

Descrizione

Questa ricerca si propone di tracciare in modo diacronico i modelli di sussistenza, economia e cambiamento ambientale in relazione ai sistemi rurali del primo millennio d.C. Il dibattito accademico sulla dinamica dell'interazione tra uomo e natura in Italia durante la transizione dalla tarda antichità all'inizio del medioevo lascia ancora oggi alcune domande aperte:

- Le precedenti strutture economiche e pratiche agricole romane furono completamente abbandonate o trovarono continuità durante questo periodo turbolento?
- Quali strategie agricole svilupparono i contadini per far fronte ai cambiamenti politici, demografici e climatici?
- È possibile identificare i fattori che hanno influenzato la distribuzione differenziale di animali e piante nei paesaggi del passato?
- Possiamo quantificare l'incertezza nella probabilità di trovare piante o animali specifici in determinate fasi cronologiche, tipi di contesto, caratteristiche geografiche, ecc.?

Gli studi precedenti sulla produzione agricola si sono basati su fonti letterarie e su indagini sul campo che hanno identificato i confini della problematica, ma è ancora assente uno studio archeologico su più fonti. Basandosi su proxy ambientali, come resti animali e vegetali, questo progetto utilizza dati provenienti da >309 siti (190 campioni botanici e 466 campioni faunistici) e un approccio bayesiano alla modellazione multilivello. In particolare, vengono utilizzati modelli bayesiani per stimare la probabilità di presenza di piante e animali, tenendo conto di diversi predittori categorici o numerici. Un'analisi specifica è posta sulla valutazione delle strategie agricole durante il passaggio dall'Impero Romano al paesaggio politicamente frammentato dell'Italia altomedievale, per valutare il ruolo dell'organizzazione politica, dell'economia, della cultura e dell'ambiente nella configurazione dei regimi agricoli e nella selezione dell'allevamento di animali e piante. L'integrazione di fonti diverse è fondamentale per gettare nuova luce sui modi di vita della classe rurale, che possono essere oscurati nei documenti testuali in quanto questi privilegiano contesti elitari e transazioni di alto livello. L'Alto Medioevo segna un periodo di cambiamenti fondamentali in cui diverse regioni geografiche (e potenzialmente microregioni) svilupparono i propri quadri politici ed economici per la produzione agricola. Le testimonianze bioarcheologiche possono aiutare a visualizzare il paesaggio in cui avvennero questi cambiamenti. I dati editi raccolti, sistematizzati in un database relazionale, sono sottoposti a nuove analisi in un contesto olistico. Ad esempio, è possibile analizzare le differenze nella dieta e nella produzione tra siti d'élite, religiosi, urbani e rurali e condurre un'analisi dei modelli regionali.

Perché un database di archeologia ambientale?

Questo studio è il primo a compilare un database di archeologia ambientale per i siti archeologici italiani datati al primo millennio d.C., fornendo dati grezzi sui resti di semi e frutti, nonché sui resti faunistici. A differenza dei lavori precedenti, che spesso forniscono solo riferimenti bibliografici o un accesso limitato ai dati, questo database offre accesso diretto ai valori grezzi. Nel caso dell'archeobotanica, Mercuri et al. (2015) e Grasso and Fiorentino (2009) sono state risorse preziose per la compilazione di un elenco completo di siti italiani con macroresti archeobotanici. Per la zooarcheologia, invece, pur non esistendo ancora un dataset specifico, Salvadori (2015) (per l'Italia medievale) e De Grossi Mazzorin and Minniti (2022) (per la città di Roma) sono punti di partenza per l'analisi. I set di dati e i codici generati in questo progetto di ricerca saranno ospitati sulla piattaforma GitHub, rendendoli apertamente disponibili ad altri ricercatori per riprodurre i risultati o esplorare nuove domande di ricerca.

Tipologia di dato	R	LR	EMA	XI sec. d.C.
Macroresti (semi)	83	64	69	18
Resti faunistici	189	203	154	69

Table 1. Stato del database. I campioni che coprono più cronologie sono stati ripetuti. **Legenda** = R (Età romana, I sec. a.C. - II sec. d.C.), LR (Età tardoromana, III-V sec. d.C.), EMA (VI-X sec. d.C.)

Perché modellare le occorrenze?

