

Biogeography of crop progenitors and wild plant resources in the terminal Pleistocene and Early Holocene of West Asia, 14.7–8.3 ka

Joe Roe^{a,b,*}, Amaia Arranz-Otaegui^c

^a University of Bern, Switzerland,

^b University of Copenhagen, Denmark,

^c University of the Basque Country, Spain,

Abstract

This paper presents the first continuous, spatially-explicit reconstructions of the palaeodistributions of 82 plant species found regularly in association with early agricultural archaeological sites in West Asia, including the progenitors of the first crops. We used machine learning to train an ecological niche model of each species based on its present-day distribution in relation to climate and environmental variables. Predictions of the potential ranges of these species at key stages of the Pleistocene–Holocene transition could then be derived from these models using downsampled data from palaeoclimate simulations. The models predict significant reductions and/or shifts in species ranges in the terminal Pleistocene and Early Holocene compared to present conditions. Many species that are found throughout the region's 'hilly flanks' today are indicated to have much more restricted distributions centered on the Levant, Cyprus and Western Anatolia. In addition, overall ranges shrunk by an average of c. 25% from the terminal Pleistocene to the Early Holocene. Modelled ranges do not reliably predict the observed occurrence of specific species at archaeological sites, but are aligned with more general trends in the relative abundance of species.

1. Introduction

The Pleistocene–Holocene transition in West Asia marked a turning point in global environmental history, as humans brought the first plants under cultivation and began modifying surrounding ecosystems to support their own subsistence. West Asia is part of the native range of a remarkable number of domesticable plant species, including wild relatives of wheat, barley, peas, lentils, and

*Corresponding author
Email address: joerroe@hey.com (Joe Roe)

other crops of global importance (Diamond, 2002). These species supported uniquely dense and complex Late Epipalaeolithic (15–11.7 ka) societies based on intensive foraging (Bar-Yosef, 1998; Maher et al., 2012) and eventually plant management and pre-domestication cultivation (Harris, 2007). As they were brought under domestication in the Pre-Pottery Neolithic period (11.7–8.5 ka), the world’s first agriculture was shaped by the ecosystems from which it emerged and was embedded in.

Decades of research in archaeobotany and zooarchaeology have reconstructed the subsistence economies of Late Epipalaeolithic and Neolithic sites in great detail. Together with studies of charcoal, pollen, soil, and a variety of palaeoclimate archives (Jones et al., 2019), they also tell us much about the environments surrounding these settlements. However, each of these sources of evidence is subject to the wide variety of taphonomic and recovery biases that are inherent in any direct record of the past. They are also, by definition, records of the (human) environment at particular times and places. Interpolating these snapshots to give a holistic picture of the regional ecologies is not straightforward – to date, it has tended to rely on non-explicit, inductive modelling. The majority are also filtered through human action, producing a mixed signal that makes it difficult to disentangle anthropic effects from the background of environmental change in this period of rapid climatic alteration.

In this paper we present a complementary, deductive approach based on ecological niche modelling. Rather than inferring environmental conditions from preserved physical evidence, we predict the ranges of individual species relevant to human subsistence based on a model of their current environmental niche and simulations of past palaeoclimate. Though hypothetical, this gives us an independent line of evidence on past ecologies that is independent of the environmental archaeological and palaeoclimatic records. Our computational approach is also readily scaled up, allowing us to model spatially-explicit palaeodistributions for a large number of species, for the whole region, under multiple past climatologies.

2. Background

Ecological niche modelling or species distribution modelling is widely used by ecologists to predict the geographic range of a species based on a set of environmental predictors (Franklin and Miller, 2009). Essentially, it involves combining records of where an organism has been observed with environmental data (climate, topography, etc.) for those locations to model the range of environmental values at which that species – its environmental niche. This model can then be used to predict the range of the organism in question either in the same or a different environment. Townsend Peterson and Soberón (2012) suggests reserving the term ‘species distribution modelling’ for when the method is used to recover the verifiable range of a species in a real and existing environment, and using ‘ecological niche modelling’ as the broader term covering hypothetical or predictive applications – a convention we follow here when referring to predictive or

‘hindcast’ models of past ranges. Within this overarching framework, ecological niche modelling encompasses a wide range of applications and a variety of potential environmental predictors, modelling approaches, and methodologies, which we will not attempt to review here.

Ecological niche modelling has long been of interest to archaeologists as both a means of exploring the biological niche of humans and for reconstructing the past environments they inhabited ([David Polly and Eronen, 2011](#); [Franklin et al., 2015](#)). In the first sense, it has been used most extensively to model the range of humans and other hominin species (e.g. [Benito et al., 2017](#); [Yousefi et al., 2020](#); [Banks et al., 2021](#); [Yaworsky et al., 2024a,b](#); [Guran et al., 2024](#)), especially in the Palaeolithic. This overlaps with what archaeologists usually call generically ‘predictive modelling’ ([Verhagen and Whitley, 2020](#))—or more precisely ‘site distribution modelling’—which is essentially the same approach as (and often borrows methodologies from) ecological niche modelling but applied to the occurrence of archaeological sites. Here what is modelled is not strictly a biological niche alone, but also aspects of human geography, taphonomy, and archaeological visibility. These applications can be distinguished from ‘palaeoecological niche modelling’, where the object of model remains, as in ecology, a non-human biological niche.

[Franklin et al. \(2015\)](#) review palaeoecological niche modelling and advocate for its greater adoption in environmental archaeology. In an early application to West Asia, [Conolly et al. \(2012\)](#) used the occurrence of wild and domestic *Bos* remains at prehistoric archaeological sites to map the evolving niche of cattle over the Pleistocene–Holocene transition. It has been used to model the availability of fauna exploited by humans at wider scales (e.g. [de Andrés-Herrero et al., 2018](#); [Yaworsky et al., 2023](#)) and, in a West Asian context, of foraged plant resources in the landscape around the Neolithic sites on the Konya Plain ([Collins et al., 2018](#)). Modelling the spread of crops has been another significant archaeological application (e.g. [Krzyszanska et al., 2022](#); [Krzyszanska, 2023](#)), though not as yet applied to West Asia.

In the majority of studies to date (palaeo)ecological niche modelling has been applied to archaeological data in an ‘inductive’ fashion, i.e. faunal and botanical remains from ancient sites are used as the occurrence dataset for training a model using either past or present environmental data. However, both the zooarchaeological and archaeobotanical records are sparse and subject to a complex array of depositional, taphonomic and recovery biases factors that , many of which are not fully understood and/or cannot be corrected for. This means that while the archaeological attestation of the presence of a species might generally be relied upon, it is highly unlikely that its absence is representative of true past distributions.

The alternative approach is to train the model using contemporary occurrence and environmental data and then use palaeoenvironmental data to ‘hindcast’ its predictions backwards in time. Like [Franklin et al. \(2015\)](#), we view the hindcasting approach as more promising, because training datasets for both occurrences

and environment are far more readily available, complete and reliable for the present than the past. There is some scepticism in the ecological niche modelling literature about the ability of such models to make accurate predictions in unknown environments (like the past, [Franklin et al., 2015](#)), but here the hindcasting approach also presents an opportunity: it reserves archaeological occurrence data as an independent dataset that can be used to assess the retrodictive performance of the model. This possibility was suggested by [Franklin et al. \(2015\)](#) but to our knowledge our study represents the first attempt to actually do so.

The major practical limitation of the hindcasting approach is that it relies on spatially explicit, high resolution palaeoenvironmental surfaces with continuous coverage of the region and periods of interest. Until recently, this has not been widely available for most applications, which is perhaps why only a minority of studies use it (cf. [Krzyszanska et al., 2022](#); [Yaworsky et al., 2023](#)). In this study, we are able to take advantage of the increasing availability of high resolution, global palaeoclimate data derived from simulation experiments with general circulation models of climate ([Brown et al., 2018, 2020](#); [Karger et al., 2023](#)).

