Understanding the spread of agriculture in the Western Mediterranean (6th-3rd millennia BC) with Machine Learning tools

Maria Elena Castiello [1,2] ✉, Emmanuele Russo[3], Héctor Martínez-Grau[4], Ana Jesus[4], Georgina Prats [4,5] & Ferran Antolín [2,4]

The first Neolithic farmers arrived in the Western Mediterranean area from the East. They established settlements in coastal areas and over time migrated to new environments, adapting to changing ecological and climatic conditions. While farming practices and settlements in the Western Mediterranean differ greatly from those known in the Eastern Mediterranean and central Europe, the extent to which these differences are connected to the local environment and climate is unclear. Here, we tackle this question by compiling data and proxies at a superregional and multi-scale level, including archaeobotanical information, radiocarbon dates and paleoclimatic models, then applying a machine learning approach to investigate the impact of ecological and climatic constraints on the first Neolithic humans and crops. This approach facilitates calculating the pace of spread of farming in the Western Mediterranean area, modelling and estimating the potential areas suitable for settlement location, and discriminating distinct types of crop cultivation under changing climatic conditions that characterized the period 5900 – 2300 cal. BC. The results of this study shed light onto the past climate variability and its influence on human distribution in the Western Mediterranean area, but also discriminate sensitive parameters for successful agricultural practices.

The spread of early farmers from SW Asia towards Western Europe is a testament to the capacity of Neolithic farmers to succeed in many different environments, substrates and climatic regions. This success is possible partly thanks to the large variety of crops available at the beginning of this process, the intensive (garden-plot-type) and mixed (closely integrated with animal management) nature of farming[1], but also the important role of wild plants in the diet[2–4]. There is abundant literature reviewing how the spread of these populations was only possible by focusing on the crops better adapted to the new climatic conditions[5–8]. However, so far no research has quantitatively assessed the ecological niche of these crops in the past, and the degree to which these farming communities and the crops they cultivated were

constrained by or adapted to climate change events that potentially happened during the Neolithic period (between 5900 and 2300 cal. BC)[9].

The North-west Mediterranean region and the western Alpine Foreland (current Switzerland) is one of the best-investigated areas regarding settlement patterns and agricultural practices in the Western Mediterranean and also one of the regions that have seen a greater improvement of the datasets in recent times. In comparison to central Europe, the area is less intensively known[10], particularly for the Early Neolithic (5900–4500 BC, our Phase 01 and Phase 02), but sites are well radiocarbon dated, partly due to the establishment of radiocarbon dates as proxies to understand the Neolithisation process[11,12] and partly

[1]Institut d'archéologie et des sciences de l'antiquité, University of Lausanne, Lausanne, Switzerland. [2]Division of Natural Sciences, German Archaeological Institute, Berlin, Germany. [3]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [4]Integrative Prehistory and Archaeological Science, Department of Environmental Sciences, University of Basel, Basel, Switzerland. [5]ARQHISTEC, Grup d'Investigació Prehistòrica (GIP-UdL), Departament de Geografia, Història i Història de l'Art, Universitat de Lleida, Lleida, Spain. ✉e-mail: MariaElena.Castiello@unil.ch

because of the long stratigraphies preserved in cave sites requiring accurate dating and modelling[13], but particularly thanks to dendrochronological dating of waterlogged sites. The Middle (4500–3500/3300 BC, our Phase 03) and Late Neolithic (3500/3300–2300 BC, our Phase 04) periods are very well researched, particularly if we consider the large amount of high-quality data coming from the pile-dwellings in current Swiss territory and the abundant research carried out in some of the other areas, partly connected to rescue archaeology interventions as human impact on the landscape of the region increases[14–16].

It is commonly agreed that the first farmers that got to the region spread along the coasts since new populations arrived through navigation, and progressively moved inland. Recent work has questioned this assumption[13]. Manen and others highlighted that early Neolithic settlements seem to have optimized adapting capacities to dwell in different environments and different types of sites: open-air sites, cave sites, pile dwellings, etc.[10]. Whether this observation translates in a particularly diverse niche or only in diverse topographical locations with similar conditions is unclear. Between the 5th and 3rd millennia strong networks develop. In this context, the settling of farmers in the Swiss Plateau (and around lakes in the Jura region) takes place, probably connected to a spread of populations from the South[12,17]. There is abundant evidence of changing technologies due to internal dynamics and external influences, as well as of the exchange of prestige goods and even small-scale migrations, particularly at the end of the Neolithic[18]. According to available archaeological evidence, migrating individuals did not necessarily move in large population waves, and they integrated into existing villages, which is well documented, for instance, in the Swiss pile dwellings[19]. Considering this, for the moment, changes in niche amplitude should not be understood as evidence of the arrival of new groups that prefer other locations. It is our hypothesis, that changes in the niche breadth imply economic changes that may or may not be connected to climatic changes. Authors observe an important use of middle/low mountain ranges in the 6th and 5th millennia BC[20,21], but the interpretations of this phenomenon differ, either as evidence of an economic specialization or of permanent occupations at middle altitudes. Higher mountain ranges would be seasonally targeted in the 4th and 3rd millennia BC[20,22].

Actually, a long-term analysis of the changing niches in the Neolithic has not been evaluated in combination with agricultural practices. This is an important research gap considering that Early Neolithic societies in the area are poorly stratified and site location will probably be mostly driven by environmental factors as well as social networks. The use of niche modelling techniques is generally not new, and our efforts here align with the most recent studies in archaeology[23–27] trying to assess the responses of human societies to climate variability, making use of the latest high-resolution climate model results. 'Habitat suitability' (HS) is (as detailed in Methods, Habitat Suitability and Niche models construction, following the work of Braunisch and others[28]) an inference based on extrapolation from archaeological site distributions, and not on a priori arguments about the suitability of different landscapes. It has the implicit assumption that archaeological sites must have been located in areas that are suitable for human habitation at a given time. Blinkhorn and others[29] have indeed tried to infer human behaviour from the archaeological record of the Late Pleistocene, while Banks and others[30] adopted more specific environmental and cultural niche modelling techniques to conclude that environmental factors did have an influence on the predisposing occupation of regions most suited to specific cultural adaptations for the prehistoric farmers. Such first quantitative attempts, and the process of exploring and explaining where and why Neolithic populations occurred and settled, and to which extent people's lives were already affected by climatic factors and constraints has rightly become a central focus of debate for an increasing number of archaeological studies, concurrently with the more pressing concern and challenge of the global climate crisis[31,32]. Computational and quantitative modelling techniques come in hand and can be of greatest benefit for archaeologists trying to address large-scale events connected with relevant modern challenges. When considering the greatest amount of data and information available today, especially when approaching the complex phenomena of the spread of agriculture that includes dozens of countries, we realize that computer tools and more advanced statistical methods are essential to combine and integrate multiple proxies and databases at different spatiotemporal scales, to test and validate several hypotheses indeed formulated to reconstruct the past.

Previous research has focused on the spread or on the understanding of expansions and declines of certain archaeological cultures. We nevertheless do not advocate for deterministic positions aiming to explain the disappearance of sites with certain pottery decorations. We are more interested in the relationship between site location and crop diversity at a given time and place since crops have a more direct relationship with weather and climate.

Among the crops available to early farmers arriving into current mainland Greece from SW Asia, we can consider naked wheat (*Triticum aestivum/durum*), emmer (*Triticum dicoccon*), einkorn (*Triticum monococcum*), Timopheev's wheat (*Triticum timopheevii*), hulled barley (*Hordeum distichon/vulgare*), naked barley (*Hordeum vulgare* var. *nudum*), lentil (*Lens culinaris*), pea (*Pisum sativum*), broad bean (*Vicia faba*), chickpea (*Cicer arietinum*), bitter vetch (*Vicia ervilia*) and flax (*Linum usitatissimum*). The characteristics of the different cultivated plants and their uses differ, as highlighted by numerous authors[33–39]. Einkorn grows well on poor soils and in cold areas and tolerates wet climates. It is mostly used to produce bulgur and similar products instead of bread-like foodstuffs. Emmer is slightly more drought-resistant than einkorn and a bit more demanding on soil quality. It is also more productive. Timopheev's wheat is the most wet-tolerant cereal of all and is also resistant to many cereal plant pathogens[40]. Barleys have a shorter growing period, and this makes them more adaptable to arid conditions, while they also grow well on poor soils. They can be turned into flour but with a low starch content, hence mostly used in porridge soups as roasted grains or added to other flours to produce bread.

Chickpea[41] is one of the crops that is first abandoned as farmers started spreading towards the European continent, although there are a few occurrences in the Iberian Peninsula[42]. Naked wheat and broad bean only seem to spread along the Mediterranean coast, but not towards central Europe until later chronologies. One final crop, opium poppy (*Papaver somniferum*) seems to have been taken into cultivation in the Western Mediterranean and then spread towards central Europe together with naked wheat[43–45]. The reason for the early abandonment of naked wheat and chickpea by farmers entering the Carpathian Basin may be due to the continental climatic conditions, which made it more difficult for these crops to thrive. Conversely, in the Western Mediterranean, naked wheat was a very important crop, and it eventually overtook the role of hulled wheat in the economy[7,33]. After the first spread of farming, further changes in the crop spectrum occurred in different regions. For instance, in the Western Mediterranean, according to previous research[46–48], a virtual replacement of naked wheat by einkorn, emmer and Timopheev's wheat occurred at ca. 4000 BC. As crops expanded from the Western Mediterranean coast towards the Alpine area (both southwest and northwest)[17] new climates had to be faced. Although this condition should have posed similar challenges to early farmers than the expansion towards the northern Balkans and the Hungarian plain, it seemed to not have affected their crop choice (with the great importance of naked wheat, naked barley and opium poppy, similar to other Mediterranean sites)[35]. Previous approaches to the reconstruction of ecological niches of Neolithic cultures[30] interpreted the potential differences in the ecological niches of early farming groups in central and southern Europe as deriving from the crops they were growing. While this may be
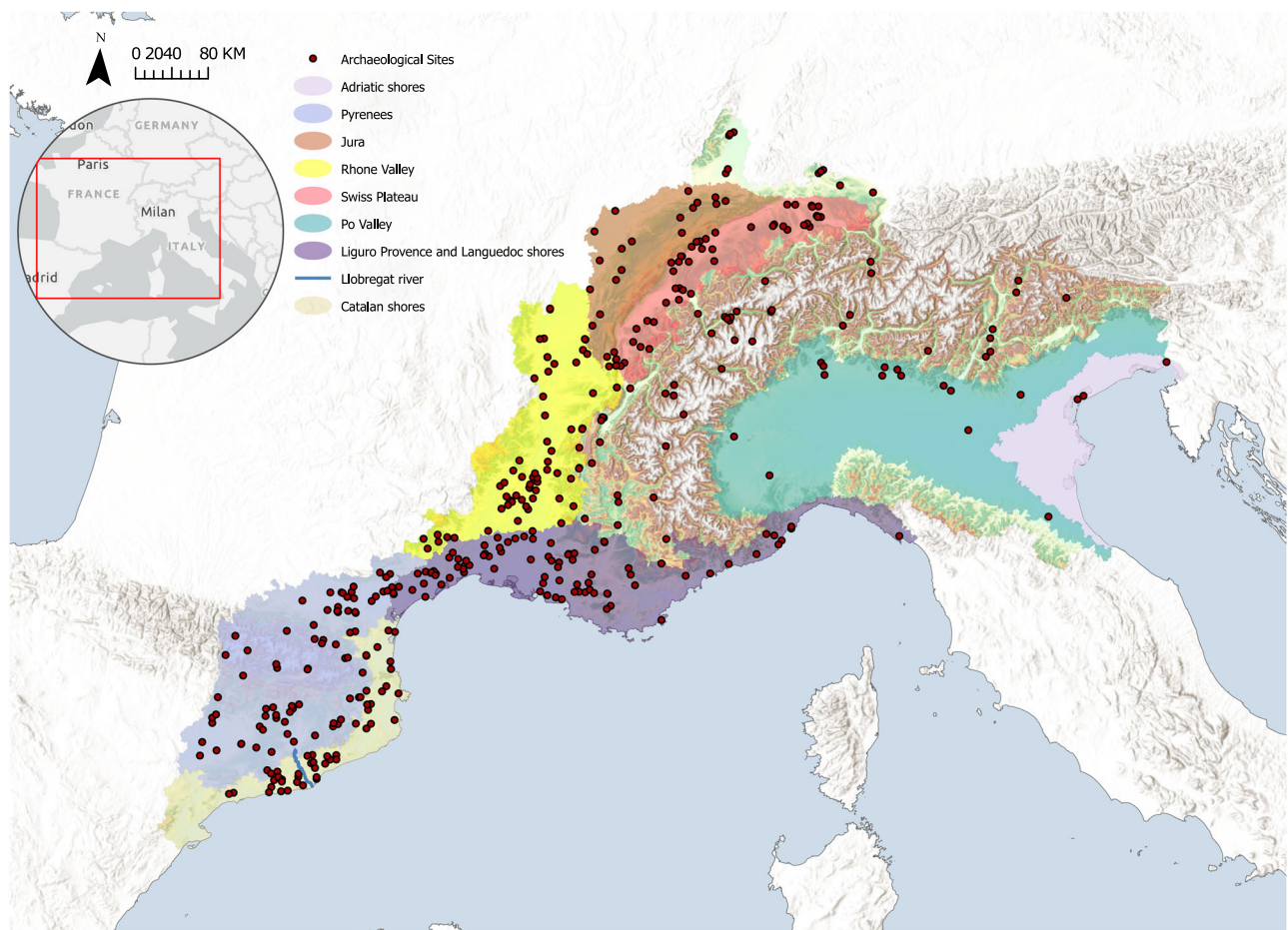
Fig. 1 | Study Area. The study area of the Western Mediterranean with the archaeological sites (red dots) and the ecologically diverse regions highlighted in different colours.

partially true, it is even more interesting to observe how farming evolved in well investigated regions and whether the niche of those crops actually expanded or reduced with time.