L'esplorazione dei dati disponibili utilizzando approcci standard è problematica a causa di vari bias presenti nei dati. Ad esempio, i campioni archeobotanici della penisola italiana sono spesso raccolti utilizzando strategie di campionamento basate su determinati indicatori (possibili fornelli, presse da uva, ecc.), che generano forti outlier con uno o pochi taxa predominanti nei campioni. Gli outlier nel dataset archeobotanico sono prevalenti ed eterogenei, ed è prudente analizzare i dati solo come presenza o assenza di piante nei campioni. I campioni zooarcheologici presentano un alto grado di variabilità nella distribuzione delle specie più comuni trovate nei siti italiani, ed è difficile identificare i fattori che contribuiscono al numero di resti faunistici in un sito. La modellazione statistica bayesiana può essere utile a gestire queste problematiche, ad esempio con i modelli binomiali, per analizzare la presenza o l'assenza di taxa nei campioni archeobotanici e modelli betabinomiali per tenere conto dell'eccessiva dispersione dei dati zooarcheologici. L'inferenza bayesiana fornisce un sistema robusto per modellare la variabilità dei dati e stimare l'incertezza dei risultati. Gli strumenti statistici utilizzati in questa ricerca sono R e Stan.

Modelli binomiali e betabinomiali

La scelta di un modello statistico e di una funzione di distribuzione della probabilità (PDF) appropriati è essenziale per analizzare la nostra variabile di risposta. La scelta del modello e della PDF dipende dalla natura della variabile di risposta e dalla domanda di ricerca. Questa ricerca utilizza principalmente due PDF nei suoi modelli:

- I **modelli binomiali**, che vengono utilizzati per analizzare dati di conteggio binari, in cui i dati possono assumere due possibili valori: successo o fallimento. Nella distribuzione binomiale:

$$y_i \sim \text{Binomial}(n_i, p_i)$$

il parametro n_i rappresenta il numero di tentativi nel campione i -esimo (ad esempio il conteggio totale di tutti i semi in un campione), e p_i rappresenta la probabilità di successo in ogni "tentativo" del campione i -esimo (= la probabilità di trovare semi di un particolare taxon nel campione, che non conosciamo e che stiamo cercando di modellare). Il conteggio y_i rappresenta il numero di successi nei n_i tentativi (= il conteggio totale di semi di un taxon in quel campione). Se utilizziamo solo la presenza/assenza, la distribuzione binomiale diventa effettivamente una distribuzione di **Bernoulli**, in cui y_i può essere 1 (se la pianta è presente nel campione) o 0 (se non lo è).

- I modelli **Betabinomiali** sono un tipo di modello binomiale in cui ogni conteggio binomiale ha la propria probabilità di successo. Questi modelli vengono utilizzati per considerare la sovradisersione nei dati di conteggio. Infatti, la distribuzione betabinomiale consente una variabilità maggiore oltre a quella prevista dalla distribuzione binomiale:

$$y_i \sim \text{Beta-Binomiale}(n_i, \bar{p}_i, \Theta)$$

dove y_i è il conteggio dell' i -esimo campione e n_i è il numero totale di prove per l' i -esimo campione. La probabilità di successo \bar{p}_i viene modellata come una distribuzione beta, una distribuzione molto flessibile, che ha un parametro Θ (se seguiamo la parametrizzazione in McElreath 2016) che descrive la media e la variazione della distribuzione.