3. Data and model

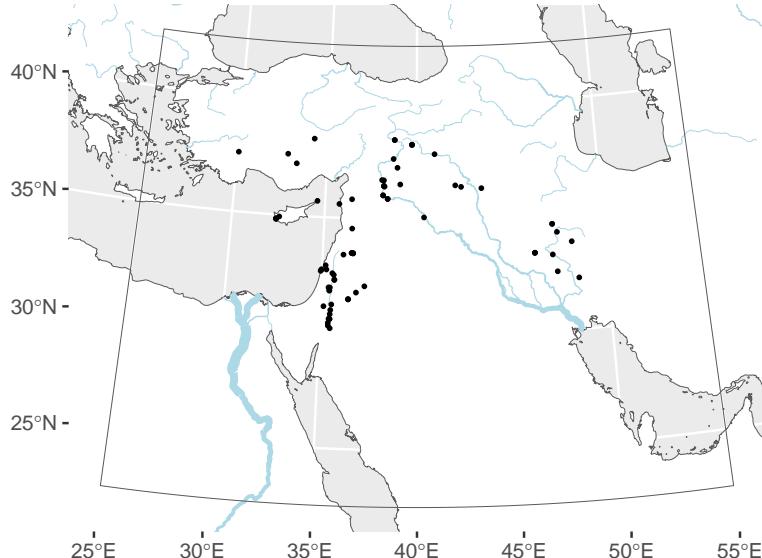


Figure 1: Late Epipalaeolithic & Pre-Pottery Neolithic archaeological sites used to generate modelled flora

The aim of our study was to model the biogeography of species relevant to human subsistence economies in West Asia (excluding the Southern Arabian peninsula, see Figure 1) during the archaeological Late Epipalaeolithic (15–11.7

ka) and Pre-Pottery Neolithic (11.7–8.3 ka) periods. Our starting point was a list of 97 taxa (Table 1) comprising the identifiable species observed at at least two Late Epipalaeolithic/Pre-Pottery Neolithic sites, according to our previous study of the regional archaeobotanical data (Arranz-Otaegui and Roe, 2023). This was based on dataset collated from three previously published regional archaeobotanical databases: ADEMNES (Riehl and Kümmel, 2005), ORIGINS (Wallace et al., 2018b), and COMPAG (Lucas and Fuller, 2018; Fuller et al., 2018; based on Colledge et al., 2004; Shennan and Conolly, 2007). We did not attempt to distinguish between the source of the remains (cf. Wallace et al., 2018a); archaeobotanical assemblages are subject to a variety of preservational and recovery biases, so by no means were all the species on our list consumed or even deliberately collected by people. However, we assume that there presence at a site of human settlement at least implies that they were part of the wider ecosystem that supported habitation there.

Taxonomic names were resolved to the canonical form specified in the GBIF Backbone Taxonomy (GBIF Secretariat, 2023). So for example occurrences for *Bolboschoenus maritimus* also include those recorded under the older nomenclature *Scirpus maritimus* (see Table 1). Domestic species meeting our inclusion criteria were substituted with their wild progenitor(s), where different.

3.1. Occurrence data

Georeferenced occurrence data for West Asia was obtained from the Global Biodiversity Information Facility (GBIF) using via its application programming interface and the R package ‘rgbif’ (Chamberlain and Boettiger, 2017; Chamberlain et al., 2024). GBIF was cleaned to removed fossil occurrences, recorded absences, and records with missing, imprecise (>1 km uncertainty), or invalid coordinates. Although niche models have reasonable predictive power even with small training samples (Stockwell and Peterson, 2002; Hernandez et al., 2006; Wisz et al., 2008), we excluded 14 taxa with less than 50 usable occurrences, following recommendations for niche models generally and Random Forest-based models specifically (Stockwell and Peterson, 2002; Luan et al., 2020). We also excluded one taxon (*Avena sterilis*) with over 50,000 occurrences, as this would have been computationally prohibitive and we were uncertain what accounted for such a disproportionately high number of records.

Occurrence data only tells us where a species is present; there is rarely definitive information on where the species is *not* found. We therefore need to generate random background points or “pseudo-absences” to feed to the model. There are several ways to do this. We follow the advice of Barbet-Massin et al. (2012) for regression-based species distribution models and use a large (10000) random sample of points, weighted equally against the presences in the regression. Valavi et al. (2022) also recommend using a large background sample for random forest models.

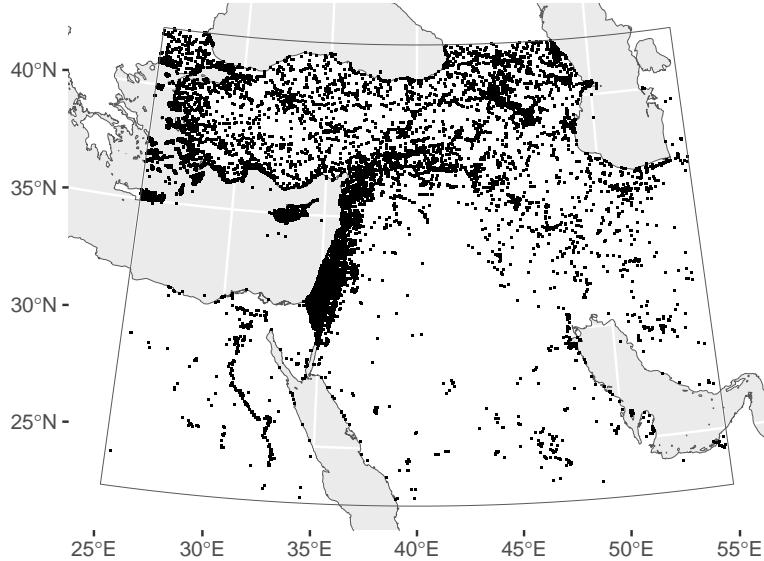


Figure 2: Georeferenced occurrence records from West Asia used to train models (N=81641)

3.2. Predictor data

We modelled the occurrence of species as a function of 24 geospatial predictor variables. These included:

- Sixteen ‘bioclimatic’ variables derived from monthly temperature and precipitation values, following standard practice for species distribution models ([Hijmans et al., 2005](#)). Contemporary bioclimatic predictor data for West Asia was extracted from the global CHELSA dataset ([Karger et al., 2017](#)), which predicts temperature and precipitation from downscaled general circulation model output at 1 km resolution.
- Terrain aspect and slope, which at high resolution perform well as proxies for solar radiation when modelling plant occurrence ([Austin and Van Niel, 2011](#); [Leempoel et al., 2015](#)); and the topographic wetness index (TWI), which serves as a proxy for soil moisture and is particularly important in modelling arid environments ([Kopecký and Čížková, 2010](#); [Campos et al., 2016](#); [Di Virgilio et al., 2018](#)). All three were derived from the SRTM15+ digital elevation model using algorithms from WhiteboxTools ([Lindsay, 2016](#)).
- Edaphic data from SoilGrids ([Hengl et al., 2014, 2017](#)), which improves model performance for plants ([Dubuis et al., 2013](#); [Mod et al., 2016](#); [Velazco et al., 2017](#)). Based on a recent assessment of the reliability of SoilGrids data for species distribution modelling ([Miller et al., 2024](#)), we used a subset of four variables relating to soil texture (clay, silt, sand) and

pH at the surface (0–5 cm depth).

For hindcasting, we used reconstructed bioclimatic data for three key climatologies generated from downscaled paleoclimate simulations from the HadCM3 general circulation model (Fordham et al., 2017; Brown et al., 2018): the Bølling–Allerød (c. 14.7–12.9 ka), the Younger Dryas (c. 12.9–11.7 ka), and the Early Holocene (11.7–8.3 ka). Terrain and soil predictors were held constant, since reconstructions of these variables in the past are not available at sufficient scale. It is unlikely that either macroscale topography or soil characteristics have altered significantly over the period of time considered here, so we assume that this does not degrade model performance, and may in fact benefit it by providing ‘anchoring’ predictors that are independent of climate change.

For training, test predictions, and archaeological predictions predictor data was left in its native projection and resolution. For hindcast palaeodistributions, it was transformed to common equal-area projection and resolution of 5 km.

3.3. Random Forest

Ecological niche modelling is a classification problem that can be approached with a wide range of statistical methods. A substantial literature exists on the relatively performance of these approaches and their respective parameterisations (reviewed in Valavi et al., 2022). Random Forest, a widely-used machine learning algorithm, is amongst the best performing methods for presence-only species distribution models, providing it is appropriately parameterised to account for the class imbalance between presence and background samples (Valavi et al., 2021, 2022). For our application, it also has the advantage of requiring little to no manual parameter tuning to achieve good predictive results, which makes it easier to model a larger numbers of taxa.

For each taxon we trained a classification model to predict occurrence (presence or absence/background) based on our 24 predictor variables (Section 3.2). We used the Random Forest algorithm implemented in the R package ‘ranger’ (Wright and Ziegler, 2017) and the ‘tidymodels’ (Kuhn and Silge, 2022) framework for data preprocessing and model selection. To avoid overfitting, we follow Valavi et al. (2021) in their recommended hyperparameters and use of down-sampling to balance presence and background samples. Models for each taxon were fit independently, with redundant zero-variance predictors excluded, and assessed based on balanced training ($\frac{3}{4}$) and test ($\frac{1}{4}$) partitions.

4. Model assessment

We trained Random Forest models for 82 taxa using contemporary occurrence data from GBIF, a random sample of background points, and the predictor variables described in Section 3.2. Substituting the “current” climate predictors for those derived from palaeoclimatic simulations (Brown et al., 2018), we could then generate hindcast predictions for reconstructed past environments in 4 key

climate periods – a total of 328 modelled palaeodistributions. Predicted distributions for individual taxa are presented in the appendix and accompanying material.