In this study, we propose an innovative multi-proxy and inter-disciplinary approach based on Machine learning (ML) algorithms, namely Random Forest (RF) and Maximum Entropy (MaxEnt) to characterize and quantify the mechanisms and the impact of climatic and environmental factors on early farmer settlement preferences and crop choices over the Neolithic period in the North-west Mediterranean region and Switzerland. We coupled a database consisting of a total of 3416 geo-referenced archaeological sites, with associated radiocarbon dates (calibrated using OxCal v. 4.4.2 and the atmospheric curve IntCal20) from *AgriChange_14Cdatabase*[49] and crop occurrences in archaeobotanical analyses obtained within the AgriChange project[50]. The study area (Fig. 1) of this research amounts in total to almost 310,000 km² of surface and offers a wide topographic and bioclimatic diversity, furthermore with ecologically diverse regions. By coupling an increased amount of published heterogeneous data and proxies at a superregional and multi-scale level, we demonstrate that climatic fluctuations, such as dropping or rising temperatures, increasing precipitation or decreasing seasonality values, were tightly intertwined with the history and distribution of the early farmers and especially their coping agricultural strategies.

Here, we show the extent to which the changes observed in crop distribution were climatically driven in this particular study area, characterizing expanding crop niches and climatic conditions across past chronologies, by coupling paleoclimatic data from the most recent high-resolution climate model results (CHELSA TraCE21k

dataset from Karger et al.[51]; see "Methods" section), with crop occurrences. The main premise of the paper is that choices regarding site location and crop cultivation made in periods of climatic change most likely reflect the adaptive strategies of prehistoric populations.

The aims of the paper are to (1) Determine whether human and crop niches changed over time in the Neolithic; (2) Establish if new niches followed climate changes but remained stable in character or if they changed in climatic/landscape factors; (3) Observe if new niches were tied to the adoption of new crops or if new crops were adopted within stable niches; (4) Generate maps of potentially suitable areas of early farmers' settlements; (5) Assess the capacity of these techniques to address key archaeological and archaeobotanical questions.

Results

Paleoclimate and environmental envelope

To evaluate Neolithic farmers and crop niches, we first provided a rapid characterization of past climate variability. We examined possible patterns in downscaled annual precipitation and temperature values with the CHELSA-TraCE21k dataset[51] (for a complete overview of the temporal evolution of the CHELSA paleoclimatic variables see Supplementary Figs. 1 and 2) which has yet to be used more widely in studies modelling archaeological or archaeobotanical occurrences for the time frame selected here, and we could thus observe important variability in the Annual Mean Temperature (Bio01) and the Temperature Seasonality (Bio04) values, especially in correspondence and between the four Phases (see the vertical dashed grey lines in Fig. 2) in which the available archaeological and archaeobotanical information was organized following previous work[49] (see Supplementary Table 1
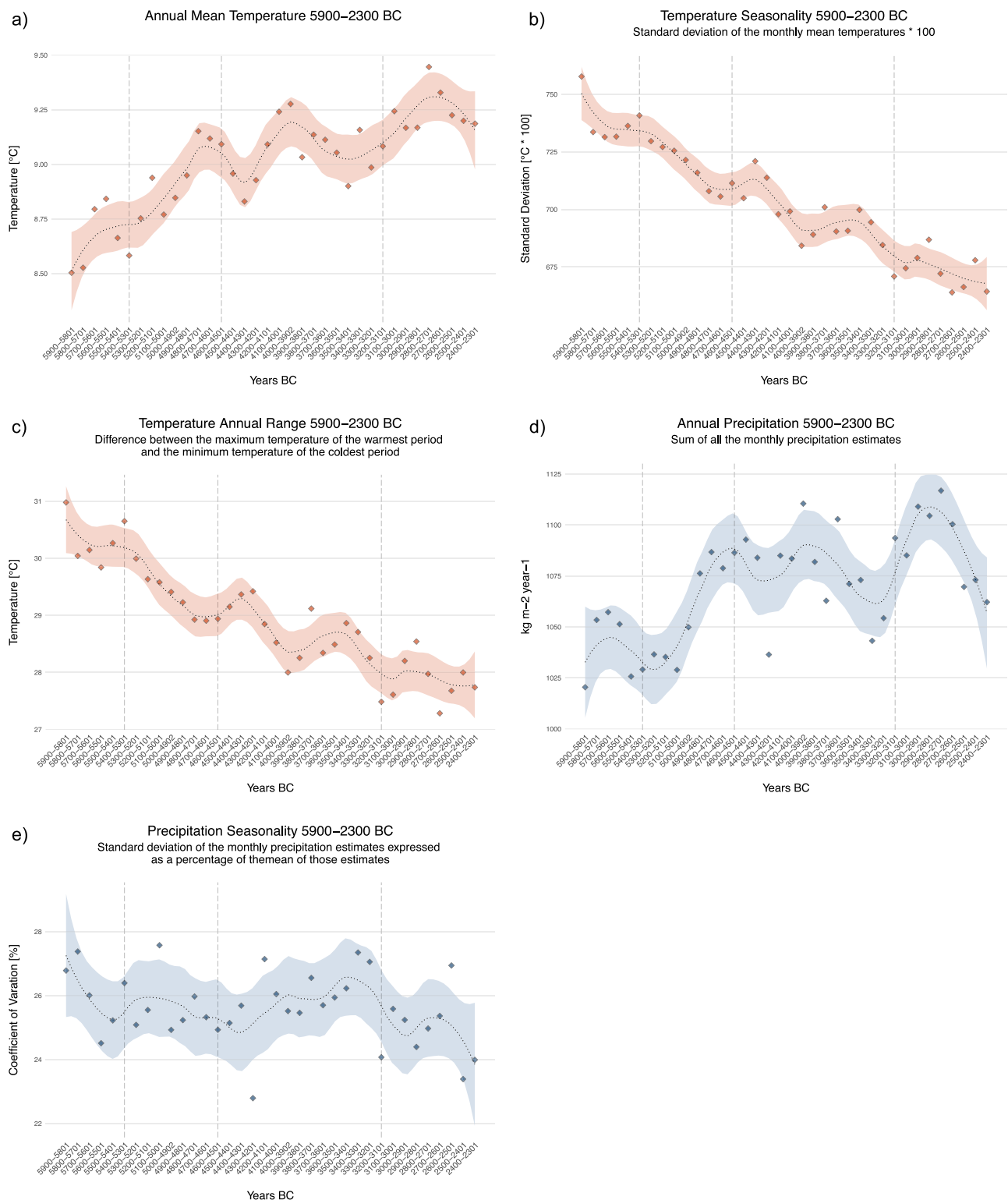
Fig. 2 | Modelled climatic oscillation in the study area over the period 5900-2300 cal. BC (derived from CHELSA-TraCE21k dataset). Evolution of Annual Mean temperature (**a**), Temperature Seasonality (**b**), Temperature Annual Range (**c**), Annual Precipitation (**d**) and Precipitation Seasonality (**e**) over the entire study area and period. The points represent the calculated mean value over the entire study area for each 100-year step. The dotted line represents the loess smooth with a span of 0.3 and the filled area the 95% confidence level interval (*geom_smooth* function from *ggplot2* package in *R*). The vertical dashed lines indicate the limits of our 4 chrono-phases. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

and Methods, Archaeological and Archaeobotanical data). These four Phases are based on crop dynamics and divide our study period according to the changing dominance of crop assemblages (for more details about the division into four Phases, see Supplementary Note 1).

In general terms, for the study area and the full period, a general increase in Annual Mean Temperature (Fig. 2a) is paired with a decrease in Temperature Seasonality (Fig. 2b) and in the Temperature Annual Range (Fig. 2c), reflecting an evolution towards a warmer and

less extreme climate for the Western Mediterranean area. Likewise, the Annual Precipitation increases over time while the Precipitation Seasonality remains generally constant, showing a less pronounced trend towards lower seasonality (Fig. 2d, e). Summers became cooler and wetter while winters became warmer with variable amounts of precipitation.

Considering these results, we would hypothesize that we should first find crops that benefit from stronger seasonality and lower temperatures, and there could be a progressive shift to crops that perhaps do not need such strong seasonality and are better prepared to withstand higher temperatures. We could also expect a spread of settlement locations to previously less favourable areas, such as higher altitudes, given a general trend to reduction of seasonality. Considering the Annual Precipitation, we observe a very dry phase around 5300-5200 BC, which coincides with the change from Phase 01 to Phase 02, and a very wet period in Phase 04. We could, in this sense hypothesize the appearance of highly drought-resistant crops in Phase 02 and wet-tolerant crops in Phase 04. We should also expect the appearance of wells in the driest phases, as a phenomenon known in other regions[52], but the evidence of wells is still sparse in our study region, perhaps due to taphonomic issues.

Given these general conditions, we then examined, more specifically, the ecological envelope for the settlements classified *per* type and phase (see Fig. 3 below; Supplementary Table 2and Supplementary Fig. 3a–d). Over the entire period, we found that the majority of sites were located in warm environments with an Annual Mean Temperature between ca. 10 and 15 °C (Fig. 3a) and Annual Precipitation values no higher than 1250 mm per year. They are especially clustered around 750 mm (Fig. 3b). A large portion of "open air" sites persistently occupied the areas in proximity to the main lakes and main rivers, especially during Phase 03 (Fig. 3c, d). Lower altitudes (below 700 m.a.s.l.) and gentler slopes were preferred by most of the sites (Fig. 3e, f; the complete list of these outputs can be found in Supplementary Fig. 3a–d).

The Neolithic farmers' niche

We developed two HS models based on two different ML algorithms: RF and MaxEnt, using paleoclimatic and environmental variables as predictor inputs to produce maps of suitable areas for Neolithic farmers' occupation, one for each of the Phases identified (Figs. 4 and 5). These maps with high (purple) and low (green) values show that suitable areas for Neolithic settlements changed considerably over time. In particular, the maps produced by the RF algorithm (Fig. 4a–d) point to a significant change between 5900-4500 BC (Phase 01 and 02) and 4499–2300 BC (Phase 03 and 04). Looking closely at Fig. 4a, during the first phase (5900–5300 BC), the most suitable areas are essentially distributed along the Liguro-Provence and Languedoc shores (Northern Mediterranean shores), the Llobregat river mouth (Southern Mediterranean shores) and the Pyrenees. Consistent high suitability is also observed in the area of the Po valley and along the Adriatic shores until the end of the second phase (4500 BC) (Fig. 4b). Starting with Phase 03 (4499–3100 BC) in Fig. 4c, the most suitable areas seem to shift towards the inland, the lower-course of the Rhone valley, the Swiss Plateau and the main Swiss lakes, as well as the Jura region, thus abandoning the Mediterranean shores (Fig. 4d). The Italian Peninsula as well as the Alpine regions present very few suitable areas during these two more recent periods.

The suitability maps produced using MaxEnt (Fig. 5) show some differences but also some similarities in many cases with the maps produced using RF. During Phase 01 (Fig. 5a), there is an absence of suitable areas in the Po valley, the Italian Peninsula, the Swiss Plateau and the higher course of the Rhone valley. With Phase 02 (Fig. 5b), the map shows again higher suitability along the Po valley, along the Pyrenees and around Lake Garda, but less suitable areas over Liguro-Provence and Languedoc shores. During Phase 03 (4499–3100 BC),

higher probabilities are found around the alpine region, the Pyrenees and the lower Rhone valley. The highest suitability scores are more widely distributed over the landscape during the last phase (Fig. 5d), and reach further inland, away from the coast, compared to earlier phases. A shift that is supported both by the MaxEnt and the RF models.

