Models of the Neolithic dispersal in Southern Asia

Kavita Gangal

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics & Statistics
Newcastle University
Newcastle upon Tyne
United Kingdom

September 2014

To Baba..

आनंदाच्या लाटा

आनंदी व्हा आनंदी रहा
आनंद वाटा आणि आनंद लुटा।।

स्वत:साठी कुणाच्यासाठी सतत कष्ट करा
श्रमण्याचे समाधान तनमनी भरा
कष्टातून उसळूदे आनंदाच्या लाटा।।१।।

नका थांबू येण्यासाठी दिवाळी दसरा
आला दिवस आनंदाने उत्साहाने भरा
दु:खावरी मात करिती आनंदाच्या लाटा।।२।।

चिंता काळजी दूर ठेवा स्वार्थ सोडुन पहा
समाधानी मनाने आनंदात रहा
आनंदात झोपा आणि आनंदात उठा।।३।।

र. वि. गांगल

Acknowledgements

First of all, I would like to thank my supervisors Prof. Anvar Shukurov and Dr. Graeme Sarson. Working with them has been a great learning experience. I am thankful for their time, inputs, advice, patience and guidance. I would like to express my gratitude towards Prof. Shukurov for his invaluable advice and support during the tough times.

I would like to thank Dr. Dorian Fuller for sharing the archaeological data on evidence of rice in the prehistoric times, Prof. Travor Watkins for the archaeological report on the site of Wadi Shu'eib. I am deeply grateful to Prof. P. P. Divakaran for his unceasing support during the tough times and otherwise. I would also like to thank Dr. Ronojoy Adhikari for his encouragement, support and interesting and useful discussions.

I thank my family for the support. I am thankful to my wonderful friends Angela, Piu, Devayani, Siddiqua, Kalyani and Samir for being there. Also a special 'thank you' to Samir for the beautiful photograph of the tree rings from the Cumbrian forest and for patiently reading the thesis chapters.

I also owe a big 'thank you' to the external examiner Prof. James Steele (Director of Archaeology, UCL, London) and the internal examiner Dr. Andrew Fletcher (Newcastle University) for a non-stressful and thought provoking viva and for giving their invaluable comments and suggestions which improved the thesis in content and quality.

Lastly, I am thankful for all the experiences I have had so far, good and bad, that have made me realise the value of life and have helped me become a better professional and more importantly, a better person.

Abstract

The Fertile Crescent in the Near East is one of the independent sources of the Neolithic. Farming and pottery making spread across Europe from the Fertile Crescent from 9,000 to 6,000 years ago at an average rate of about 1 km/yr. The Neolithic in the Near East and the Indian subcontinent is far less explored in terms of absolute (^{14}C) dates as compared to the European Neolithic. The Neolithic chronology in the Indian subcontinent is mainly established with comparative dating and few ^{14}C dates are available from the region. Hence, though a strong causal connection between the Neolithic of the Near East and the Indus valley has been suggested, any detailed study has been challenging.

The different nature of the archaeological and ^{14}C dates and their uncertainties make it difficult to combine them in any quantitative analysis of the Neolithic. We overcome this by allocating different uncertainties and thus compile a comprehensive database of Early Neolithic dates in South Asia.

Using the earliest Neolithic arrival time(s) for each site, the globally averaged Neolithic dispersal speed from the Near East to the Indian subcontinent is calculated to be $U = 0.65 \pm 0.1$ km/yr.

Further inspection of the data shows that the Neolithic sites were restricted to the Fertile Crescent until about 7,000 BCE, and only later spread along the northern and southern borders of modern Iran. Analysis identifies two distinct routes of the Neolithic dispersal, one from the northern Zagros and the other from the southern Zagros, with significantly different dispersal speeds (about 0.6 km/yr for the southern route and 2 km/yr for the northern route). Furthermore, when combined with the Indus sites, the data suggests that the Neolithic plausibly reached the Indian subcontinent from the Northern Zagros.

The Neolithic of the Near East is studied further by using a two dimensional spatial diffusion model that incorporates palaeo-vegetation and topography data. Further expanding the scope of the diffusion model, the spread of incipient farming in the Indian subcontinent is studied. Depending on the environment, different staple food crops dominate in different regions (e.g. wheat in north-western India and rice in eastern India); and the model considers the spread of these different staple crops allowing for the altitude and water requirement for these crops.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Neolithic

At the end of the last glacial period (the Ice Age) about 12,000 years ago, environmental conditions became more hospitable for the survival and growth of living species. Human subsistence strategies also evolved towards a more settled lifestyle during this epoch. This time period at the end of the Stone Age is referred to as the 'Neolithic' period. The term is derived from the Greek word 'neolithicos' meaning 'New Stone Age'. A gradual inclination towards sedentism, that formed the basis of modern complex societies, is one of the salient features of this period. The Neolithic period is marked by a technological leap in human subsistence, and is thus of prime importance in the study of human prehistory.

There are different interpretations of the term 'Neolithic'. It was first proposed to describe the final phase of the Stone Age by Sir John Lubbock in 1865. Through time this interpretation has changed, ranging from the term being strictly used for the beginning of agriculture [16, 38][90, p. 151] to an all-encompassing culture of networking, sharing and trading [102, p. 19]. The modern-day definition sees the Neolithic as a multifaceted phenomenon comprising not just a single trait but an ensemble of (often related) features, e.g. the appearance of agriculture, pottery making, animal husbandry and pastoralism, and sedentism. Although the individual traits were neither simultaneously developed nor

Figure 1.1: Current political map from the Mediterranean Sea (the Near East) to the Indus Valley in the Indian subcontinent.

adopted together everywhere, they do appear to be closely linked [25, 112, 120].

The Neolithic in Europe has been studied in more detail than anywhere else in the world, with more than 700 sites dated with absolute techniques [108]. Though strong connections are posed between the Neolithic of the Near East and the Indian subcontinent (see section 1.3), these regions have not been as extensively studied with absolute dating (see Fig. 1.1 and Fig. 1.2 for maps). Only about 150–200 Neolithic sites in the Near East have been radiocarbon-dated. Studying the spread of the Neolithic from the Fertile Crescent to the Indus Valley has therefore been a challenging problem, solution to which has been elusive. In this study we aim to use mathematical and statistical modelling to explain the complex neolithisation process in this region from the Fertile Crescent in the Near East to the Indus Valley in the Indian subcontinent.

1.2 The Neolithic Spread

The Neolithic did not emerge independently at each location across the world. There are few places where the Neolithic developed indigenously. From these regions, the Neolithic gradually started spreading, at the beginning of the Holocene at about 12,000 years ago.

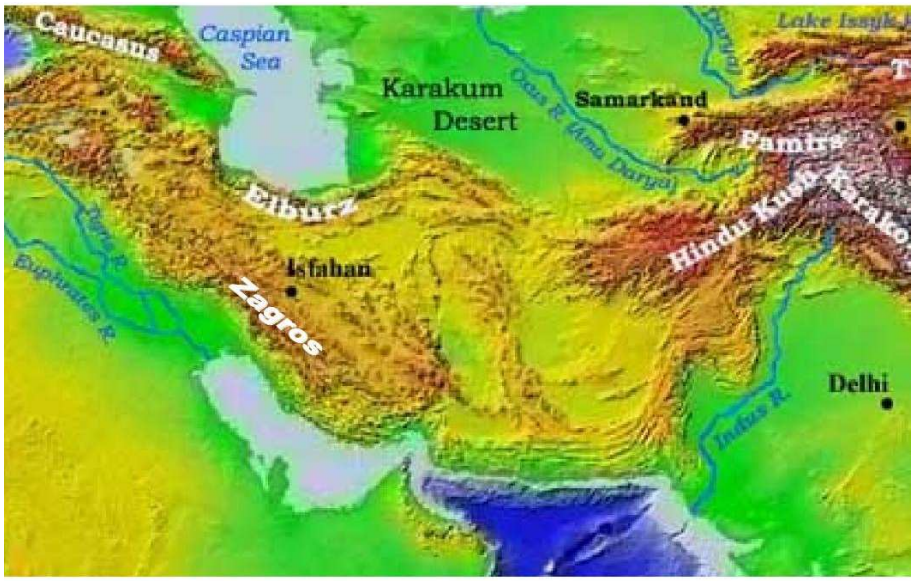


Figure 1.2: The physical map of the region under study showing the important geographical features.

The Fertile Crescent in the Near East, called 'the cradle of civilisation' [38], is the oldest of such sources. The first analysis of the absolute dates obtained for the Neolithic in Europe revealed a linear relationship between the Neolithic arrival dates and distances from a conventional source of Jericho in the Near East, and gave the Neolithic spread speed of 1 km/yr [16].

As mentioned earlier, the Neolithic traits were not all adopted at the same time but nevertheless were closely linked. Locally, the whole Neolithic package was gradually adopted over time. Early farming is an important trait in this Neolithic bundle, that has inbuilt sedentism. Vavilov's detailed study of the modern-day vegetation across the world shows localised regions (Vavilov's centers of diversity), that were original centers for domestication of plants [135]. His theory on the centers of origin of cultivated plants states that "*plants were not domesticated somewhere in the world at random but there are regions where the domestication started.*" Even now a variety of wild relatives of the domesticated crops are found in the Vavilov centres. This theory of independent crop origins implies the spatio-temporal initiation of incipient farming and thus effectively, possible source regions for the Neolithic origination. Figure 1.3, shows these source regions.

Figure 1.3: Centres of plant domestication as suggested by Vavilov's study of modern-day vegetation (http://giovanni-1.vn.gp/vavilov-nikolai.html).

1.2.1 Mechanisms of the Neolithic spread

The archaeological dates (absolute and relative) available for the Neolithic facilitate the calculation of the speed of the Neolithic spread. But these data alone cannot give information about the mechanisms behind the Neolithic spread. There are two main ways in which the Neolithic propagation could occur. One is the mass movement of populations known as the 'demic expansion', and the other is the 'cultural transmission' that includes transfer of knowledge, ideas, and plants and animals, rather than a movement of population. In the demic expansion view, population pressure is considered a prime cause of the Neolithic movement [122], whereas the cultural transmission involves the adoption of cultural traits not necessarily associated with massive long-range travel of individuals [140].

Gordon Childe writes about the European Neolithic [38, Ch. 1], "*Whatever part Mesolithic folk may have formed in Neolithic populations, the flocks of sheep and the seeds of grain on which the new economy was based were not carried by wind or intertribal barter, but brought by actual immigrant shepherds and cultivators*". His clear mention of

farmers and herders moving to a region and taking the knowledge of farming and herding with them shows his preference for the demic expansion view.

Some studies attach greater significance to the indigenous adoption of agriculture, driven by contacts between invading farmers and local foragers [112, 140]. According to Salem, the rapid spread of Palestinian pottery is "a result of cultural transmission, cultural change and social interactions" [121].

In Greece some regions were occupied during the Neolithic period, without any earlier Mesolithic populations present in the area. Similarly, Bulgaria, southern Yugoslavia, and other parts of south-eastern Europe lack evidence of indigenous Mesolithic populations, and the only plausible explanation for occupation of those areas during the Neolithic period is the demic diffusion. It is thus clear that some demic expansion occurred during the Neolithic period.

An attempt has been made to explain the spread of language families in terms of demic expansions [4]. The spread of farming technology through Europe was analysed by Ammerman and Cavalli-Sforza using the idea of demic expansion [16].

Some scholars however have acknowledged the importance of both demic and cultural expansions in the mechanisms of Neolithic spread [17]. According to them, the mechanism of Neolithic spread was a suitable admixture of the two processes. Fort [61] has attempted to combine these two approaches in a single model. Mathematically both processes can be modelled with nearly the same equations involving the diffusion operator, though the parameters in each case will be different, with different interpretations.

1.2.2 Overview of previous studies

The Neolithic has been explored in great detail for Europe. One of the recent analyses of the Neolithic dispersal in Europe includes 735 sites [108] and a ^{14}C database for the European Neolithic contains about 640 dates for the earliest Neolithic alone [69]. There have been a number of studies to explore the European Neolithic propagation processes.

Edmonson (1961) discussed first the concept of Neolithic diffusion rate for a population. According to his model, if a technology is considered to have been invented at the center

of a circle and if people are uniformly distributed on the periphery of the circle, then the invention will reach every person at the same time. He assumed that the population is homogeneous and people behave identically. This is similar to the isotropic diffusion with traits diffusing at a constant rate through space. He studied six different traits including projectile points, ground stone axes, pottery, domesticated plants, domesticated animals and alphabetic writing. Figure 1.4 shows Edmonson's analysis on two main traits, pottery and copper usage. Along with calculation of the Neolithic diffusion rate of 1.15 miles/yr or 1.85 km/yr (which is nearly twice the current estimate), Edmonson also noted that the estimate lies within a single order of magnitude of current speed estimates. This is a remarkable estimate given the limited data available to him [54]. This is the first time Neolithic spread was studied as a diffusive phenomenon.

In 1965, Clark presented the first visualisation of the spatio-temporal trends of the Neolithic transition for Europe on a map [39]. The data were binned into relatively broad age ranges involving only three bins, yet it showed distinct trends (see Fig. 1.5). The oldest settlements are in the south-eastern region, over the time span of 5200 to 2800 BC (in and around the Fertile Crescent), and as time passes they progress in the north-western direction. Clark also suggested that the Neolithic was confined to an area of about 10 degrees latitude between Greece and Iran until about 5,200 BCE. This crude picture is still consistent with the radiocarbon records that are now available.