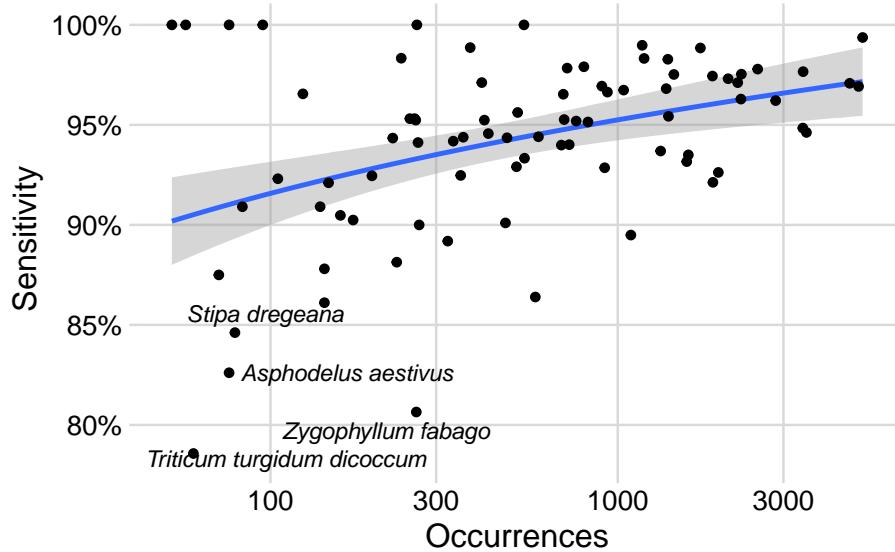


Figure 3: Model sensitivity by number of training occurrences

We assessed the predictive performance of the fitted niche models in the contemporary environment based on the reserved test partition Table 1. Model accuracy (proportion of correctly classified presence and background samples) ranged between 73% and 99%, with an average of 93%. Sensitivity (proportion of correctly classified presence samples) ranged between 79% and 100%, with an average of 94%. The area under the models’ receiver operating characteristic curves (ROC-AUC) was on average 0.978 ± 0.061 . Model sensitivity is loosely correlated with the number of occurrences available for training (Figure 3), with the worst-performing models all having less than 300 recorded occurrences: *Asphodelus aestivus*, *Triticum turgidum dicoccum*, *Zygophyllum fabago*, and *Stipa dregeana*. Test metrics and ROC curves for the individual models are included in the appendix.

The ability of the hindcast models to predict the occurrence of specific species at archaeological sites is worse, with 10% of presences in archaeobotanical assemblages successfully predicted. Model sensitivity (proportion of correctly predicted presences) in relation to the archaeological data is on average 0.06 ± 0.13 Table 1. A full assessment of the hindcasting performance of the individual models can be found in the appendix.

Table 1: Summary of species modelled

Taxon	Occurrences			Model		
	Arch.	Cur.	Acc.	ROC-AUC	Sens.	Sens. (arch.)
<i>Adonis dentata</i>	2	715	0.99	1.00	0.98	0.00
<i>Aegilops crassa</i>	4	143	0.82	0.94	0.86	0.00
<i>Aegilops speltoides</i>	—	701	0.92	0.98	0.95	—
<i>Aegilops tauschii</i>	—	725	0.89	0.98	0.94	—
<i>Aizoanthemopsis hispanica</i> ¹	4	515	0.98	1.00	0.96	0.00
<i>Ammi majus</i>	4	900	0.98	0.99	0.97	0.00
<i>Androsace maxima</i>	17	231	0.88	0.95	0.88	0.12
<i>Arenaria serpyllifolia</i>	3	139	0.84	0.95	0.91	0.00
<i>Arnebia decumbens</i>	21	324	0.95	0.95	0.89	0.00
<i>Arnebia linearifolia</i>	13	225	0.98	0.99	0.94	0.00
<i>Asphodelus aestivus</i>	2	76	0.97	0.97	0.83	0.00
<i>Atriplex prostrata</i>	4	266	0.98	1.00	0.94	0.00
<i>Avena fatua</i>	2	579	0.86	0.93	0.86	0.00
<i>Avena sterilis</i>	5	53244 ²	—	—	—	—
<i>Bassia arabica</i>	4	264	0.98	1.00	1.00	0.00
<i>Bassia prostrata</i>	2	46 ²	—	—	—	—
<i>Bolboschoenus glaucus</i>	5	27 ²	—	—	—	—
<i>Bolboschoenus maritimus</i> ³	31	1040	0.98	0.99	0.97	0.03
<i>Brachypodium distachyon</i>	5	2270	0.98	0.99	0.98	0.00
<i>Bromus sterilis</i>	3	759	0.97	0.99	0.95	0.00
<i>Buglossoides arvensis</i>	23	260	0.87	0.97	0.95	0.09
<i>Buglossoides tenuiflora</i>	26	252	0.96	0.99	0.95	0.08
<i>Camphorosma monspeliacana</i>	2	38 ²	—	—	—	—
<i>Capparis spinosa</i>	4	4938	0.97	0.99	0.97	0.00
<i>Capsella bursa-pastoris</i>	2	1329	0.94	0.99	0.94	0.00
<i>Carex divisa</i>	9	336	0.95	0.98	0.94	0.00
<i>Celtis tournefortii</i>	2	14 ²	—	—	—	—
<i>Ceratonia siliqua</i>	2	5072	0.98	1.00	0.99	0.50
<i>Chenopodium album</i>	5	480	0.93	0.98	0.94	0.00
<i>Cicer reticulatum</i> ⁴	3	52	0.97	1.00	1.00	0.00
<i>Citrullus colocynthis</i>	4	475	0.89	0.96	0.90	0.00
<i>Crithopsis delileana</i>	3	406	0.98	1.00	0.97	0.00
<i>Erodium ciconium</i>	2	359	0.92	0.99	0.94	0.00
<i>Euclidium syriacum</i>	3	32 ²	—	—	—	—
<i>Ficus carica</i>	8	3495	0.94	0.99	0.95	0.25
<i>Fumaria densiflora</i>	4	413	0.97	0.99	0.95	0.00
<i>Gypsophila elegans</i>	3	76	0.92	0.99	1.00	0.00
<i>Gypsophila pilosa</i>	5	173	0.93	0.97	0.90	0.00
<i>Gypsophila vaccaria</i> ⁵	8	511	0.96	0.99	0.93	0.00
<i>Halothamnus hierochunticus</i>	2	57	0.98	1.00	1.00	0.00
<i>Helianthemum ledifolium</i>	2	697	0.97	0.99	0.97	0.00
<i>Helianthemum salicifolium</i>	2	1400	0.97	0.99	0.95	0.00
<i>Henrardia pubescens</i>	3	1 ²	—	—	—	—
<i>Hordeum bulbosum</i>	4	2850	0.97	0.99	0.96	0.00
<i>Hordeum murinum</i>	5	1949	0.94	0.98	0.93	0.00
<i>Hordeum spontaneum</i> ⁶	110	4656	0.96	0.99	0.97	0.20
<i>Juglans regia</i>	2	821	0.85	0.96	0.95	0.00
<i>Krascheninnikovia ceratoides</i>	2	42 ²	—	—	—	—
<i>Lathyrus aphaca</i>	4	2079	0.94	0.99	0.97	0.00
<i>Lathyrus cicera</i>	2	591	0.89	0.98	0.94	0.50

<i>Lathyrus oleraceus</i> ⁷	8	1580	0.89	0.97	0.93	0.12
<i>Lathyrus sativus</i>	4	196	0.79	0.95	0.92	0.00
<i>Lepidium perfoliatum</i>	3	83	0.77	0.94	0.91	0.00
<i>Lepidium ruderale</i>	2	37 ²	—	—	—	—
<i>Linum bienne</i> ⁸	14	376	0.96	1.00	0.99	0.00
<i>Lolium rigidum</i>	5	2529	0.98	1.00	0.98	0.00
<i>Lolium temulentum</i>	3	159	0.94	0.96	0.90	0.00
<i>Medicago astroites</i> ⁹	15	143	0.89	0.96	0.88	0.33
<i>Medicago minima</i>	2	1379	0.92	0.99	0.97	0.50
<i>Medicago radiata</i>	20	799	0.93	0.99	0.98	0.20
<i>Minuartia rubella</i>	2	0 ²	—	—	—	—
<i>Moltkia angustifolia</i>	2	11 ²	—	—	—	—
<i>Moltkia coerulea</i>	2	71	0.86	0.97	0.88	0.00
<i>Neotorularia torulosa</i> ¹⁰	2	539	0.98	0.98	0.93	0.00
<i>Peganum harmala</i>	2	917	0.92	0.97	0.93	0.00
<i>Phalaris paradoxa</i>	3	1191	0.99	1.00	0.98	0.00
<i>Phragmites australis</i>	4	3412	0.97	0.99	0.95	0.00
<i>Pistacia atlantica</i>	6	1874	0.98	1.00	0.97	0.17
<i>Pistacia terebinthus</i> ¹¹	2	3419	0.98	1.00	0.98	0.50
<i>Plantago lagopus</i>	2	1730	0.98	1.00	0.99	0.00
<i>Poa bulbosa</i>	5	1880	0.95	0.98	0.92	0.00
<i>Polygonum arenarium arenarium</i> ¹²	7	5 ²	—	—	—	—
<i>Polygonum corrigioloides</i>	6	4 ²	—	—	—	—
<i>Prosopis farcta</i>	5	2263	0.98	0.99	0.96	0.00
<i>Quercus ithaburensis</i>	2	2216	0.99	1.00	0.97	0.00
<i>Rumex pulcher</i>	6	1394	0.98	1.00	0.98	0.00
<i>Salsola kali</i>	6	537	0.98	1.00	1.00	0.00
<i>Salvia absconditiflora</i> ¹³	3	105	0.94	0.98	0.92	0.00
<i>Secale cereale</i>	7	268	0.81	0.94	0.90	0.00
<i>Secale strictum</i> ¹⁴	3	124	0.76	0.94	0.97	0.00
<i>Stipa dregeana</i>	2	79	0.95	0.96	0.85	0.00
<i>Suaeda fruticosa</i>	3	238	0.99	1.00	0.98	0.00
<i>Taeniatherum caput-medusae</i>	5	262	0.81	0.96	0.95	0.00
<i>Tribulus terrestris</i>	2	1091	0.90	0.96	0.89	0.00
<i>Triticum aestivum compactum</i>	2	147	0.82	0.96	0.92	0.00
<i>Triticum aestivum spelta</i> ¹⁵	2	16 ²	—	—	—	—
<i>Triticum durum</i>	3	95	0.91	0.99	1.00	0.00
<i>Triticum monococcum aegilopoides</i> ¹⁶	62	1176	0.88	0.99	0.99	0.08
<i>Triticum turgidum dicoccum</i> ¹⁷	60	60	0.73	0.86	0.79	0.18
<i>Triticum urartu</i>	—	424	0.93	0.99	0.95	—
<i>Verbena officinalis</i>	3	934	0.96	0.99	0.97	0.00
<i>Vicia ervilia</i>	26	688	0.87	0.96	0.94	0.12
<i>Vicia faba</i>	7	1597	0.90	0.98	0.94	0.43
<i>Vicia narbonensis</i>	3	1451	0.93	0.99	0.98	0.33
<i>Vicia orientalis</i> ¹⁸	16	353	0.84	0.95	0.92	0.19
<i>Vitis sylvestris</i>	3	12 ²	—	—	—	—
<i>Zygophyllum fabago</i>	3	263	0.89	0.93	0.81	0.00