The significance of each variable used in the two models is presented as *Variable Importance Ranking* (Supplementary Fig. 4) and *Partial Dependence plots* (Supplementary Fig. 5a–d) for RF. The *Variable Importance Ranking* for RF (using Mean Decrease Accuracy) is a statistical measure computed by looking at how much removing a variable decreases the model accuracy[53]. The *Partial Dependence plots* are a graphical indication of the influence (or marginal effect) of the specific class/range of values on the computed probability of site location[54].

For MaxEnt, the *Response Plots* (Supplementary Fig. 6a–d) indicate the role of a particular variable depicted as a response curve showing the predicted relative occurrence rate (suitable areas) against the values of that predictor (variable). As stated by Hong and others[53], "a key advantage of variable importance measures, as compared to univariate screening methods, is that they cover the impact of each predictor variable individually as well as in multivariate interactions with other predictor variables".

As a result of the RF computations, the Mean Temperature Diurnal Range (Bio02) and the distance to the main lakes (D_Lakes) are the two most important variables in predicting high suitability areas during the first Phase 01, with a peak of suitability prediction at a Temperature Diurnal Range of 6 °C and far away from the main lakes above 100 km. During Phase 02, Bio02 is associated with the elevation variable (DEM), and together they represent the two most important factors in defining the best ecological settings. Similar to Phase 01, the peak of suitability prediction of Bio02 is maintained between 6 °C and 8 °C. Regarding elevation, the highest positive dependence is shown below 500 m.a.s.l. During Phase 03 and 04, the distance to the main lakes plays the most relevant role in predicting the highest suitability areas, with the highest suitability in close proximity to them, associated with the Mean Temperature of Wettest Quarter (Bio08) during Phase 03 with values over 10 °C and to elevation during Phase 04, with positive values between 200 and 700 m.a.s.l. (see Supplementary Figs. 4 and 5a–d).

While MaxEnt provides more generalized response curves, the variable *Response Plots* show very similar results for these variables. The highest prediction appears again at a Mean Temperature Diurnal Range (Bio02) of 6 °C and gently decreases with increasing Bio02 values (during Phase 01 and 02). Similarly, the peak in prediction lies below 500 m.a.s.l. in Phase 02 with high values (> 0.5) up to 700 m.a.s.l. in Phase 04. While in Phase 01, the highest predictions correspond to a distance to lakes between 100 and 150 km, in Phase 03 and 04 they lie in close proximity to them (see Supplementary Fig. 6a–d).

The shift further inland and towards higher altitudes observed above when comparing the suitability maps computed by the two models has been also statistically inspected as the distribution of the predicted most suitable areas (with suitability ≥ 0.75) over the climatic and environmental features. The RF model shows that between 5900 and 4500 BC, the most suitable areas are those with an Annual Mean Temperature (Bio01) around 13 °C and mainly located at very low altitudes (with median values at 193 for Phase 01 and 210 m.a.s.l. for Phase 02 and a peak in the density of occurrences below these values). These preferences changed with the onset of Phase 03, when the best areas spread over a larger range both considering altitude and mean annual temperature. Median altitudes increase to 323 m.a.s.l (Phase 03) and 392 m.a.s.l. (Phase 04) with a mean annual temperature showing much larger ranges than in the previous phases (see Fig. 6a, b). The MaxEnt model shows very similar predictions, with a peak of high suitability areas at very low altitudes during Phase 01 and
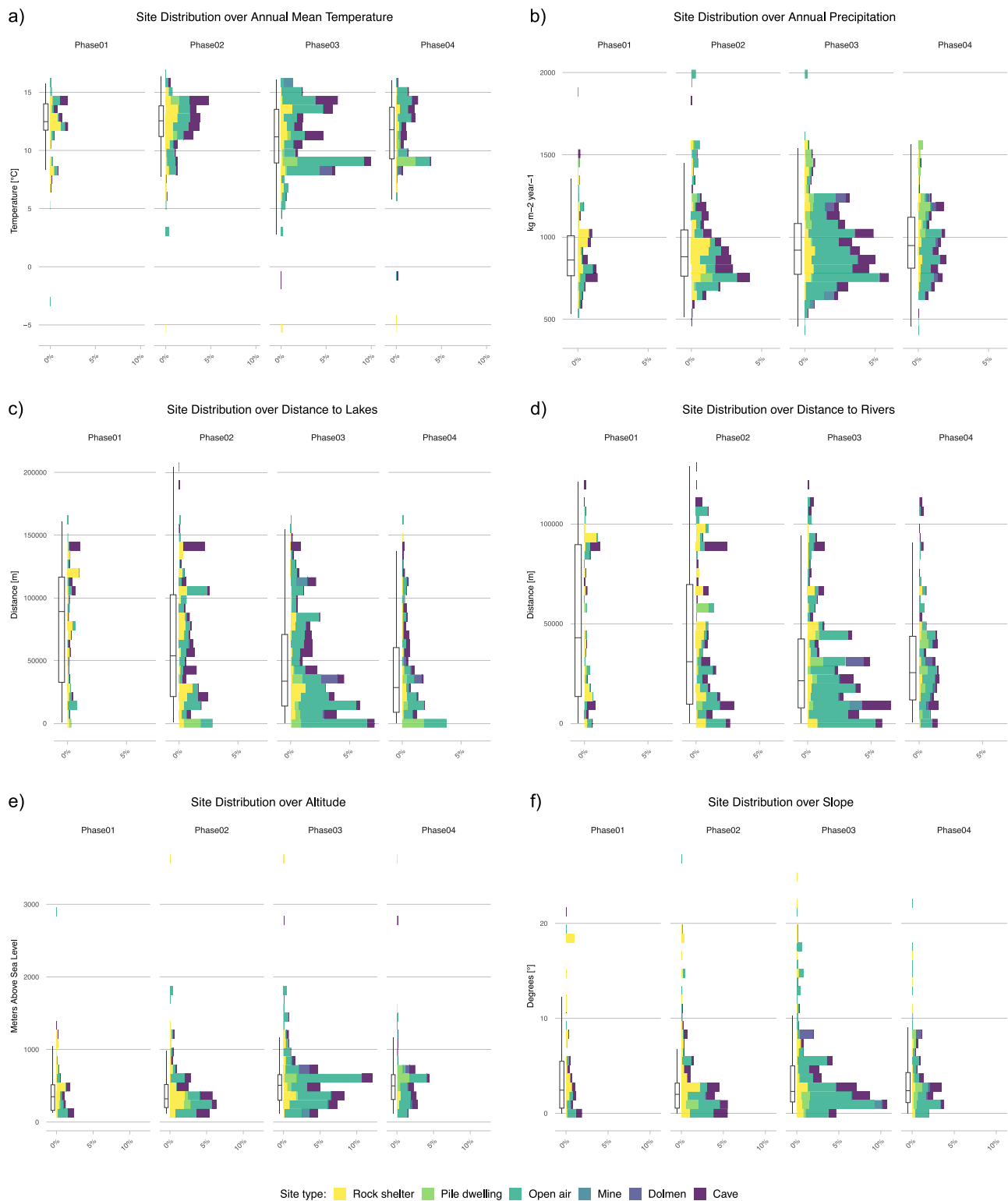
Fig. 3 | Distribution of Site types over paleoclimate and environmental variables per Phase. Each panel shows the overall distribution (box plot) of all sites and the distribution of site types (histograms) over Annual Mean Temperature (**a**), Annual Precipitation (**b**), Distance to lakes (**c**), Distance to rivers (**d**), Altitude (**e**) and Slope (**f**) per Phase. The histograms display 30 equally wide bins showing the percentage distribution of site types over the variables. The boxplots show medians, first and third quartiles (hinges), and minimum and maximum values no further than 1.5*IQR from the hinge where IQR is the inter-quartile range (whiskers). Details about sample numbers are provided in Supplementary Table 2. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

similar interquartile ranges (boxes) with slightly higher median values during Phase 01 and 02. A similar altitude shift between Phase 02 and Phase 03 is observed when the median altitude of the most suitable areas increases from 304 m.a.s.l. to 360 m.a.s.l and their distribution spreads over a larger range (see Fig. 6c, d). Indeed, it was around 4500 BC that we could also observe a shift in the type of preferred landscapes. While during the first two phases, the most suitable areas were mainly located at the lowest altitude, with warmer temperatures and
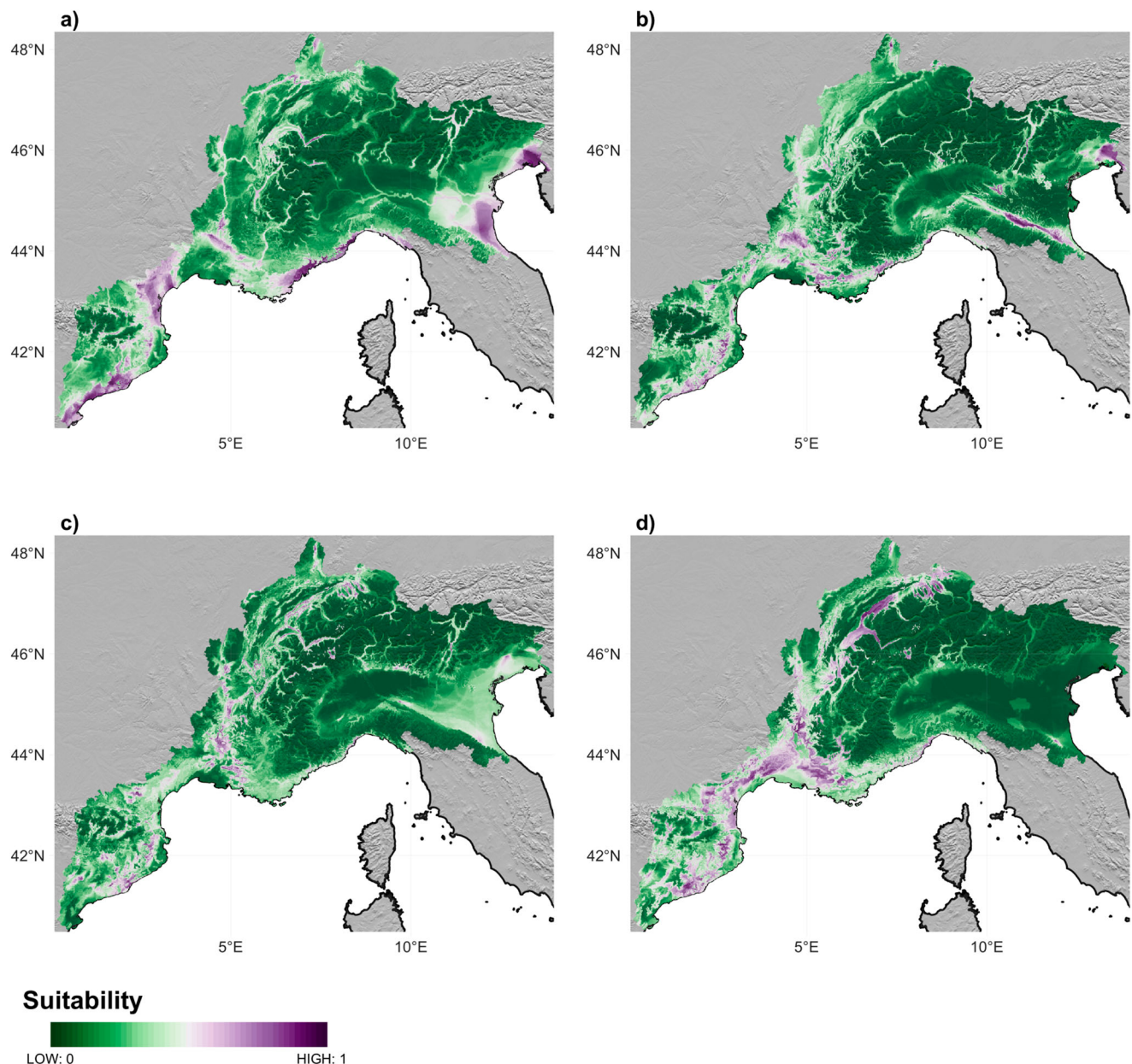
Fig. 4 | Suitability maps resulting from the Random Forest models for the four Phases. Phase 01 is shown in panel (**a**), Phase 02 in panel (**b**), Phase 03 in panel (**c**) and Phase 04 in panel (**d**). High suitability areas are coloured in purple, while low suitability areas are coloured in green. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

low precipitation values, in time they shifted towards higher elevations, in colder and more humid environments. (for a complete overview of the distribution of the predicted areas over all variables see Supplementary Fig. 7).

**Agricultural variability and Crop niches**

We also analysed the ecological niche of each of the main crops over the entire period. These might differ individually from the general trends shown by the sites since not all the crops were present at all sites during all phases (see Fig. 7a).

