervals, *E. hemionus* and *D. dama mesopotamica* have 12, *P. leo* has 10, *C. fiber* has 9, *E. ferus* has 5, *B. bonasus* has 3, and *E. hydruntinus* has 2 intervals between each record. *B. bonasus* and *E. hydruntinus* have an insufficient number of intervals to calculate skewness and kurtosis.

E. ferus (n=5) seems left skewed on the histogram with a negative value of -0.223 and platykurtic. Other species' skewness change between 0.877 and 2.107. They are mainly skewed to the right. *C. fiber* and *P. leo* also have negative kurtosis and they are platykurtic. Other species' kurtosis changed between 1.559 and 4.754.

4.2.1.1.2 Q-Q Plots and Probability Plot Correlation Coefficient Test Results

The Probability Plot Correlation Coefficient Test of Goodness-of-Fit (PPCC Test) was used to test the uniformity of occurrence record data. r and p values of the PPCC test were calculated with the *ppccTest* R package. Test results were calculated with 10000 Monte Carlo replicates. Many hypotheses test alternatives are available; however, the PPCC test has better statistical power with small sample-sized ordered occurrence records data (Vogel, 2008). Therefore, the PPCC test is preferred. Vogel et al. performed the PPCC test to understand the distribution of occurrence records. Greater r values mean there is little evidence to reject the null hypothesis, and the distribution is uniform. The p-value gives the probability that r is the smallest possible value if the distribution is uniform. This test only shows us that the population is not uniform; however, there is no certainty to accept that the sample comes from a uniform population (Vogel, 2008).

Strauss and Sadler's method assumes occurrence records have a uniform distribution. To test this assumption, plotting quantiles of ordered observed values on the y-axis and ordered expected values on the x-axis as a theoretical Quantile Quantile plot is a useful graphical tool (Wang, 2009). To draw the Q-Q plots, the *qqtest* package, the self-calibrating qqplot tool, was used in R. *qqplots* generate 1000 samples from the test distribution and represents their distribution on a shaded area on the plot (Oldford, 2016).

Probability Plot Correlation Coefficient Test for Absolute Data

Probability plot correlation coefficient test results for absolute dated occurrence records are represented in Table 11. The test statistic (r value), occurrence records (n), and p-value of test statistic result was evaluated. The null hypothesis and alternative hypothesis are defined below.

H_0: Distribution is uniform.

H_1: Distribution is not uniform.

p-value < 0.05 means H_0 is rejected, and the distribution of the sample is not uniform. If p-value>0.05 H_0 is not rejected. Results of the PPCC test are interpreted above with estimated extinction times of species.

Table 9 Probability Plot Correlation Coefficient Test for Non-Uniformity in Absolute Data. Not Reject means H_0 is not rejected.

Absolute Data	*r value*	*n*	*p-value*	H_0
Bison bonasus	114.6219	7	0.0405	Reject
Bos primigenius	88.5338	31	< 2.2e-16	Reject
Bos primigenius-15	169.1667	15	0.0001	Reject
Dama dama mesopotamica	74.4307	3	0.2996	Not Reject
Equus hydruntinus	19.011	6	0.8122	Not Reject
Equus ferus	2.4811	12	1.0000	Not Reject
Equus hemionus	33.8497	15	0.1962	Not Reject
Castor fiber	36.6344	15	0.1659	Not Reject
Panthera leo	124.7643	6	0.0407	Reject

The null hypothesis is rejected for *B. bonasus, B. primigenius, B. primigenius* with 15 samples, and *P. leo*. That means these have not come from a uniform population. For the other species, the null hypothesis is not rejected. That means these can be rooted in a uniformly distributed population. *D. d. mesopotamica* (n=3) has an extremely low number of absolute dated occurrence records, and this situation reduces statistical significance.

Q-Q plots show the distribution of the occurrence records to the expected distribution (Figure 9). *E. hydruntinus* and *E. ferus* have occurrence records close to the median and in the quantile dashed line. However, *E. hydruntinus*, with a low sample size seems lightly right skewed. The last occurrence records of *P. leo* and *C. fiber* are historical records and do not meet the expectations of the Q-Q plot. *B. bonasus*, *B.*

primigenius, and *B. primienius* with n=15, have heavy tails. *D. d. mesopotamica* has only 3 samples. *E. hemionus* has gaps between values. *C. fiber* and *P. leo* have spikes of identical values. These values are historical records that are close to each other than archaeological records.

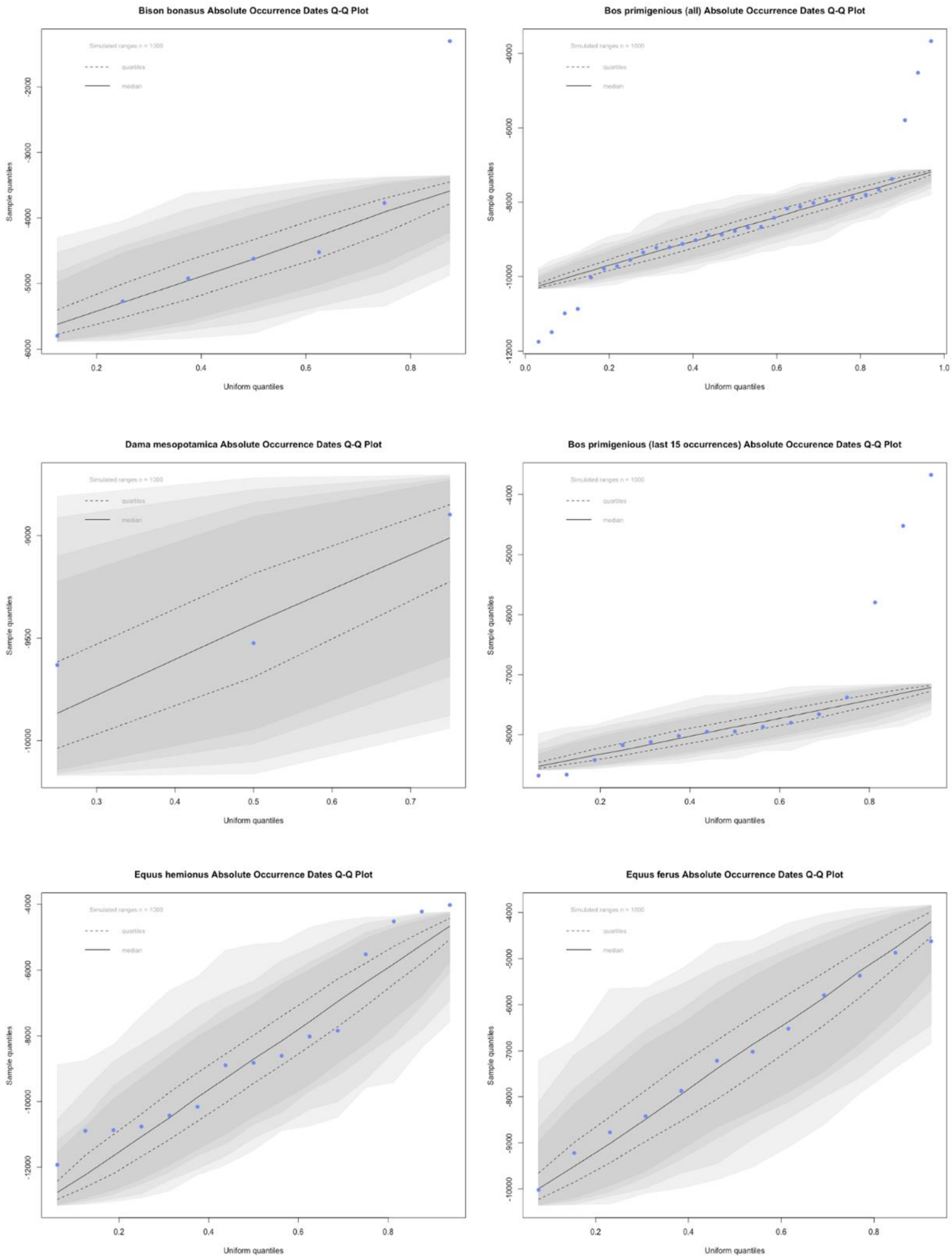

Figure 7 Q-Q plots of absolute dating occurrence records by species.