- ¹Including *Aizoon hispanicum*
²Excluded from modelling due to sample size
³Including *Scirpus maritimus*
⁴Including *Cicer arietinum*
⁵Including *Vaccaria pyramidata*
⁶Including *Hordeum vulgare*
⁷Including *Pisum sativum* and *Pisum elatius*
⁸Including *Linum usitatissimum*
⁹Including *Trigonella astroites*
¹⁰Including *Torularia torulosa*
¹¹Including *Pistacia palaestina*
¹²Including *Polygonum venantianum*
¹³Including *Salvia cryptantha*
¹⁴Including *Secale montanum*
¹⁵Including *Triticum spelta*
¹⁶Including *Triticum monococcum* and *Triticum boeoticum*
¹⁷Including *Triticum aestivum*, *Triticum dicoccum*, and *Triticum dicoccoides*
¹⁸Including *Lens culinaris*

5. Discussion

5.1. Reduction in range sizes over the Pleistocene/Holocene boundary

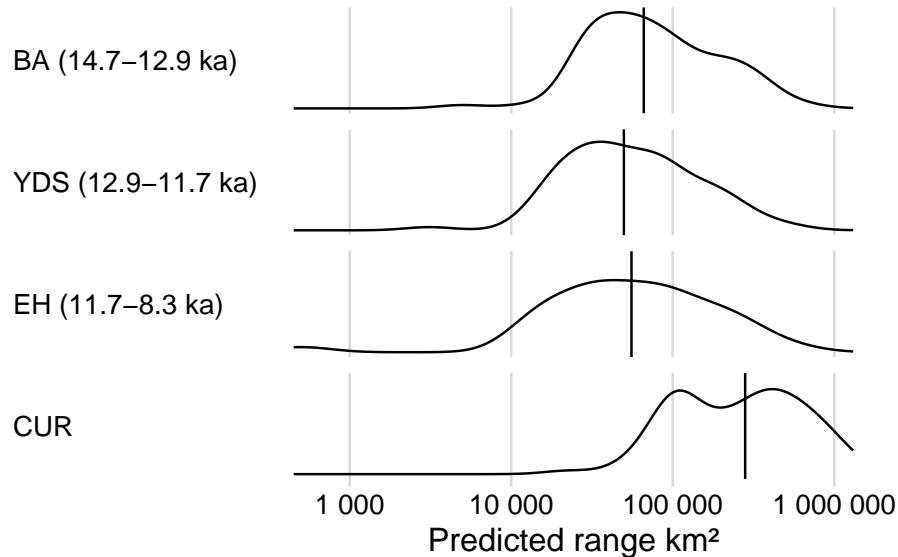


Figure 4: Distribution of predicted species ranges by period. Dashed lines indicate the median range.

Our reconstructed palaeodistributions (shown in full in the appendix) indicate that the majority of species have experienced a significant reduction and/or shift in their ranges since the Early Holocene. 75 of 82 species have reduced ranges; 66

of more than 10% or more. Though the magnitude of the reduction from prehistory to the present likely reflects a degree of overfitting in the model (discussed further below), fluctuations in modelled range size between the Bølling–Allerød (14.7–12.9 ka), Younger Dryas (12.9–11.7 ka), and Early Holocene (11.7–8.3 ka) are more directly comparable (Figure 4). The average range of modelled species was 16% in the Early Holocene compared to the Bølling–Allerød, and 25% during the Younger Dryas (i.e. ranges recovered slightly between the Younger Dryas and Early Holocene). This perhaps indicates that although this period is considered one of climatic amelioration globally (Jones et al., 2019), the colder conditions of the Pleistocene may have supported more extensive plant-based economies in West Asia specifically.

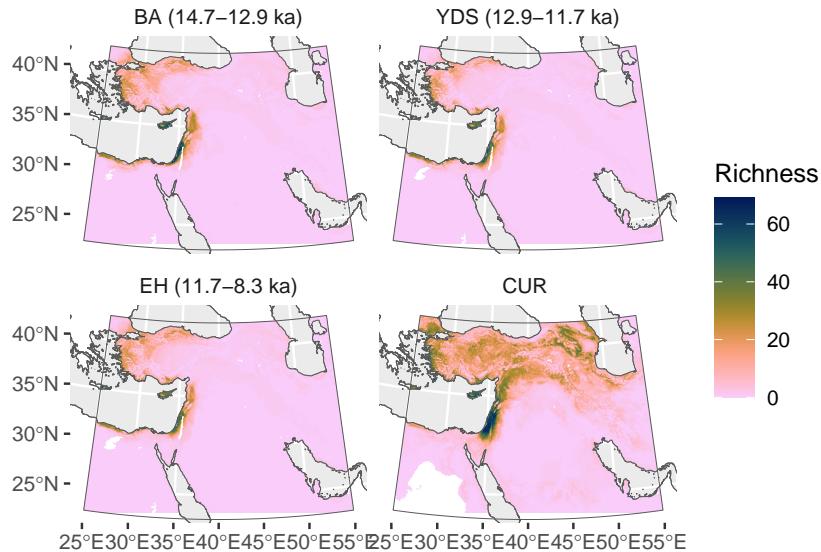


Figure 5: Predicted species richness (sum of predicted ranges) by period

Many taxa that occur (or are predicted to occur) across the ‘hilly flanks’ today—including most crop progenitors—are reconstructed to have had a significantly more restricted distribution in the terminal Pleistocene/Early Holocene (Figure 5). These include *Ficus carica* (fig); *Hordeum* spp. (wild barleys); *Lathyrus aphaca* and *L. sativus* (both marginally edible legumes); *Triticum aestivum compactum* (in the N. Levant), *T. monococcum aegilopoides*, *T. durum*, and *Triticum urartu* (but not the other wheat progenitor, *T. turgidum dicoccum* – see below); *Aegilops speltoides*, but not *Aegilops tauschii* (goatgrasses); and *Vicia* spp. (vetches), including *Vicia faba* (broad beans). Most of Anatolia, Northern Mesopotamia, and the Zagros Mountains in particular disappear from the predicted ranges of these species, leaving the Levant and to a lesser extent the Aegean and Cyprus as refugia.

Our results for the Levant are consistent with the current understanding of this region as developing early intensive foraging economies (the Natufian culture, Bar-Yosef, 1998) and as a centre of origin of agriculture (Zeder, 2011). Within the Levant, many species show moderate reductions in ranges over the Pleistocene/Holocene boundary, retreating from the Badia/transjordan region (e.g. barley, fig. Figure 8).

Loss of the Northern Mesopotamia–Anatolia region from the predicted ranges of crop progenitors is interesting in light of the ‘golden triangle’ hypothesis (Lev-Yadun et al., 2000; Kozłowski and Aurenche, 2005; Abbo et al., 2010), which puts this region at the centre of the development of agriculture. Multiple lines of archaeological evidence have emerged that point away from this hypothesis and towards a more geographically diverse origin (Asouti, 2006; Fuller et al., 2011; Arranz-Otaegui et al., 2016), but it remains the area with some of the earliest clear evidence of domestication (Zohary et al., 2012; Kabukcu et al., 2021; Ulaş et al., 2024). Comparative genetics also points to the Northern Mesopotamia region as the centre of diversity of many crops (e.g. Haas et al., 2019). But since these studies are based on modern genomes, if the wild range of these plants has, as our modelling suggests, shifted since the Early Holocene, the apparent centre of diversity may have shifted with them. Our reconstructions are consistent with the late arrival of intensive plant-based foraging economies in this region (cf. the Natufian of the Levant), and more broadly there need not be a link between the core *wild* range of a plant and the core area of its domestication. A scenario in which cultivation emerged at the edges of the ranges of valuable plant resources—as a means of extend their natural niche—is also plausible.