Figure 8 reproduces the distribution of crop occurrences over Annual Mean Temperature and Mean Annual Precipitation. It shows that for the entire period analysed, naked wheat clustered around higher temperatures, between 12 °C and 15 °C, and annual precipitation values of ca. 800 mm per year (see also Supplementary Fig. 8a, k), spreading in warmer and more arid environments. Glume

wheat (einkorn, emmer and Timopheev's wheat), similar to naked wheat, cluster with preference in areas with relatively high Annual Mean Temperatures (ca. 14 °C to 15 °C) and mean annual precipitation around 800–880 mm per year. Poppy and apple/pear instead are found in more humid and cooler environments, with precipitation values around 1000–1250 mm and temperature values of less than 10 °C. Pea and hazel, and to a lesser extent lentils and oak, are found both in warmer and drier as well as in humid and cooler conditions.

The results (Fig. 9) show that mild winters with lower precipitation values were suitable for barley, glume wheat, naked wheat, and oak (clusters around temperature values of the coldest quarter above 5 °C and precipitation of the coldest quarter around 200–250 mm, see also Supplementary Fig. 8j,r). In places with harsher winters (temperature values of the coldest quarter below 0 °C and precipitation of the coldest quarter above 250 mm), wild crops such as apple/pear and
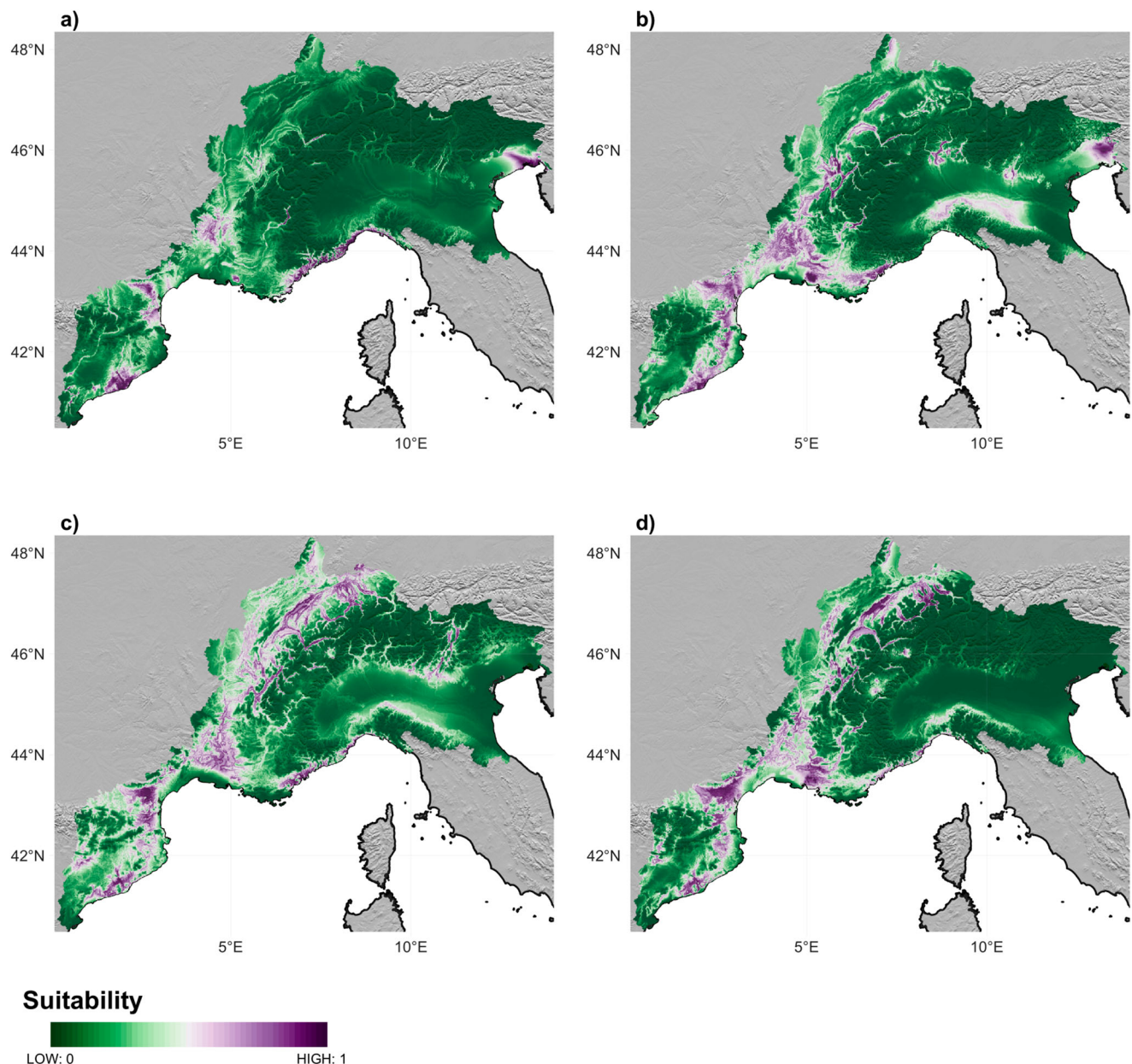
Fig. 5 | Suitability maps resulting from the MaxEnt models for the four Phases. Phase 01 is shown in panel (**a**), Phase 02 in panel (**b**), Phase 03 in panel (**c**) and Phase 04 in panel (**d**). High suitability areas are coloured in purple, while low suitability areas are coloured in green. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

hazel were complemented with oil seeds (flax and opium poppy). Lentils and peas are found in both types of environments.

Discussion

The main objective of this research was to identify changes in human and crop niches and their interrelationship with climatic changes across the Western Mediterranean area, based on a big database and a data-driven approach using machine learning techniques. While computational and quantitative modelling techniques offer an exciting new opportunity to create and empirically test more explicit models of past human ecological dynamics and agriculture development strategies, one must keep in mind that these approaches carry also some limitations. Accessing and manipulating large datasets across different sources requires, on the one side, significant computational power and processing time and on the other side, complementary and multi-disciplinary expert knowledge. Moreover, the inherent bias within the

archaeological and archaeobotanical datasets represents today a challenge for many modelling exercises, in particular here where the available site and crop samples with associated C14 dates are likely under-representative for some of the regions included in the study area (e.g., the Catalan coast or the regions of northern Italy and the Alps), and for some of the periods identified (Phase 04 in particular). The state of archaeobotanical research may indeed influence the suitability models of some of the crops. This is the case of the opium poppy, for instance, a crop of Mediterranean origin that spread to central Europe together with other Neolithic elements (and possibly populations) most likely from the Western Mediterranean area[43,44]. Our analyses could suggest its great suitability to wet and cold climates, but actually, they show how quickly the plant was able to adapt to and thrive in new climatic conditions under cultivation. The spatial and temporal bias may be due to preservation issues due to a different distribution of research efforts on the territories or to the fact that
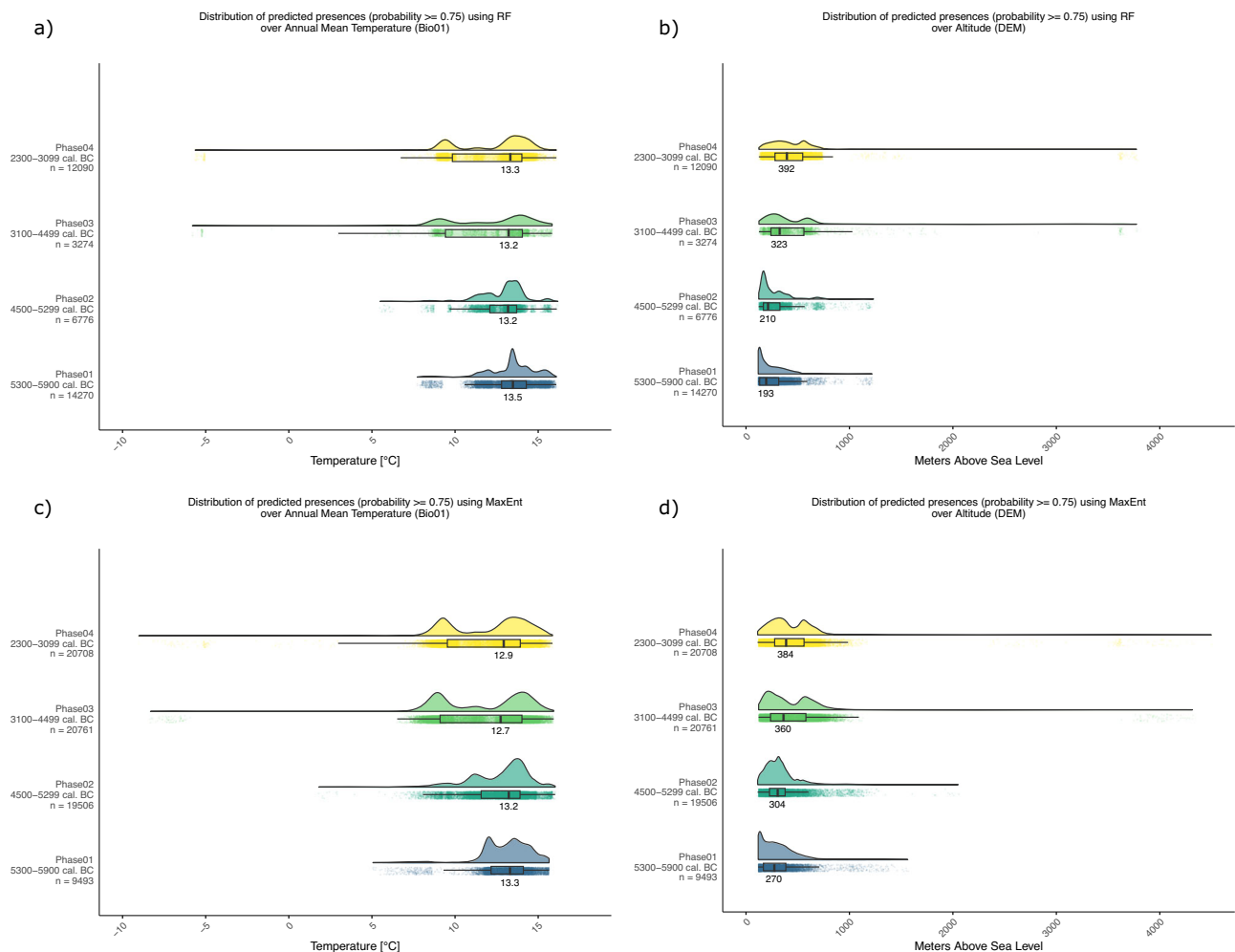
Fig. 6 | Comparison of the distribution of predicted high suitability areas using both models over Annual Mean Temperature and Elevation. a shows the distribution of predicted high suitability areas (≥ 0.75) modelled using Random Forest (RF) over Annual Mean Temperature (Bio01). **b** shows the distribution of predicted high suitability areas (≥ 0.75) modelled using Random Forest (RF) over Elevation (DEM). **c** shows the distribution of predicted high suitability areas (≥ 0.75) modelled using Maximum Entropy (MaxEnt) over Annual Mean Temperature (Bio01). **d** shows the distribution of predicted high suitability areas (≥ 0.75) modelled using Maximum Entropy (MaxEnt) over Elevation (DEM). Ridge lines on each panel show kernel density estimates, and the boxplots show medians, first and third quartiles (hinges), minimum and maximum values no further than 1.5*IQR from the hinge where IQR is the inter-quartile range (whiskers). Points show the jittered distribution of data points. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

archaeological sites have a greater chance of being discovered if they are repeatedly used in time. Lastly, our study can also be partly limited by the absence of soil-related variables used in the modelling procedure, such as indices of soil types, texture, and vegetation distribution[55]. All of these drawbacks need to be considered in the discussion of the results generated.

In this study, we applied both spatial statistics analysis and machine learning/species distribution modelling comparison techniques using RF and MaxEnt. One can consider these two algorithms as complementary methods. RF is an ensemble learning method based on decision trees, constructing multiple decision trees during training that outputs the average prediction of the individual trees. MaxEnt, however, is a probabilistic modelling approach that aims to find the distribution of maximum entropy given a set of constraints. RF tends to capture complex interactions between variables, while MaxEnt might provide more generalised response curves. We believe that comparing these two techniques allowed us to take advantage of the strength of the two models and to be more confident in the comparable predictions.