Ammerman and Cavalli-Sforza [16] calculated the rate of spread of early farming to be $U \sim 1\,\mathrm{km/yr}$ on average across Europe noting significant regional variations (e.g. along the Danube and Rhine corridors). In this analysis the interpretation of the Neolithic as 'early farming' is employed. The two terms (Neolithic and early farming) are then used interchangeably throughout the study. Thus, unlike Edmonson, the study leaves out many sites that exhibit Neolithic traits other than agriculture (e.g. pottery), and is based on a single trait (evidence for cereals). The study primarily focuses on Western Europe, which exhibited plenty of radiometric evidence for cereals. Two main highlights of this study are the use of the earliest arrival dates of the Neolithic at a site, and locating an effective source for the Neolithic spread. The effective source is merely a reference point rather than

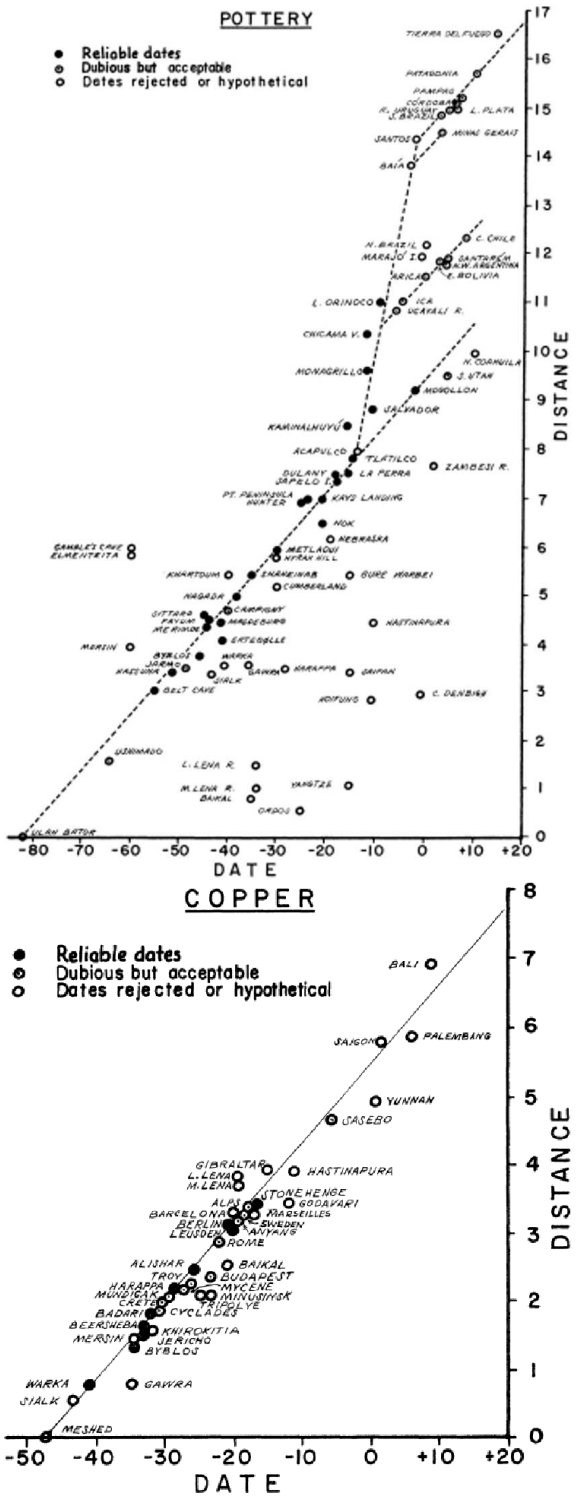


Figure 1.4: Edmonson's analysis of radiocarbon dates for pottery (above) and copper usage (below) [54].

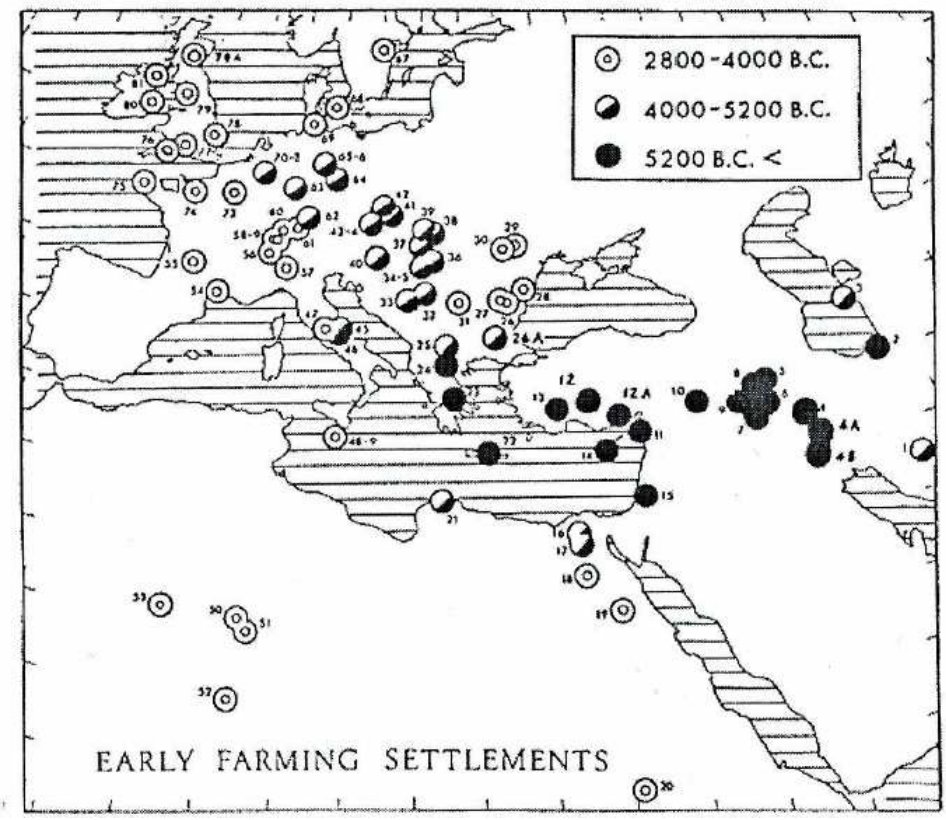


Figure 1.5: Spatio-temporal trends for Neolithic transition for Europe from [39]. The data is divided in to three time periods as shown.

an actual centre of the Neolithic. For finding this effective virtual source, the European domain was divided into grid cells and the distance to each site was calculated taking each of the grid cells as a possible source location. These distances and the Neolithic times were then analysed using linear regression, and Jericho in the western Fertile Crescent near the Mediterranean Sea was chosen as the best effective source. Figure 1.6 illustrates the progress of the Neolithic across Europe at a constant speed. A very important concept, demic diffusion (diffusion due to population growth and displacement), was first introduced by the authors in this study. An updated estimate of the propagation speed in the western Mediterranean coastal area has a significantly larger value (10–20 km/yr) [144]. A more recent study gives the Neolithic spread speed of 0.6–1.3 km/yr at the 95% confidence level [108] and confirms the average speed estimate obtained by Ammerman and Cavalli-Sforza.

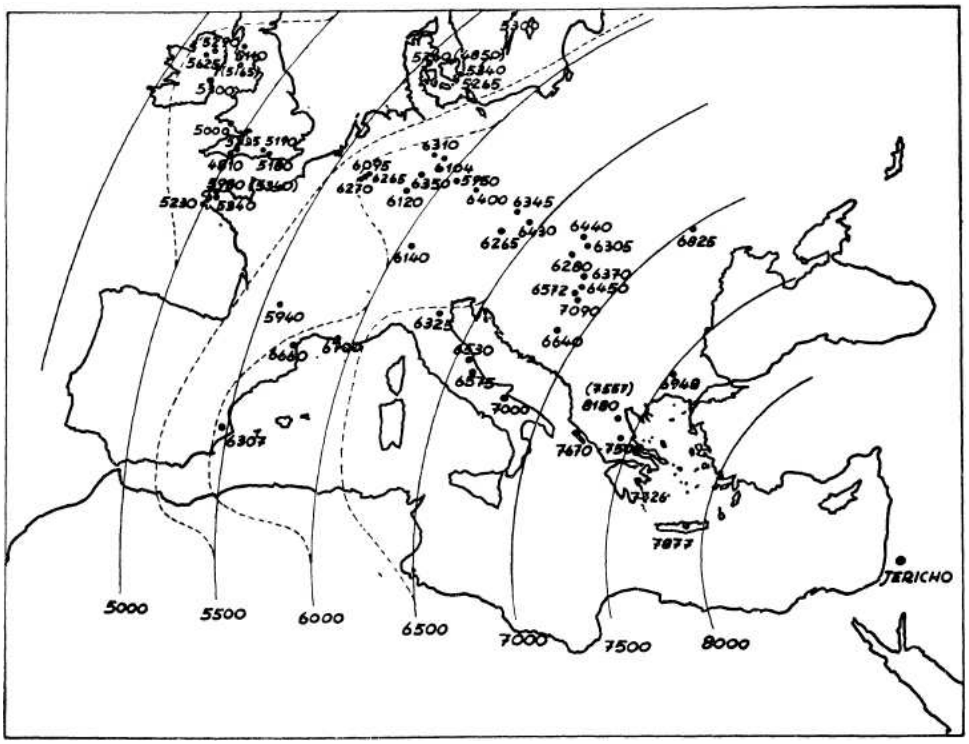


Figure 1.6: Spread of early farming across Europe from Jericho with a constant speed of 1 km/yr [16]. The arcs are plotted at every five hundred years.

The coarse-grained, large-scale pictures obtained from the aforementioned analyses apply at spatial scales of order hundreds of kilometers and time intervals of hundreds of years, without precluding significant variations in the rate and direction of the dispersal at smaller spatio-temporal scales [28, 58, 62]. In particular, a 'leap-frog' colonization (which could be especially important in coastal and riverine areas [46, 120, 144]), involving directed, relatively rapid movements over distances of order 100 km or less, is fully consistent with this global picture. Likewise, as evidenced by sufficiently realistic models [4, 22, 46, 106], such a spread does not need to be unidirectional or uniform (see [60] for a review).

In order to study the mechanisms of the Neolithic spread, more complex models have been employed since the 1970's. A mathematical model based on the growth and diffusion of population, the Fisher-Kolmogorov-Petrovskii-Piskunov (FKPP) model, is a significant and simple model in population dynamics. This model was used by Fisher in 1937 to

describe the wave of advance of an advantageous gene [59]. The Neolithic of Europe was first studied in detail with the FKPP model, also known as the 'wave of advance' model [17]. The model equation has the form

$$\frac{\partial N}{\partial t} = \gamma N \left(1 - \frac{N}{K}\right) + \nabla \cdot (\nu \nabla N). \tag{1.1}$$

Where N is the population density, γ is the growth rate, K is the carrying capacity and ν is the (generally, spatially varying) diffusion coefficient. The first term on the RHS is the logistic growth rate and the second term describes the population diffusion. See section 3.2.2 for more details.

The occurrence of agriculture was an important event that drastically changed the way of living for humans. Adoption of agriculture was not just a subsistence change, but it also brought about a cultural change and introduced sedentism. The process of conversion of hunter gatherers to farmers was at the heart of this change. Cohen has modelled this process using the wave of advance model for a single population [41]. The wave of advance model given by Eq. (1.1), was generalised to include food production-dependent carrying capacity, and population density-dependent birth and death rates. The model also included additional parameters (e.g. fertility, area utilised by populations). Front propagation and boundary formation are important features of this model based on the following equation:

$$\frac{\partial n}{\partial t} = \frac{D}{N_s} \nabla \cdot \ N_s \nabla n + \left(\frac{1}{\tau} \frac{n_1 + n}{n_0 + n}\right) n(1 - n) \tag{1.2}$$

Where N_s is the carrying capacity, $n = N/N_s$ is the relative population density, D is the diffusivity, $1/\tau$ is the initial growth rate, and n_0 and n_1 parameters dependent on the ratio of required to available food production and birth and death rates. The first term on the RHS describes the diffusion and the second term is the logistic growth factor with a population-dependent growth rate $\gamma = \tau^{-1}\,(n_1 + n)/(n_0 + n)$ (see [41] for details).

Another view of the Neolithic spread involves the occupation of the hunter-gatherer

lands by farmers. Aoki et al. [18] proposed a multi-population model involving three populations: hunter-gatherers, farmers and converts (hunter-gatherers converted to farming). These interacting populations are modelled with three coupled reaction-diffusion equations

$$\frac{\partial F}{\partial t} = r_f F\left[1 - \frac{(F+C)}{K}\right] + D\frac{\partial^2 F}{\partial x^2}, \tag{1.3}$$

$$\frac{\partial C}{\partial t} = r_c C\left[1 - \frac{(F+C)}{K}\right] + D\frac{\partial^2 C}{\partial x^2} + e(F+C)H, \tag{1.4}$$

$$\frac{\partial H}{\partial t} = r_h H\left[1 - \frac{H}{L}\right] + D\frac{\partial^2 H}{\partial x^2} - e(F+C)H. \tag{1.5}$$

Here D is the universal diffusion coefficient, K and L are the carrying capacities of the farmers and hunters respectively. F, H, C are the population densities of farmers, hunters and converts. The population growth rates for the three populations are r_f, r_h, r_c respectively. The parameter e represents the conversion rate of hunters to converts. The model was later extended to allow for environmental heterogeneities, and population-dependent parameters and was applied to the spread of the Neolithic in the Indian subcontinent [106].

A three population model was modified and combined with Cohen's model in [4]. The model showed that, for some parameter choices, spatial boundaries are formed between the farmers and the converts beyond which the farmers cannot travel into the converts' territory, so that farming spreads only with the spread of the converts into the hunter-gatherer domain. The model was applied to Europe, the Indian subcontinent and Africa, and an attempt was made to explain the genetic traits and the occurrence of the LBK pottery in Europe, the four language families in the Indian subcontinent, and the spread of the Bantu language family through central and southern Africa.

A wave of advance model was developed to study the Paleo-Indian dispersal into North America [130]. The model was implemented using the two-dimensional FKPP equation, and was numerically solved on a grid. This model studied the effects of the environment on the wave of advance by incorporating spatially varying carrying capacity and uses palaeo-environmental reconstructions as input. As a result of the variable carrying capacity, a patchy population distribution is achieved which cannot be reproduced with a model that

does not take the environmental variation into account. The model accurately reproduces the greatest occupation density in the eastern woodland habitats by using the spatially varying carrying capacity. The authors also note that the diffusivity of population must be a function of position and time, and suggest that the spread might have followed major river valleys, but do not include this factor in their model.

The role of waterways in case of the European Neolithic has been studied in [46]. The model studies the rapid advancement of the LBK pottery along the Danube–Rhine corridor and of the Impressed Ware tradition along the Mediterranean coastline. The authors employ anisotropic diffusion leading to an advection term in the FKPP equation,

$$\frac{\partial N}{\partial t} + \boldsymbol{V} \cdot \nabla N = \gamma N \left(1 - \frac{N}{K}\right) + \nabla \cdot (\nu \nabla N) - N \nabla \cdot \boldsymbol{V}. \tag{1.6}$$

The second term on the LHS is the advection term that gives the preferential movement in a particular direction, with the advection velocity $\boldsymbol{V}$. The model considers only one population and reproduces the accelerated spread of the LBK pottery and the agricultural colonisation of Britain and southern Scandinavia.