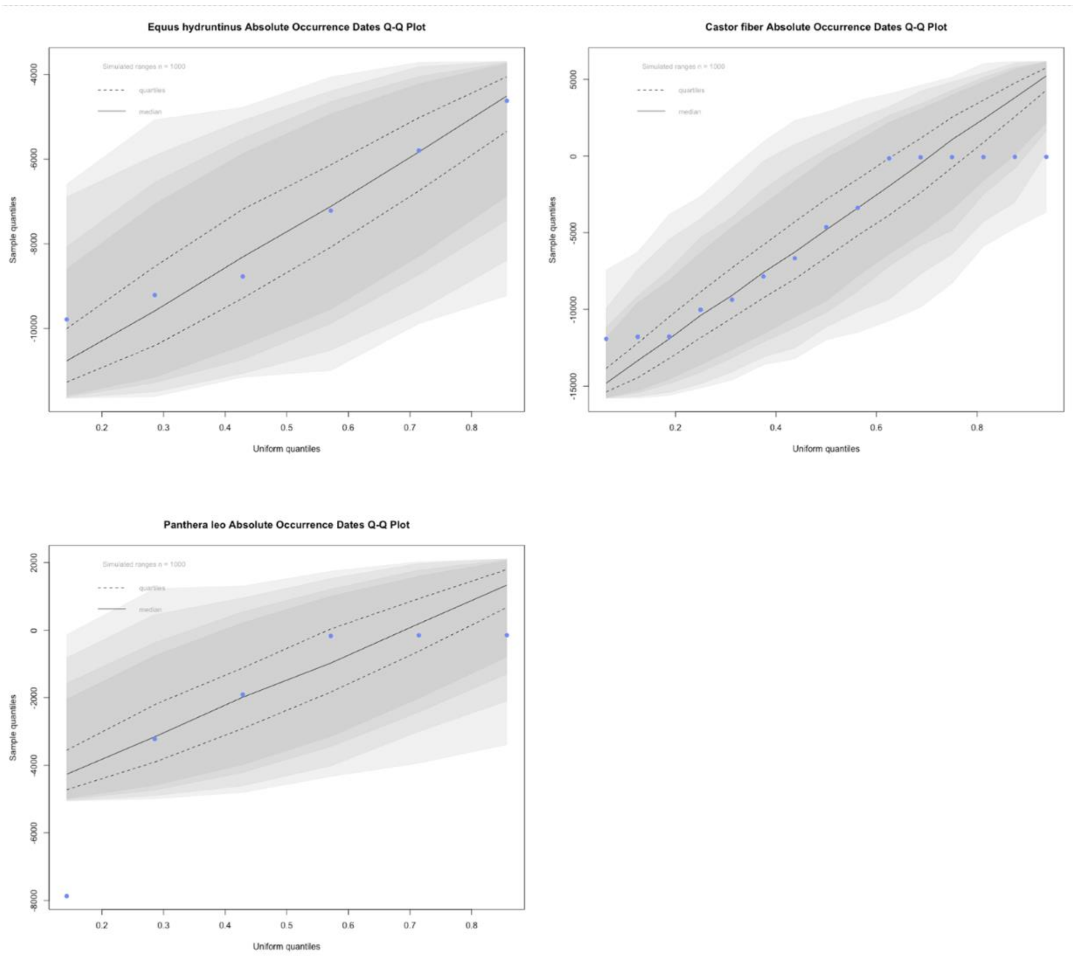

Figure 9 (Cont'd)

Probability Plot Correlation Coefficient Test for Relative Data

Probability plot correlation coefficient test results are represented in Table 12. The test statistic (r value), occurrence records (n), and p-value of test statistic result were evaluated. The null hypothesis and alternative hypothesis are defined below.

H_0: Distribution is uniform.

H_1: Distribution is not uniform.

p-value < 0.05 means H_0 is rejected, and the distribution of the sample is not uniform. If p-value>0.05 H_0 is not rejected. Results of the PPCC test are interpreted above with estimated extinction times of species.

Table 10 Probability Plot Correlation Coefficient Test results for relative data.

Relative	*r value*	*n*	*p-value*	*H_0*
Bison bonasus	131.0928	3	0.0148	Reject
Bos primigenius	26.0438	18	0.2389	Not Reject
Bos primigenius-15	32.4096	15	0.2197	Not Reject
Dama dama mesopotamica	23.6654	12	0.4972	Not Reject
Equus hydruntinus	-4.2188e-12	2	0.9436	Not Reject
Equus ferus	18.9249	5	0.8268	Not Reject
Equus hemionus	24.1744	12	0.4852	Not Reject
Castor fiber	77.0202	9	0.0734	Not Reject
Panthera leo	49.2946	10	0.187	Not Reject

Only the relative dated *B. bonasus* PPCC test rejected the null hypothesis and that indicates this sample is not uniform. Other species' relative dated samples can come from uniformly distributed populations.

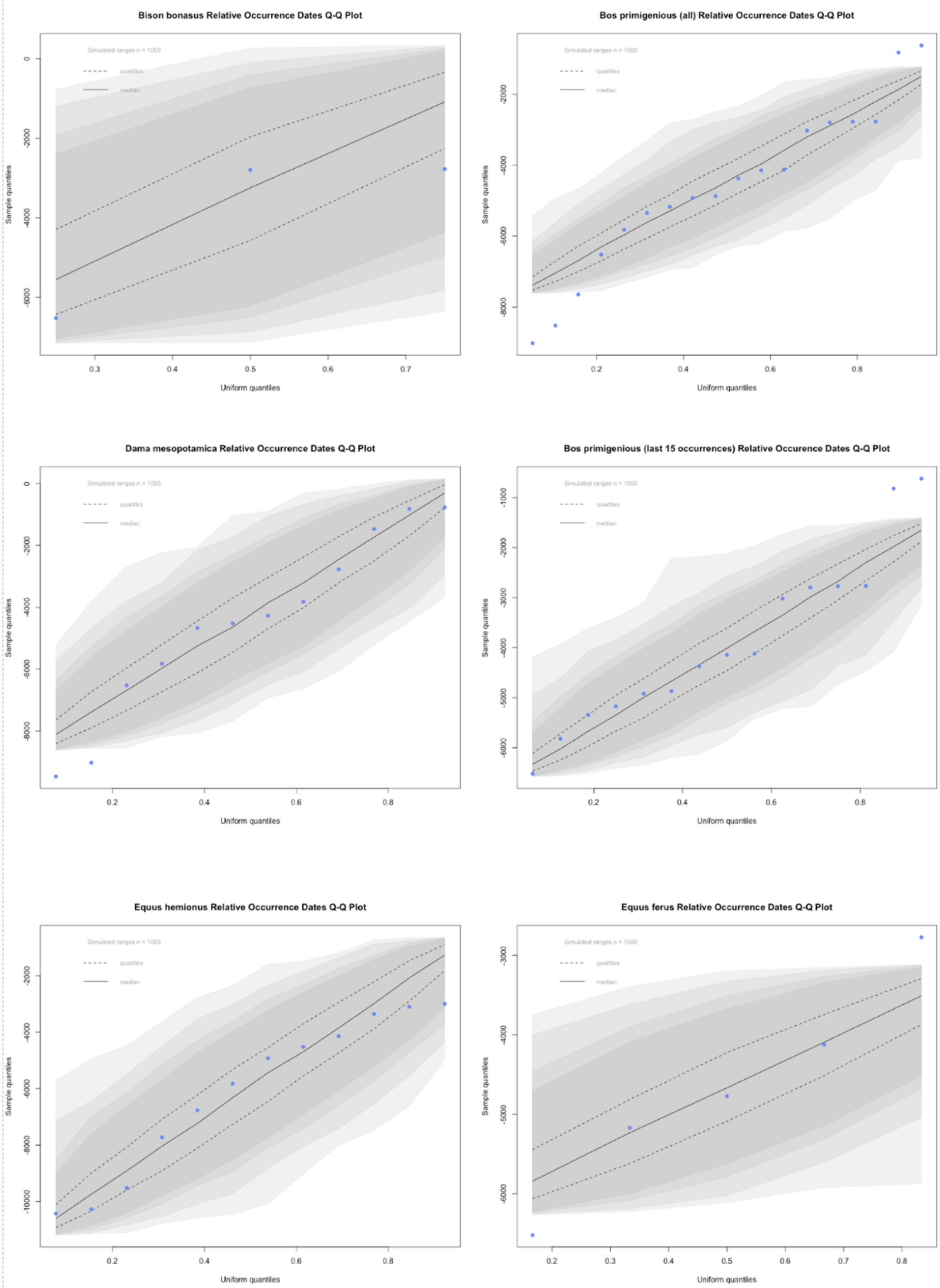

Figure 8 Q-Q plots of relative dating occurrence records by species.