The combined analyses performed in this paper prove very useful to answer questions regarding agricultural decision-making and spreading in the past. Phase 01 presents a relatively narrow niche. The

strong seasonality and the colder climatic conditions favoured glume wheat as a predominant taxa, possibly due to their resilience capacity, within broadly diverse crop assemblages. These first two Phases, Phase 01 and Phase 02 also witness a cluster of settlements suitability areas mainly distributed along the Mediterranean shores. In Phase 03, a warmer and more humid phase allowed farmers to migrate to drier internal zones and to colder temperatures, where they mostly took the set of crops from the previous phase. The optimal climate conditions during this Phase, along with the ease of processing of these specific taxa prior to consumption might have led the farmer communities to prefer these among others[56]. Several factors might have influenced this decision, such as a higher population density or a possible population boom, as observed in several archaeological proxies from the area[57] (and in Europe)[58]. So not only could these cereals have been suited for the climate, but also the most productive and interesting to sustain a growing population inhabiting a broader niche. With the general increase in temperature and decrease in seasonality (see Fig. 2b), the climatic niche for barley, naked and glume wheat shifted inland towards higher altitudes over time. This might have allowed farmer communities to expand and occupy new territories. Similar observations were made in the natural environment of the alpine and
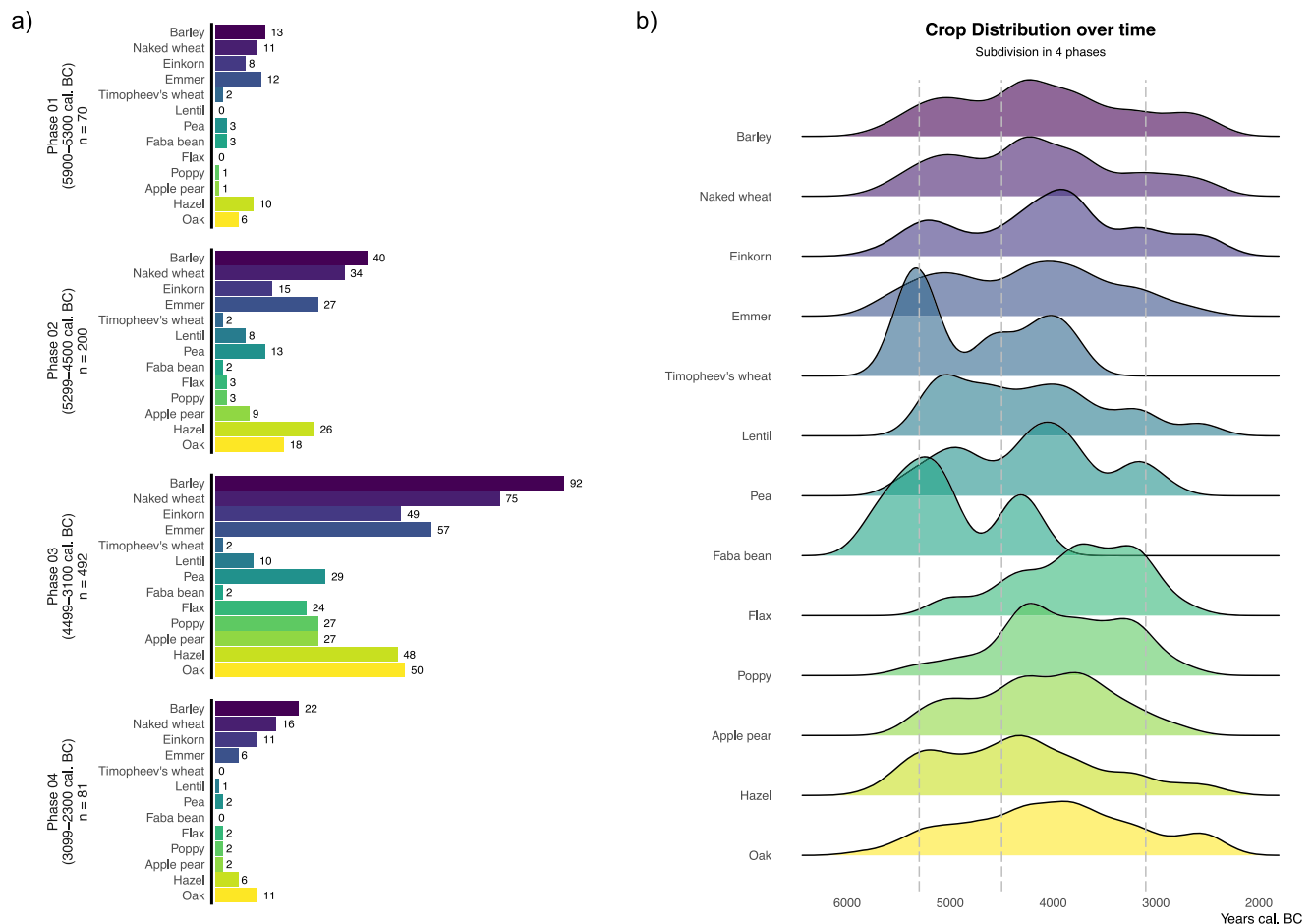
Fig. 7 | Quantified crop occurrences per phase (a) and density of crop occurrences over time (b). a The bars show the number of occurrences in the database for each type of crop per chronological phase. **b** Coloured areas show Kernel Density estimates of crop occurrences over the entire period for each crop. The vertical dashed lines indicate the limits of our 4 chrono-phases. Details about sample numbers are provided in Supplementary Table 3. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

subalpine areas of the central Pyrenees during the first half of the Holocene period[20]. Some authors had interpreted these sites at higher altitudes as evidence of increased site specialization[21], but our results would support an expansion into the highlands of year-round farming sites during a period of climatic amelioration, which would have a more or less similar economy and would be tightly networked in comparison to previous phases. The networks during this period have been extensively analysed using many different types of artifacts (i.e., pottery[59,60]). Our results suggest that farmers responded to climate change by moving to higher altitudes (broadening their niche) and not as a shift to specialization in particular products, but rather to maintain the resilience of the whole mixed farming network. During Phase 04, wetter conditions prompted farmers to adapt and favour wet-tolerant crops. In colder and wetter environments, wild crops such as oil seeds flax, opium poppy, pea or apple/pear might have been more regularly added to the communities' diets in order to compensate for the difficulty of growing naked wheats as in earlier periods. Emmer seems to partially replace free-threshing wheat in wetter and colder areas.

The results highlight in particular three relevant moments of change at a large scale and of high relevance: (1) A first one in correspondence of 5300 BC approx., when the climate became drier in the NW Mediterranean area, and thus the early farmers had the possibility to choose the most appropriate crops to grow and keep growing at specific locations. Until this moment, glume wheat was the predominant taxa within broadly diverse crop assemblages and the settlements clustered mainly along the Mediterranean shores, but during

Phase 02, naked wheat and (naked) barley become more important, and glume wheat become residual (as also observed in other works[7]); (2) This situation started to change around 4000 BC, Specifically, the peak of naked wheat occurrences (Fig. 7b) happens at a moment when Mean Annual Temperature and Precipitation are at their highest (in Phase 03). After this moment, glume wheats, particularly einkorn, gain importance. It is possible also that not only climate variations but also the emergence of main storage pests, such as the grain weevil, might have played a role in the clear shift of the crop spectrum, particularly visible in the Mediterranean regions, as recently suggested by other works[47,56]. The spread of these pests could have been facilitated by the active networks that functioned at the time. (3) Finally, a last moment of major change can be identified around 3100 BC, when the wetter climate in the alpine Foreland prompted the farmers to adapt to the new conditions and to make different choices in order to face the new climatic/environmental conditions. For example, we see that oil seeds like flax, opium poppy, pea or apple/pear, have been found in colder and wetter environments than barley, glume wheats and naked wheats. These wild crops might have been more regularly added to the communities' diets in order to compensate for the difficulty of growing naked wheat in such environments (see also Steiner et al.[61]). Changes towards more widespread cultivation of einkorn and emmer have also been pointed out for regions such as Southern France[62] and Switzerland[35]. Our analyses indicate that these changes could have been, to some extent, an adaptation to new climatic conditions, although social factors cannot be excluded.
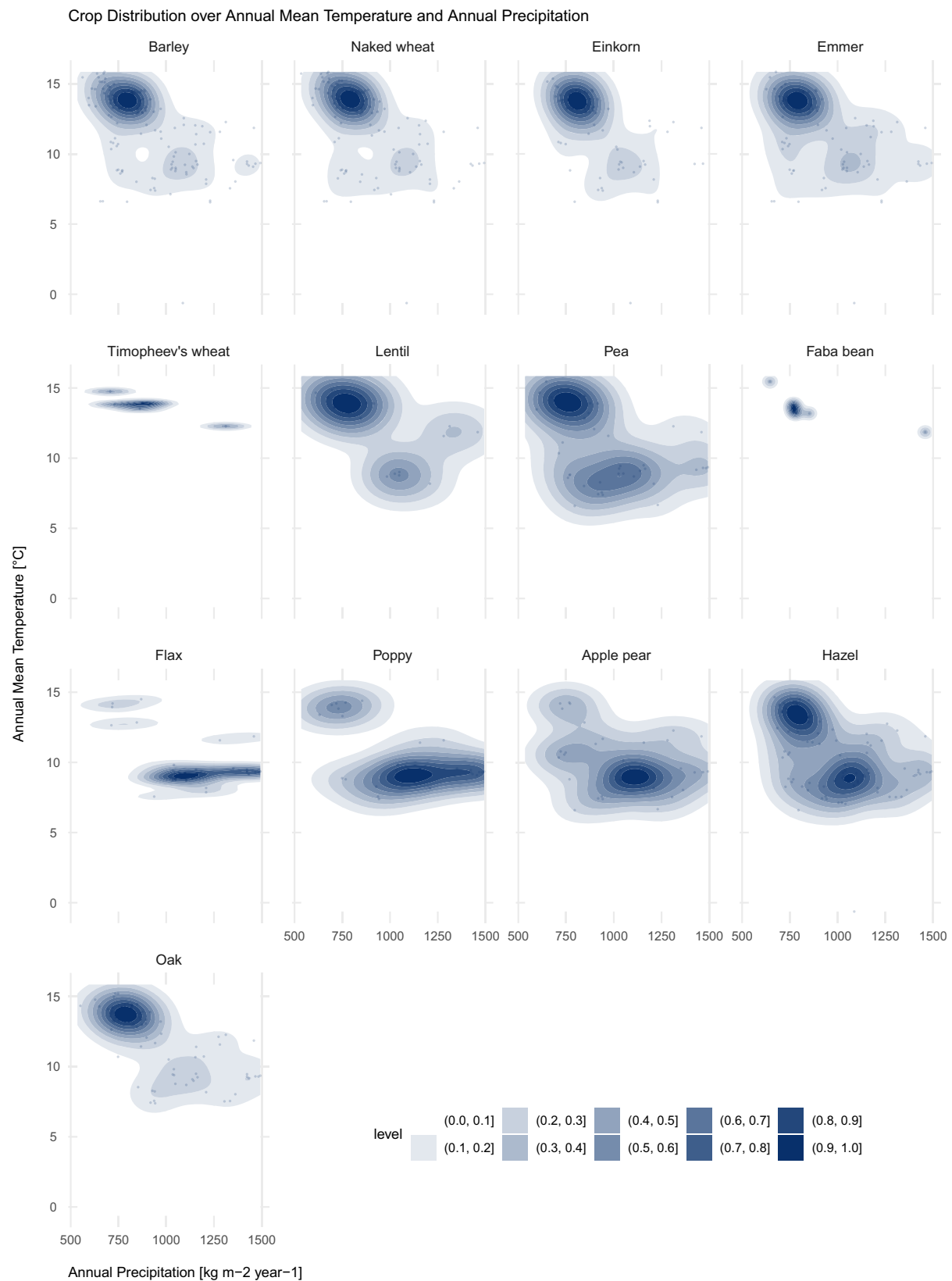
Fig. 8 | Distribution of crop occurrences over Annual Mean Temperature and Annual Precipitation. Blue areas in each panel show 2D kernel density estimates of the distribution of crop occurrences over Annual Mean Temperature (*y*-axis) and Annual Precipitation (*x*-axis) over the study period. Darker blue indicates higher density and light blue indicates low density. Details about sample numbers are provided in Supplementary Table 3. Source data can be found at https://doi.org/10.5281/zenodo.14253277.