The near-absence of the Zagros in any predicted ranges is also surprising, given mounting evidence that domestication took place just as early in the eastern *Mashriq* as it did in the west (Zeder, 2024). We consider that the most likely explanation for this is that our flora does not include the species that were most important to plant subsistence in the east. Archaeobotanical data on Neolithic sites in the Zagros is limited (compared to the Levant in particular) due to a hiatus in field research there from the 1980s to early 2010s (Zeder, 2024). Recent research (Whitlam et al., 2018; González Carretero et al., 2023) indicates that plant subsistence in this region was based on a distinct set of species, compared to the Levant and Anatolia.

Cyprus and the Aegean are not conventionally considered part of the primary zone of domestication but rather amongst the first regions that acquired agriculture from West Asia. Our analysis complicates this picture, as it indicates that the wild ranges of many crop progenitors included these regions. Early examples of several domesticates are recorded at sites on Cyprus, Western Anatolia and Greece (Arranz-Otaegui and Roe, 2023), and the Aegean region was probably connected to West Asia by a land bridge via Anatolia until the Early Holocene (Aksu and Hiscott, 2022). Were these area part of the same broader ‘interaction sphere’ that produce Neolithic agriculture in West Asia?

Exceptions to the dominant trend of range reduction include *Cicer reticula-*

tum (wild chickpea), which has a relatively stable range centered on Northern Mesopotamia; and *Triticum turgidum dicoccum* (wild emmer wheat), which is predicted to have two limited ranges centered around the Black Sea Coast of Anatolia and the Palmyra basin. In the latter case, neither of these areas are part of the predicted modern distribution of wild emmer (centered around the Caucasus and Northern Mesopotamia), but it would be consistent with archaeological evidence for early cultivation at sites in the Upper Euphrates (Willcox, 2024).

5.2. Biogeography of crop progenitors

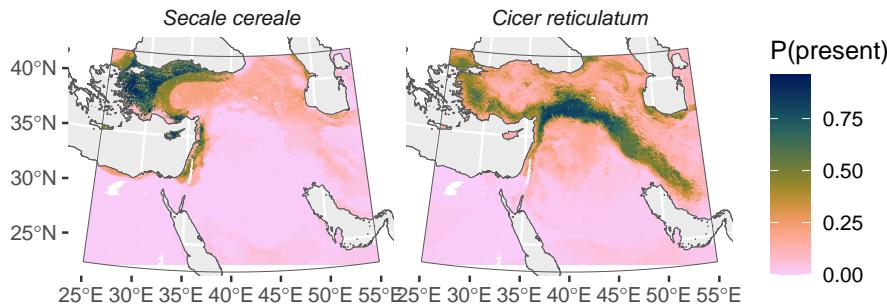


Figure 6: Predicted palaeodistribution of wild chickpea and rye in the Early Holocene (11.7–8.2 ka)

Almost all the cereal and legume crop progenitors we modelled are predicted to have only been found in the Levant during the terminal Pleistocene and Early Holocene (see appendix). Part of this may be due to the fact that both our initial flora and training occurrence dataset have a strong bias towards the southern Levant, but strikingly the modelled *current* ranges of these plants do tend to include Anatolia and the Zagros, so this cannot be the only factor. One notable exception is the wild ancestor of chickpea (*Cicer reticulatum*), which is predicted to have a distribution centered on Northern Mesopotamia but encompassing much of the 'hilly flanks' (except the southern Levant, Figure 6). Another is rye (*Secale cereale*), which is inferred to be primarily Western Anatolian (Figure 6). This is perhaps relevant to rye's unusual domestication history, as a crop of West Asian origin that was apparently not first cultivated until much later than the 'founder crops', in Europe (Schreiber et al., 2021).

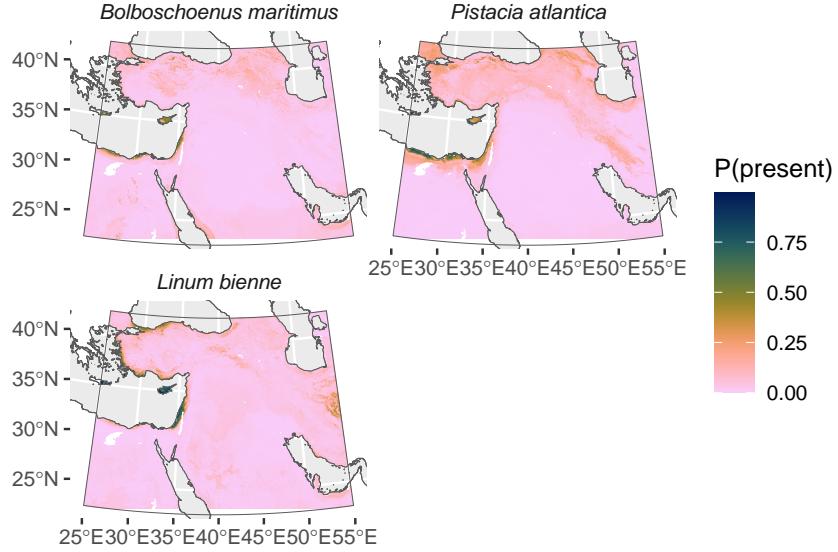


Figure 7: Predicted palaeodistribution of flax, pistachio and clubrush in the Early Holocene (11.7–8.2 ka)

Flax (*Linum bienne*) is predicted to have had a highly concentrated distribution in Cyprus and along the Mediterranean coast of the southern Levant. This is consistent with its low ubiquity in archaeobotanical assemblages (Arranz-Otaegui and Roe, 2023), despite conventionally being considered a ‘founder crop’, and presumably implies that its domestication was similarly geographically constrained. It is the only unambiguous crop progenitor with such a restricted range, though pistachio (*Pistacia atlantica*) and clubrush (*Bolboschoenus maritimus*) are similarly constrained to the Mediterranean coast (and North Africa, in the case of pistachio).

Wild barley (*Hordeum spontaneum*) and its relatives show a contraction of their predicted ranges from the Pleistocene to the Holocene, concurrent with it being brought into cultivation (Figure 8). It also sees a marked decline in the archaeobotanical record from the Early PPNA/Early PPNB (where it was amongst the most common taxa) to the Late PPNB and Late Neolithic (Arranz-Otaegui and Roe, 2023). Pistachio (*Pistacia atlantica*) shows similar trends, but it is less certain that this species was cultivated in the Neolithic. Conversely, the various wild wheat species native to West Asia show almost no response to Pleistocene/Holocene climate change, even within the Levant, and in the archaeobotanical record wheat displays the opposite trend to barley and pistachio – becoming gradually more abundant through the course of the Neolithic and dominant by its end. We hypothesise that climate-linked range shifts were a factor driving these changes in the apparent economic importance of different crops: the decreasing availability of barley and pistachio in the terminal Pleistocene

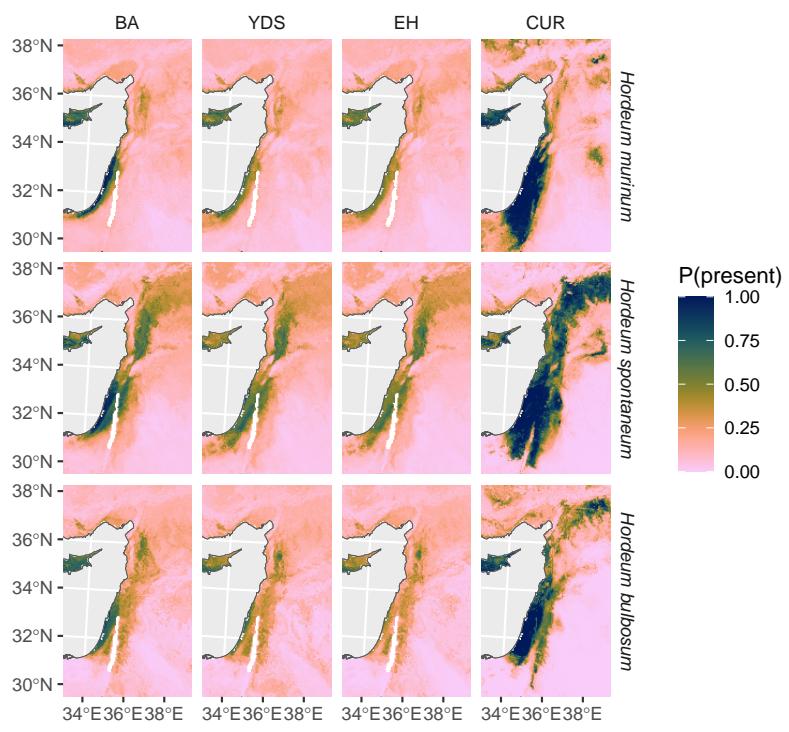


Figure 8: Predicted palaeodistribution of wild barleys in the Levant

may have prompted cultivation as a strategy to retain access to them, though in the long run it made them less attractive as staple crops compared to the more resilient wheat. However, these dynamics are not seen in the majority of crop progenitors.