We describe in Chapter 6 two FKPP-based models employing spatially varying diffusivity for the region from the Fertile Crescent in the Near East to the Indus Valley in the Indian subcontinent.

1.3 The Neolithic of the Near East

The Fertile Crescent region in the Near East, bounded by the Mediterranean in the west and the Zagros Mountains in the east, is the oldest of the Neolithic centres of origin, showing the presence of Neolithic traits before 10,000 BCE. The Neolithic sites in the region have been extensively dated with absolute radiometric dating techniques. Most of the Neolithic sites in Iran also have absolute dates. In contrast, the chronologies in more than two thousand Indus Valley sites in the Indian subcontinent are based on relative dating. Therefore, correlating the Neolithic chronologies in the two regions has been difficult despite a strong evidence for their connections.

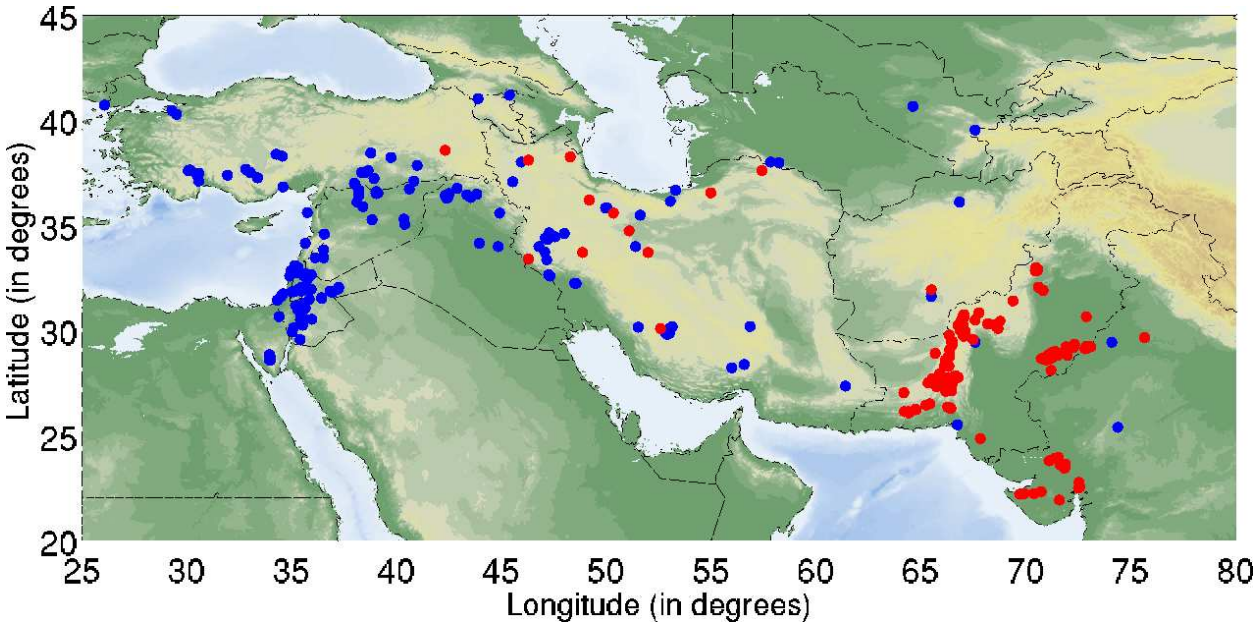


Figure 1.7: The Early Neolithic sites (10,000 BCE to 3,800 BCE) used in our analysis. Sites shown with blue symbols have ^{14}C dates available, and those in red are archaeologically dated. Modern national borders are shown dashed.

There are several lines of evidence that support the idea of a connection between the Neolithic in the Near East and that in the Indian subcontinent. The prehistoric site of Mehrgarh in Baluchistan (modern Pakistan) is the earliest Neolithic site in the north-west Indian subcontinent, dated as early as 8500 BCE [110]. Neolithic domesticated crops in Mehrgarh include more than 90% barley and a small amount of wheat. There is good evidence for the local domestication of barley and zebu cattle at Mehrgarh [42, 83]. However, the wheat varieties are suggested to be of Near-Eastern origin, as the modern distribution of wild varieties of wheat is limited to Northern Levant and Southern Turkey [66]. A detailed satellite study of a few archaeological sites in the Baluchistan and Khybar Pakhtunkhwa regions also suggests similarities in the early phases of farming with sites in Western Asia [107]. Pottery prepared by sequential slab construction, circular fire pits filled with burnt pebbles, and large granaries are common to both Mehrgarh and many Mesopotamian sites [71]. The postures of the skeletal remains in graves at Mehrgarh bear strong resemblance to those at Ali Kosh in the Zagros Mountains of southern Iran [83]. Clay figurines found in Mehrgarh resemble those discovered at Zaghe on the Qazvin plain

south of the Elburz range in Iran (the 7th millennium BCE) and Jeitun in Turkmenistan (the 6th millennium BCE) [82]. Strong arguments have been made for the Near-Eastern origin of some domesticated plants and herd animals at Jeitun in Turkmenistan (pp. 225–227 in [75]).

The Near East is separated from the Indus Valley by the arid plateaus, ridges and deserts of Iran and Afghanistan, where rainfall agriculture is possible only in the foothills and *cul-de-sac* valleys [78]. Nevertheless, this area was not an insurmountable obstacle for the dispersal of the Neolithic. The route south of the Caspian sea is a part of the Silk Road, some sections of which were in use from at least 3,000 BCE, connecting Badakhshan (north-eastern Afghanistan and south-eastern Tajikistan) with Western Asia, Egypt and India [92]. Similarly, the section from Badakhshan to the Mesopotamian plains (the Great Khorasan Road) was apparently functioning by 4,000 BCE and numerous prehistoric sites are located along it, whose assemblages are dominated by the Cheshmeh-Ali (Tehran Plain) ceramic technology, forms and designs [78]. Striking similarities in figurines and pottery styles, and mud-brick shapes, between widely separated early Neolithic sites in the Zagros Mountains of north-western Iran (Jarmo and Sarab), the Deh Luran Plain in southwestern Iran (Tappeh Ali Kosh and Chogha Sefid), Susiana (Chogha Bonut and Chogha Mish), the Iranian Central Plateau (Tappeh-Sang-e Chakhmaq), and Turkmenistan (Jeitun) suggest a common incipient culture [12]. The Neolithic dispersal across South Asia plausibly involved population migration ([48] and [75], pp. 231–233). This possibility is also supported by Y-chromosome and mtDNA analyses [113, 114].

The existence of these material and cultural links between the Neolithic sites in the Near East and those in the Indian subcontinent is the motivation behind this study. We explore this region in greater detail and try to link the two chronologies. We compile a database using the absolute and the relative dates in this region and analyse these data. Figure 1.7 shows the locations of the archaeological sites in our dataset.

1.4 Thesis Outline

Chapter 2: The qualitative and the quantitative dating techniques used in archaeology are discussed in this chapter. The qualitative dating technique involves the relative or indirect dating that is based on comparisons of material cultures between nearby regions. The technique gives a broad occupation period of a deposition layer at an excavated site. Along with a discussion of various quantitative (direct) dating techniques, which provide absolute dates, we provide the science behind the first and the most common absolute dating technique, radiocarbon dating.

Chapter 3: There are various mathematical methods used in modelling population dynamics systems. We give here an overview of cellular automata (CA), microsimulation models (MSM), agent-based models and network models. We also describe the traditional mathematical methods that are used extensively in the present study: linear regression, and population dynamics using differential equations.

Chapter 4: The area from the Fertile Crescent (near the Mediterranean Sea) in the west to the Indian subcontinent in the east (latitude: 20–45°N longitude: 25–80°E) is our study domain (see Fig. 1.7). We combine the absolute and the relative dates (by assigning suitable uncertainties to the latter) and analyse the composite dataset by introducing two novel statistical methods. We find the average speed of the Neolithic, from the conventional centre of 'Gesher' near the Mediterranean to the Indus Valley to be $U = 0.65 \pm 0.1$ km/yr.

Chapter 5: The geographically diverse terrain of the region of study contains productive lands (e.g. the Fertile Crescent in West Asia), as well as arid and barren lands (e.g. the Dasht-e Kavir and Dasht-e Lut in Iran). With such an inhomogeneous terrain it is not possible to have a uniform Neolithic spread throughout the region. We hence investigate this further and find that there are two likely routes for the Neolithic propagation, one via north Zagros, northern Iran (possibly a precursor to the Silk Road), and the other via southern Iran through the Fars and Kerman provinces. The two routes have distinctly different spread speeds; the Neolithic spread via the northern route had a speed $U_N = 1.9$

km/yr, and via the southern route at $U_S = 0.66$km/yr.

Chapter 6: The FKPP equation is the basis of a deterministic continuous mathematical model used in various subjects, including population dynamics, to study the growth and movement of a population. The diverse terrain of our region of interest has different ecological palaeovegetation zones. Utilising this information about these ecological zones, we present a mathematical model based on the FKPP equation that incorporates variable diffusion coefficients. The results obtained from preliminary runs of the model corroborate our conclusions from chapters 4 and 5. Another model based on [4] is given in a separate section that studies competition between two crop complexes in the Indian subcontinent. We also address the pre-Indo-Aryan language distribution in northern India.

Appendices A and B: The radiometric and archaeological dates used in the analysis are given in Appendix A, and the histograms of the Gaussian temporal clusters for sites in our combined database are given in Appendix B.

Chapter 2

Dating in Archaeology

Archaeology is a discipline which studies human history and prehistory with the help of cultural records and artefacts left behind by ancient people and unearthed during excavations. In order to understand the temporal development of an excavated site, it is essential that the events that took place at that site be arranged in an appropriate chronological order.

History is the study of the human past after the invention of writing systems. Often the written sources (e.g. inscriptions, manuscripts, administrative and legal records, numismatics, etc.) involve a direct mention of, or an indirect pointer to, the date when the artefact was created. Existence of such markers makes it easier to sequentially arrange the respective historical events. Records of astronomical events (e.g. records of solar and lunar eclipses, supernovae, etc.) may also help to determine the sequence of events.

The study of prehistory involves the investigation of time periods before the invention of scripts. In this, there are no written records to comfortably and confidently establish the chronologies for the prehistoric societies. Scholars of prehistoric archaeology have thus had to develop dating methods that would be helpful in sequencing events using excavated cultural artefacts such as lithic tools, figurines, pottery sherds, etc.

In order to establish a cultural sequence or stratigraphy of the occupation at a site, both

qualitative and quantitative types of dating techniques are used. Qualitative dating techniques are based on context-dependent comparisons of material cultures at different sites. Quantitative or absolute dating involves chemical analyses of the material remains found in the layers or strata of the archaeological dig. The information compiled in this Chapter is mainly from the books: [9, 84, 136].

2.1 Relative (Comparative) Dating

Before the 1950's i.e. prior to the invention of the quantitative dating techniques like radiocarbon dating, the task of sequencing archaeological events relied upon intelligent guesswork based on a body of archaeological information. This relative dating does not necessarily give the absolute ages of events in a region, but it does establish the relative chronological order. It remains useful especially when the deposition layer lacks any datable organic material. Relative dating in archaeology is similar to the relative dating in geology and employs the following principles [84]:

1. *The principle of original horizontality:* Sedimentation or soil deposits are in the form of horizontal layers, unless disturbed.

2. *The principle of superposition:* The soil deposition layer near the surface is newer than the layers away from the surface, and thus (usually) the material remains found in the layer near the surface are newer than the ones found in the deeper layers.

3. *The principle of lateral continuity:* If similarity is observed in the sediment deposits that are separated by deformities like a depression or a valley or other erosional features, then the separate features could have been continuous at earlier times. (This principle allows for a comparison between two layers at different locations).

4. *The principles of cross cutting and inclusions:* Any feature that cuts through the deposition layers must be younger than the layers, and any objects that are buried in a layer must be older than when the layer was formed.

Archaeological dating is strongly context dependent. In addition to the aforementioned principles from geological studies, archaeologists also use the following techniques to establish chronologies in a region.

Stratigraphic Dating is the oldest indirect dating technique in an archaeologist's arsenal that has survived the test of time and is still widely used, especially to date those layers where no suitable organic material can be located for absolute dating. Stratigraphic dating provides the age of archaeological materials by their association with the geological deposits. It uses the aforementioned principle of superposition. The deeper the deposition layer from the surface, the older it is, unless disturbed. The soil deposits are differentiated on the basis of variations in colour, materials and texture. The objects and artefacts found in a single layer usually are contemporaneous and those found in layers below and above are older and younger, respectively. Thus, a relative chronological sequence can be formed by carefully noting the stratigraphical information. The strata at a location are carefully drawn and recorded and then the profile from one site is compared with the profile from surrounding sites in order to establish a cultural sequence for the region. The obvious pitfall is artefacts are sometimes trapped in depressions, natural (e.g. deep pits, depressions) or man-made (e.g. wells). These objects would be dated as older than they actually should be.

Another indirect dating method is the **seriation technique**. Human behaviour changes with time; this change is reflected in the artefacts. The degree of change depends on how much the human behaviour changes. Seriation is the method that arranges artefacts from different sites in a chronological sequence dependent on the stylistic variations. It can be used for dating pottery and lithic tools along with other artefacts.

The records of changes in the technology of the material artefacts play a very important role in determining the relative ages as well. The three age system (Stone Age, Bronze Age, Iron Age), introduced by the Danish archaeologist Christian Thomsen, is an example of age determination based on the changes in technology. A similar example is the further division of the Stone Age into Palaeolithic, Mesolithic and Neolithic based on observations

of changes in the stone tool technology.

Along with human-made artefacts, biological markers are also useful in relative age determination. The Fossil or Faunal succession, palynology (the study of fossilised pollens in lake deposits or sediments), floral time charts are a few relevant techniques.