Distribuzione dei dati

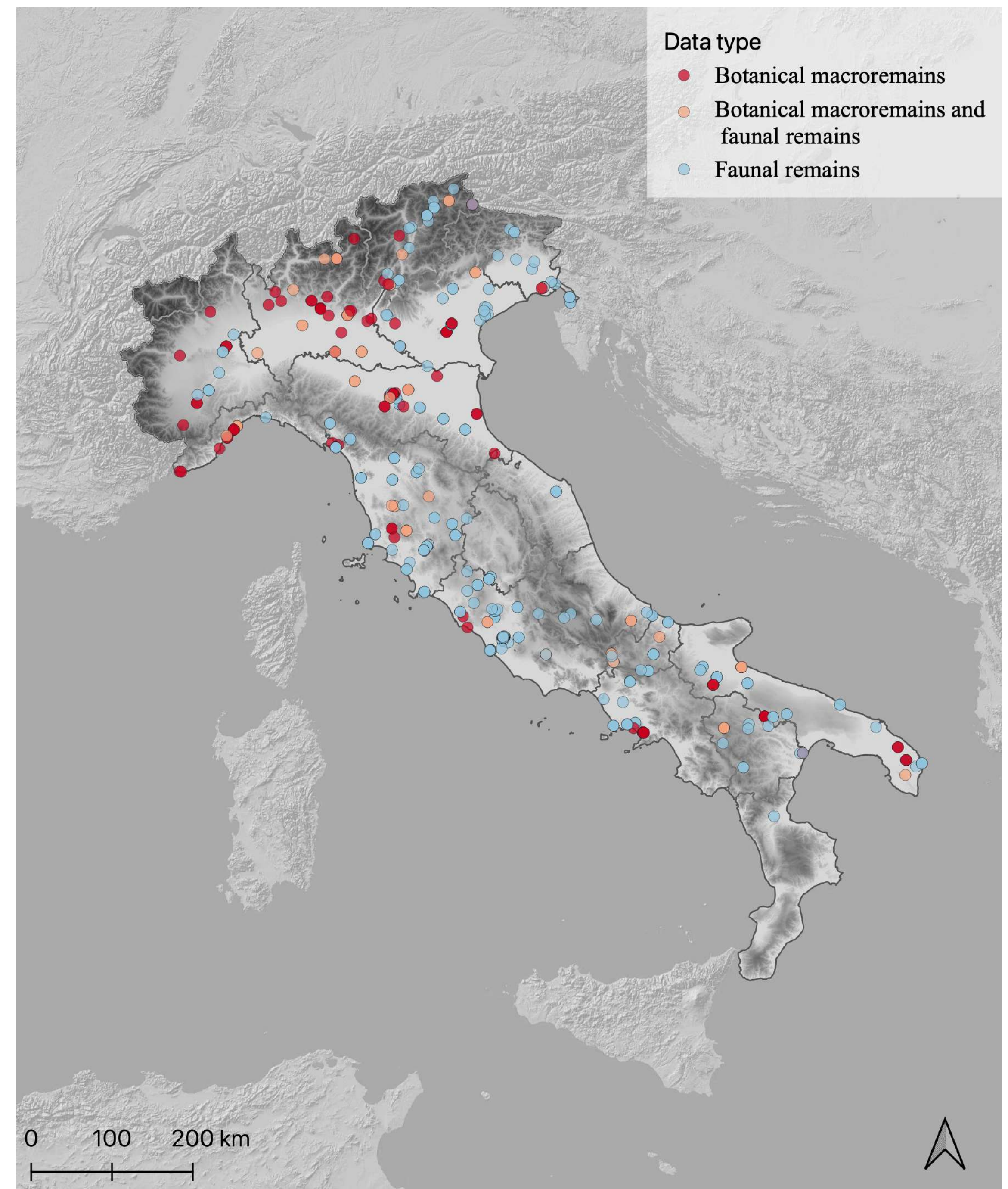


Figure 1. Distribuzione dei siti del I millennio d.C. nel database. I punti più scuri segnalano la disponibilità di un maggior numero di campioni per un determinato sito.

Cos'è la modellazione bayesiana?

"L'analisi bayesiana dei dati assume una domanda sotto forma di modello e utilizza la logica per produrre una risposta sotto forma di distribuzioni di probabilità" (McElreath 2016). In breve, se abbiamo una conoscenza o una convinzione precedente (*a priori*) su un evento, possiamo usare il teorema di Bayes per aggiornare la nostra convinzione alla luce di nuovi dati:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

La probabilità $P(A|B)$ che un'ipotesi A sia vera, data l'evidenza osservata B , è proporzionale alla probabilità dell'evidenza data l'ipotesi $P(B|A)$, moltiplicata per la probabilità precedente dell'ipotesi $P(A)$. Questa viene divisa per la probabilità dell'evidenza $P(B)$ stessa, che normalizza il risultato in modo che la somma delle probabilità sia uno. La distribuzione di probabilità risultante è chiamata distribuzione di probabilità *a posteriori*, che quantifica la nostra convinzione aggiornata sull'ipotesi data l'evidenza osservata. Più di un'ipotesi è plausibile e la teoria della probabilità bayesiana permette di quantificare l'**incertezza** e di confrontare diverse ipotesi (piuttosto che confrontare la nostra ipotesi con un'ipotesi nulla) in base alle loro distribuzioni di probabilità posteriori.