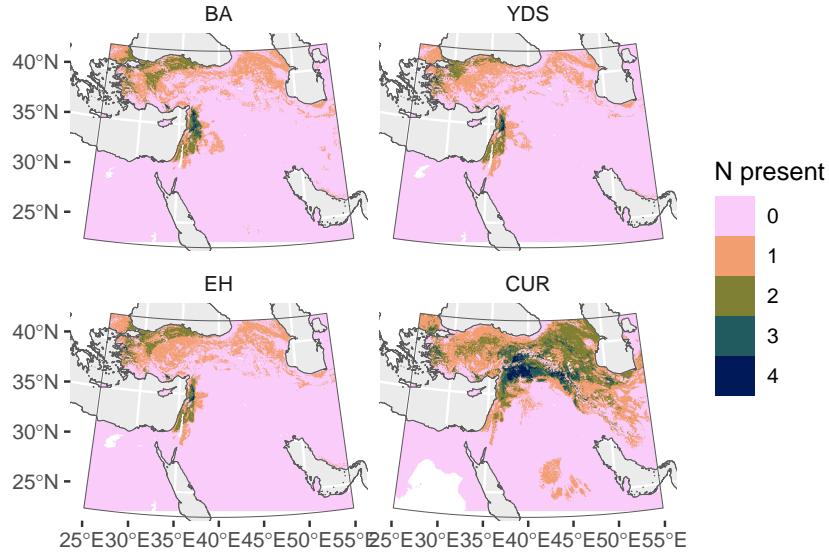


Figure 9: Combined predicted palaeodistributions of bread wheat progenitors

Bread wheat (*Triticum aestivum*), the most common wheat cultivar today, has a complex ancestry that involves two recent hybridisation events (Levy and Feldman, 2022): most recently between emmer (*Triticum turgidum dicoccum*) and a goatgrass (*Aegilops tauschii*) c. 9 ka, and before that, in emmer, between wild red einkorn (*T. urartu*) and another goatgrass (*Aegilops speltoides*). Although there is fairly wide overlap between these species today, according to our modelling the only place all four (or even just three) are predicted to have coincided or been found in close proximity to each other in the time frame of domestication is a narrow area around the Orontes river in the northern Levant (Figure 9). This area is c. 200 km from the only two sites in our archaeobotanical database where either both species of wheat or both species of wheat and the goatgrasses are found together: Tell Abu Hureyra and El Kown II (Arranz-Otaegui and Roe, 2023).¹ Taken together this suggests an origin of domesticated bread wheat, otherwise only loosely geographically constrained to the Levant–Upper Euphrates corridor (Levy and Feldman, 2022), is likely to occur in this vicinity.

¹We are grateful to Benjamin Nowak for pointing this out to us.

5.3. Hindcast models do not predict archaeobotanical composition

The failure of our hindcast models to predict the occurrence of species in archaeobotanical assemblages has several possible explanations. Since they do accurately predict the test dataset, a likely culprit is overfitting of the models to the present environment. This implies that the modelled palaeodistributions should be seen as conservative estimates or a minimal range. Another obvious flaw in our methodology is that the time slices used for palaeoclimatic reconstruction are very broad—each covering around two millennia—and therefore potentially unrepresentative of the environment around sites at the specific time at which they were occupied. The variable quality of the archaeological test dataset, especially in terms of chronology, is also a plausible factor.

At the same time, we cannot rule out more substantive reasons for the discrepancy between predicted and observed archaeological occurrences. The niches of the modelled species could have changed since the Early Holocene, which would not be captured in a model trained purely on modern specimens. Human economic choices—mobility, foraging strategies, cultivation, etc.—could also produce archaeobotanical assemblages whose composition depart significantly from that of the surrounding local flora. Further refinement of the methodology for hindcast palaeoecological niche models, for example using more finely resolved palaeoclimate sequences (e.g. Karger et al., 2023), hyperparameter tuning to avoid overfitting, and improved archaeological datasets, would help disentangle these potential explanations.

It is consistent with more “macro” trends such as the reduced range of *Hordeum* and *Pistacia* correlating with its reduced abundance in the archaeobotanical record.

6. Conclusion

We present the first continuous, spatially explicit models of the palaeodistributions of 82 plant species found regularly in association with Late Epipalaeolithic and Early Neolithic sites in West Asia. This deductive approach—modelling the niche of a species based on its occurrence in relation to environmental factors today, and using this together with palaeoclimate simulations to infer its past distribution—represents a new line of evidence on the archaeoecology of the world’s first agricultural societies. It provides a complementary picture to that gleaned from environmental archaeology and climatological archives because it is independent of the taphonomic, anthropic, and recovery-related processes that affect these records.

“All models are wrong” (Box, 1976), but the performance of our models on independent test datasets give confidence in its predictions, which are plausible and consistent with broad-scale patterns in the archaeological record. Modelling a large number of species using machine learning, the substantial occurrence datasets available for the present day, and a hindcasting approach to past

distributions also represents a significant advance in the methodology of palaeoecological niche modelling. This approach is enabled by the availability of high quality, global open datasets in ecology (GBIF, 2025; GBIF Secretariat, 2023), earth science (Farr et al., 2007), and climatology (Karger et al., 2017; Brown et al., 2018). Unfortunately in relation to these fields open data in archaeology lags conspicuously behind. Though we have benefitted from the relatively long tradition of compiling archaeobotanical data in our region of study (Colledge et al., 2004; Shennan and Conolly, 2007; Riehl and Kümmel, 2005; Lucas and Fuller, 2018; Fuller et al., 2018; Wallace et al., 2018a,b), further development of open, comprehensive and up-to-date ‘backbone’ datasets on site locations and chronologies is needed to advance archaeoecological modelling to the same level.

7. Data availability

The data and R code used to produced is archived with Zenodo at <https://doi.org/10.5281/zenodo.14629984>

References

- Abbo, S., Lev-Yadun, S., Gopher, A., 2010. Agricultural Origins: Centers and Noncenters; A Near Eastern Reappraisal. *Critical Reviews in Plant Sciences* 29, 317–328. doi:[10.1080/07352689.2010.502823](https://doi.org/10.1080/07352689.2010.502823).
- Aksu, A.E., Hiscott, R.N., 2022. Persistent Holocene outflow from the Black Sea to the eastern Mediterranean Sea still contradicts the Noah’s Flood Hypothesis: A review of 1997–2021 evidence and a regional paleoceanographic synthesis for the latest Pleistocene–Holocene. *Earth-Science Reviews* 227, 103960. doi:[10.1016/j.earscirev.2022.103960](https://doi.org/10.1016/j.earscirev.2022.103960).
- Arranz-Otaegui, A., Colledge, S., Zapata, L., Teira-Mayolini, L.C., Ibáñez, J.J., 2016. Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. *Proceedings of the National Academy of Sciences* 113, 14001–14006. doi:[10.1073/pnas.1612797113](https://doi.org/10.1073/pnas.1612797113).
- Arranz-Otaegui, A., Roe, J., 2023. Revisiting the concept of the ‘Neolithic Founder Crops’ in southwest Asia. *Vegetation History and Archaeobotany* doi:[10.1007/s00334-023-00917-1](https://doi.org/10.1007/s00334-023-00917-1).
- Asouti, E., 2006. Beyond the Pre-Pottery Neolithic B interaction sphere. *Journal of World Prehistory* 20, 87–126. doi:[10.1007/s10963-007-9008-1](https://doi.org/10.1007/s10963-007-9008-1).
- Austin, M.P., Van Niel, K.P., 2011. Improving species distribution models for climate change studies: Variable selection and scale. *Journal of Biogeography* 38, 1–8. doi:[10.1111/j.1365-2699.2010.02416.x](https://doi.org/10.1111/j.1365-2699.2010.02416.x).
- Banks, W.E., Moncel, M.H., Raynal, J.P., Cobos, M.E., Romero-Alvarez, D., Woillez, M.N., Faivre, J.P., Gravina, B., d'Errico, F., Locht, J.L., Santos, F., 2021. An ecological niche shift for Neanderthal populations in Western

Europe 70,000 years ago. *Scientific Reports* 11, 5346. doi:[10.1038/s41598-021-84805-6](https://doi.org/10.1038/s41598-021-84805-6).

Bar-Yosef, O., 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. *Evolutionary Anthropology* 6, 159–177. doi:[10.1002/\(SICI\)1520-6505\(1998\)6:5<159::AID-EVAN4>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<159::AID-EVAN4>3.0.CO;2-7).

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: How, where and how many?: How to use pseudo-absences in niche modelling? *Methods Ecol. Evol.* 3, 327–338. doi:[10.1111/j.2041-210X.2011.00172.x](https://doi.org/10.1111/j.2041-210X.2011.00172.x).