2.1.1 Relative dating pitfalls

Relative dating relies upon a qualitative comparison of the material culture of an archaeological site with the material cultures of the sites in the nearby region. Along with the stratigraphy, the chronology at a site depends on layer-by-layer comparisons between the sites. If a layer is dated incorrectly (e.g. because of the discovery of a younger artefact in an older layer or vice versa), that reflects in the dating of that site as well as the other nearby sites, and results in the overall chronology of the region being established incorrectly. The seriation technique uses reliably identified change in the typology of the artefacts found in different layers. Often fashions change, and instead of evolving, an artefact may very well be completely replaced by another artefact. The artefact distribution thus might not be unimodal. The other problem is the possibility that an artefact presents its design for a long time, without any modifications (e.g. knives in early medieval Europe apparently do not have any chronological variation). In all such cases, creating a reliable typology is difficult. Establishing uncertainties in relative dating is yet another issue. Both the dating of artefacts and the associated uncertainties depend on the subjective judgement.

2.2 Quantitative Dating

The advent of radiocarbon dating by Libby in the late 1940's, has had prominent impact on many areas and especially on archaeology. Radiocarbon (^{14}C) dating was the first of the plethora of absolute dating techniques. Often the dates suggested by the relative dating techniques have been revised to affect both historic and prehistoric archaeologies.

Stonehenge, dated with the relative dating technique, was believed to be occupied from 3000 BCE. Recent ^{14}C analysis showed that there was a semi-permanent settlement at Stonehenge from about 7,500 BCE and continuous occupation through the later millennia [9].

2.2.1 Radioactive isotopes

The process of age determination of a material by detecting the amount of radioactivity in it is known as the radiometric dating. Atomic nuclei consist of protons and neutrons. Atomic number, the characteristic of an element, is the number of protons in the nucleus and atomic mass is the total number of nucleons (protons and neutrons). Atoms with nuclei that have the same atomic number (and thus same chemical properties) but different atomic masses are called isotopes of the same element. Many elements have more than one isotope. If the nucleus of an isotope is unstable, it is radioactive. Hydrogen, for example, has three isotopes with atomic mass numbers 1, 2 and 3. Out of these three, the first two are stable and only the third is unstable and is thus radioactive. The unstable nucleus of the radioactive isotope decays by ejecting energetic particles (positively charged α particles, negatively charged β particles and/or gamma radiation) in order to gain stability. During the decay, the element may be converted into one or more lighter element(s) by nuclear fission.

2.2.2 Radiocarbon (^{14}C) dating

Carbon is the principal component of any organic material. Hence any material with organic content can be dated with the radiocarbon dating method. Carbon has three naturally occurring isotopes, the abundant isotope ^{12}C, rare ^{13}C and the rarest ^{14}C. Out of these three, the ^{12}C and ^{13}C are stable and only ^{14}C is radioactive.

Radioactive carbon is constantly produced in the Earth's atmosphere, in the lower stratosphere and upper troposphere. Cosmic rays generate neutrons that interact with nitrogen nuclei in the atmosphere to produce radiocarbon: $\mathrm{n} + {}^{14}_{7}\mathrm{N} \longrightarrow {}^{14}_{6}\mathrm{C} + \mathrm{p}$.

The radiocarbon is chemically indistinguishable from the two non-radioactive isotopes and rapidly combines with oxygen to form carbon dioxide CO_2. Radiocarbon in the form of CO_2 is absorbed by plants via photosynthesis and by other living life forms through diet and thus enters the biological cycle. Radiocarbon decays by emission of an electron and an antineutrino and is converted back into nitrogen: ${}^{14}_{6}\mathrm{C} \longrightarrow {}^{14}_{7}\mathrm{N} + \mathrm{e}^- + \overline{\nu}_e$.

Due to radioactive decay, the number of radiocarbon atoms in the body of a living organism reduces with time. But the continuous production of radiocarbon in the atmosphere replenishes it continuously. Thus a living organism has the same fraction of radiocarbon as the biosphere. Once the organism is dead, the exchange of the radiocarbon with the surroundings stops and the number of radiocarbon atoms in the organism can only be reduced via radioactive decay.

The reduction rate for a radioactive material is governed by the differential equation:

$$\frac{dN}{dt} = -\lambda\, N, \tag{2.1}$$

Where, N is the number of radioactive carbon atoms and λ is the decay constant, which leads to

$$N = N_0\, e^{-\lambda\, t}, \tag{2.2}$$

where, N_0 is the initial number of the radiocarbon atoms present when the organism was alive and t is the time that has passed since the death of the organism, i.e. the age of the organism,

$$t = \tau\, \ln \frac{N}{N_0}, \qquad \tau = \frac{1}{\lambda}, \qquad t_{1/2} = \tau\, \ln(2) \tag{2.3}$$

with τ the mean life time of radiocarbon (8267 years) and $t_{1/2}$ is the half-life time of radiocarbon (5730 years).

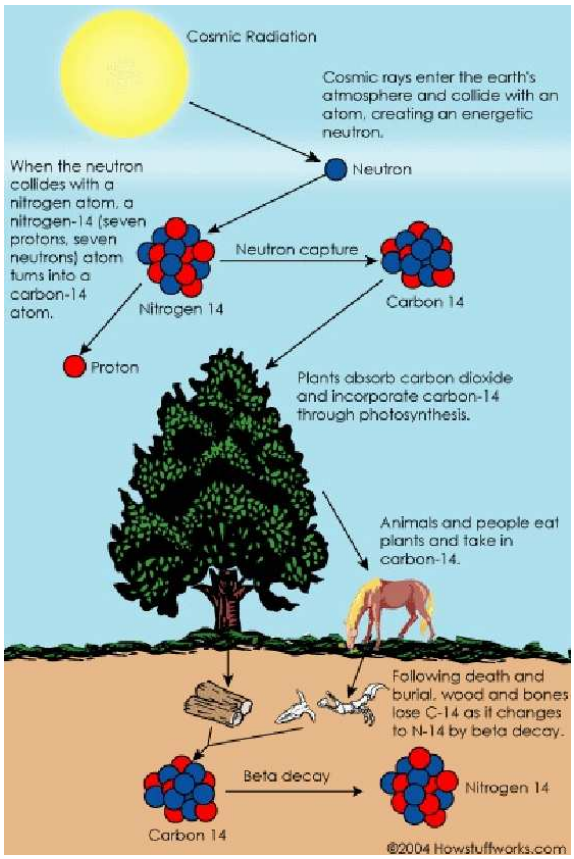


Figure 2.1: The diagram shows the working of the carbon-14 cycle [80].

Radiocarbon measurement

In early days of radiocarbon dating, the amount of radiocarbon in a sample was measured by detecting the decay of individual ^{14}C atoms. This process of measuring the number of beta decays per unit time per unit mass of the sample is known as beta counting. Libby first used his own version of the Geiger counter for the beta counting, but this was later replaced by advanced counting instruments like gas proportional counters and liquid scintillation counters. An advanced method, accelerator mass spectrometry (AMS), introduced in the 70's, measures the number of ^{14}C and ^{12}C atoms directly instead of the activity of the sample. The AMS dating technique has allowed dating of very small quantities of material (milligrams) like a single charred grain, which is not possible with the beta counting. The AMS dating is also much faster than the beta counting method. It takes merely a few hours for the AMS method while beta counting could take several

days.

Uncertainties in the radiocarbon measurements

There are various possible sources for errors, random and systematic, in quantifying the radioactive decay. This is a random process, i.e. the specific time of decay of a radioactive nuclide cannot be determined with certainty (e.g. ten repeated measurements of the same bulk cellulose sample gave date values between 4442–4542 BP with different error estimates [123]). Thus the process has inbuilt uncertainty and the random component of uncertainty can be reduced by repeating measurements and taking averages and the average provides a better (less uncertain) estimate of the true date of the sample. The systematic component of uncertainty usually occurs because of the instruments used for measurements (the instrumentation error); due to either instrument limitations and/or incorrect handling. It is possible that different laboratories measure different values of the mean date and uncertainty for the same sample, dependent on the instruments used. Thus the same sample can give different mean dates with different uncertainties. A better technique (e.g. AMS dating is better than beta counting) can reduce the systematic error in measurements considerably.

Every laboratory gives a 'quoted error' with the measured sample date. This quoted error is calculated as

$$\sigma(t) = \frac{1}{\lambda}\sqrt{\frac{\sigma(A_t)^2}{{A_t}^2} + \frac{\sigma(A_0)^2}{A_0^2}},$$

where t is sample age, λ is the decay constant, A_t is the activity of the sample material, and A_0 is the current equilibrium living activity [123]. When the beta decay is assumed to be Poisson noise, the quoted error is simply given by

$$\sigma = N^{\frac{1}{2}}.$$

This statistical uncertainty is the basic and the minimum possible error in a radiocarbon measurement. Apart from the random and laboratory errors, there are other possible

sources of errors (e.g. contamination by younger or older radiocarbon, the reservoir effect). To accommodate these errors, an error multiplier is used and as a result the quoted uncertainty is often larger than the standard error.

Many archaeological sites have a series of radiocarbon dates, and associated with each there is a quoted laboratory error. This laboratory error thus characterises the accuracy of the measurement of a sample radioactivity and not the uncertainty of the true age of the archaeological site. To represent the true uncertainty, along with counting and calibration, the errors introduced by disturbed stratigraphy and reservoir effects, etc. also should be taken into account [51]. In this study, to determine the uncertainties in the oldest continuous Neolithic occupation at a site we employ clustering analysis using Gaussian mixture modelling (see Chapter 4 for details).

Calibration

Around 1955, it was pointed out by Libby that the ratio of ^{14}C to ^{12}C does not remain constant but changes over the course of time, and there is a difference between the radiocarbon age and the calendar age of the sampled material. Various causes, of natural and human origin, alter this ratio significantly. Changes in the geomagnetic field and the geomagnetic polar reversals affect the deflection of the flux of cosmic rays from the earth's atmosphere, resulting in varied ^{14}C levels. Low solar wind activity has lead to weak magnetic activity in the past, leading to increase in the cosmic ray flux that in turn increased the production of the ^{14}C thus increasing the percentage of the radiocarbon in the atmosphere. Similarly a high solar wind activity would cause a decrease in the radiocarbon abundance. The proportion of the radiocarbon in the atmosphere also increases because of nuclear testing. Natural reservoir effects like ocean mixing, and human-origin effects like the burning of fossil fuel (the Suess effect), change the percentage of the principle carbon isotope (^{12}C) in the atmosphere.

Because of these variations in the fraction of the radiocarbon in the atmosphere, a correction factor needs to be applied to obtain the true calendar age of the sampled material.

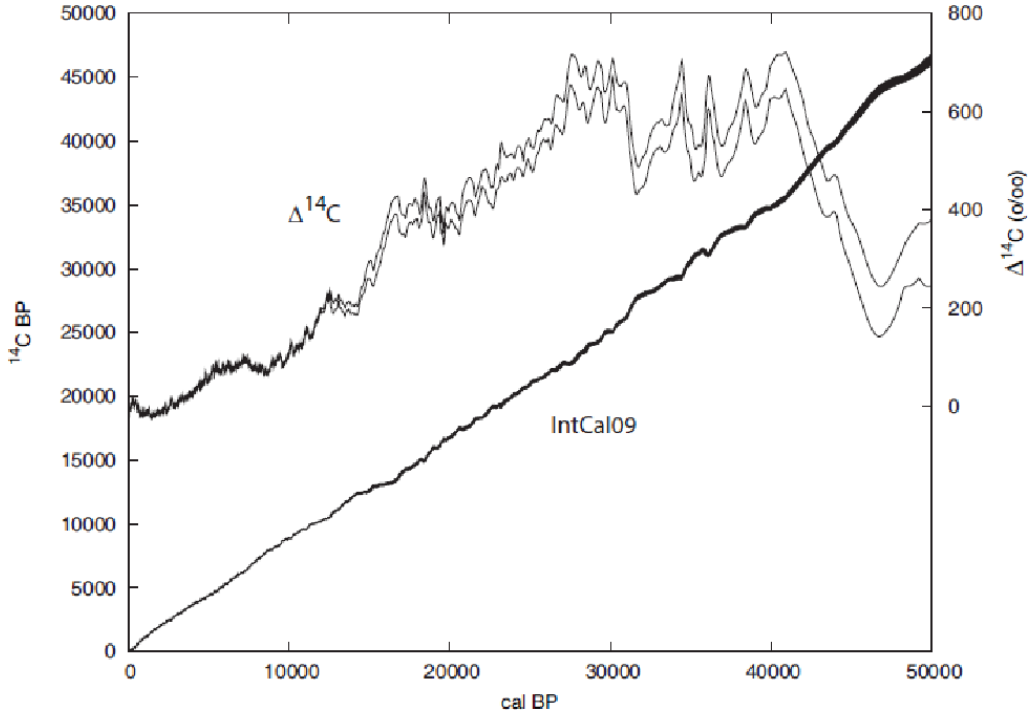


Figure 2.2: The Intcal09 calibration curve along with the Δ^{14}C curve [116]. The ^{14}C age of a sample, obtained from the laboratory, is read from the vertical axis and the corresponding calibrated age is then read from the horizontal axis. The Δ^{14}C curve gives the fractionation-corrected depletion in the ^{14}C age.

The first of such correction curves, the calibration curve, using the tree ring sequences of the bristlecone pine, was published by Hans Suess in 1967. Since then other calibration curves have been published using different methods. Out of these, the IntCal curve is the recent leading series started in 1998 with IntCal98, the current one being IntCal13.

Several software packages like BCal, CalPal, CALIB and OxCal are available for obtaining the calendar age from the radiocarbon age. In this study OxCal (versions 4.0 and later) was used along with the calibration curve IntCal09. Figure 2.2 shows this curve: the radiocarbon date is on the vertical axis and the corresponding calendar date is read from the horizontal axis.

With the radiocarbon dating, it is possible to reliably date organic samples that are younger than 40,000 yr BP. For older samples, other methods are used as described below.

2.2.3 Radiometric dating: other isotopes

Radiocarbon dating is the first quantitative radiometric dating technique that facilitated the absolute age estimation of organic material. Although very useful and popular, it has its limitations: viz. it can only date material that contains carbon, and the time span of dating is limited to at the most 70,000 yr BP and later. Organic material older than this cannot be dated reliably with this method. Recently, many other radiometric dating methods have become available for dating older organic and non-organic materials. A few of the important techniques are discussed below.