Quali sono alcune applicazioni?

L'inferenza bayesiana è sempre più utilizzata dagli scienziati come alternativa ai test di significatività delle ipotesi nulle. In archeologia, le sue prime applicazioni includono la calibrazione delle date al radiocarbonio e la modellazione della cronologia. Più recentemente, è stato utilizzato anche per modellare occorrenze e dati spaziali (Otárola-Castillo, Torquato, and Buck 2023). Un classico esempio concreto è l'uso della teoria bayesiana della probabilità nella seconda guerra mondiale da parte degli Alleati per decifrare il codice tedesco Enigma. Il team di decifratore, guidato da Alan Turing, utilizzò l'analisi bayesiana per migliorare l'efficienza dei propri metodi di decifrazione. Utilizzarono le conoscenze pregresse della lingua tedesca e dei pattern di codice, combinate con i messaggi osservati, per aggiornare la probabilità delle possibili chiavi dei messaggi. Questo permise loro di identificare più rapidamente la chiave corretta e di decifrare i messaggi che aiutarono gli Alleati a vincere la guerra (McGrayne 2011).

Bibliografia

- De Grossi Mazzorin, Jacopo and Claudia Minniti (2022). *Gli animali a Roma. Tre millenni di interazione con l'uomo*. All'Insegna del Giglio. 344 pp. ISBN: 978-88-928-5107-8.
- Grasso, Anna Maria and Girolamo Fiorentino (2009). "Studi Archeobotanici per l'Italia Medievale: Una Sintesi". In: *V Congresso Nazionale Di Archeologia Medievale. Palazzo Della Dogana, Salone Del Tribunale (Foggia); Palazzo Dei Celestini, Auditorium (Manfredonia); 30 Settembre-3 Ottobre 2009*. Ed. by Giuliano Volpe and Pasquale Favia. Vol. 30. Firenze: All'Insegna del Giglio, pp. 120-126.
- McElreath, Richard (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Texts in Statistical Science Series 122. Boca Raton: CRC Press/Taylor & Francis Group. 469 pp. ISBN: 978-1-4822-5344-3.
- McGrayne, Sharon Bertsch (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy*. New Haven [Conn.]: Yale University Press. 320 pp. ISBN: 978-0-300-16969-0.
- Mercuri, Anna Maria et al. (2015). "Pollen and Macroremains from Holocene Archaeological Sites: A Dataset for the Understanding of the Bio-Cultural Diversity of the Italian Landscape". In: *Review of Palaeobotany and Palynology* 218, pp. 250-266. ISSN: 00346667. DOI: 10.1016/j.revpalbo.2014.05.010.
- Otárola-Castillo, Erik, Melissa G. Torquato, and Caitlin E. Buck (Mar. 24, 2023). "The Bayesian Inferential Paradigm in Archaeology". In: *Handbook of Archaeological Sciences*. Ed. by A. Mark Pollard, Ruth Ann Armitage, and Cheryl A. Makarewicz. 1st ed. Wiley, pp. 1193-1209. ISBN: 978-1-119-59204-4 978-1-119-59211-2. DOI: 10.1002/9781119592112.ch59. (Visited on 2023).
- Salvadori (2015). *Uomini e animali nel Medioevo: ricerche archeozoologiche in Italia, tra analisi di laboratorio e censimento dell'edito*. Saarbrücken: Edizioni accademiche italiane. ISBN: 978-3-639-65749-4.

Overview

This research aims to diachronically trace the patterns of subsistence, economy, and environmental change in relation to regional patterns of Italian peasantry in the 1st millennium CE. Scholarly debate on the dynamics of human-nature interaction in Italy during the transition from Late Antiquity to the early Middle Ages still leaves today some open questions:

- Were the former Roman economic structures and farming practices completely abandoned or did they find continuity during this turbulent time?
- Which agricultural strategies did peasants develop to cope with political, demographic and climatic change?
- Can factors that influenced the differential distribution of animals and plants in past landscapes be identified?
- Can we quantify the uncertainty in the probability of finding specific plants or animals in certain chronological phases, context types, geographical features, etc.?