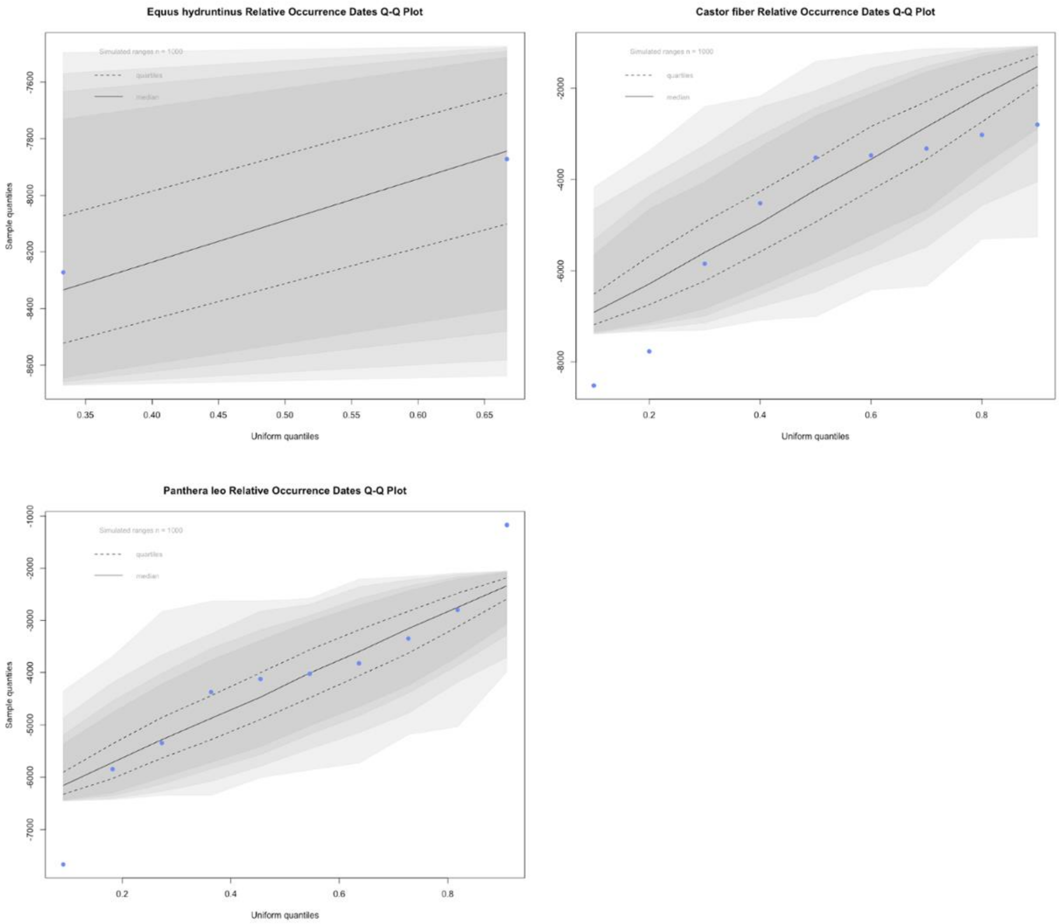

Figure 10 (Cont'd)

Some points deviated considerably from the median line of almost every species. *E. hemionus* is the closest to expected uniform distribution. *B. bonasus* (n=3) and *E. hydruntinus* (n=2) has very low sample size. *B. primigenius* has tails, and there are gaps between values. *B. primigenius* (n=15) has also a tail similar to the whole *B. primigenius* sample. *D. d. mesopotamica* has a light tail at the lower end. *E. hemionus* is lightly left skewed. *E. ferus* tailed but has only 5 samples. *C. fiber* seems left skewed and heavily tailed. *P. leo* also tailed at both ends.

4.2.1.2 Results of Strauss and Sadler Method for Extinction Time Estimation

Strauss and Sadler's method is one of the most fundamental methods for paleontological studies to estimate the extinction date of fossils. Extinction times of the extinct Anatolian large mammal species were calculated using Equation 16, which is defined as the best unbiased point estimator by Strauss and Sadler.

4.2.1.2.1 Absolute Dated Data

The estimated extinction times of the selected species are presented in Table 7. Estimation is carried out with a point estimation equation. Estimated Extinction time (Years difference from Present) means, how many years before or after this species go extinct. Negative values represent years before the present, and positive values represent years after the present. Lower CI and Upper CI are the boundaries of confidence intervals of estimated points with a 95% confidence level as a year difference from the present. Visual representations of estimated extinction dates and the occurrence records are given in Figure 5. In Figure 5, absolute dated occurrence records are given as black dots on the timeline. The pink dot represents the extinction date estimated by the Strauss and Sadler method. Each different colored line indicates a different species. A red dashed vertical line is the present.

Table 11 Estimated extinction time for the species with Strauss and Sadler method for absolute dated data results table.

Species	*Sample Size*	*Estimated Extinction Date (BC or AD)*	*Estimated Extinction time (Years difference from Present)*	*a (p=0.95)*	*Lower CI (Years difference from Present)*	*Upper CI (Years difference from Present)*
Bison bonasus	7	1469 AD	-553	0,8340	-1303	2444.996
Bos primigenius	31	1381 BC	-3403	0.131	-3672	-2613.52
Dama dama mesopotamica	3	6508 BC	-8530	4.84	-8897	-5344.44
Equus hydruntinus	6	1567 BC	-3589	1.065	-4622	878.725
Equus hemiounus	15	1436 BC	-3458	0.3	-4022	-901.5
Equus ferus	12	2109 BC	-4131	0.395	-4622	-3002
Castor fiber	15	4988 AD	2966	0.3	-49	3512.9
Panthera leo	6	3419 AD	1397	1.065	-148	8078.06
Bos primigenius-15	*15*	1292 BC	*-3314*	0.3	-3672	-2169

C. fiber and *P. leo* samples are mixed with historical and archaeological data, and their unrealistic estimation can be caused by this situation. PPCC test results show the distribution of *P. leo* is not distributed uniformly. This may explain the unrealistically estimated extinction date of *P. leo*. PPCC test result of *C. fiber* does not reject uniform distribution. However, its estimated extinction date does not reflect reality. Because *C.fiber* is an extinct species in Anatolia recently. The kurtosis value of *C. fiber* sample is 0.624 and the skewness is 1.381. *C. fiber* is skewed, and this is not expected from a symmetric uniform distribution. *C. fiber* does not meet the assumption of Strauss and Sadler method and therefore these unrealistic extinction estimation results are expected.

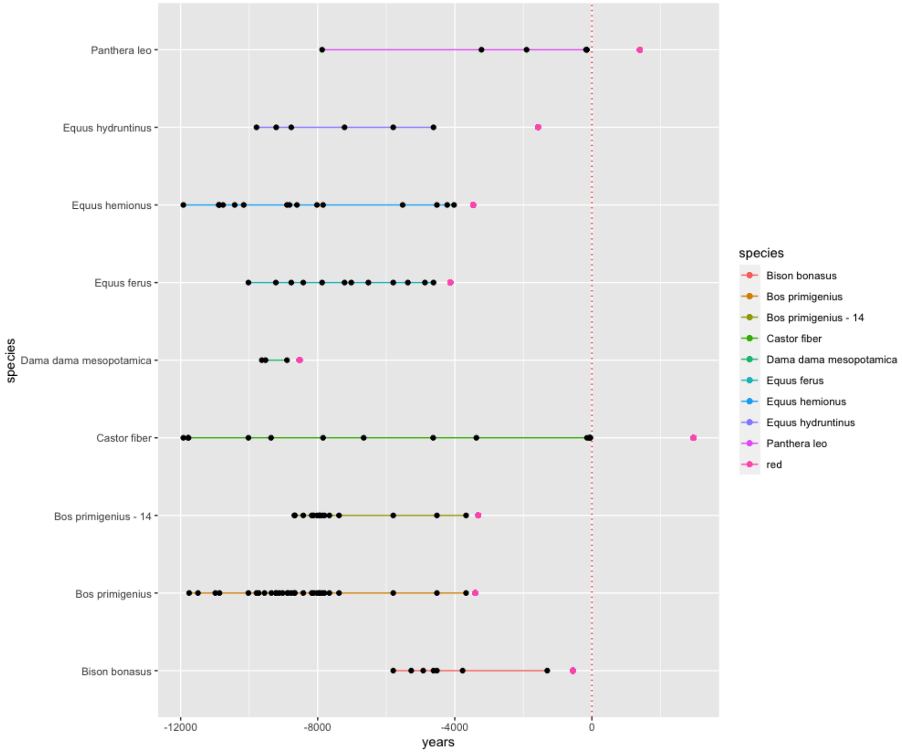

Figure 9 Estimated extinction times of absolute dated occurrence records using Strauss and Sadler method.