Potassium-argon (K-Ar) dating is a method commonly used to date material samples from mica, feldspar, hornblende etc. that are believed to be older than 20,000 yr BP. This technique is based on the decay of radioactive potassium (^{40}K) into stable argon (^{40}Ar) that is trapped within the rocks. With this method samples from a very wide temporal range can be dated reliably. The technique however, requires samples that have not been heated to or beyond 125°C, as argon escapes the rock beyond this temperature. Another radiometric dating method, the radium-strontium dating, can be used to cross-check the dates obtained by the K-Ar dating because strontium can withstand mild heating.

Ionium-thorium (Io-Th) dating is yet another radiometric technique, that is primarily used to date deep-sea sediments. Thorium (Th) is one of the by-products of uranium (^{238}U) decay. The two radioactive isotopes of thorium ^{232}Th and ^{230}Th (earlier known as ionium) have half-lives of 14.5 billion and 75,200 years respectively. ^{238}U is soluble in water but the decay product, thorium, is insoluble and hence is deposited in the sediments. The ratio of ^{232}Th and ^{230}Th can thus be used for the sediment age determination.

Uranium-lead (U-Pb) dating is a less common dating technique. ^{238}U follows two parallel decay paths giving rise to various radiometric dating techniques. The technique can be used to date very old samples, even rocks that date to the Earth's formation. In fact, American geochemist C. Patterson gave one of the first correct estimates of the Earth's age using this method in mid-1950's.

Fission track dating can be used to date minerals and glass that contain uranium. This is a simple but robust method. ^{238}U undergoes spontaneous fission and leaves fission tracks inside glass or mineral crystal. Instead of measuring the fission product isotopes, the method counts the number of fission tracks to estimate the radioactivity and, thus, the age.

2.2.4 Other quantitative dating techniques

Ice cores, coral reefs, stalagmites, lake deposits, fossilised pollen and spores, as well as long living trees like bristle cone pine, etc. store information about the past climate events.

Dendrochronology

Dendrochronology, the method of dating by analysis of tree ring patterns, was developed by astronomer A. E. Douglass in early nineteenth century. Trees like the bristle cone pine are the oldest living species and have a slow growth rate. These trees can live for more than 5,000 years and each year produce one ring near the bark. Thus by counting the rings formed, it is possible to accurately obtain a tree's age (see Fig. 2.3). Trees in the same region have similar growth patterns, and hence it is possible to establish tree ring sequences based on different trees (felled and living) in the same region. Dendrochronology is useful in the dating of wood used in constructing houses. With the help of the worldwide tree dataset, along with the data from the marine sediments, a calibration curve, IntCal04, has been constructed with calibrated dates up to 26,000 yr BP. The recent IntCal09 and IntCal13 curves supersede the IntCal04 curve and with the IntCal13 curve, the calibrated dates now can be reliably determined up to about 50,000 yr BP.

Electromagnetic dating

Archaeomagnetic (palaeomagnetic) dating is the most popular electromagnetic dating technique. Archaeomagnetism is the study of the remnant magnetic field in archaeo-

Figure 2.3: Left: section of felled tree in a Cumbrian forest. Right: Dating of tree rings [15].

logical samples. The remnant magnetic field is preserved in archaeological material that contains magnetite and is cooled below the Curie temperature. Many settlements used pottery and bricks fired in kilns. By analysing these palaeomagnetic fields, the age of the artefact can be calculated. The technique is used to date sites within the last 10,000 years.

Electron spin resonance dating is used to date quartz, calcites, enamels, shells, flints and fossils. Imperfections in crystalline structures capture free electrons and holes, and act as natural dosimeters. This property is used in calculating their ages. The dating range is between a few thousand to a couple of million years. This technique is most commonly applied to date tooth enamels.

Thermoluminescence (TL) dating is a very common dating technique, and objects several hundred to several thousand years old can be dated with this method. When the sample object is heated to incandescent temperatures, the electrons trapped in the lattice imperfections gain enough energy to escape the traps, and once they recombine with holes they emit photons of the trapping energy. By measurement of the thermoluminescence, the age of the object can be calculated.

Chemical dating

Obsidian hydration dating is a geochemical method used in dating artefacts made with obsidian. Obsidian, a volcanic glass, was used in making tools, arrowheads etc. When exposed to air, it absorbs water at a well-defined rate. From observing the thickness of the narrow band formed due to the hydration, and with the rate of hydration, the age of the object can be calculated.

Amino acid racemization is based on the fact that all biological tissues contain amino-acids. The amino-acids (except glycine) have two different orientations, dextrorotary (right) and laevorotary (left), that are mirror images of each other. In a living organism, all the amino acids are in the left or laevorotary orientation and the ratio of right to left orientations is close to zero. With the death of the organism the ratio of right to left tends to unity at a constant rate, in a process known as racemization. Measurement of this ratio allows determination of the time passed since the death of the tissue.

Bone nitrogen dating is a relative dating technique. When bones are buried in soil, there is a gradual reduction in their nitrogen content at a constant rate. In the same surrounding conditions, nitrogen is reduced at the same rate in all bones. By measuring the nitrogen content in different bones, it can be determined whether or not the bones were buried at the same time. A relative chronology of the bones can thus be constructed.

Chapter 3

Modelling

A model is a representation of a system. An ideal model would capture all the features of the system under consideration. In practice though, a model is considered good if it captures the most important features of the system. Computer modelling and simulation techniques find use in widely varying fields throughout the natural sciences and social sciences. These virtual laboratories have become handy, important, cost effective and indispensable modern day tools in understanding and visualising real world data and phenomena. There are numerous types of physical representational models and conceptual models that can be used to describe the real world.

3.1 Non-Traditional Mathematical Models

Unlike the top-down mathematical modelling methods where properties of a system are modelled using a set of fundamental equations, non-traditional mathematical approaches, such as cellular automata, microsimulations and agent-based models, rely on the emergence of structures with (usually simple) rules employed at the level of the smallest possible unit in a system.

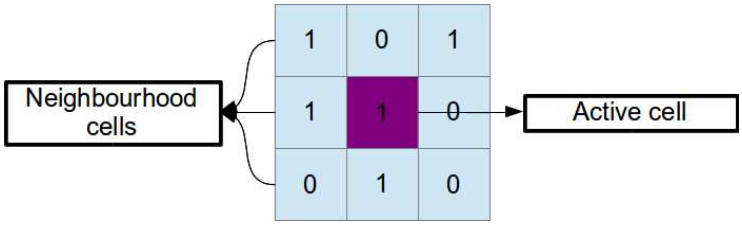


Figure 3.1: A cellular automata grid cell in purple with its neighbourhood (the Moore neighbourhood) in blue. Each cell can have one of the two states, 0 or 1, and the number inside the cell shows whether the cell is in state 0 or 1.

3.1.1 Cellular automata and microsimulations

The concept of cellular automata (CA) was developed in Los Alamos National Laboratory by Stanislaw Ulam and John von Neumann in the 1940s. CA is an individual-based modelling approach that employs a bottom-up strategy. It is a discrete dynamic system that is characterised by local neighbourhood interactions [43]. A CA typically employs a regular grid of cells (Fig. 3.1). Each cell can be in a finite number of states and at a given time a cell can exist in only one of them depending on the set of rules provided. A neighbourhood is defined for every cell. There is no unique way of defining a neighbourhood, but usually the cells that are in close proximity are included. The von Neumann neighbourhood and the Moore neighbourhood are two common prescriptions. The locations of a cell and its neighbourhood remain stationary over time. The state that a cell can acquire at a given time depends on the states of the neighbourhood cells. Thus, when evolved in time, each cell changes its state according to predefined rules that involve the states of neighbourhood cells; a pattern emerges when CA is allowed to evolve for sufficient time.

An useful description of the CA can be found in [124, Ch. 7]. The 'Game of Life', developed by British mathematician John Conway in 1970, is an excellent example of a cellular automata that requires no input other than the initial conditions [5]. The system then evolves on its own through time showing the construction and destruction of cells on the grid. An interactive demonstration of the CA can be downloaded from [142].

Just like the CA, **microsimulation modelling** (MSM) is an individual-based modelling

methodology, but in the MSM the behaviour of a system is not a consequence of interactions between lower-level components. It instead studies how a predefined policy would affect the evolution of the individual lower level constituents [43].

3.1.2 Agent-based modelling

The Agent-based modelling (ABM) is another individual-based modelling approach. An agent in agent-based modelling does not have a precise definition, but several features discussed below are common to agents in different systems [43, 115].

Agents are autonomous. They interact with other agents and exchange information in order to make independent decisions without need for any centralised control. Interaction with other agents does not affect the autonomy of an agent. **Agents can be heterogeneous.** A group of lower-level agents can be amalgamated to form a higher-level agent that can have its own characteristics (see Fig. 3.2). For example, biological cells are independent autonomous units with individual characteristic properties. When these cells combine to make up an individual (e.g. a human), that individual in itself is an autonomous entity with its own properties (e.g. age, height, weight in case of a human). **Agents are often goal-directed**. They can behave pro-actively in order to achieve a set goal. Agents can be made **perceptive** by providing prior knowledge. They can be constructed to be aware of agents other than themselves. Agents can also have the ability to make a **rational choice** and **share information** with other agents. Agents can **change over time** and different agents can have different life cycles possibly including birth and death. They can be made **stationary or mobile** in the simulation space of the model. Agents can also be designed to have **adaptive behaviour**, to have memory and to make choices based on the previous experiences.

Though an agent in ABM is similar to a cell in CA and a small area in MSM, its autonomy sets ABM apart from CA and MSM. An agent interacts with other agents and its surroundings, similar to a cell in the CA interacting with cells in its neighbourhood, but unlike in a CA it can make a decision independent of the neighbouring agents. An

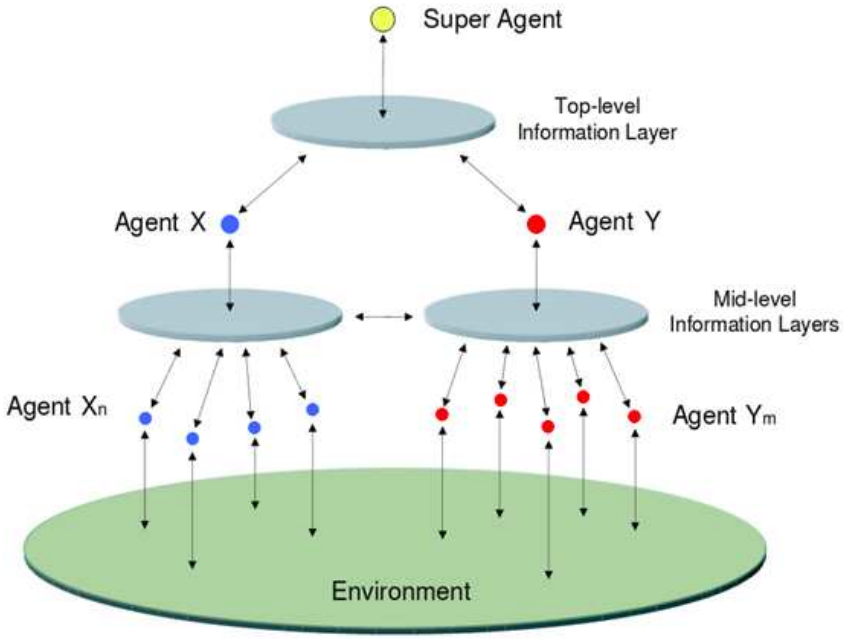


Figure 3.2: Heterogeneous, three level hierarchical structure of an ABM is shown in this figure. Two-way interactions are shown to exist between the lower level and upper level agents. (source: [6])

agent's interaction range and mobility are limited by its memory. Pattern formation is a consequence of the collective behaviour of agents in a system. Thus the independent decisions taken at the level of an individual affect the decisions of the collective as well.

An agent can represent any type of unit. Continuous approaches to modelling smooth out the fluctuations in a model but ABM does not. Agents are capable of exhibiting complex behaviour, learning and adapting to the situation. They may provide a way to test social theories that are difficult to describe by traditional mathematical models.

One of the disadvantages of the ABM approach is that it can be computationally expensive. The computation time depends on the degree of complexity of the model. The problem needs to be carefully designed with an appropriate abstraction level. (If the abstraction is too coarse, the model will be simplistic and if it is too detailed it becomes computationally taxing.) Guidelines to build a good ABM are provided in [3, 19]. The nature of the system being modelled is also sometimes a concern. Living beings exhibit complex

psychology. Their decision making is not always rational and is difficult to quantify, justify and incorporate in an ABM. Since ABM might involve varying degrees of accuracy, it is not always possible to quantify the results. ABM implementations are also very sensitive to initial conditions and small variations in them and/or interaction rules can change the model output drastically.

The structure of an object oriented programming (OOP) language facilitates construction of an ABM. Higher level languages like C++ and Java are frequently used for building ABM frameworks. NetLogo is a special language that has been designed for implementation of ABMs [132]. The tool kits in Java like Repast, SWARM and MASON also help make the complex coding easier.

ABMs are used in a variety of subject areas, e.g. biological sciences (models of infectious diseases [143], bacterial colony growth [91], spatial pattern formation and cooperative dilemmas [127]), history (alliance formation during the Second World War), economics, modern urban planning, share-market analyses (price variations in trading), politics (understanding political stability, voting behaviours in elections), archaeology (modelling ancient civilisations [20, 86]), geospatial research, etc. An interesting application includes analysing social networks of terrorist groups [101].

3.1.3 Network models

Networks are encountered in many subjects. Computer networks, the world-wide web, neural networks, traffic networks, demand, production and supply networks, trading networks, social networks (e.g. facebook, twitter) and biological networks are all examples of important networks. A network is considered to be efficient if it exchanges maximum (ideally infinite) information at minimum cost [29].