Past work on agricultural production has been based upon literary sources and field surveys which identify boundaries; a multi-source archaeological study is absent from the discourse. Drawing on environmental proxies, such as animal and plant remains, this project uses data from **>309 sites** (190 botanical and 466 faunal samples) and a Bayesian approach to multilevel modelling. In particular, Bayesian models are utilized to estimate the probability of the presence of plants and animals, while taking into account several categorical or numerical predictors. Specific emphasis is placed on evaluating agricultural strategies during the shift from the Roman Empire to the politically fragmented landscape of early medieval Italy, to assess the role of political organization, economy, culture, and environment in the configuration of agricultural regimes and animal/plant husbandry selection. The integration of different sources is fundamental to casting new light on peasants' lifeways, which may be obscured in the textual records, privileging elite contexts and high-status transactions. The Early Middle Ages mark a period of fundamental change when different geographical regions (and potentially micro-regions) developed their own political and economic frameworks for agricultural production. Bioarchaeological evidence can help visualise the landscape in which these changes were taking place, as archive collections of samples are subjected to new analyses in a holistic context, through the database I have constructed. For example, one can analyse differences in diet and production between elite, religious, urban and rural sites, and conduct regional pattern analysis.

Why an Environmental Archaeology database?

This study is the first to compile an environmental archaeology database for Italian archaeological sites dated to the first millennium CE, providing raw data on seed and fruit remains, as well as faunal remains. Unlike previous works, which often provide only bibliographical references or limited access to raw data, this database offers unprecedented access to raw counts. In the case of archaeobotany, Mercuri et al. (2015) and Grasso and Fiorentino (2009) were invaluable resources for compiling a comprehensive list of Italian sites with archaeobotanical macroremains. Meanwhile, for zooarchaeology, while a specific dataset does not yet exist, Salvadori (2015) (for Medieval Italy) and De Grossi Mazzorin and Minniti (2022) (for the city of Rome) are starting points for analysis. The data sets and codes generated in this research project will be hosted on the GitHub platform, making them openly available for other researchers to reproduce the results or explore new research questions.

Data Type	Roman	Late Roman	Early Medieval	11th c.
Macroremains (seeds)	83	64	69	18
Faunal remains	189	203	154	69

Table 1. Database status. Samples that span more chronologies have been repeated.

Why do we model counts?

Exploring the available data using standard approaches is challenging due to various biases in the data. For example, archaeobotanical samples from the Italian peninsula are often collected using feature-based sampling strategies, leading to strong outliers with one or a few taxa dominating the samples. The outliers are so prevalent and diverse that the data can only be analysed as presence or absence of plants in the samples. Zooarchaeological samples have a high degree of variability in the distribution of the most common species found on Italian sites, and it is difficult to identify the factors that contribute to the number of faunal remains at a site. To address these challenges, I utilise Bayesian statistical modeling, including binomial models to analyse the presence or absence of taxa in archaeobotanical samples and betabinomial models to account for overdispersion in the zooarchaeological data. Bayesian inference provides a robust framework for modeling variability in the data and estimating uncertainty in the results. The statistical tools used in this research are **R** and **Stan**.

Binomial and Betabinomial models

Choosing an appropriate statistical model and probability distribution function (PDF) is essential in analysing our response variable. The choice of model and PDF depend on the nature of the response variable and the research question. This research mainly uses two PDFs in its models:

- **Binomial models**, which are used to analyse binary count data, where data can take two possible values: success or failure. In the binomial distribution:

$$y_i \sim \text{Binomial}(n_i, p_i)$$

the parameter n_i represents the number of trials of the i -th sample (e.g. the total counts of all seeds in a sample), and p_i represents the probability of success in each "trial" of the i -th sample (= the probability of finding seeds of a particular taxon in the sample, which we do *not* know and we are trying to model). The count y_i represents the number of successes in the n_i trials (= the total counts of seeds of one taxon in that sample). If we only use presence/absence, the binomial distribution is effectively a **Bernoulli** distribution, where y_i can either be 1 (if the plant is present in the sample) or 0 (if it is not).