B. bonasus, *B. primigenius*, and reduced sample sized *B. primigenius* (n=15) are not uniform. PPCC test results of them rejected uniform distribution. Thus, their estimated extinction times calculated with Strauss and Sadler method can be controversial.

D. d. mesopotamica sample only have 3 occurrence record and only 2 intervals between them. Therefore, both distribution analysis and extinction analysis are not sensible.

PPCC test statistics of *E. hemionus* show that this sample can be uniform. However, the skewness and kurtosis of this sample do not meet uniform distribution assumptions. *E. hemionus* is skewed (skewness=1,83) and leptokurtic (kurtosis=3,484) with long tails.

E. hydruntinus's and *E. ferus*'s PPCC test results do not reject uniform distribution. Also, these species samples' have skewness around zero (-0.337 and 0.72 respectively). Moreover, their kurtosis are negative (-2.745 and -0.568 respectively). These samples do not violate the Strauss and Sadler method assumptions and their extinction estimations can be accepted as the most unbiased point estimations.

The estimated extinction time for *E. hydruntinus* is 1567 BC, with a 95% confidence level and confidence intervals ranging from 2000 BC as the lower bound to 2901 AD as the upper bound. However, due to the small sample size of only 6 occurrence records, the margins of the confidence intervals are very large, resulting in reduced precision.

The extinction time of *E. ferus* is estimated as 2109 BC with confidence intervals of 2600 BC lower bound and 980 BC upper bound (with 95% confidence.). The gap between the upper and lower bounds of confidence intervals is more realistic than *E. hydruntinus*. This may be the result of the larger sample size (n=12).

Consequently, the estimated extinction times of *E. hydruntinus* (1436 BC) and *E. ferus* (2019 BC) are appropriate results in terms of the assumptions of the model. Other species occurrence record data do not meet the assumptions of the model.

4.2.1.2.2 Relative Dated Data

Almost all relative dated samples have a bigger mean error (ε) value than the mean of intervals between records (î) (Table 4.). That means the upper and lower dating boundaries (dating error gives these points) of each occurrence record can overlap with each other's, and there is no data to solve these overlapping time regions. Therefore, extinction estimation results of relative data are not reliable and accurate.

The estimated extinction times of the selected species with relative dated samples are represented in Table 8. Estimated Extinction time (Years difference from Present) means, how many years before or after this species go extinct. Negative values represent years before the present, positive values represent years after the present. Lower CI and Upper CI are the boundaries of confidence intervals of estimated points with a 95% confidence level as a year difference from the present. Estimated extinction times and the occurrence records of species are given on timeline for each species in Figure 6. Black dots are occurrence records; pink dots are estimated extinction dates. The vertical red dashed line is present.

Table 12 Estimated extinction time for the species analyzed using Strauss and Sadler method for relative dated data.

Species	*Sample Size*	*Date (BC or AD)*	*Estimated Extinction Date (Years difference from Present)*	*a (p2=0.95)*	*Lower CI (Years difference from Present)*	*Upper CI (Years difference from Present)*
Bison bonasus	3	1125 AD	-897	4.84	-2772	15378
Bos primigenius	18	1894 AD	-128	0.241	-622	1402
Dama dama mesopotamica	12	2040 AD	18	0.395	-772	2663
Equus hydruntinus	2	5450 BC	-7472	29	-7872	3728
Equus hemiounus	12	300 BC	-2322	0.395	-2997	-64
Equus ferus	5	188 AD	-1835	1.465	-2772	2722
Castor fiber	9	139 BC	-2161	0.58	-2797	524
Panthera leo	10	1572 AD	-450	0.5	-1172	2078
Bos primigenius-15	15	1821 AD	-201	0.3	-622	1898

Relative dated occurrence record data of *B. Bonasus* (n=3) is not uniform according to PPCC test statistics. Therefore, its result for extinction estimation is questionable.

B. primigenius distribution shape is platykurtic (kurtosis: 2.544) and highly skewed (skewness: 1.548) distribution. The PPCC test is not rejected uniformity; however this sample is not distributed uniformly. *B. primigenius* with 15 samples is leptokurtic (kurtosis: 4.754) and highly skewed (2.107). Skewness and kurtosis show that this sample does not uniform, and extinction estimation is not sensible.

The extinction time estimation for *D. d. mesopotamica* is AD 2040. This estimation is not realistic because this species is extinct in Anatolia. The PPCC test is not rejected uniform distribution for *D. d. mesopotamica* species. Skewness (1.529) and Kurtosis (2.889) is indicating this sample does not uniform. This sample does not meet the assumption of Strauss and Sadler's method, and the estimated extinction time is not valid.

E. hydruntinus species only have 2 samples. Both extinction estimation and other analyses are not remarkable.

The extinction estimation time for *E. hemionus* is 300 BC. The PPCC test is not rejected uniformity for this sample. The shape of the distribution is platykurtic (kurtosis: 1.559); however, this sample is not symmetric (skewness: 1.049). When Q-Q plot and histogram of the relative dated *E. hemionus* sample are examined, it can easily be seen that only the last two occurrence records deviate from the whole distribution.

E. ferus extinction time is estimated as AD 188. The PPCC test is not reject uniformity and this sample close to a symmetrical platykurtic shape (skewness: -0.223, kurtosis: -4.726). This estimation is notable for the relative dating sample of *E. ferus*.

C. fiber's and *P. leo*'s extinction time estimation times, 139 BC and AD 1572 respectively, are not realistic. Both species have absolute dated records from the 20th century. These results are caused by incomplete relative dated occurrence records.

As a result, only estimated extinction times of *E. hemionus* (300 BC) and E. *ferus* (188 AD) are significant for Strauss and Sadler method with their relative dated samples. Other species occurrence record data are not meet assumptions of the model.

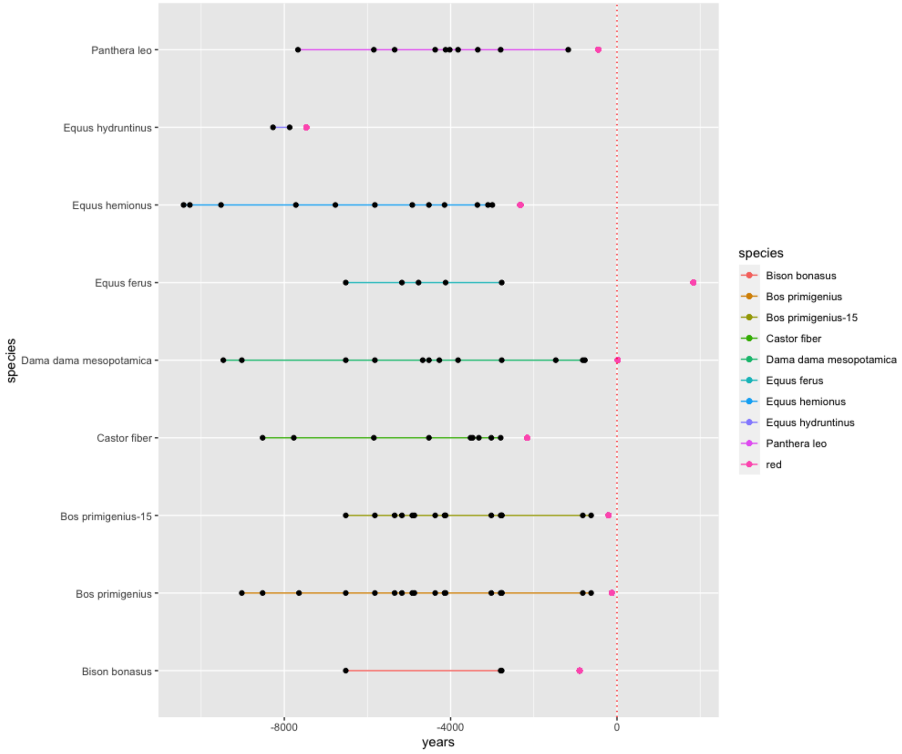

Figure 10 Representation of estimated extinction times of relative dated samples using Strauss and Sadler method and occurrence records of species on a timeline.