Figure 3.3 shows a general demand-supply network that aims to reduce the effective cost of transport through the network. The production source, receivers and intermediate entities are called nodes (circles in Fig. 3.3) while the links connecting them are called the 'arcs' of the network (directed arrows in Fig. 3.3). Between nodes 1 and 2, the arc arrow is

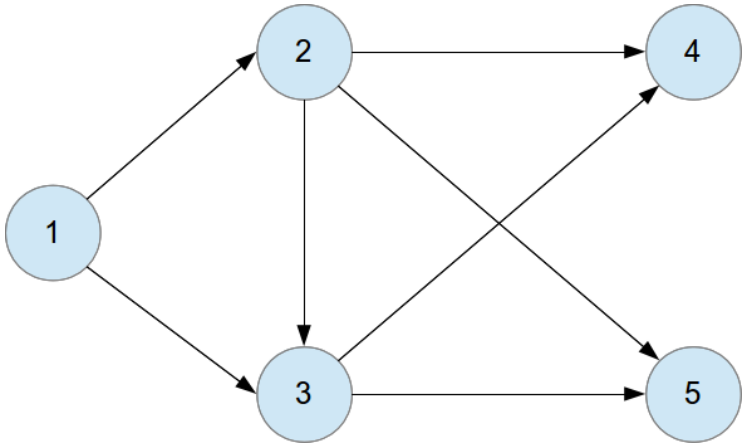


Figure 3.3: A network is shown with five nodes and transport links (arcs) between the nodes. Node 1 is the source node, nodes 4 and 5 are the receiver or sink nodes and nodes 2 and 3 are the intermediate nodes.

pointed from node 1 to node 2, thus implying that information can be sent from node 1 to node 2 but not vice versa. Thus node 1 is a source and node 2 is a receiver. A two-way information flow is shown by the arc arrows pointing towards both nodes. Thus from Fig. 3.3, it can be seen that node 1 acts as a source / production node for the receiver nodes 4 and 5, along with intermediate exchange nodes 2 and 3.

The objective of this general demand-supply network analysis is to find a cost-efficient way to send goods from the source node to the receiver nodes via intermediate nodes. In terms of the amount (x_{ij}) of goods sent from the source (node i) to the receiver (node j) via arc i–j, the flow balance equations at each node state the conservation of flow and are given by:

$$\begin{pmatrix}\text{Net supply}\\ \text{at a node}\end{pmatrix} = \begin{pmatrix}\text{Flow out of}\\ \text{a node}\end{pmatrix} - \begin{pmatrix}\text{Flow into}\\ \text{a node}\end{pmatrix}.$$

The minimum cost flow problem with n such nodes can be solved by minimizing the quantity

$$z = \sum_i \sum_j c_{ij} x_{ij},$$

subject to the flow balance condition:

$$\sum_j x_{ij} - \sum_k x_{ki} = b_i \quad (i = 1, 2, ..., n), \qquad l_{ij} \leq x_{ij} \leq u_{ij},$$

where l_{ij} and u_{ij} are the lower and upper bounds on arc flow. The summations are performed over the arcs of the network. This problem is solved iteratively starting from an initial state and tending to an optimal solution (see [29, Ch. 8]).

Other than the generic network analysis problem, aiming to find an optimum network configuration with 'minimisation of cost', there are few special network models. A network-flow model without any 'intermediate' location nodes, is discussed in the **transport** problem. The **maximal flow** problem studies how maximum material can be transported from a source node to a sink node (there are no costs associated with the flow). Finding the **shortest path** of the flow is one of the problems that has been worked upon extensively. A description of these problems and different solution methods can be found in [29].

3.2 Mathematical Models of spread

Mathematical modelling is not only of prime importance in natural sciences (in particular, physics), but also in social sciences like economics, sociology and political sciences. An appropriate mathematical representation of a system can be helpful in understanding and studying the effects of various system parameters. Systems whose behaviour in the real world cannot be tested by repeated experiments, can be understood better by repeating virtual experiments with the help of computational mathematical modelling.

3.2.1 Linear regression

Linear regression is the basic and the simplest mathematical model that can be fitted to data. We use linear regression to calculate the average speed of the Neolithic spread using observed Neolithic times (T) and the geodesic distances (D) of archaeological sites from a source (see Ch. 4). In the case of two variables (one explanatory and the other predicted), linear regression fits a polynomial of degree one to the data and is called simple linear

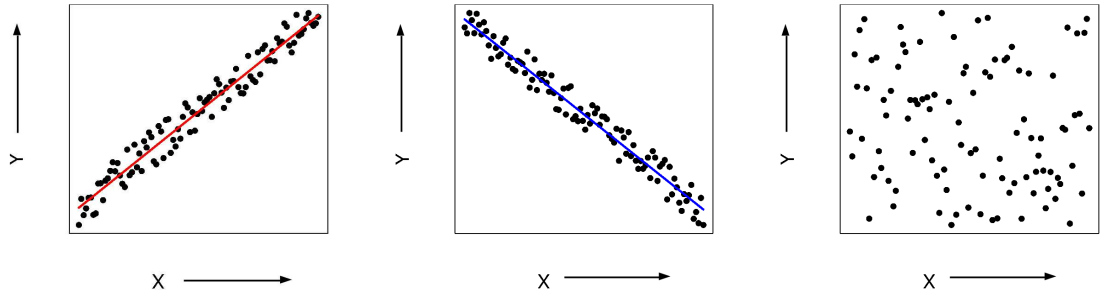


Figure 3.4: The figure on left shows positive correlation, the one in the middle shows negative correlation and figure on right shows no correlation, between the two variables X and Y.

regression (as opposed to multiple linear regression that too fits polynomial of degree one and involves one predicted variable, but works with more than one explanatory variable). For a dataset composed of two variables, X and Y, the fitted functions are given by: $Y = m_1X + c_1$ if the regression is carried out for Y, and $X = m_2Y + c_2$ if the regression is carried out for X. In the first case, X is the independent explanatory variable and Y is the dependent or predicted variable. The coefficients m_1 and c_1 are constants. Similarly, in the second equation Y is the independent variable and X is the dependent variable with m_2 and c_2 being constants. The linear regression would be valid only if the correlation coefficient is high enough (whether positive or negative), (see Fig. 3.4).

The difference between observed and fitted values for the predicted variable is a measure of the goodness of the fit. In an ideal fit, the fitted line will pass through every single data point and the difference between the fitted and the observed values of the predicted variable will be zero. In real-world datasets, due to errors in measurement and potentially due to other (deterministic) sources of variance that are independent of the predictor variable, the line does not pass through every point.

The best-fit line is thus calculated by minimising the squared difference between observed and fitted values $\sum (Y_\mathrm{o} - Y_\mathrm{f})^2$, where Y_o is the observed value and Y_f is the fitted value in the first case, and similarly in the second case. The regression function is chosen on the basis of whether errors in X, or in Y, are to be minimised. This method is known as the

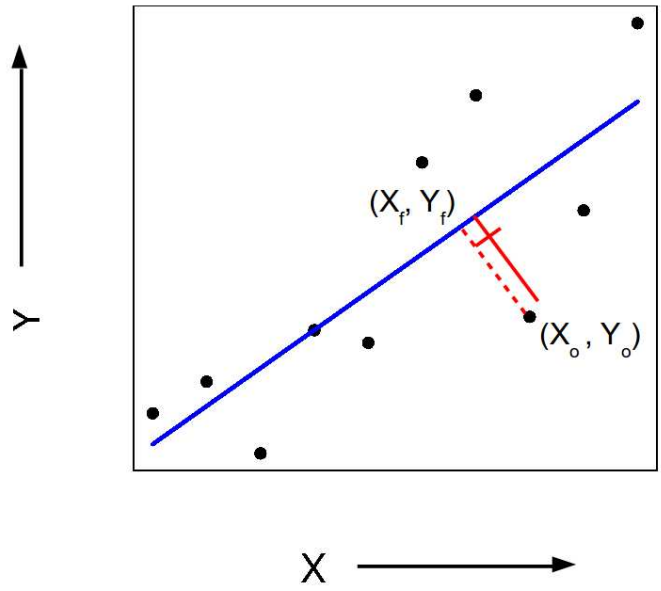


Figure 3.5: The red dashed line orthogonal to the fitted blue line shows the distance of an observed data point from the fitted line. The TLS method minimises the sum of squares of the orthogonal distances of each data point from the fitted line.

Ordinary Least Squares (OLS) method. In some situations however, there are inherent errors in both variables (e.g. Neolithic time of sites (T) and their respective distances (D) from a source). It is then appropriate to minimise errors in both variables. In such cases the method of Total Least Squares (TLS) is used [70]. In the TLS method, the function $Y = mX + c$ is fitted by minimising the sum of square of orthogonal distance of each data point from the fitted line given by $\sum (X_{\rm o} - X_{\rm f})^2 + \sum (Y_o - Y_f)^2$ (shown by the dashed red line in Fig. 3.5). We have also used a 'fracture' or 'piecewise linear' model to explore the possibility of different Neolithic spread speeds at different distances from the source, the details of which are given in Ch. 4.

3.2.2 Population dynamics

A population is a group of individuals of the same species living in the same geographic area. Population dynamics is a branch of mathematical biology and life sciences that involves studies of changes in population size and how it is influenced by various biological and environmental processes and technological changes. The main processes that bring about changes in population sizes are birth, death, immigration and emigration. Population dynamics thus studies the effect of these processes on the growth and decline of populations. Here we consider continuous population dynamics models. The net change in population at a point in space is given by the balance equation:

$$\frac{\partial N}{\partial t} = \text{births} - \text{deaths} + \text{net migration} \tag{3.1}$$

where, N is the population density [99, Ch. 1].

Malthusian and logistic growth models

A simple exponential growth model suggested by Malthus in 1798 is one of the earliest population growth models that form the basis of present-day population dynamics. This basic model is still in use especially in modelling bacterial and insect population growth. The Malthusian model of population growth is obtained by writing Eq. (3.1) without the migration term as:

$$\frac{dN}{dt} = bN - dN = \gamma N \qquad \Rightarrow \quad N(t) = N_0 e^{\gamma t}. \tag{3.2}$$

Here, b and d are the per capita birth and death rates respectively and are positive constants, γ is the net growth rate. N_0 is the population at time $t = 0$. If $b > d$, i.e. $\gamma > 0$, the population increases exponentially without any upper bound. If $b < d$, i.e. $\gamma < 0$, then the population dies out. Figure 3.6 shows the population evolution for three different growth rates. The model thus shows that the population grows exponentially

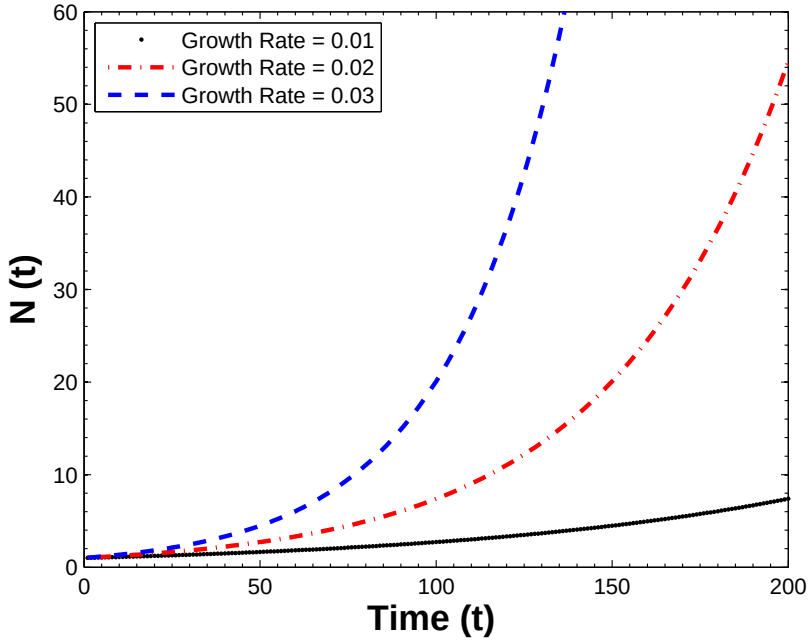


Figure 3.6: Three curves plotted with black, red, and blue colours show the Malthusian (exponential) population growths for $\gamma = 0.01$, $\gamma = 0.02$ and $\gamma = 0.03$ respectively.

unless restricted by some effects which are absent from Eq. (3.2). When applied to the bacterial colonies grown in a medium, bacteria follow the same rapid exponential growth curve until there is no longer sufficient space of nutrients for growth. Similarly when there are resources human populations grow exponentially and exponential growth cannot continue indefinitely if the resources are limited. The population then reaches a saturation level. This is a feature of the logistic growth model:

$$\frac{dN}{dt} = \gamma N \left(1 - \frac{N}{K}\right), \tag{3.3}$$

where γ and K are positive constants, with K known as the carrying capacity of the environment. The carrying capacity is a measure of the ability of the environment to support a population. Solution of Eq. (3.3) is given by,

$$N(t) = \frac{N_0 K e^{\gamma t}}{[K + N_0(e^{\gamma t} - 1)]} \quad \to K \quad \text{as} \quad t \to \infty. \tag{3.4}$$

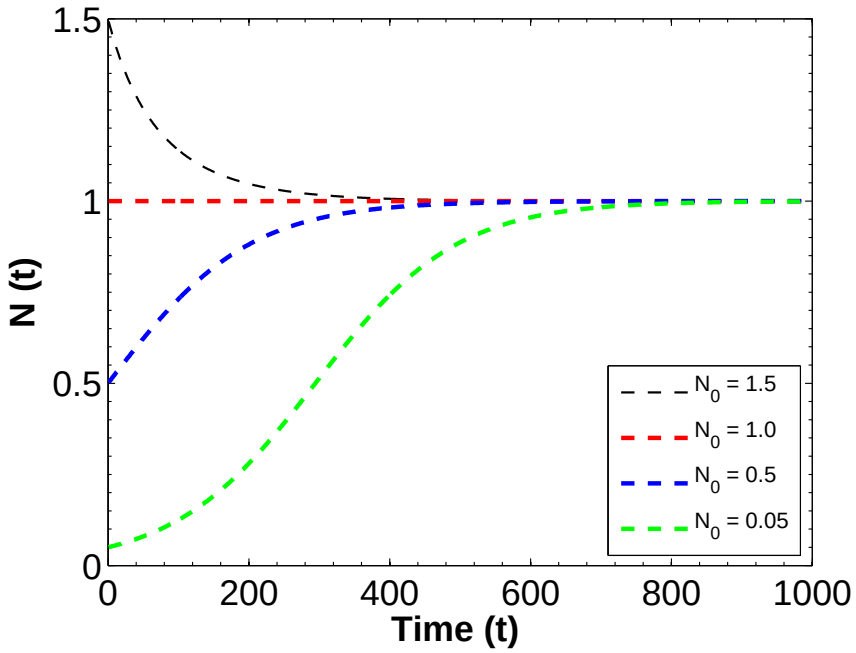


Figure 3.7: Logistic population growths for $\gamma = 0.02$, $K = 1$, and $N_0 = 1.5$, 1, 0.5 *and* 0.05 respectively. All curves show population saturation whether the initial population N_0 was larger or smaller than the carrying capacity K. For an initial population $< K/2$ the curve is shaped like a sigmoid.