- **Betabinomial models**, which are a type of binomial model where each binomial count has its own probability of success. These models are used to account for overdispersion in count data. In fact, the betabinomial distribution allows for extra variation beyond what is expected from the binomial distribution:

$$y_i \sim \text{Beta-Binomial}(n_i, \bar{p}_i, \Theta)$$

where y_i is the count of the i -th sample and n_i is the total number of trials for the i -th sample. The probability of success \bar{p}_i is modeled as a beta distribution, a very flexible distribution, which has a parameter Θ (if we follow the parameterization in McElreath 2016) that describes the mean and variation of the distribution.

Data distribution

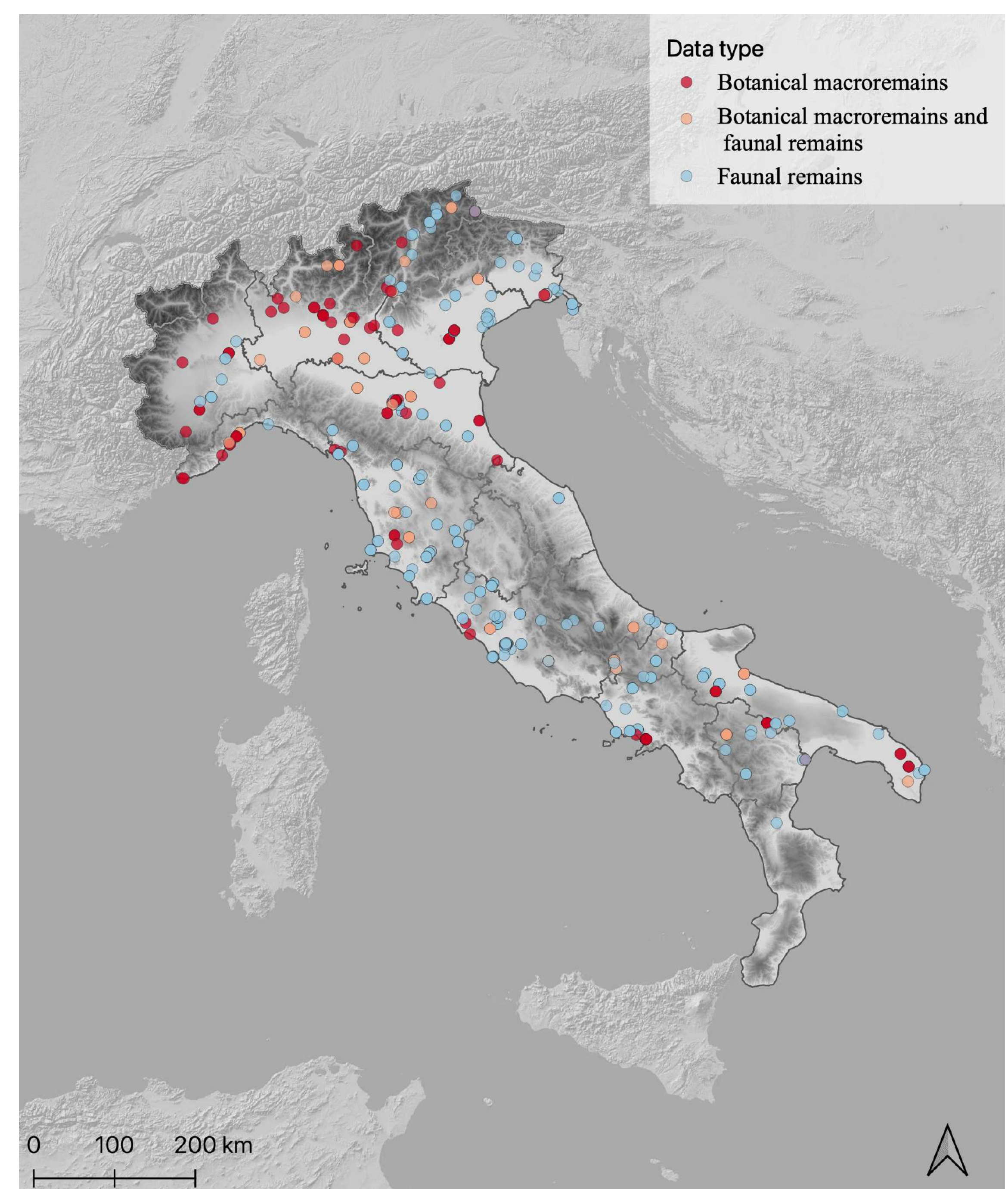


Figure 1. Distribution of 1st millennium CE sites in the database. Darker points indicate that more samples are available for one particular site.