4.2.2 Robert and Solow Method for Extinction Time Estimation

Roberts and Solow's OLE method is one of the most popular methods for both ecological and paleontological studies. Also, this method is preferred for archaeological studies. The main feature of this method is that it is one of the least strict methods. Distributional assumptions are not required. Also, there is no assumption about sampling methods.

4.2.2.1 Results of Absolute Dated Data

Estimated extinction times for the selected species' absolute dated samples are presented in Table 13 with lower and upper confidence intervals. Negative values indicate years before the present, and positive values indicate years after the present. In Figure 10, absolute dated occurrence records are given as black dots on the timeline. The pink dot represents the estimated extinction date. Each different colored line indicates a different species. A red dashed vertical line is the present

Table 13 Estimated extinction time for the species with OLE method for absolute dated data results table.

Species	*n*	*Lower CI*	*Upper CI*	*The interval between 2022 and the estimated extinction date*	*Estimated Extinction Date (Year, BC or AD)*
Bison bonasus	7	400	16686	2375	4397 AD
Bos primigenius	32	-3371	-258	-2673	651 BC
Dama dama mesopotamica	3	-6955	6535422	-5453	3431 BC
Equus hydruntinus	6	-4099	7783	-2157	135 BC
Equus hemiounus	15	-3959	-1342	-3247	1225 BC
Equus ferus	11	-4591	-2647	-4111	2089 BC
Castor fiber	15	-49	104	-34	1988 AD
Panthera leo	6	-148	816	-99	1923 AD
Bos primigenius-15	15	-2595	4225	-1734	288 AD

Except for *B. bonasus*, estimated extinction dates are realistic because all species are extinct. The OLE method gives confidence intervals for estimated extinction dates however these intervals can be enormous depending on sample size. Therefore, the confidence intervals for estimated times are not reliable.

The extinction estimation result of *D. d. mesopotamica* will not be reliable. Since *D. d. mesopotamica* only has 3 samples. Many authors point out that too small a sample size will give much wider confidence intervals and not reliably estimated extinction time.

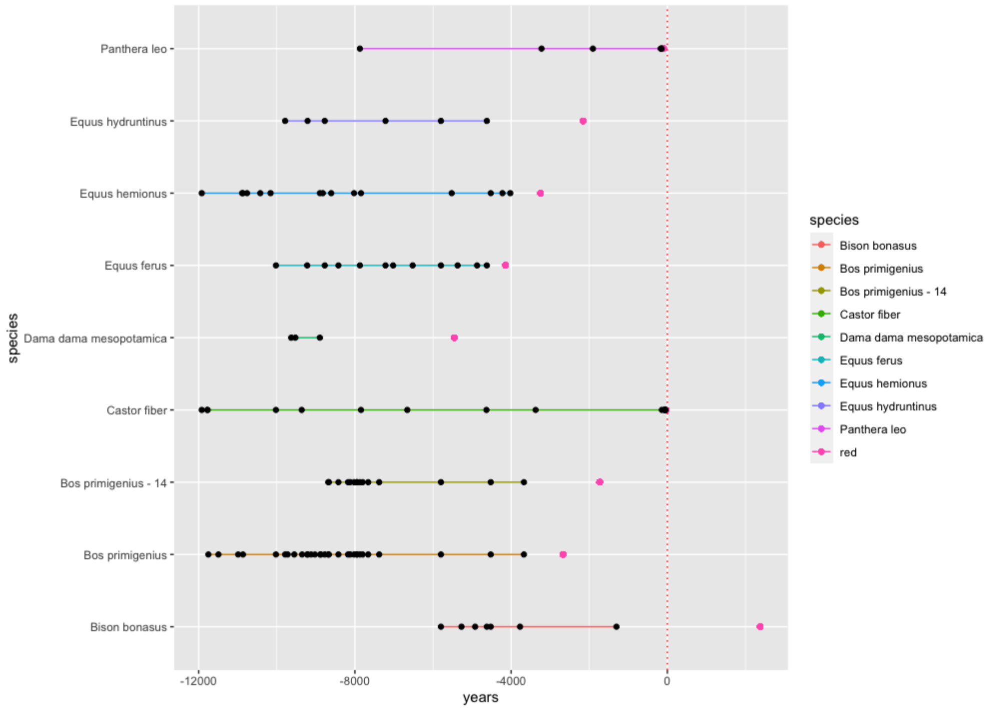

FIGURE 10. Representation of estimated extinction times of absolute dated samples with the OLE method and occurrence records of species on a timeline.

4.2.2.2 Results of Relative Dated Data

Estimated extinction times from the selected species' relative dated samples are presented in Table 14 with lower and upper confidence intervals. Negative values indicate years before present, positive values indicate years after the present. Figure 11. is the visual presentation of the occurrence records and estimated extinction times. The x-axis is years, the y-y axis is species. Black dots are occurrence records, pink dots are estimated extinction dates. The red dashed line is present.

As mentioned before, the error average of the relative dating data is larger than the occurrence records interval average. For this reason, analyses made with these data are not reliable and accurate. However, these analyses still needed. Because some of the most recent samples did not date with absolute methods but their analyses are important because they are the last records of the extinct taxa.

Table 14 Estimated extinction time for the species with OLE method for relative dating data results table.

Species	*Sample Size*	*Lower CI*	*Upper CI*	*The interval between 2022 and the estimated extinction date*	*Estimated Extinction Date (BC or AD)*
Bison bonasus	3	-2772	1046793	-2506	484 BC
Bos primigenius	18	-578	1485	-6	2016 AD
Dama dama mesopotamica	12	-763	1089	-376	1646 AD
Equus hydruntinus	2	-7872	-8272	NaN	NaN
Equus hemiounus	12	-2991	-1540	-2708	686 BC
Equus ferus	5	-2398	9404	-807	1215 AD
Castor fiber	9	-2794	-1389	-2576	554 BC
Panthera leo	10	-1022	3539	70	2092 AD
Bos primigenius-15	15	-520	2089	183	2205 AD

Estimated extinction dates of extinct *P. leo* and extinct *B. primigenius* with 15 samples in the future. The extinction date for *E. hydruntinus* with only 2 samples cannot be estimated.

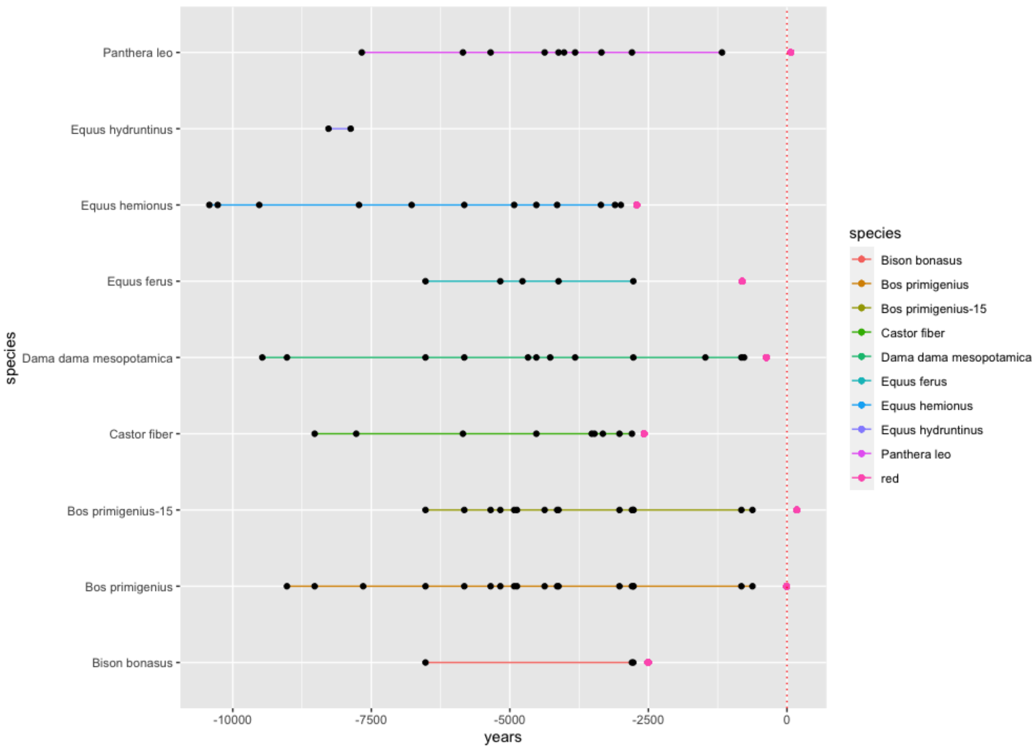

Figure 11 Representation of estimated extinction times of relative dated samples with the OLE method and occurrence records of species on a timeline.