Benito, B.M., Svenning, J.C., Kellberg-Nielsen, T., Riede, F., Gil-Romera, G., Mailund, T., Kjaergaard, P.C., Sandel, B.S., 2017. The ecological niche and distribution of Neanderthals during the Last Interglacial. *J. Biogeogr.* 44, 51–61. doi:[10.1111/jbi.12845](https://doi.org/10.1111/jbi.12845).

Box, G.E.P., 1976. Science and Statistics. *J. Am. Stat. Assoc.* 71, 791–799. doi:[10.1080/01621459.1976.10480949](https://doi.org/10.1080/01621459.1976.10480949).

Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., Haywood, A.M., 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci Data* 5, 180254. doi:[10.1038/sdata.2018.254](https://doi.org/10.1038/sdata.2018.254).

Brown, S.C., Wigley, T.M.L., Otto-Bliesner, B.L., Fordham, D.A., 2020. StableClim, continuous projections of climate stability from 21000 BP to 2100 CE at multiple spatial scales. *Scientific Data* 7, 335. doi:[10.1038/s41597-020-00663-3](https://doi.org/10.1038/s41597-020-00663-3).

Campos, V.E., Cappa, F.M., Viviana, F.M., Giannoni, S.M., 2016. Using remotely sensed data to model suitable habitats for tree species in a desert environment. *Journal of Vegetation Science* 27, 200–210. doi:[10.1111/jvs.12328](https://doi.org/10.1111/jvs.12328).

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., Ram, K., 2024. rgbif: Interface to the Global Biodiversity Information Facility API. URL: <https://CRAN.R-project.org/package=rgbif>. r package version 3.8.1.

Chamberlain, S., Boettiger, C., 2017. R python, and ruby clients for gbif species occurrence data. PeerJ PrePrints URL: <https://doi.org/10.7287/peerj.preprints.3304v1>.

Colledge, S., Conolly, J., Shennan, S., 2004. Archaeobotanical Evidence for the Spread of Farming in the Eastern Mediterranean. *Current Anthropology* 45, S35–S58. doi:[10.1086/422086](https://doi.org/10.1086/422086).

Collins, C., Asouti, E., Grove, M., Kabukcu, C., Bradley, L., Chiverrell, R., 2018. Understanding resource choice at the transition from foraging to farming: An application of palaeodistribution modelling to the Neolithic of the Konya Plain, south-central Anatolia, Turkey. *J. Archaeol. Sci.* 96, 57–72. doi:[10.1016/j.jas.2018.02.003](https://doi.org/10.1016/j.jas.2018.02.003).

- Conolly, J., Manning, K., Colledge, S., Dobney, K., Shennan, S., 2012. Species distribution modelling of ancient cattle from early Neolithic sites in SW Asia and Europe. *Holocene* 22, 997–1010. doi:[10.1177/0959683612437871](https://doi.org/10.1177/0959683612437871).
- David Polly, P., Eronen, J.T., 2011. Mammal Associations in the Pleistocene of Britain: Implications of Ecological Niche Modelling and a Method for Reconstructing Palaeoclimate, in: Ashton, N., Lewis, S.G., Stringer, C. (Eds.), *Developments in Quaternary Sciences*. Elsevier. volume 14, pp. 279–304. doi:[10.1016/B978-0-444-53597-9.00015-7](https://doi.org/10.1016/B978-0-444-53597-9.00015-7).
- de Andrés-Herrero, M., Becker, D., Weniger, G.C., 2018. Reconstruction of LGM faunal patterns using Species Distribution Modelling. The archaeological record of the Solutrean in Iberia. *Quat. Int.* 485, 199–208. doi:[10.1016/j.quaint.2017.10.042](https://doi.org/10.1016/j.quaint.2017.10.042).
- Di Virgilio, G., Wardell-Johnson, G.W., Robinson, T.P., Temple-Smith, D., Hesford, J., 2018. Characterising fine-scale variation in plant species richness and endemism across topographically complex, semi-arid landscapes. *Journal of Arid Environments* 156, 59–68. doi:[10.1016/j.jaridenv.2018.04.005](https://doi.org/10.1016/j.jaridenv.2018.04.005).
- Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707. doi:[10.1038/nature01019](https://doi.org/10.1038/nature01019).
- Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P., Guisan, A., 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science* 24, 593–606. doi:[10.1111/jvs.12002](https://doi.org/10.1111/jvs.12002).
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Rev. Geophys.* 45, RG2004. doi:[10.1029/2005RG000183](https://doi.org/10.1029/2005RG000183).
- Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T.M.L., Otto-Bliesner, B.L., Chan, K.C., Brook, B.W., 2017. PaleoView: A tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography* 40, 1348–1358. doi:[10.1111/ecog.03031](https://doi.org/10.1111/ecog.03031).
- Franklin, J., Miller, J.A., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.
- Franklin, J., Potts, A.J., Fisher, E.C., Cowling, R.M., Marean, C.W., 2015. Paleodistribution modeling in archaeology and paleoanthropology. *Quat. Sci. Rev.* 110, 1–14. doi:[10.1016/j.quascirev.2014.12.015](https://doi.org/10.1016/j.quascirev.2014.12.015).
- Fuller, D.Q., Lucas, L., Carretero, L.G., Stevens, C., 2018. From intermediate economies to agriculture: Trends in wild food use, domestication and cultivation among early villages in Southwest Asia. *Paléorient* 44, 59–74. arXiv:26595375.

- Fuller, D.Q., Wilcox, G., Allaby, R.G., 2011. Cultivation and domestication had multiple origins: Arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol.* 43, 628–652. doi:[10.1080/00438243.2011.624747](https://doi.org/10.1080/00438243.2011.624747).
- GBIF, 2025. What is GBIF? <https://www.gbif.org/what-is-gbif>.
- GBIF Secretariat, 2023. GBIF Backbone Taxonomy. doi:[10.15468/390MEI](https://doi.org/10.15468/390MEI).
- González Carretero, L., Lucas, L., Stevens, C., Fuller, D.Q., 2023. Investigating early agriculture, plant use and culinary practices at Neolithic Jarmo (Iraqi Kurdistan). *Journal of Archaeological Science: Reports* 52, 104264. doi:[10.1016/j.jasrep.2023.104264](https://doi.org/10.1016/j.jasrep.2023.104264).
- Guran, S.H., Yousefi, M., Kafash, A., Ghasidian, E., 2024. Reconstructing contact and a potential interbreeding geographical zone between Neanderthals and anatomically modern humans. *Scientific Reports* 14, 20475. doi:[10.1038/s41598-024-70206-y](https://doi.org/10.1038/s41598-024-70206-y).
- Haas, M., Schreiber, M., Mascher, M., 2019. Domestication and crop evolution of wheat and barley: Genes, genomics, and future directions. *Journal of Integrative Plant Biology* 61, 204–225. doi:[10.1111/jipb.12737](https://doi.org/10.1111/jipb.12737).
- Harris, D.R., 2007. Agriculture, Cultivation and Domestication: Exploring the Conceptual Framework of Early Food Production, in: Rethinking Agriculture. Routledge.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km—global soil information based on automated mapping. *PLoS One* 9, e105992. doi:[10.1371/journal.pone.0105992](https://doi.org/10.1371/journal.pone.0105992).
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12, e0169748. doi:[10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748).
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0906-7590.2006.04700.x>, doi:[10.1111/j.0906-7590.2006.04700.x](https://doi.org/10.1111/j.0906-7590.2006.04700.x).
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi:[10.1002/joc.1276](https://doi.org/10.1002/joc.1276).

- Jones, M.D., Abu-Jaber, N., AlShdaifat, A., Baird, D., Cook, B.I., Cuthbert, M.O., Dean, J.R., Djamali, M., Eastwood, W., Fleitmann, D., Haywood, A., Kwiecien, O., Larsen, J., Maher, L.A., Metcalfe, S.E., Parker, A., Petrie, C.A., Primmer, N., Richter, T., Roberts, N., Roe, J., Tindall, J.C., Ünal-İmer, E., Weeks, L., 2019. 20,000 years of societal vulnerability and adaptation to climate change in southwest Asia. *WIREs Water* 6, e1330. doi:[10.1002/wat2.1330](https://doi.org/10.1002/wat2.1330).
- Kabukcu, C., Asouti, E., Pöllath, N., Peters, J., Karul, N., 2021. Pathways to plant domestication in Southeast Anatolia based on new data from aceramic Neolithic Gusir Höyük. *Scientific Reports* 11, 2112. doi:[10.1038/s41598-021-81757-9](https://doi.org/10.1038/s41598-021-81757-9).
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. *Sci Data* 4, 170122. doi:[10.1038/sdata.2017.122](https://doi.org/10.1038/sdata.2017.122).
- Karger, D.N., Nobis, M.P., Normand, S., Graham, C.H., Zimmermann, N.E., 2023. CHELSA-TraCE21k – high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Climate of the Past* 19, 439–456. doi:[10.5194/cp-19-439-2023](https://doi.org/10.5194/cp-19-439-2023).
- Kopecký, M., Čížková, v.S.e., 2010. Using topographic wetness index in vegetation ecology: Does the algorithm matter? *Appl. Veg. Sci.* 13, 450–459. doi:[10.1111/j.1654-109X.2010.01083.x](https://doi.org/10.1111/j.1654-109X.2010.01083.x).
- Kozłowski, S.K., Aurenche, O., 2005. Territories, Boundaries and Cultures in the Neolithic Near East. Archaeopress.
- Krzyzanska, M., 2023. Modelling spatio-temporal changes in the ecological niches of major domesticated crops in China: Application of Species Distribution Modelling. Ph.D. thesis. University of Cambridge.
- Krzyzanska, M., Hunt, H.V., Crema, E.R., Jones, M.K., 2022. Modelling the potential ecological niche of domesticated buckwheat in China: Archaeological evidence, environmental constraints and climate change. *Vegetation History and Archaeobotany* 31, 331–345. doi:[10.1007/s00334-021-00856-9](https://doi.org/10.1007/s00334-021-00856-9).
- Kuhn, M., Silge, J., 2022. Tidy Modeling with R: A Framework for Modeling in the Tidyverse. O'Reilly.
- Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S., 2015. Very high-resolution digital elevation models: Are multi-scale derived variables ecologically relevant? *Methods in Ecology and Evolution* 6, 1373–1383. doi:[10.1111/2041-210X.12427](https://doi.org/10.1111/2041-210X.12427).
- Lev-Yadun, S., Gopher, A., Abbo, S., 2000. The Cradle of Agriculture. *Science* 288, 1602–1603. doi:[10.1126/science.288.5471.1602](https://doi.org/10.1126/science.288.5471.1602).