What is Bayesian modeling?

"Bayesian data analysis takes a question in the form of a model and uses logic to produce an answer in the form of probability distributions" (McElreath 2016). In short, if we have *prior* knowledge or belief about an event, we can use Bayes' rule to update our belief in light of new data:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

The probability $P(A|B)$ of a hypothesis A being true, given the observed evidence B , is proportional to the likelihood of the evidence given the hypothesis $P(B|A)$, multiplied by the prior probability of the hypothesis $P(A)$. This is divided by the probability of the evidence $P(B)$ itself, which normalizes the result so that the probabilities sum to one. The resulting probability distribution is called the *posterior* probability distribution, which quantifies our updated belief in the hypothesis given the observed evidence. More than one hypothesis is plausible, and Bayesian probability theory allows to quantify **uncertainty** and compare different hypotheses (rather than comparing our hypothesis against a null hypothesis) based on their posterior probability distributions.

What are some applications?

Bayesian inference is increasingly being used by scientists as an alternative to null-hypothesis significance testing. In archaeology, its first applications included the calibration of radiocarbon dates and modeling chronology. More recently, it has also been used to model counts and spatial data (Otárola-Castillo, Torquato, and Buck 2023). A classic real-life example is the use of Bayesian probability theory in WW2 by the Allies to crack the German Enigma code. The code-breaking team, led by Alan Turing, used Bayesian analysis to improve the efficiency of their deciphering methods. They used prior knowledge of the German language and code patterns, combined with observed messages, to update the probability of possible message keys. This allowed them to more quickly identify the correct key and decrypt messages that helped the Allies win the war (McGrayne 2011).

References

- De Grossi Mazzorin, Jacopo and Claudia Minniti (June 30, 2022). *Gli animali a Roma. Tre millenni di interazione con l'uomo*. All'Insegna del Giglio. 344 pp. ISBN: 978-88-928-5107-8.
- Grasso, Anna Maria and Girolamo Fiorentino (2009). "Studi Archeobotanici per l'Italia Medievale: Una Sintesi". In: *V Congresso Nazionale Di Archeologia Medievale. Palazzo Della Dogana, Salone Del Tribunale (Foggia); Palazzo Dei Celestini, Auditorium (Manfredonia); 30 Settembre-3 Ottobre 2009*. Ed. by Giuliano Volpe and Pasquale Favia. Vol. 30. Firenze: All'Insegna del Giglio, pp. 120-126.
- McElreath, Richard (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Texts in Statistical Science Series 122. Boca Raton: CRC Press/Taylor & Francis Group. 469 pp. ISBN: 978-1-4822-5344-3.
- McGrayne, Sharon Bertsch (2011). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy*. New Haven [Conn.]: Yale University Press. 320 pp. ISBN: 978-0-300-16969-0.
- Mercuri, Anna Maria et al. (2015). "Pollen and Macroremains from Holocene Archaeological Sites: A Dataset for the Understanding of the Bio-Cultural Diversity of the Italian Landscape". In: *Review of Palaeobotany and Palynology* 218, pp. 250-266. ISSN: 00346667. DOI: 10.1016/j.revpa.2014.05.010.
- Otárola-Castillo, Erik, Melissa G. Torquato, and Caitlin E. Buck (Mar. 24, 2023). "The Bayesian Inferential Paradigm in Archaeology". In: *Handbook of Archaeological Sciences*. Ed. by A. Mark Pollard, Ruth Ann Armitage, and Cheryl A. Makarewicz. 1st ed. Wiley, pp. 1193-1209. ISBN: 978-1-119-59204-4 978-1-119-59211-2. DOI: 10.1002/9781119592112.ch59. (Visited on 2023).
- Salvadori (2015). *Uomini e animali nel Medioevo: ricerche archeozoologiche in Italia, tra analisi di laboratorio e censimento dell'edito*. Saarbrücken: Edizioni accademiche italiane. ISBN: 978-3-639-65749-4.