Fig. 9 | Distribution of crop occurrences over Mean Temperature and Precipitation of Coldest Quarter. Blue areas in each panel show 2D kernel density estimates of the distribution of crop occurrences over Annual Mean Temperature (*y*-axis) and Annual Precipitation (*x*-axis) over the study period. Darker blue indicates higher density and light blue indicates low density. Details about sample numbers are provided in Supplementary Table 3. Source data can be found at https://doi.org/10.5281/zenodo.14253277.
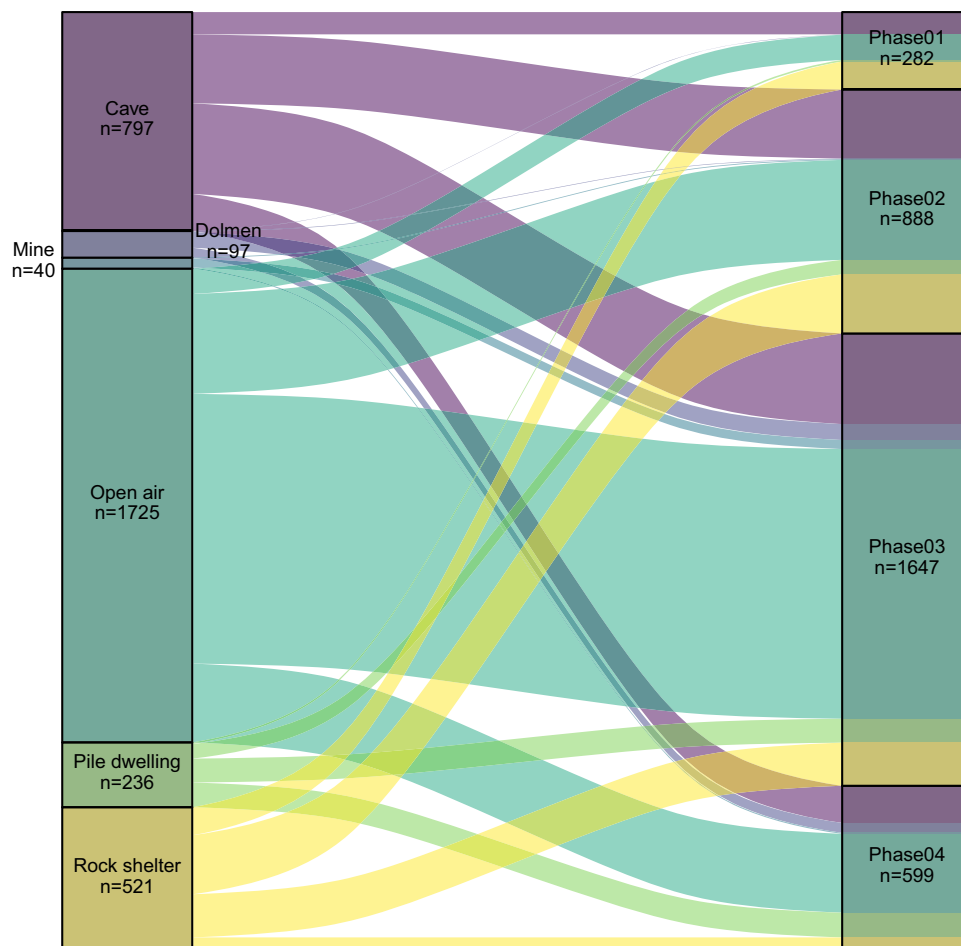
Fig. 10 | Distribution of sites per type and phase. The Alluvial diagram shows the distribution of sites per site type (left) and per phase (right). Source data can be found at https://doi.org/10.5281/zenodo.14253277.

Comparably to recent works that made use of a similar methodology, for more specific ecological analyses[63,64], we have identified correlations between temperature and precipitation increases/decreases and the overall settlement distribution patterns and crop diversification for the period 5900-2300 BC cal. The results we obtained are consistent with previous studies[64,65] that highlighted considerable challenges related to climate and climatic changes being among the key drivers of Neolithic human dynamics and agricultural adaptation and resilience strategies[30,66], which furthermore suggest how small-scale communities could have developed different adaptive or resilient strategies to face climatic-based limitations in a given time and space[67].

We conclude that human niche changes and crop changes in the Neolithic period of the region are independent of each other. Crop niches prove to have a plasticity beyond initial expectations from present-day crops and human niche breadth expansion is possible without crop changes. Crop changes resulting from short-term and long-term climatic oscillations were instead detected. As archaeobotanical and climate datasets gain higher resolution, paleo-climate and paleo-environmental modelling approaches become the road to be travelled for the future to reconstruct ancient crop ecologies and, hence, to refine uniformitarian inferences.

Methods

Archaeological and archaeobotanical data

A database consisting of a total of 3416 geo-referenced archaeological sites, with associated radiocarbon dates (calibrated using OxCal v. 4.4.2 and the atmospheric curve IntCal20) falling within the timeframe considered in this research was initially drawn from *AgriChange_14Cdatabase*[49]. This open-access dataset contains a unique collection of inventoried sites located between the Upper Rhine, the Po, the Rhone and the Ebro valleys, obtained from a combination of own fieldwork and the published literature. It includes six different types of archaeological sites (we grouped *Chasm* type in the *Rock shelter* category), among which *Open air* sites with multiple occupation phases account for over half of the occurrences, followed by *Cave* sites, *Rock shelters* and *Pile-dwellings* (see Fig. 10 and Supplementary Table 2). The chronological phases used in this paper follow previous work[49,68,69] and correspond to the main socioeconomic dynamics in the region.

The fields extracted from this database and retained for the present research concern the specificity of the sites (ID, name and type of the site, their geographic coordinates (GCS WGS 1984 using decimal degrees as angular unit)) and the dating information (the corresponding calibrated *mean*). When sites were occupied multiple times, and samples lay in different stratigraphic units, these were treated as individual sites with their respective environmental conditions.

In addition, a second database[33,70], collecting archaeobotanical information for 843 archaeological sites, was integrated and processed. Thus, the archaeobotanical dataset assembled for this study provides a synthesis of the recorded presence of taxa (only charred remains were considered) at a site level for the region of interest, with a specific reference to the seeds and the crop typologies described (the term "crop typology" refers here to economically important,

potentially managed, plants and not only to traditional crops). We merged this information in a unique database, and hence, all available crop records were georeferenced and attributed to a spatiotemporal dimension (Fig. 7 and Supplementary Table 3).

Paleoclimate variables

The selection of paleoclimate records was essentially based on a set of criteria, including high dating reliability, high time resolution and a spatial extent that could cover the entire region in the exam. Specifically, we use a set of 18 reconstructed climatic variables (see Supplementary Table 4) from the mid-to-late Holocene for land surface areas, selected based on their relevance and interpretability for our research. These variables are retrieved from the open access dataset *CHELSA-TraCE21k v.1.0*, downscaled[51,71] and derived from the CCSM3_TraCE21k model to a 30 arcsec resolution using the CHELSA V1.2 algorithm[71], which covers time steps of 100 years for the last 21'000 years, with minimum and maximum temperatures, surface precipitation, and paleo-orography information. It contributes to creating a paleoclimatic envelope that best matches the spatiotemporal distribution of our settlements and crops. The data were read as GeoTiff using the *raster* package[72] in *R*[73]

Environmental variables

In addition to the paleoclimate variables, more specific environmental variables (see Supplementary Table 4) were computed and used to describe landscape characteristics, such as the digital elevation model (DEM) expressing the elevation and derived from the same *CHELSA-TraCE21k v.1.0* dataset. Although the elevation may not be an optimal proxy for describing the landscape, as it may introduce bias in the modelling procedure and catalyse predicted areas[74], we nevertheless decided to include it here to evaluate if adaptation to height might have played a role in the distribution of the early Neolithic farmers and their specific subsistence activities as done in other works[75]. We further derived the slope, which defines the steepness of the terrain, and the terrain ruggedness index (TRI), which is the mean of the absolute differences between the value of a cell and the value of its eight surrounding cells, using the *terrain* function in the *terra* package[76], and calculated the proximity to important water resources[77] as reconstructed permanent lakes and rivers in a GIS environment (see the tool *Euclidean distance* in ArcGIS 10.8 – ESRI).

Habitat Suitability and Niche models construction

The methodology developed in this study is primarily based on machine learning techniques and on species distribution modelling (SDM) approaches, linking known species localities with predictor variables to assess patterns of species occurrence and habitat suitability[78,79]. Such methods, well established in ecology and biogeography[80], evolution and more recently in conservation biology and climate change research[78,81–83], have only recently seen first applications in archaeology and, more specifically in research related to archaeobotanical studies, as modelling tools for analysing occurrence data and for predicting human habitat suitability[29,66]. Especially, extensive literature supports the use of advanced quantitative and machine learning methods to reconstruct empirical ecological settings of ancient human populations and their subsistence activities[25,84–86], as well as to examine potential environmental and climatic changes and their ecological consequences in modern and future scenarios.

Although there is a variety of algorithms with different levels of complexity for SDMs[87], only a limited number of algorithms are being applied in archaeology (often Maximum Entropy alone) and thus we decided to test and compare the algorithms of *Random Forest* (RF)[88] and *Maximum Entropy* (MaxEnt)[89] and to explore and discuss their results, as also suggested by several authors[80,90].

RF is an ensemble decision tree method of ML based on classification and regression tree algorithms. It is widely recognized for its capacity to produce good predictive models with few sparse data[91]. Further, it can tolerate noise overfitting and can handle a large number of predictors and their interactions[92–95], although, it is still in its infancy in archaeological studies, with few applications to different research branches[96–98]. This technique has been applied here to investigate the relationships between a categorical dependent variable (Neolithic settlements) to a set of predisposing factors such as ecological and climatic variables (topography, paleoclimatic data, etc.), to identify if and to what extent the ecological setting of the study area acts as a predictor in the determination of the dependent variable (settlements) and by this means contributes to the shaping of an eco-cultural niche.

MaxEnt builds upon the principle of maximum entropy and is used to approximate a target probability distribution of data occurrences that is closest to uniform and subject to environmental/climate constraints[99]. Being a generative approach rather than discriminative, it shows several advantages when the amount of training data is limited and when only presence/occurrences data is available[100]. For its high-performance rate, MaxEnt is mainly preferred in SDMs in the field of ecology and palaeoecology and has repeatedly shown to be an invaluable tool in a wide number of applications[101–103], but yet only a few and very recent applications of this specific algorithm can be enumerated in archaeology[75,104,105]. Both algorithms can utilize continuous and categorical data and can incorporate interactions between different variables. They are considered among the leading data-mining ML methods for their high accuracy prediction[93].

Our modelling pipeline conceptually follows the research workflow defined in the most recent literature[65,100,106], as well as by research exercises for modelling population distribution, agriculture and crop niches on different spatiotemporal scales[25,66].

Data preparation was performed both in *ArcGIS 10.8, ArcGISPro* and in *R*. The model calibration, final model computations, post-modelling analyses, and assessment were performed in *R*.

We used 18 downscaled Bio (paleoclimate) variables extracted for the study area and limited to the period 5900-2300 BC to spot climate trends for the specific region and period[29]. We then combined this climatic information, along with the computed five environmental variables (elevation, slope, TRI, distance to main rivers, distance to main lakes) to classify the local ecological envelope for each site type and location.

The paleoclimatic and environmental variables retained as model inputs in the subsequent modelling procedure were further used to build two HS Models (based on the *randomForest*[107] and *maxnet*[108] packages available at *cran.r-project.org/web/packages/randomForest/* and *cran.r-project.org/package=maxnet*).

For each of the four phases, we imported the occurrence data (archaeological sites) and the predictor variables (paleoclimate and environmental), defining the spatial resolution and extent of the analysis to a 30-arcsecond cell size.

To avoid spatial autocorrelation affecting the predictors and the model results[18] and to reduce the dimensionality of the initial pool of 18 paleoclimates and five environmental variables, we used the non-parametric Spearman correlation coefficient test for each variable set (paleoclimatic and environmental). From the list of the paleoclimate and environmental variables (as shown in Supplementary Table 4), we retained those predictors with correlation values of $r < 0.25$. Where several variables form highly correlated clusters ($r > 0.75$), we selected the one predictor from each cluster least correlated to all other variables. Slope and elevation are correlated by less than 0.01 more than the threshold of 0.75. Nonetheless, as the dataset contains high altitude sites, we decided to keep both variables, as we considered them to be important factors to detect relations between site locations and their environmental surroundings, at play here both in past human mobility choices and adaptations strategies, thus retaining 9 paleoclimate and 4 environmental variables as input in the modelling process. These include Mean Diurnal Range (Bio02), Temperature

Seasonality (Bio04), Mean Temperature of Warmest Month (Bio05), Mean Temperature of Wettest Quarter (Bio08), Mean Temperature of Driest Quarter (Bio09), Precipitation of Wettest Month (Bio13), Precipitation Seasonality (Bio15), Precipitation of Driest Quarter (Bio17) and Precipitation of Coldest Quarter (Bio19), as well as Distance from Lakes (D_Lakes), Distance from Rivers (Dist_Riv), Slope (Slope) and Altitude (DEM) (see Supplementary Fig. 9).

We performed both model fittings with spatial split-block cross-validation techniques. The data are split into *k*-independent subsets, and for each subset, the model is trained with *k*-1 subsets and evaluated on the *k*th subset[109]. The optimal block size was selected based on the spatial autocorrelation structure of the predictors using the *spatialAutoRange* function in the *R* package *blockCV*[110] and was set at 170 km X 170 km.

We ran the RF model in classification mode, and because our dataset does not contain confirmed absences, we generated a set of random pseudo-absences over the landscape, equal in number to the occurrence dataset. Bootstrapped comparisons were repeated 1000 times to ensure that sites and their surroundings were treated as a whole, and to generate a distribution of AUC values using different training data sets, while four predictors were chosen at each split.

Unlike RF, MaxEnt uses presence-only data to predict the distribution of species based on the theory of maximum entropy. We run it with presences only and a group of 10,000 random background points, as this amount of background data is considered large enough to provide an accurate representation of the study area, while a larger background sample increases computation time without improving modelling performance[111,112]. The model prediction function was set on *cloglog* transform[108], it is the most appropriate method for estimating probability of presence and since it gives a better result over logistic when bias correction is used[108]. Other settings were left at their default values.