Figure 3.7 shows the logistic growth curves for different initial populations with $K = 1$. If the initial population N_0 is less than the carrying capacity K, the growth rate is positive and the population grows and saturates at K. For an initial population density lower than $K/2$ the curve has a sigmoid shape. If N_0 is larger than K, the growth rate is negative and the population decays to K. This model is more realistic than the Malthusian one.

Diffusion

The word 'diffusion' is derived from the Latin word, 'diffundere' and means ' to spread out'. It denotes the spread of a quantity from a higher concentration to a lower concentration. Mixing of two gases or oxygen entering into the blood are examples of diffusion.

Population diffusion is the process in which a population spreads out into surrounding areas (either empty or less populated), by displacing, replacing, or intermixing with a

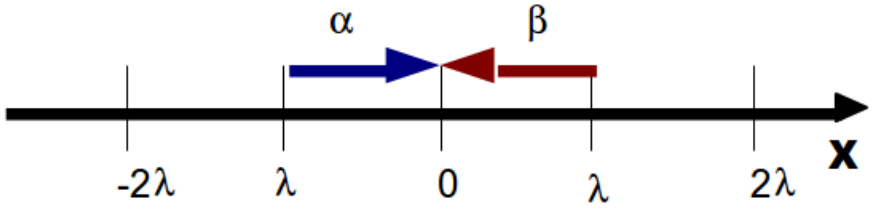


Figure 3.8: Random walk in one dimension with equal step length λ in both directions. Probability of a step in positive x direction is α and that in negative x direction is β. For $\alpha = \beta$, the random walk is isotropic.

pre-existing population in that area. The process underlying the diffusion is a random walk of individuals in a large population or molecules in a gas etc. A random succession of steps, without any directional preference, is an isotropic random walk. Figure 3.8 is a schematic diagram of a random walk in one-dimension. Consider an isotropic random walk for a population of individuals, with all individuals at $x = 0$ at time $t = 0$. At each time step the individuals move either in the positive or the negative x-direction with an equal probability. The distance λ covered in each step is a constant. The population density at a point on the line after a few time steps is the total number of people that stay at that point having arrived from either left or right:

$$N(x,t) = \alpha\, N(x-\lambda, t-\tau) + \beta N(x+\lambda, t-\tau).$$

where, N is the population density (number of individuals per unit length), λ is the step length and τ is the time between the movements. In two dimensions, similarly,

$$\begin{aligned} N(x,y,t) = \alpha\, N(x-\lambda, y, t-\tau) + \beta N(x+\lambda, y, t-\tau) \\ +\rho\, N(x, y-\lambda, t-\tau) + \xi N(x, y+\lambda, t-\tau). \end{aligned} \tag{3.5}$$

where, α, β, ρ, and ξ are the probabilities of moving in the positive x, negative x, positive y and negative y directions, respectively.

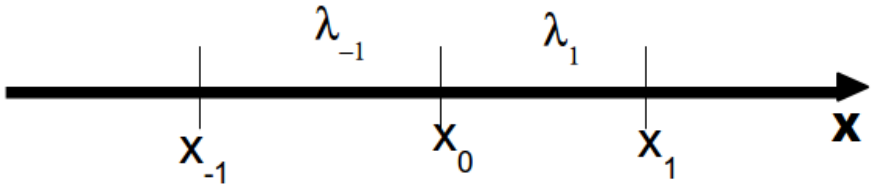


Figure 3.9: The figure shows random walk with different step lengths. Step lengths are functions of position.

Equation (3.5) can be expanded into Taylor series discarding terms of order λ^3 and above,

$$N(x \pm \lambda, y, t-\tau) = N - \tau\frac{\partial N}{\partial t} \pm \lambda\frac{\partial N}{\partial x} + \frac{\tau^2}{2}\frac{\partial^2 N}{\partial t^2} \mp \tau\lambda\frac{\partial^2 N}{\partial x \partial t} + \frac{\lambda^2}{2}\frac{\partial^2 N}{\partial x^2},$$
$$N(x, y \pm \lambda, t-\tau) = N - \tau\frac{\partial N}{\partial t} \pm \lambda\frac{\partial N}{\partial y} + \frac{\tau^2}{2}\frac{\partial^2 N}{\partial t^2} \mp \tau\lambda\frac{\partial^2 N}{\partial y \partial t} + \frac{\lambda^2}{2}\frac{\partial^2 N}{\partial y^2}. \tag{3.6}$$

Since an individual cannot be stationary at any time, $\alpha + \beta + \rho + \xi = 1$. If the random walk is isotropic, all directions are assigned equal probabilities i.e. $\alpha = \beta = \rho = \xi = 1/4$. Thus, we obtain

$$\frac{\partial N}{\partial t} = \frac{\tau}{2}\frac{\partial^2 N}{\partial t^2} + \frac{\lambda^2}{4\tau}\frac{\partial^2 N}{\partial x^2} + \frac{\lambda^2}{4\tau}\frac{\partial^2 N}{\partial y^2}.$$

Taking the limit $\tau \to 0$ but such that $\lambda^2/\tau = \text{const}$, i.e. $\lambda^2 = O(\tau)$, we obtain

$$\frac{\partial N}{\partial t} = \frac{\lambda^2}{4\tau}\left(\frac{\partial^2 N}{\partial x^2} + \frac{\partial^2 N}{\partial y^2}\right) = \nu\nabla^2 N. \tag{3.7}$$

Equation (3.7) is the diffusion equation. This is the simplest case of isotropic diffusion and can be easily generalised. A one-dimensional random walk with position-dependent step size $\lambda = f(x, y)$, is illustrated in Fig. 3.9. The balance equation for population at a point (x_0, y_0) now has the form

$$\begin{aligned} N(x_0, y_0, t) = \alpha\, N(x_{-1}, y_0, t-\tau) + \beta N(x_1, y_0, t-\tau) \\ +\rho\, N(x_0, y_{-1}, t-\tau) + \xi N(x_0, y_1, t-\tau). \end{aligned} \tag{3.8}$$

Since the process is isotropic, $\alpha = \beta = \rho = \xi = 1/4$. From Fig. 3.9, it can be seen that:

$$x_1 = x_0 + f(x_0, y_0), \qquad x_{-1} = x_0 - f(x_0, y_0),$$

Applying the same for the y-direction, we have

$$y_1 = y_0 + f(x_0, y_0), \qquad y_{-1} = y_0 - f(x_0, y_0).$$

Substituting this in Eq. (3.8) yields:

$$\frac{\partial N}{\partial t} = \nabla\nu \cdot \nabla N + \frac{\tau}{2}\frac{\partial^2 N}{\partial t^2} - \frac{\tau}{2}\nabla\nu \cdot \nabla\frac{\partial N}{\partial t} + \nu\nabla^2 N + O(\lambda^4)$$

where as before, $\lambda^2/4\tau = \text{const}$. Taking $\tau \to 0$ we obtain the diffusion equation with variable diffusivity ν:

$$\begin{aligned} \frac{\partial N}{\partial t} &= \nabla\nu \cdot \nabla N + \nu\nabla^2 N \\ &= \nabla \cdot (\nu \nabla N)\,. \end{aligned} \tag{3.9}$$

For a constant diffusivity we recover Eq. (3.7).

The FKPP equation

The reaction-diffusion equation has the form

$$\frac{\partial q}{\partial t} = \nu\,\nabla^2 q + R(q), \qquad \nu = \text{const}.$$

This equation describes the evolution of a quantity subject to diffusion (with diffusivity ν) and production or decay at a net rate $R(q)$. Using $R(q)$ from the logistic population model and assuming inhomogeneous diffusivity, we obtain the Fisher-Kolmogorov-Petrovskii-Piskunov (FKPP) equation, proposed and used by Kolmogorov, Petrovskii and Piskunov to model the evolution of bacteria populations and by Fisher to study the diffu-

sion of advantageous genes [59]. The FKPP equation is:

$$\frac{\partial N}{\partial t} = \gamma N \left(1 - \frac{N}{K}\right) + \nabla \cdot (\nu \nabla N)\,. \tag{3.10}$$

Development of a propagation front is a salient feature of solutions to Eq. (3.10): the front position is defined as a point (or curve) where N is equal to a given constant. The front advances at a constant speed in a homogeneous environment [99] with a speed $U = 2\sqrt{\gamma\nu}$ in one dimension. The width of the front is given by $\lambda/\sqrt{\gamma\tau}$ [87].

Competing populations

If two or more populations occupy the same area, they can interact and compete for the available resources. The nature of the interactions and the intensity of the competition determines whether the populations can coexist peacefully. For two competing populations N_1 and N_2, the governing equations are:

$$\frac{dN_1}{dt} = \gamma_1 N_1 \left(1 - \frac{N_1}{K_1} - b_{12}\frac{N_2}{K_1}\right) + \nabla \cdot (\nu \nabla N_1), \tag{3.11}$$

$$\frac{dN_2}{dt} = \gamma_2 N_2 \left(1 - \frac{N_2}{K_2} - b_{21}\frac{N_1}{K_2}\right) + \nabla \cdot (\nu \nabla N_2). \tag{3.12}$$

where γ_1 and γ_2 are the growth rates and K_1 and K_2 are the respective carrying capacities for the populations at densities N_1 and N_2; b_{12} is a measure of the competition of the population 2 affecting population 1 and b_{21} is the reverse competition effect. Solutions of these equations can be classified as follows

1. $b_{12}K_2/K_1 < 1$ and $b_{21}K_1/K_2 < 1$: *Both populations can coexist.* For example if K_1 and K_2 are nearly equal and the competition parameters b_{12} and b_{21} are not very large, then the two populations adjust and settle at a lower population density.

2. ($b_{12}K_2/K_1 > 1$ and $b_{21}K_1/K_2 > 1$) OR ($b_{12}K_2/K_1 < 1$ and $b_{21}K_1/K_2 > 1$) OR ($b_{12}K_2/K_1 > 1$ and $b_{21}K_1/K_2 < 1$) : *Only one of the two populations would survive.* At early times the populations coexist but, the population with initial advantage

eventually wins and the other population dies.

Finite difference approximations

In order to solve a system of differential equations numerically, they should be discretised on a mesh (grid) $x_{i+1} = x_i + \Delta x_i, \quad \mathrm{i} = 1, 2, 3, ...n$. There are many methods, to approximate a differential equation, at varying orders of the truncation errors. A function f can be expanded at x_{i+1} into Taylor series about point x_i, with only the first order terms retained, to yield

$$f(x_{i+1}) = f(x_i + \Delta x_i) = f(x_i) + f'\Delta x_i + O({\Delta x_i}^2). \tag{3.13}$$

The approximation produces truncation errors of the order $O(\Delta x_i)$. This leads to

$$\left.\frac{df}{dx}\right|_{x_i} = \frac{f_{i+1} - f_i}{\Delta x_i} + O(\Delta x). \tag{3.14}$$

Similarly

$$f(x_{i-1}) = f(x_i - \Delta x_i) = f(x_i) - f'\Delta x_i + O({\Delta x_i}^2), \tag{3.15}$$

$$\left.\frac{df}{dx}\right|_{x_i} = \frac{f_i - f_{i-1}}{\Delta x_i} + O(\Delta x). \tag{3.16}$$

Both the forward and backward finite-difference expressions use information from only two points and have the first-order accuracy. The central difference expression can be obtained using both (3.13) and (3.15):

$$\left.\frac{df}{dx}\right|_{x_i} = \frac{f_{i+1} - f_{i-1}}{2\Delta x_i} + O(\Delta x^2), \tag{3.17}$$

where, $f_{(i+1)} = f(x_{i+1})$ and $f_{(i-1)} = f(x_{i-1})$. The truncation error is of $O(\Delta x^2)$ in this case.

Diffusion in spherical polar coordinates

In applications to global population changes, the equations of population dynamics, should be solved on the Earth's surface. Hence we use the spherical polar form of the equations. The diffusion equation can be written as

$$\frac{\partial N}{\partial t} = -\nabla \cdot (-\nu \nabla N) = -\nabla \cdot \boldsymbol{J}, \tag{3.18}$$

where $\boldsymbol{J} = -\nu \nabla N$ is the population current density. The rate of change of population is thus related to the divergence of $\boldsymbol{J} = -\nu \nabla N$. In spherical polar coordinates,

$$\nabla N = \frac{1}{r_e} \frac{\partial N}{\partial \theta} \hat{\boldsymbol{\theta}} + \frac{1}{r_e \cos\theta} \frac{\partial N}{\partial \phi} \hat{\boldsymbol{\phi}}, \tag{3.19}$$

where r_e is the radius of the Earth, θ is the co-latitude and ϕ is the longitude. For Eq. (3.19), the centered difference scheme has the form

$$\nabla N \approx \frac{1}{r_e} \frac{\Delta N}{\Delta \theta} \hat{\boldsymbol{\theta}} + \frac{1}{r_e \cos\theta} \frac{\Delta N}{\Delta \phi} \hat{\boldsymbol{\phi}}, \tag{3.20}$$

$$\nabla N \Big|_{\theta_i \phi_i} \approx \frac{1}{r_e} \frac{(N_{i+1,j} - N_{i-1,j})}{2\Delta\theta_i} \hat{\boldsymbol{\theta}} + \frac{1}{r_e \cos\theta} \frac{(N_{i,j+1} - N_{i,j-1})}{2\Delta\phi_j} \hat{\boldsymbol{\phi}}. \tag{3.21}$$

The spatial resolutions in zonal (Δx,$\Delta\phi$) and meridional (Δy,$\Delta\theta$) directions are given by

$$\Delta x = r_e \cos\theta \Delta\phi, \quad \text{and} \quad \Delta y = r_e \Delta\theta \tag{3.22}$$

The spatial resolution Δx reduces from the equator to the poles and vanishes at the poles. Since the domain of this study is away from the poles, this does not pose a problem.