CHAPTER 5

CONCLUSION

5.1 Extinct Species List

16 mammal species have gone extinct in the last 10 thousand years in Anatolia listed below. More than 50% of the extinct species belong to ungulates. Many species of extinct ungulates were ancestors of domesticated ungulates (cattle, horse, sheep, donkey).

In addition, some species are difficult to identify in archaeological finds. *D. d. mesopotamica, G. sungutturosa*, Equus species, Ovis/Capra species have some morphologically similar bones.

Extinct Large Mammal Species in Anatolia

1 *Dama dama mesopotamica*

2 *Ovis ammon*

3 *Bos primigenius*

4 *Bison bonasus*

5 *Gazella subgutturosa*

6 *Equus hemionus*

7 *Equus africanus*

8 *Equus hydruntinus*

9 *Equus ferus*

10 *Lepus capensis*

11 *Elaphas maximus*

12 *Vulpes corsac*

13 *Panthera leo*

14 *Acinonyx jubatus venaticus*

15 *Panthera tigris virgata*

16 *Castor fiber*

5.2 Overall Estimated Extinction Times

Only the 8 species have sufficient number of occurrence records to estimate the extinction date which are *Bison bonasus, Bos primigenius, Dama dama messopotamica, Equus hemionus, Equus hydruntinus, Equus ferus, Castor fiber*, and *Panthera leo*. To ensure consistency, their occurrence records data were divided into two groups according to the dating method of records, relative dated, and absolute dated. The relative dated data did not have good enough resolution because of the larger dating error. All data were analyzed with Strauss and Sadler's method and Roberts and Solow's OLE method to estimate species extinction times in Anatolia. Strauss and Sadler's method had a distributional assumption and performed other statistical analyses on the sample to understand if data meet this assumption. Few of species samples met this assumption. Roberts and Solow's OLE method is not needed distributional assumptions.

Roberts and Solow's OLE method performed better with absolute dated data and gave realistic extinction estimation times, listed below.

Estimated Extinction year of Large Mammal Species with the only absolute dated sample:

1) *Castor fiber* (Beaver) - “AD 1988”

2) *Panthera leo persica* (Lion) - “AD 1923”

3) *Equus hemionus* (Onager) - “1225 BC”

4) *Equus hydruntinus* (European Wild Ass) - “135 BC”

5) *Equus ferus* (Wild Horse) - “2089 BC”

6) *Bos primigenius* (Wild Ox) - “651 BC”

7) *Dama dama mesopotamica* (Persian fallow deer) - “3431 BC”

8) *Bison bonasus* (European Bison) - “AD 4397”

Extinction estimation results are compared for two methods in Table 15. Not only estimated extinction times but also the last occurrence records of each specimen are compared. The last occurrence records of *B. Primigenius, D. D. Mesopotamica, E. hemionus, E. ferus,* and *B. primigenius* with 15 records are dated with relative dating methods, and these occurrence records beyond these estimated extinction time with absolute dated samples. For this reason, detailed estimation results for each species are separately evaluated below.

Species	*Absolute dated sample's last record and dating error*		*Relative dated sample's last record and dating error*		*Absolute dated sample's extinction estimaiton results*		*Reltive dated sample's extinction estimation results*	
	Last record	Dating Error of last record	Last record	Dating Error of last record	Strauss Sadler method	OLE method	Strauss Sadler method	OLE method
Bison bonasus	719 AD	101	750 BC	900	1469 AD	4397 AD	1125 AD	484 BC
Bos primigenius	1650 BC	300	1400 AD	200	1381 BC	651 BC	1894 AD	2016 AD
Dama dama mesopotamica	6875 BC	390	1250 AD	100	6508 BC	3431 BC	2040 AD	1646 AD
Equus hydruntinus	2600 BC	140	5850 BC	700	1567 BC	135 BC	5450 BC	NaN
Equus hemiounus	2000 BC	0	975 BC	186	1436 BC	1225 BC	300 BC	686 BC
Equus ferus	2600 BC	140	750 BC	900	2109 BC	2089 BC	188 AD	1215 AD
Castor fiber	1973 AD	0	775 BC	450	4988 AD	1988 AD	139 BC	554 BC
Panthera leo	1874 AD	0	850 AD	100	3419 AD	1923 AD	1572 AD	2092 AD
Bos primigenius-15	1650 BC	300	1400 AD	200	1292 BC	288 AD	1821 AD	2205 AD

Table 15. Extinction estimation results for both two method and the last occurrence records of relative dated and absolute dated samples of each species.

5.2.1 *Bos primigenius*

B. primigenius has 31 absolute dated and 18 relative dated occurrence records. The last relative dated record is AD 1400 and the last absolute dated record is 1650 BC. There is a quite big difference between the last records of absolute dated and relative dated samples. Because of that, the results of extinction estimations with the absolute dated sample are not match up with the real situation which dates are 1381 BC with Strauss and Sadler method and 651 BC with the OLE method because the last occurrence record is approximately from AD 1400.

Relative dated samples were also analyzed with extinction estimation methods. OLE method is far from the current situation and gave AD 2016 for extinction time. Strauss and Sadler's method estimated the extinction time as AD 1894. PPCC test does not reject the uniformity of the sample however *B. primigenius* relative dated sample is not symmetric and has tails on the Q-Q plot.

The estimated extinction time of a relative dated sample of *B. primigenius* with Strauss and Sadler method is plausible but it is not a precise estimated extinction time.

Reduced sample sized *B. primigenious* (n=15) also gives approximately similar extinction estimation results. Still, the main problem is the dating method of specimens in this reduced sample.

5.2.2 *Dama dama mesopotamica*

D. d. mesopotamica has 3 absolute dated and 12 relative dated occurrence records. The absolute dated sample is not enough to perform extinction estimation analysis. Estimated extinction times for relative dated data are AD 2040 with Strauss and Sadler method and AD 1646 with the OLE method.

AD 1646, the OLE method's extinction time estimation, seems reasonable however dating method of the sample is not precise and this cause lower accuracy.

5.2.3 Equus hydruntinus

E. hydruntinus has 6 absolute dated occurrence records and only 2 relative dated occurrence records. The last record of this sample is 2600 BC with an absolute dated occurrence record. Estimated extinction times for absolute dated sample by methods are 1567 BC (with 2600 BC and 2900 AD lower and upper confidence intervals) with Strauss and Sadler method and 135 BC (with 2077 BC lower interval and AD 9805 upper interval, with a 95% confidence level) with OLE method. This sample met the assumptions of the Strauss and Sadler method.

Consequently, Strauss and Sadler's estimation results with the absolute dated sample of *E. hydruntinus* can be considered more plausible which is 1567 BC with 2600 BC and 2900 AD lower and upper confidence intervals (with a 95% confidence level) than results of the OLE method which has wider confidence intervals.

5.2.4 Equus hemionus

E. hemionus has 15 absolute dated occurrence records and 12 relative dated occurrence records. This species' occurrence records suffer the same problem as *B. primigenius*. The last absolute dated sample is from 2000 BC and the last relative dated sample is from 975 BC. Estimated extinction times with an absolute dated sample are 1436 BC with Strauss and Sadler method and 1225 BC with the OLE method. However, these estimations are not fit with the real world. In other words, these estimations are false because of the incomplete absolute dated data.