- Levy, A.A., Feldman, M., 2022. Evolution and origin of bread wheat. *The Plant Cell* 34, 2549–2567. doi:[10.1093/plcell/koac130](https://doi.org/10.1093/plcell/koac130).
- Lindsay, J.B., 2016. Whitebox GAT: A case study in geomorphometric analysis. *Comput. Geosci.* 95, 75–84. doi:[10.1016/j.cageo.2016.07.003](https://doi.org/10.1016/j.cageo.2016.07.003).
- Luan, J., Zhang, C., Xu, B., Xue, Y., Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research* 227, 105534. doi:[10.1016/j.fishres.2020.105534](https://doi.org/10.1016/j.fishres.2020.105534).
- Lucas, L., Fuller, D., 2018. Dataset: From intermediate economies to agriculture: Trends in wild food use, domestication and cultivation among early villages in southwest Asia.
- Maher, L.A., Richter, T., Stock, J.T., 2012. The Pre-Natufian Epipaleolithic: Long-term Behavioral Trends in the Levant. *Evolutionary Anthropology: Issues, News, and Reviews* 21, 69–81. doi:[10.1002/evan.21307](https://doi.org/10.1002/evan.21307).
- Miller, T., Blackwood, C.B., Case, A.L., 2024. Assessing the utility of Soil-Grids250 for biogeographic inference of plant populations. *Ecology and Evolution* 14, e10986. doi:[10.1002/ece3.10986](https://doi.org/10.1002/ece3.10986).
- Mod, H.K., Scherrer, D., Luoto, M., Guisan, A., 2016. What we use is not what we know: Environmental predictors in plant distribution models. *Journal of Vegetation Science* 27, 1308–1322. doi:[10.1111/jvs.12444](https://doi.org/10.1111/jvs.12444).
- Riehl, S., Kümmel, C., 2005. Archaeobotanical Database Of Eastern Mediterranean And Near Eastern Sites (ADEMNES).
- Schreiber, M., Özkan, H., Komatsuda, T., Mascher, M., 2021. Evolution and Domestication of Rye, in: Rabanus-Wallace, M.T., Stein, N. (Eds.), *The Rye Genome*. Springer International Publishing, Cham, pp. 85–100. doi:[10.1007/978-3-030-83383-1_6](https://doi.org/10.1007/978-3-030-83383-1_6).
- Shennan, S.J., Conolly, J., 2007. Dataset: The Origin and Spread of Neolithic Plant Economies in the Near East and Europe.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148, 1–13. URL: <https://www.sciencedirect.com/science/article/pii/S030438000100388X>. doi:[10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X).
- Townsend Peterson, A., Soberón, J., 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *NatCon* 10, 102–107. doi:[10.4322/natcon.2012.019](https://doi.org/10.4322/natcon.2012.019).
- Ulaş, B., Abbo, S., Gopher, A., 2024. Drawing diffusion patterns of Neolithic agriculture in Anatolia. *Review of Palaeobotany and Palynology* 322, 105057. doi:[10.1016/j.revpalbo.2024.105057](https://doi.org/10.1016/j.revpalbo.2024.105057).

- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2021. Modelling species presence-only data with random forests. *Ecography* 44, 1731–1742. doi:[10.1111/ecog.05615](https://doi.org/10.1111/ecog.05615).
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecological Monographs* 92, e01486. doi:[10.1002/ecm.1486](https://doi.org/10.1002/ecm.1486).
- Velazco, S.J.E., Galvão, F., Villalobos, F., De Marco Júnior, P., 2017. Using worldwide edaphic data to model plant species niches: An assessment at a continental extent. *PLoS One* 12, e0186025. doi:[10.1371/journal.pone.0186025](https://doi.org/10.1371/journal.pone.0186025), arXiv:[29049298](https://arxiv.org/abs/29049298).
- Verhagen, P., Whitley, T.G., 2020. Predictive Spatial Modelling, in: Archaeological Spatial Analysis. Routledge.
- Wallace, M., Jones, G., Charles, M., Forster, E., Stillman, E., Bonhomme, V., Livarda, A., Osborne, C.P., Rees, M., Frenck, G., Preece, C., 2018a. Re-analysis of archaeobotanical remains from pre- and early agricultural sites provides no evidence for a narrowing of the wild plant food spectrum during the origins of agriculture in southwest Asia. *Veg. Hist. Archaeobot.* doi:[10.1007/s00334-018-0702-y](https://doi.org/10.1007/s00334-018-0702-y).
- Wallace, M., Livarda, A., Charles, M., Jones, G., 2018b. Origins of agriculture: Archaeobotanical database. doi:[10.5284/1046750](https://doi.org/10.5284/1046750).
- Whitlam, J., Bogaard, A., Matthews, R., Matthews, W., Mohammadifar, Y., Ilkhani, H., Charles, M., 2018. Pre-agricultural plant management in the uplands of the central Zagros: The archaeobotanical evidence from Sheikh-e Abad. *Vegetation History and Archaeobotany* 27, 817–831. doi:[10.1007/s00334-018-0675-x](https://doi.org/10.1007/s00334-018-0675-x).
- Willcox, G., 2024. Sowing, harvesting and tilling at the end of the Pleistocene/beginning of the Holocene in northern Syria: A reassessment of cereal and pulse exploitation. *Vegetation History and Archaeobotany* doi:[10.1007/s00334-023-00984-4](https://doi.org/10.1007/s00334-023-00984-4).
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Group, N.P.S.D.W., 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14, 763–773. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1472-4642.2008.00482.x>, doi:[10.1111/j.1472-4642.2008.00482.x](https://doi.org/10.1111/j.1472-4642.2008.00482.x).
- Wright, M.N., Ziegler, A., 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 77, 1–17. doi:[10.18637/jss.v077.i01](https://doi.org/10.18637/jss.v077.i01).

- Yaworsky, P.M., Hussain, S.T., Riede, F., 2023. Climate-driven habitat shifts of high-ranked prey species structure Late Upper Paleolithic hunting. *Scientific Reports* 13, 4238. doi:[10.1038/s41598-023-31085-x](https://doi.org/10.1038/s41598-023-31085-x).
- Yaworsky, P.M., Hussain, S.T., Riede, F., 2024a. The effects of climate and population on human land use patterns in Europe from 22ka to 9ka ago. *Quaternary Science Reviews* 344, 108956. doi:[10.1016/j.quascirev.2024.108956](https://doi.org/10.1016/j.quascirev.2024.108956).
- Yaworsky, P.M., Nielsen, E.S., Nielsen, T.K., 2024b. The Neanderthal niche space of Western Eurasia 145 ka to 30 ka ago. *Scientific Reports* 14, 7788. doi:[10.1038/s41598-024-57490-4](https://doi.org/10.1038/s41598-024-57490-4).
- Yousefi, M., Heydari-Guran, S., Kafash, A., Ghasidian, E., 2020. Species distribution models advance our knowledge of the Neanderthals' paleoecology on the Iranian Plateau. *Scientific Reports* 10, 14248. doi:[10.1038/s41598-020-71166-9](https://doi.org/10.1038/s41598-020-71166-9).
- Zeder, M.A., 2011. The Origins of Agriculture in the Near East. *Curr. Anthropol.* 52, S221–S235. doi:[10.1086/659307](https://doi.org/10.1086/659307), arXiv:[10.1086/659307](https://arxiv.org/abs/10.1086/659307).
- Zeder, M.A., 2024. Out of the Shadows: Reestablishing the Eastern Fertile Crescent as a Center of Agricultural Origins: Part 1. *Journal of Archaeological Research* doi:[10.1007/s10814-024-09195-5](https://doi.org/10.1007/s10814-024-09195-5).
- Zohary, D., Weiss, E., Hopf, M., 2012. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. 4th ed ed., Oxford University Press, Oxford.