Eventually, we obtained eight suitability maps, one for each Phase, in which each raster cell was assigned with the relative index of occurrence of the archaeological settlement (ranging from 0 for predicted absence to 1 for predicted presence – low to high suitability). We further generated variable importance rankings, partial dependence and response plots to reveal the importance of each predictor variable in the prediction of site occurrence.

Model validation and performance assessment
In this study and in spatial data analysis more broadly, test dataset observations are often situated near training dataset observations. This spatial proximity can cause an overestimation of model performance due to spatial autocorrelation, where nearby observations share similar attributes.[92] To address this issue, training and testing datasets should be geographically separated. Spatial *k*-fold cross-validation is a common method to achieve this. The dataset is divided into *k* non-overlapping groups, the model is trained on *k*-1 groups, and performance is evaluated on the remaining group. This process is repeated *k* times, and the error estimates are averaged to provide an overall performance metric. For this study, we employed 5-fold spatial cross-validation, implemented with the blockCV package[109].

Model performance for both algorithms was assessed on test data as the *Area Under the Receiver-operating characteristic curve* (AUC), which represents the capacity of the models to distinguish between the presence and absence or background points. This measure is independent of threshold selection, making it a powerful tool for assessing model performance[113]. The accuracy of the two models was assessed through the AUC curves, which are presented in the Supplementary Fig. 10. This graphical method illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR), both represented as percentages of the total cases. True positives (TP) and true negatives (TN) refer to correctly identified outcomes, while false positives (FP) occur when a prediction incorrectly labels an outcome as a presence when it is actually an absence or background point.

Conversely, false negatives (FN) happen when an outcome is wrongly classified as an absence or background point instead of a presence. AUC can generally range from 0 to 1, where a score of 1 indicates perfect discrimination and values < 0.5 indicate a predictive performance worse than random. AUC values for RF models range between 0.64 and 0.78 accuracy on the testing dataset, while the AUC values for MaxEnt models range between 0.69 and 0.86 on the testing dataset, indicating that RF model predictions are slightly more accurate than MaxEnt for Phase 02 and 03 and MaxEnt model predictions are more accurate for Phase 01 and 04.

Statistics
R 4.1.2 version was used for all statistical and ML modelling analyses. Parameters such as sample size, number of replicates are reported in the Text, Figures and Figure Legends.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The source archaeological data used in this study are available at: https://doi.org/10.5334/joad.72. The paleoclimatic dataset used in this study was published by: Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., Zimmermann, N. E. (2020). CHELSA-TraCE21k: Downscaled transient temperature and precipitation data since the last glacial maximum. EnviDat. https://doi.org/10.16904/envidat.211 and is available at: https://envicloud.wsl.ch/#/?bucket=https%3A%2F%2Fos.zhdk.cloud.switch.ch%2Fchelsav1%2F&prefix=chelsa_trace%2F. The computed analyses, archaeological and archaeobotanical (crop) datasets produced in this study are available on the corresponding author's GitHub at https://github.com/MaCasti21/Nat-Comm_Castiello_2024. Data to generate all figures can be found on the following Zenodo repository: Castiello, M.E., et al., Understanding the spread of agriculture in the Western Mediterranean (6th-3rd millennia BC) with Machine Learning tools, Nat-Comm_Castiello_2024, https://doi.org/10.5281/zenodo.14253277, 2024.

Code availability
Code to reproduce the results and generate all figures can be found on the following Zenodo repository: Castiello, M.E., et al., Understanding the spread of agriculture in the Western Mediterranean (6th-3rd millennia BC) with Machine Learning tools, Nat-Comm_Castiello_2024, https://doi.org/10.5281/zenodo.14253277, 2024 and on the corresponding author's GitHub at https://github.com/MaCasti21/Nat-Comm_Castiello_2024.

References
1.   Bogaard, A. *Neolithic Farming in Central Europe*. (Routledge, 2004).
2.   Antolín, F. *Local, Intensive and Diverse? Early Farmers and Plant Economy in the North-East of the Iberian Peninsula (5500–2300 cal BC)*. (Barkhuis, 2016).
3.   Antolín, F. et al. Quantitative approximation to large-seeded wild fruit use in a late Neolithic lake dwelling. The case study of layer 13 of Parkhaus-Opéra in Zürich (Central Switzerland). *Quat. Int.* **404**, 56–68 (2016).
4.   Antolín, F. & Jacomet, S. Wild fruit use among early farmers in the Neolithic (5400–2300 cal bc) in the north-east of the Iberian Peninsula: an intensive practice? *Veg. Hist. Archaeobot.* **24**, 19–33 (2015).
5.   Colledge, S. The evolution of Neolithic farming from SW Asian origins to NW european limits. *Eur. J. Archaeol.* **8**, 137–156 (2005).
6.   Colledge, S. & Conolly, J. *The Origins and Spread of Domestic Plants in Southwest Asia and Europe*. (Left Coast, 2007).

7. de Vareilles, A. et al. One sea but many routes to Sail: the early maritime dispersal of Neolithic crops from the Aegean to the western Mediterranean. *J. Archaeol. Sci. Rep.* **29**, 102140 (2020).

8. McClatchie, M. et al. Neolithic farming in north-western Europe: archaeobotanical evidence from Ireland. *J. Archaeol. Sci.* **51**, 206–215 (2014).

9. Vidal-Cordasco, M. & Nuevo-López, A. Difference in ecological niche breadth between Mesolithic and Early Neolithic groups in Iberia. *J. Archaeol. Sci. Rep.* **35**, 102728 (2021).

10. Manen, C. et al. Territoriality and settlement in Southern France in the Early Neolithic: Diversity as a strategy? **7**, 923–938 (2021).

11. Zilhao, J. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc. Natl. Acad. Sci. USA* **98**, 14180–14185 (2001).

12. Martínez-Grau, H. et al. Global processes, regional dynamics? Radiocarbon data as a proxy for social dynamics at the end of the Mesolithic and during the Early Neolithic in the NW of the Mediterranean and Switzerland (c. 6200–4600 cal BC). *Doc. Praehist.* **47**, 170–191 (2020).

13. Manen, C. et al. The neolithic transition in the Western Mediterranean: a complex and non-linear diffusion process—the radiocarbon record revisited. *Radiocarbon* **61**, 531–571 (2019).

14. Lemercier, O. in *4e millénaire. La transition du Néolithique moyen au Néolithique final dans le sud-est dee la France et les régions voisines* Vol. 27 (eds O. Lemercier, R. Furestier, & É. Blaise) 17–44 (Monographies d'Archéologie Méditerranéenne, 2010).

15. Burri-Wyser, E. & Jammet-Reynal, L. in *Chronologie de la Préhistoire Récente Dans le Sud de la France. Acquis 1992-2012. Actualité de la Recherche. Actes des 10e Rencontres Méridionales de Préhistoire Récente, Porticcio, 18 au 20 octobre 2012* (eds Ingrid Sénépart et al.) 75–86 (Archives d'Écologieee Préhistorique, 2012).

16. Stöckli, W. *Urgeschichte der Schweiz im Überblick (15000 v.Chr. –Christi Geburt). Die Konstruktion einer Urgeschichte*. Vol. 54 (Archäologie Schweiz Antiqua, 2016).

17. Antolín, F. et al. Neolithic occupations (c. 5200-3400 cal BC) at Isolino Virginia (Lake Varese, Italy) and the onset of the pile-dwelling phenomenon around the Alps. *J. Archaeol. Sci. Rep.* **42**, 103375 (2022).

18. Lemercier, O. et al. in *Implantations Humaines en Milieu Littoral Méditerranéen: Facteurs d'installation et Processus d'appropriation de l'espace (Préhistoire, Antiquité, Moyen Âge)* (eds L. Mercuri, R. González Villaescusa, & F. Bertoncello) 191–203 (Éditions APDCA, 2014).

19. Heitz, C. in *Mobility and Pottery Production. Archaeological & Anthropological Perspectives* (eds C. Heitz & R. Stapfer) 257–292 (Sidestone Press, 2017).

20. Gassiot, E., Rodríguez-Antón, D., Burjachs, F., Antolín, F. & Ballesteros, A. Poblamiento, explotación y entorno natural de los estadios alpinos y subalpinos del Pirineo central durante la primera mitad del Holoceno. *Cuatern. Geomorfol.* **26**, 29–45 (2012).

21. Bréhard, S., Beeching, A. & Vigne, J.-D. Shepherds, cowherds and site function on middle Neolithic sites of the Rhône valley: An archaeozoological approach to the organization of territories and societies. *J. Anthropol. Archaeol.* **29**, 179–188 (2010).

22. Cunill, R., Soriano, J. M., Bal, M. C., Pèlachs, A. & Pérez-Obiol, R. Holocene treeline changes on the south slope of the Pyrenees: a pedoanthracological analysis. *Veg. Hist. Archaeobot.* **21**, 373–384 (2012).

23. Pedersen, J. B., Assmann, J. J., Normand, S., Karger, D. N. & Riede, F. Climate niche modeling reveals the fate of pioneering late Pleistocene populations in Northern Europe. *Curr. Anthropol.* **64**, 599–608 (2023).

24. Yaworsky, P. M., Hussain, S. T. & Riede, F. Climate-driven habitat shifts of high-ranked prey species structure Late Upper Paleolithic hunting. *Sci. Rep.* **13**, 4238 (2023).

25. Krzyzanska, M., Hunt, H. V., Crema, E. R. & Jones, M. K. Modelling the potential ecological niche of domesticated buckwheat in China: archaeological evidence, environmental constraints and climate change. *Veg. Hist. Archaeobot.* **31**, 331–345 (2022).

26. Burke, A. et al. Risky business: The impact of climate and climate variability on human population dynamics in Western Europe during the Last Glacial Maximum. *Quat. Sci. Rev.* **164**, 217–229 (2017).

27. Weide, A. et al. A new functional ecological model reveals the nature of early plant management in southwest Asia. *Nat. Plants* **8**, 623–634 (2022).

28. Braunisch, V. et al. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* **36**, 971–983 (2013).

29. Blinkhorn, J., Timbrell, L., Grove, M. & Scerri, E. M. L. Evaluating refugia in recent human evolution in Africa. *Philos. Trans. R. Soc. B Biol. Sci.* **377**, 20200485 (2022).

30. Banks, W. E., Antunes, N., Rigaud, S. & Francesco, D. E. Ecological constraints on the first prehistoric farmers in Europe. *J. Archaeol. Sci.* **40**, 2746–2753 (2013).

31. Burke, A. et al. The archaeology of climate change: The case for cultural diversity. *Proc. Natl. Acad. Sci. USA* **118**, e2108537118 (2021).

32. Altschul, J. H. et al. To understand how migrations affect human securities, look to the past. *Proc. Natl. Acad. Sci. USA* **117**, 20342–20345 (2020).

33. Antolín, F., Bouby, L., Martin, L., Rottoli, M. & Jesus, A. Archaeobotanical evidence of plant food consumption among early farmers (5700–4500 BC) in the Western Mediterranean Region. *Food Hist.* **19**, 235–253 (2021).

34. Hajnalová, M. & Dreslerová, D. Ethnobotany of einkorn and emmer in Romania and Slovakia: Towards interpretation of archaeological evidence. *Památky Archeol.* **CI**, 169–202 (2010).

35. Jacomet, S., Brombacher, C. & Dick, M. *Archäobotanik am Zürichsee. Ackerbau, Sammelwirtschaft und Umwelt von neolithischen und bronzezeitlichen Seeufersiedlungen im Raum Zürich. Ergebnisse von Untersuchungen pflanzlicher Makroreste der Jahre 1979–1988*. Vol. 7 (Orell Füssli Verlag, 1989).

36. Percival, J. *The Wheat Plant*. (Duckworth, 1974 (Reprint von 1921)).

37. Peña-Chocarro, L. *Prehistoric Agriculture in Southern Spain during the Neolithic and the Bronze Age. The application of ethnographic models*. Vol. 818 (Archaeopress, 1999).

38. Rovira, N. *Agricultura y gestión de los recursos vegetales en el sureste de la Península Iberíca durante la prehistoria reciente*, Universitat Pompeu Fabra, (2007).

39. Shands, H. L. & Dickson, A. D. Barley: Botany, production, harvesting, processing, utilization and economics. *Econ. Bot.* **7**, 3–26 (1953).

40. Badaeva, E. D., Filatenko, A. A. & Badaev, N. S. Cytogenetic investigation of Triticum timopheevii (Zhuk.) Zhuk. and related species using the C-banding technique. *Theor. Appl. Genet.* **89**, 622–628 (1994).