Extinction times are also estimated with relative dated samples. Results are 300 BC with the Strauss and Sadler method and 686 BC with the OLE method. The mean dating error is 683 years for relative dated sample. Relative dated sample of *E. hemionus* is close to expected uniform distribution. Therefore extinction estimation result of Strauss and Sadler method is reasonable in terms of the assumption of this method.

5.2.5 *Equus ferus*

E. ferus has 12 absolute dating and 5 relative dating occurrence records. The last record of the *E. ferus* is 750 BC with a relative dated sample and 2600 BC with an absolute dated sample. Estimated extinction times for absolute dated occurrence records 2109 BC with Strauss and Sadler method and 2089 BC with the OLE method. Both absolute dated extinction estimation results are far from the relative dated the last sample and not realistic. Relative dated extinction time estimation results are AD 188 with Strauss and Sadler method and AD 1215 with the OLE method. These results are more realistic than absolute dated results somehow. However, these results do not have a clear resolution about the extinction of *E. ferus* because only 5 relative dated sample gives these estimations. The mean dating error of the relative dated sample is 1300±496.99 (Standard Error) years (Table 4.). Samples' dating eras are overlaped and not clear resolution.

As a result, the absolute dated sample is consisting of the oldest *E. ferus* occurrence records. All the younger occurrence records, which are essential for the extinction estimation, are dated with relative dating methods. In this case, it is not possible to make a precise estimation of the extinction date of this species.

5.2.6 *Castor fiber*

C. fiber has 15 absolute dated and 9 relative dated occurrence records. *C. fiber* also has historical occurrence records which are precisely dated from the 20th century (AD 1973 is the last absolute dated occurrence record). The relative dated sample consists of older specimens than the absolute dated sample. Therefore, it was sufficient to analyze with only absolute dated samples.

Estimated extinction times for *C. fiber* are AD 4988 with Strauss and Sadler method and AD 1988 with the OLE method. Absolute dated *C. fiber* sample is not uniform thus Stauss and Sadler's method is not suitable for *C. fiber*.

Consequently, the results of the estimated extinction time with the OLE method is valid and seem realistic for *C. fiber*. The estimated extinction time for *C. fiber* is AD 1988 with upper and lower confidences, respectively AD 1973 and AD 2126 (with a 95% confidence level).

5.2.7 *Panthera leo*

P. leo has 6 absolute dated and 10 relative dated occurrence records. *P. leo* also has historical occurrence records in the absolute dated sample.

This sample is not uniform, and Strauss and Sadler's method does not perform well. The estimated extinction time with Strauss and Sadler method is AD 3419 and not feasible. OLE method's extinction estimation is realistic and more reliable. OLE methods result is AD 1923.

As a result, the estimated extinction time for P. *leo* is AD 1923 with lower and upper confidence intervals AD 1874 and AD 2838 respectively (with a 95% confidence level).

5.2.8 *Bison bonasus*

B. bonasus has 7 absolute dated and 3 relative dated occurrence record points. Extinction time estimation results of absolute dated occurrence records of *B. bonasus* are AD 1469 with the Strauss and Sadler method and AD 4397 with the OLE method.

The result of the Straus-Sadler method seems more reliable than the OLE method's result. However, the sample has not met the assumption of the Straus-Sadler method and this sample is not uniform. Therefore, this result (AD 1469) is questionable.

The result of the OLE method gives an unrealistic date, AD 4397. That is well known this species is extinct in Anatolia.

When the occurrence records are examined in detail, it is seen that the last sample is the sample obtained from the Yeni Kapı Marmaray Excavation dated to AD 720 at the Theodosius Harbour in Byzantine times (Onar, 2017). The interval between the last and the previous record is 2470 years. This difference is greater than the average interval length (î=749) of occurrence records of the absolute dated *B. bonasus* sample.

B. bonasus sample from Yeni Kapı Marmaray Excavations analyzed detailed. mtDNA of this sample has been sequenced by Onar et al. and the whole mtDNA sequence was publicly published at GenBank (GenBank accession number: KX773459).

Mitochondrial DNA (mtDNA) is maternally inherited and gave information about the ancestors of organisms. Haplogroup analysis is an approach to understand ancestors from the mtDNA of organisms. It is possible if there are other samples of *B. bonasus* from different populations to say where is the origin population of the Yeni Kapı Marmaray sample.

There are many modern and ancient *B. bonasus* mtDNA sequences in the GenBank database from West Europe to Caucasia. Neov et al. sequenced mtDNA (D-loop region) and analyzed haplogroups of ancient *B. bonasus* specimens unearthed from Bulgaria dated to 1778 BC (2021). Their haplogroup analysis is comprised of 122 different individuals' mtDNA sequences, mostly from ancient samples, from GenBank Database.

B. bonasus has two main genetically distinct lineages. The first lineage lived in The Late Pleistocene (Bb1 or Clade X). The other has lived Late Pleistocene to the present day, Bb2 lineage. Bb2 lineage has two genetically and geographically different clades. The first one is Bb2/1 which has distribution from the North Sea to Caucasia. This group inhabited Poland, Russia, and Georgia. The other one is Bb2/2 from the Alpine part of Europe. Bb2/2 linage is extinct (Neov, 2021).

B. bonasus specimen obtained from the Yeni Kapı Marmaray belonged to the Bb2/1 haplogroup (Neov, 2021). The Bb2/1 haplogroup was geographically distributed in Poland, Russia, and Georgia. Since there is no mtDNA sequence of any *B. bonasus* specimens from Anatolia in the databases, it was not possible to compare them with our specimen.

Considering the results of this haplogroup study and the sample from the ancient Theodosius Harbor (the Yeni Kapı Marmaray sample), the dates obtained by the OLE method are more consistent when we exclude the last occurrence record from extinction estimation studies, assuming that the last *B. bonasus* specimen from the Marmaray Yeni Kapı Excavation was transported to Theodosius Harbor for trade or similar reasons.

When the Marmaray Yeni Kapı occurrence record is excluded from OLE analysis, the estimated extinction time is 872 BC with 1587 BC lower and AD 1398 upper confidence interval (with a 95% confidence level).

It is also consistent with the relative dated *B. bonasus* occurrence records. The last relative dated occurrence record has come from the Büyüktepe Höyük Iron Age period which is 1000 BC to 550 BC.

5.3 Importance of Absolute Dating and more reliable species identification for Extinction Time Estimation studies

71 specimens were dated with the relative dating method. In other words, these samples were evaluated within a certain date range according to the archaeological context they were found in the excavations. This dating method does not give the exact date with narrow dating error of the samples like absolute dating methods such as radiocarbon. In fact, due to the width of the dating range of the relative dating approach, these data are located at overlapping regions on the timeline of each other. In this case, using them reliably in extinction time estimation studies is not plausible.

The biggest problem in our data and analysis has been experienced due to the presence of samples that have not been dated with absolute methods. Because some of the samples dated with the absolute dating method fall far behind the samples dated with relative dating methods. In this case, the extinction estimation obtained by absolute dating samples may precede the relative dated samples.

To overcome this problem, it is essential that archaeometrical dating methods need to become more widespread and samples from old excavations, which were not dated with absolute methods, are also dated and recompiled for further analysis.

In addition, different methods must become widespread for more precise and accurate identification of archeozoological specimens. Molecular and chemical identification methods are become important, especially in the accurate differentiation of morphologically similar species. For example, it is impossible to determine morphologically whether the *Gazella subgutturosa* or *Gazella marica*, which recently revealed that species are different.

Moreover, molecular phylogenetic studies can help to understand ancient wild populations and their origins. *Bison bonasus* case in this study points to the importance of molecular studies.

5.4 Challenges Faced in Obtaining and Analyzing Data

During the course of this research, several challenges were encountered. One of the major obstacles was the difficulty in locating and verifying suitable data sources. Many of the data sources were quite old and required extensive effort to ensure their accuracy and relevance.

Additionally, the number of occurrence records for wild animal species was found to be extremely limited, which presented a significant challenge for data analysis. To mitigate this issue, we included samples with very small sample sizes in the analysis, which inevitably led to decreased precision.

Moreover, analyzing the distribution of occurrence records was challenging due to the limited number of records. To address this issue, we used various methods to evaluate the distribution of data.

CHAPTER 6

DISCUSSION AND FURTHER STUDIES

6.1 Extinction Time Estimation with archaeological animal remains

Extinction time estimation with archaeological occurrence records is very similar to paleontological samples. However, there are some particular issues. Archaeological occurrence records from archaeological sites are shaped by human impact. Generally, species moved to settlement from a close environment by humans. Archaeological occurrence records are more like sporadic occurrence records than the non-discrete stratigraphic section from a paleontological site. Besides, taphonomic processes are highly variable between different eras.

Another big difference between paleontological and archaeological records is that many paleontological species are gone extinct much more in older times, and they did not have the possibility of being extant. However, archaeological samples can also include some species that can be extant today or recently. This makes a difference between samples and occurrence date distributions. For example, *Castor fiber* has much more records from the 19th and 20th centuries. But these samples are not enough to analyze only themselves. Analyzing whole archaeological and historical data can give much more perspective about this species.

6.2 Further Studies

To further understand the effects of climate change and human populations on fauna, it is recommended to conduct an analysis of animal occurrence records with their respective habitats and past climate. Spatial and temporal distribution analyses of animals in Anatolia can provide insights into these issues.

Although this study was conducted solely in the Anatolian region, it is important to note that this region is significantly influenced by surrounding biogeographical regions. As such, expanding the study to include data from Georgia, Armenia, Iran, Iraq, Syria, and the western part of the Thrace region in Turkey can enhance its efficacy.

Moreover, for models that require data distribution to adhere to certain assumptions, there is room for improvement. The data utilized in this study consists of animals hunted by humans, rather than naturally occurring data. Nonetheless, considering that domestication has led to reduced hunting over time, there is an opportunity to re-model this data in a more robust manner.

REFERENCES

Alroy, J. (2014) A simple Bayesian method of inferring extinction. Paleobiology, 40, 584-607.

Alvarez LW, Alvarez W, Asaro F, Michel HV. (1980) Extraterrestrial cause for the Cretaceous – Tertiary extinction. Science 208, 1095 – 1108.

Burgman, M. A., R. C. Grimson, and S. Ferson. 1995. Inferring threat from scientific collections. Conservation Biology 9: 923–928.

Du, A., Rowan, J., Wang, S. C., Wood, B. A., & Alemseged, Z. (2020). Statistical estimates of hominin origination and extinction dates: A case study examining the Australopithecus anamensis–afarensis lineage. Journal of human evolution, 138, 102688.

Holland, S. M. (2003). Confidence limits on fossil ranges that account for facies changes. Paleobiology, 29(4), 468-479.

Bradshaw, J. A. (2015). Uncertainties in dating constrain model choice for inferring extinction time from fossil records. Quaternary Science Reviews, 112:128–137.

Kadowaki, K. (2019). Stochastic Processes. Oxford University Press.

Key, A., Roberts, D., & Jarić, I. (2021). Reconstructing the full temporal range of archaeological phenomena from sparse data. Journal of Archaeological Science, 135, 105479.

Kodikara, S., Demirhan, H., Wang, Y., & Stone, L. (2021). Inferring extinction date of a species using non-homogeneous Poisson processes with a change-point. Methods in Ecology and Evolution, 12(3), 530-538.

Lee, T.E. (2014) A simple numerical tool to infer whether a species is extinct. Methods in Ecology and Evolution, 5, 791–796.

Marshall, C. R. (1990). Confidence intervals on stratigraphic ranges. Paleobiology, 16(1), 1-10.

Marshall, C. R. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. Paleobiology 23:165–173.

McCarthy, M. A. 1998. Identifying declining and threatened species with museum data. Biological Conservation 83:9–17.

McInerny, G. J., D. L. Roberts, A. J. Davy, and P. J. Cribb. 2006. Significance of sighting rate in inferring extinction and threat. Conservation Biology 20:562–567.

Neov, B., Spassov, N., Hristova, L., Hristov, P., & Radoslavov, G. (2021). New data on the evolutionary history of the European bison (Bison bonasus) based on subfossil remains from Southeastern Europe. Ecology and evolution, 11(6), 2842-2848.s

Onar, V., Soubrier, J., Toker, N. Y., van Loenen, A., Llamas, B., Siddiq, A. B., Pasicka E. & Tokarska, M. (2017). Did the historical range of the European bison (Bison bonasus L.) extend further south?—a new finding from the Yenikapı Metro and Marmaray excavation, Turkey. Mammal Research, 62(1), 103-109.

Patzkowsky, M. E., & Holland, S. M. (2012). Stratigraphic paleobiology. In Stratigraphic Paleobiology. University of Chicago Press.

Paul C.R.C (1982) The adequacy of the fossil record reconsidered, In: Patterson, C., Joysey, K. A., & Friday, A. E. (1982). Problems of phylogenetic reconstruction. Systematics association special, 21, 21-74.

Pohlert, T., Pohlert, M. T., Uniform, C., Logistic, G. L., & VGAM, S. (2020). Package 'ppcc'.

R Core Team (2021). R: A language and environment for statistical ## computing. R Foundation for Statistical Computing, Vienna, Austria.

Reznick, D. N., & Ricklefs, R. E. (2009). Darwin's bridge between microevolution and macroevolution. Nature, 457(7231), 837-842.

Rivadeneira, M.M., Hunt, G. & Roy, K. (2009) The use of sighting records to infer species extinctions: an evaluation of different methods. Ecology, 90, 1291–1300.

Roberts, D. L., and A. R. Solow. 2003. Flightless birds: When did the dodo become extinct? Nature 426:245–245.

Robson, D. S., and J. H. Whitlock. 1964. Estimation of truncation point. Biometrika 51:33–39.

Saltré, F., Brook, B. W., Rodríguez-Rey, M., Cooper, A., Johnson, C. N., Turney, C. S., & Bradshaw, C. J. (2015). Uncertainties in dating constrain model choice for inferring extinction time from fossil records. *Quaternary Science Reviews*, *112*, 128-137.

Schueth, J. D., Keller, K., Bralower, T. J., & Patzkowsky, M. E. (2014). The Probable Datum Method (PDM): a technique for estimating the age of origination nannoplankton THE PROBABLE DATUM METHOD. Paleobiology, 40(4), 541-559.

Shaffer, M. L. (1981). Minimum population sizes for species conservation. BioScience, 31(2), 131-134.

Smith, A. W., & Stewart Thomson, K. (1994). Systematics and the fossil record: documenting evolutionary patterns. Nature, 372(6505), 419-420.

Solow, A. R. (1993a). Inferring extinction from sighting data. Ecology 74:962–964.

Solow, A. R. (1993b). Inferring extinction in a declining population. Journal of Mathematical Biology 32:79–82.

Solow, A. R. (2005). Inferring extinction from a sighting record. Mathematical Biosciences 195:47–55.

Solow, A. R., and A. R. Beet. (2014). On uncertain sightings and inference about extinction. Conservation Biology 28:1119–1123.

Solow, A. R., and D. L. Roberts. (2003). A nonparametric test for extinction based on a sighting record. Ecology 84:1329– 1332.

Solow, A.R., Smith, W., Burgman, M.A., Rout, T.M., Wintle, B.A. & Roberts, D.L. (2012) Uncertain sightings and the extinction of the ivory-billed woodpecker Biology, 26, 180–184.

Strauss, D., and P. M. Sadler. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology 21:411–421.

Thompson, C. J., Kodikara, S., Burgman, M. A., Demirhan, H., & Stone, L. (2019). Bayesian updating to estimate extinction from sequential observation data. Biological Conservation, 229, 26-29.

Thompson, C. J., T. E. Lee, L. Stone, M. A. McCarthy, and M. A. Burgman. 2013. Inferring extinction risks from sighting records. Journal of Theoretical Biology 338:16–22.

Wang SC, Everson PJ, Zhou HJ, Park D, Chudzicki DJ. 2016Adaptive credible intervals on stratigraphic ranges when recovery potential is unknown. Paleobiology, 42, 240–256.

APPENDICES

A. Species occurrence records

Species occurrence records data will be uploaded to the GitHub repo.

https://github.com/amaeksi

B. R codes for analyses

R codes will be uploaded to the github repo.

https://github.com/amaeksi