41. Marinova, E. & Popova, T. Cicer arietinum (chick pea) in the Neolithic and Chalcolithic of Bulgaria: implications for cultural contacts with the neighbouring regions? *Veg. Hist. Archaeoabot.* **17**, 73–80 (2008).

42. Antolín, F. & Schäfer, M. Insect pests of pulse crops and their management in Neolithic Europe. *Environ. Archaeol.* **29**, 20–33 (2024).

43. Jesus, A. et al. A morphometric approach to track opium poppy domestication. *Sci. Rep.* **11**, 9778 (2021).

44. Salavert, A. et al. Direct dating reveals the early history of opium poppy in western Europe. *Sci. Rep.* **10**, 20263 (2020).

45. Salavert, A., Martin, L., Antolín, F. & Zazzo, A. The opium poppy in Europe: exploring its origin and dispersal during the Neolithic. *Antiquity* **92**, e1 (2018).

46. Jesus, A., Prats, G., Follmann, F., Jacomet, S. & Antolín, F. Middle Neolithic farming of open-air sites in SE France: new insights from archaeobotanical investigations of three wells found at Les Bagnoles (L'Isle-sur-la-Sorgue, Dépt. Vaucluse, France). *Veg. Hist. Archaeobot.* **30**, 445–461 (2021).

47. Antolín, F. et al. An archaeobotanical and stable isotope approach to changing agricultural practices in the NW Mediterranean region around 4000 BC. *Holocene* **34**, 239–254 (2024).

48. Martin, L. et al. in *Le Chasséen, des Chasséens... Retour sur une Culture Nationale et ses Parallèles, Sepulcres de Fossa, Cortaillod, Lagozza. Colloque International de Paris, 18–20 novembre 2014* (eds T. Perrin, P. Chambon, J. F. G. Bao, & G. Goude) 259–272 (Archives d'Écologie Préhistorique, 2016).

49. Martínez-Grau, H., Morell-Rovira, B. & Antolín, F. Radiocarbon dates associated to neolithic contexts (Ca. 5900–2000 Cal BC) from the Northwestern Mediterranean Arch to the high rhine area. *J. Open Archaeol. Data* **9**, 1–10, (2021).

50. Antolín, F. et al. The AgriChange project: an integrated on-site approach to agricultural and land-use change during the Neolithic in Western Europe. *PAGES N. Glob. Chang. Mag.* **26**, 26–27 (2018).

51. Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H. & Zimmermann, N. E. CHELSA-TraCE21k – high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Clim. Past* **19**, 439–456 (2023).

52. Tegel, W., Elburg, R., Hakelberg, D., Stäuble, H. & Büntgen, U. Early Neolithic Water Wells Reveal the World's Oldest Wood Architecture. *PLOS ONE* **7**, e51374 (2012).

53. Hong, H., Xiaoling, G. & Hua, Y. in *7th IEEE International Conference on Software Engineering and Service Science (ICSESS).* 219–224 (2016).

54. Molnar, C. *Interpretable Machine Learning* (2022).

55. Contreras, D. A. et al. From paleoclimate variables to prehistoric agriculture: Using a process-based agro-ecosystem model to simulate the impacts of Holocene climate change on potential agricultural productivity in Provence, France. *Quat. Int.* **501**, 303–316 (2019).

56. Häberle, S. et al. Small animals, big impact? early farmers and Pre- and Post-harvest pests from the middle Neolithic site of les bagnoles in the South-East of France (L'Isle-sur-la-Sorgue, Vaucluse, Provence-Alpes-Côte-d'Azur). *Animals* **12**, 1511 (2022).

57. Barton, C. M., Ullah, I. I. & Bergin, S. Land use, water and Mediterranean landscapes: modelling long-term dynamics of complex socio-ecological systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **368**, 5275–5297 (2010).

58. Timpson, A. et al. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. *J. Archaeol. Sci.* **52**, 549–557 (2014).

59. Borrello, M. A. & Van Willigen, S. Lagozza et Chasséen: Insertion chronologique et culturelle des céramiques de la Lombardie occidentale et du Sud-est de la France. *Sibrium Cent. di Stud. Preistorici e Archaeol. di Varese* **26**, 90–111 (2012).

60. Bouby, L. et al. Early Neolithic (ca. 5850-4500 cal BC) agricultural diffusion in the Western Mediterranean: An update of archaeobotanical data in SW France. *PLOS ONE* **15**, e0230731 (2020).

61. Steiner, B. L. et al. Archaeobotanical and isotopic analyses of waterlogged remains from the Neolithic pile-dwelling site of Zug-Riedmatt (Switzerland): Resilience strategies of a plant economy in a changing local environment. *PLOS ONE* **17**, e0274361 (2022).

62. Bouby, L., Philippe, M. & Núria, R. Late Neolithic plant subsistence and farming activities on the southern margins of the Massif Central (France). *Holocene* **30**, 599–617 (2020).

63. Sánchez Goñi, M. F. et al. The expansion of Central and Northern European Neolithic populations was associated with a multi-century warm winter and wetter climate. *Holocene* **26**, 1188–1199 (2016).

64. Timmermann, A. et al. Climate effects on archaic human habitats and species successions. *Nature* **604**, 495–501 (2022).

65. Hua, X., Wiens, J. J., Associate Editor: Uta, B. & Editor: Troy, D. How does climate influence speciation? *Am. Nat.* **182**, 1–12 (2013).

66. Burke, A., Riel-Salvatore, J. & Barton, C. M. Human response to habitat suitability during the Last Glacial Maximum in Western Europe. *J. Quat. Sci.* **33**, 335–345 (2018).

67. Ordonez, A. & Riede, F. Changes in limiting factors for forager population dynamics in Europe across the last glacial-interglacial transition. *Nat. Commun.* **13**, 5140 (2022).

68. Hafner, A. & Suter, P. J. *Das Neolithikum in der Schweiz* (2003).

69. Oms, F. X. et al. The Neolithic in Northeast Iberia: Chronocultural phases and 14C. *Radiocarbon* **58**, 291–309 (2016).

70. Jesus, A. *Crop Dynamics in the Neolithic Period in the NW Mediterranean Area and the Swiss Plateau The Role of Opium Poppy (P. somniferum/setigerum)* PhD thesis, University of Basel, (2021).

71. Karger, D. N. et al. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).

72. Hijmans, R. J. *raster: Geographic Data Analysis and Modeling. R package version 3.5-2.* (2021).

73. *R Core Team*, https://r-project.org (2021).

74. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147–186 (2000).

75. Rodríguez, J., Willmes, C., Sommer, C. & Mateos, A. Sustainable human population density in Western Europe between 560.000 and 360.000 years ago. *Sci. Rep.* **12**, 6907 (2022).

76. Hijmans, R. J. *terra: Spatial Data Analysis. R package version 1.5-17,* <https://CRAN.R-project.org/package=terra > (2022).

77. Osborne, P. E., Alonso, J. C. & Bryant, R. G. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *J. Appl. Ecol.* **38**, 458–471 (2001).

78. Segurado, P. & Araújo, M. B. An evaluation of methods for modelling species distributions. *J. Biogeogr.* **31**, 1555–1568 (2004).

79. Bocinsky, R. K. & Kohler, T. A. A 2000-year reconstruction of the rain-fed maize agricultural niche in the US Southwest. *Nat. Commun.* **5**, 5618 (2014).

80. Nogués-Bravo, D. Predicting the past distribution of species climatic niches. *Glob. Ecol. Biogeogr.* **18**, 521–531 (2009).

81. Araújo, M. B. et al. Standards for distribution models in biodiversity assessments. *Sci. Adv.* **5**, eaat4858 (2019).

82. Drew, C. A., Wiersma, Y. F. & Huettmann, F. *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications.* (Springer New York, 2011).

83. Franklin, J., Potts, A. J., Fisher, E. C., Cowling, R. M. & Marean, C. W. Paleodistribution modeling in archaeology and paleoanthropology. *Quat. Sci. Rev.* **110**, 1–14 (2015).

84. Banks, W. E. et al. An ecological niche shift for Neanderthal populations in Western Europe 70,000 years ago. *Sci. Rep.* **11**, 5346 (2021).

85. Bergin, S. & Pardo-Gordó, S. *Simulating Transitions to Agriculture in Prehistory.* (Springer International Publishing, 2022).

86. Timbrell, L., Grove, M., Manica, A., Rucina, S. & Blinkhorn, J. A spatiotemporally explicit paleoenvironmental framework for the Middle Stone Age of eastern Africa. *Sci. Rep.* **12**, 3689 (2022).

87. Guisan, A., Thuiller, W. & Zimmermann, N. E. *Habitat Suitability and Distribution Models: With Applications in R.* (Cambridge University Press, 2017).

88. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

89. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 (2006).

90. Zhao, Z., Xiao, N., Shen, M. & Li, J. Comparison between optimized MaxEnt and random forest modeling in predicting potential distribution: A case study with Quasipaa boulengeri in China. *Sci. Total Environ.* **842**, 156867 (2022).

91. Mi C., Huettmann F., Guo Y., Han X. & L., W. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* **5**, https://doi.org/10.7717/peerj.2849 (2017).

92. Castiello, M. E. & Tonini, M. An explorative application of random forest algorithm for archaeological predictive modeling. A swiss case study. *J. Comput. Appl. Archaeol.* **4**, 110–125 (2021).

93. Cutler, D. R. et al. Random forests for classification in ecology. *Ecology* **88**, 2783–2792 (2007).

94. Han, X., Guo, Y., Mi, C., Huettmann, F. & Wen, L. Machine learning model analysis of breeding habitats for the black-necked crane in central Asian uplands under anthropogenic pressures. *Sci. Rep.* **7**, 6114 (2017).

95. Vignoles, A. Guide francophone pour la modélisation de niches écologiques. *Biodivers. Inform.* **17**, 67–95 (2022).

96. Castiello, M. E. *Computational and Machine Learning Tools for Archaeological Site Modeling*. 296 (Springer 2022).

97. Jones, P. J., Williamson, G. J., Bowman, D. M. J. S. & Lefroy, E. C. Mapping Tasmania's cultural landscapes: Using habitat suitability modelling of archaeological sites as a landscape history tool. *J. Biogeogr.* **46**, 2570–2582 (2019).

98. Roalkvam, I. Algorithmic classification and statistical modelling of coastal settlement patterns in mesolithic South-Eastern Norway. *J. Comput. Appl. Archaeol.* **3**, 288–307 (2020).

99. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).

100. Phillips, S. J. et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).

101. Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol., Evol. Syst.* **40**, 677–697 (2009).

102. Galletti, C. S., Ridder, E., Falconer, S. E. & Fall, P. L. Maxent modeling of ancient and modern agricultural terraces in the Troodos foothills, Cyprus. *Appl. Geogr.* **39**, 46–56 (2013).

103. Qin, A. et al. Maxent modeling for predicting impacts of climate change on the potential distribution of Thuja sutchuenensis Franch., an extremely endangered conifer from southwestern China. *Glob. Ecol. Conserv.* **10**, 139–146 (2017).

104. Conolly, J., Manning, K., Colledge, S., Dobney, K. & Shennan, S. Species distribution modelling of ancient cattle from early Neolithic sites in SW Asia and Europe. *Holocene* **22**, 997–1010 (2012).

105. Demján, P. et al. Long time-series ecological niche modelling using archaeological settlement data: Tracing the origins of present-day landscape. *Appl. Geogr.* **141**, 102669 (2022).

106. Gibert, C. et al. Climate-inferred distribution estimates of mid-to-late Pliocene hominins. *Glob. Planet. Change* **210**, 103756 (2022).

107. Liaw, A. & Wiener, M. Classification and regression by random-Forest. *R. N.* **2**, 18–22 (2002).

108. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. Opening the black box: an open-source release of Maxent. *Ecography* **40**, 887–893 (2017).

109. Merow, C., Smith, M. J. & Silander, J. A. Jr A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**, 1058–1069 (2013).

110. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10**, 225–232 (2019).

111. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175 (2008).

112. Anderson, R. P. & Gonzalez, I. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecol. Model.* **222**, 2796–2811 (2011).

113. Hosseini, N., Ghorbanpour, M. & Mostafavi, H. Habitat potential modelling and the effect of climate change on the current and future distribution of three Thymus species in Iran using MaxEnt. *Sci. Rep.* **14**, 3641 (2024).

Acknowledgements

Author contributions

M.E.C.: Conceptualisation; Methodology; Formal analysis; Visualisation; Writing original draft; Manuscript Review and Editing. E.R.: Climatic analyses: Conceptualisation and Methodology of the climate analysis; Review original draft. H.M.G., A.J., G.P. and F.A.: Archaeological and archaeobotanical data Collection and Processing. F.A.: Obtained funding; Supervised the study; Writing original draft; Manuscript Review and Editing. All authors reviewed the manuscript and provided feedback.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55541-y.

Correspondence and requests for materials should be addressed to Maria Elena Castiello.

Peer review information *Nature Communications* thanks Mark Vander Linden, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.