The spatial domain on which the numerical solutions are to be obtained, is divided into cells with uniform angular size. Each cell is identified by unique indices (i,j), where i increases eastwards and j increases northwards. Using central differences, the population

fluxes at the eastern and northern boundaries of the cells are then given by

$$J_{i+\frac{1}{2},j} = -\nu_{i+\frac{1}{2},j} \frac{N_{i+1,j} - N_{i,j}}{\Delta x_j}, \tag{3.23}$$

$$J_{i,j+\frac{1}{2}} = -\nu_{i,j+\frac{1}{2}} \frac{N_{i,j+1} - N_{i,j}}{\Delta y_i}. \tag{3.24}$$

Using the first-order approximation for the time derivative the final discretised form of the diffusion equation is given by:

$$\frac{{N^{t+1}}_{i,j} - {N^t}_{i,j}}{\Delta t} = -\frac{1}{\Delta x_j \, \Delta y_i} \left[\left(\Delta y_i \, J_{i+\frac{1}{2},j} - \Delta y_i \, J_{i-\frac{1}{2},j} \right) + \left(\Delta {x_{j+\frac{1}{2}}} J_{i,j+\frac{1}{2}} - \Delta {x_{j-\frac{1}{2}}} J_{i,j-\frac{1}{2}} \right) \right] \tag{3.25}$$

The stability criterion for the diffusive part of the equation, $(\nu \Delta t)/(\Delta x_j \Delta y_i) < 1$ ensures that the flux cannot travel further than the grid size in one time step.

In the present work we employ the models described in section 3.2. The standard linear regression fits using ordinary least squares (OLS), from section 3.2.1 are used for the upper envelope fits in Sec. 4.2.1. The FKPP model described in section 3.2.2 is applied in 6.1 using the finite difference numerical approximation with spherical polar form of the equations. The competing populations model is used in 6.2 and same numerical scheme as the FKPP model (i.e. finite difference with spherical polar form) is applied.

Chapter 4

Neolithic spread from the Near East to the Indian subcontinent

A propagating front of a spreading population (or of any other diffusing quantity) is the location where the population density (or other relevant quantity) reaches a certain fixed value, arbitrary but much smaller than the maximum density behind the front. In homogeneous systems controlled by the reaction-diffusion equation fronts propagate at a constant speed. Spread at a constant speed has been identified as a salient feature of the Neolithic dispersal in Europe, which suggests that the reaction-diffusion equation provide a suitable basis for the mathematical modelling of the Neolithic dispersal. Here we consider the application of such models to South Asia.

We first quantify the Neolithic dispersal across South Asia by exploring the connection between the age T_0 of the advent of the Neolithic and the distance D from its plausible source(s) in the Near East assuming a constant speed U:

$$T_0(D) = T_* + D/U, \tag{4.1}$$

where $T_0(D)$ is the time of the first appearance of the Neolithic at a distance D from

the source of the spread and T_* is the time when the spread started. Such an extremely coarse-grained analysis is a necessary step before any more detailed work which would include regional variations in the speed and direction of the spread.

A difficulty inherent in any study of the spread of incipient agriculture is the identification of the time of the first appearance of the Neolithic at a given location. Even a firmly established earliest evidence of the Neolithic at an archaeological site does not necessarily correspond to the arrival of the Neolithic to the wider local area, since that site might have been occupied at a later time, rather than by the first Neolithic farmers in the region. And the earliest Neolithic layer has not always been discovered (and then dated) with confidence. This problem is less prominent in the case of the better explored European Neolithic, but becomes acute in Asia.

Thus, the earliest Neolithic dates available tell us that the Neolithic appeared in that region *not later* than the available dates suggest. In terms of the dependence of the earliest known Neolithic date T on the distance to the source of the dispersal D, this implies that all the data points must lie below the line $T_0(D)$ in the (D,T) plane, where $T_0(D)$ is the true (generally, unknown) arrival date at a distance D (assuming that earlier dates are plotted higher, as in the Fig. 4.14). In other words, the line $T_0(D)$ is the upper *envelope* of the data points in the (D, T)-plane: ideally, no correctly identified and accurately dated Neolithic data point can lie above this line. Earlier authors presumed that the dates available (most often, obtained after careful selection) do represent the true 'first arrival' time and then fitted a certain dependence $T_0(D)$. On the contrary, we explicitly allow for the fact that, even after the most careful selection, one cannot guarantee that the true arrival time to a given distance has been identified: we seek an upper envelope for the data points in the (D, T)-plane.

However, any age determination of an archaeological site contains random and systematic errors which most often are difficult or impossible to estimate (even in the case of ^{14}C dates). These uncertainties can place a data point above the curve $T_0(D)$. More importantly, any local acceleration of the spread can also produce a data point lying above the

globally averaged dependence $T_0(D)$, by producing a local Neolithic arrival time that is earlier than the average value of T_0 at the relevant distance D. Therefore, our determination of the envelope representing the globally averaged arrival time T_0 must rely on statistical procedures. We analyse a compilation of ^{14}C and archaeological age determinations for the early Neolithic sites in South Asia to reveal and quantify the spatio-temporal continuity of the Neolithic dispersal in Southern Asia.

4.1 Data Selection and Treatment

Since only the first arrival date of the Neolithic at a site matters in the present context, we need to identify the earliest Neolithic date at each of the sites considered, for which either archaeological or radiocarbon dates are available.

We use the archaeological age determinations from Appendix A of [110] for the Indian subcontinent (the definitions of the archaeological phases are from [68], see also Table A4 in Appendix A), together with archaeological records from the Middle and Near East taken from various sources. A complete date list can be found in tables A5 and A6 in Appendix A. For sites only dated with comparative dating (i.e. in terms of archaeological stages), we use the starting date of the relevant time period in our analysis. We have compiled the ^{14}C dates from 160 Early Neolithic sites in West and South Asia [1, 55, 96, 110]. Many of the ^{14}C dates from the Arabian peninsula [53] are also presented in the CONTEXT database [1]. The data are shown in Fig. 1.7. The sites shown in blue filled circles are the sites dated with radiometric methods and the sites shown in red filled circles are dated with comparative dating techniques. For various reasons a few dates had to be discarded. A list of these dates, along with laboratory numbers and reasons for discarding them, is given in the Table A3 in Appendix A. (A histogram of this combined dataset is given in Fig. 4.1 and the distribution of dates within each of the bins is shown in Fig. 4.5.) Our combined database contains 160 sites with ^{14}Cdates and 229 sites with relative dates. This sparsity of the data in Asia significantly complicates the analysis. For comparison, a recent analysis of the Neolithic dispersal in Europe involved 735 sites [108]; a ^{14}C database

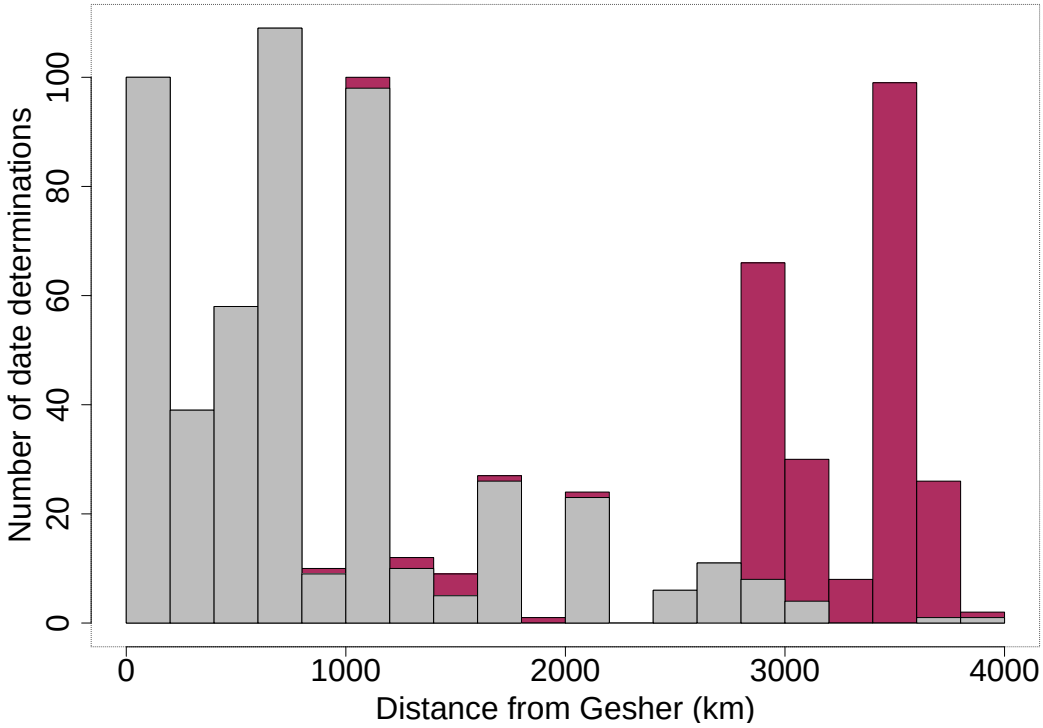


Figure 4.1: The histogram of the number of date determinations per distance bin, of width $\Delta D = 200\,\mathrm{km}$, with respect to the source at Gesher. The maroon shading represent archaeological dates belonging to the Neolithic, whereas the ^{14}C Neolithic dates are shown with gray shading.

for the European Neolithic contains about 640 dates for the earliest Neolithic alone [69].

Primary sources do not always agree about the attribution of a site to the Early Neolithic. For example, a number of sites classified as Chalcolithic, by their authors and then in the CONTEXT database (http://context-database.uni-koeln.de), are included into the list of Neolithic sites by Marshall [96]. We considered both attributions. We excluded all ^{14}C dates marked as doubtful or cited without rating in the CONTEXT database. Likewise, we omitted the "unreliable" ^{14}C dates in the list of Marshall [96], but retained those that have standard deviation in excess of 150 yr since our statistical procedures have their own ways of treating errors.

We performed our statistical analysis with and without the dates from Marshall's list that are not classified as Neolithic in the CONTEXT database, to satisfy ourselves that the results do not change significantly. Our final results are based on the largest data set

available to us, i.e., that including Marshall's list.

4.1.1 The earliest Neolithic dates

Many (131) sites have multiple ^{14}C date determinations and their treatment depends on the nature of the site and the dates (see also [51]). If there are reasons to believe that a group of dates are contemporary and only differ because of random errors, then all such dates can be used towards the analysis. We used a statistical clustering analysis to isolate, where possible, a distinct group (cluster) of the earliest Neolithic dates (using the MCLUST Gaussian mixture model of R, as described below). A similar treatment has been incorporated into OxCal 4.2 (http://c14.arch.ox.ac.uk/embed.php?File=oxcal.html) [30] as a part of the grouping analysis [31], in particular where a group of events within a sequence that can be considered as randomly sampled from a Gaussian distribution (the `Sigma_Boundary` option of OxCal). We identify the most plausible earliest Neolithic date(s) for each site using the following criteria (further details can be found in Figures 4.2 and 4.3; see also [47]):

1. For a site with a single ^{14}C date (29 such sites), we use this date (unless it is discarded for any other reasons, such as dubious context, questionable attribution, etc.).

2. For sites that have a statistically significant number of ^{14}C dates[1], we applied a statistical Gaussian mixture model to isolate (where possible) a well-fit temporal cluster of the oldest dates. The dates within this cluster are then used in the subsequent analysis. The conditions for a well-fit cluster are given below.

3. If a well-fit cluster cannot be identified, then the mean of those dates which lie within 350 years of the site's oldest date are used in the analysis. (If the earliest and second

[1] A χ^2 test is a statistical test used widely and reliably by many researchers, however, care must be taken in order to get sensible inferences from the test'. For a χ^2 test, at least 5 data points are required ideally in every interval to obtain meaningful results from the test [137, pp 535]. There is also a lower limit for the total number data points in a sample which states that for a one-sample χ^2 test at least 10 data points [137, pp549] and for a two-sample test at least 30 data points are required [137, pp549]. For smaller samples other statistical tests (e.g. K-S test , M-W test) can be used.

We required the Gaussian mixture model to fulfil a similar requirement to the χ^2 test, only applying it to a cluster that contains at least 3 dates.

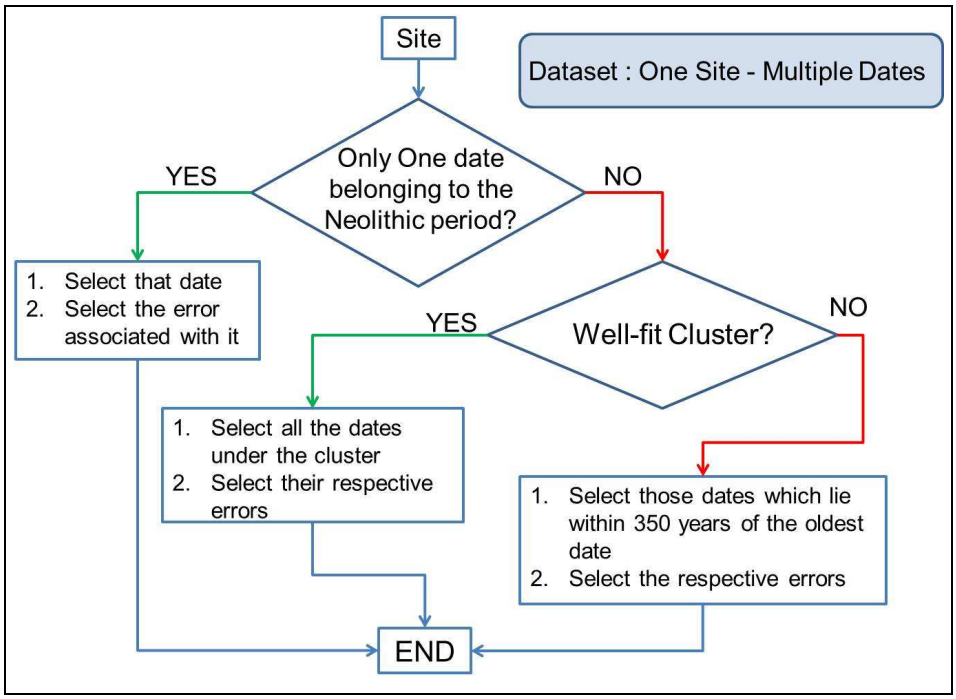


Figure 4.2: A flowchart of the selection procedure for the earliest Neolithic ^{14}C date(s) at a site. One site, multiple dates. See the text for further details.

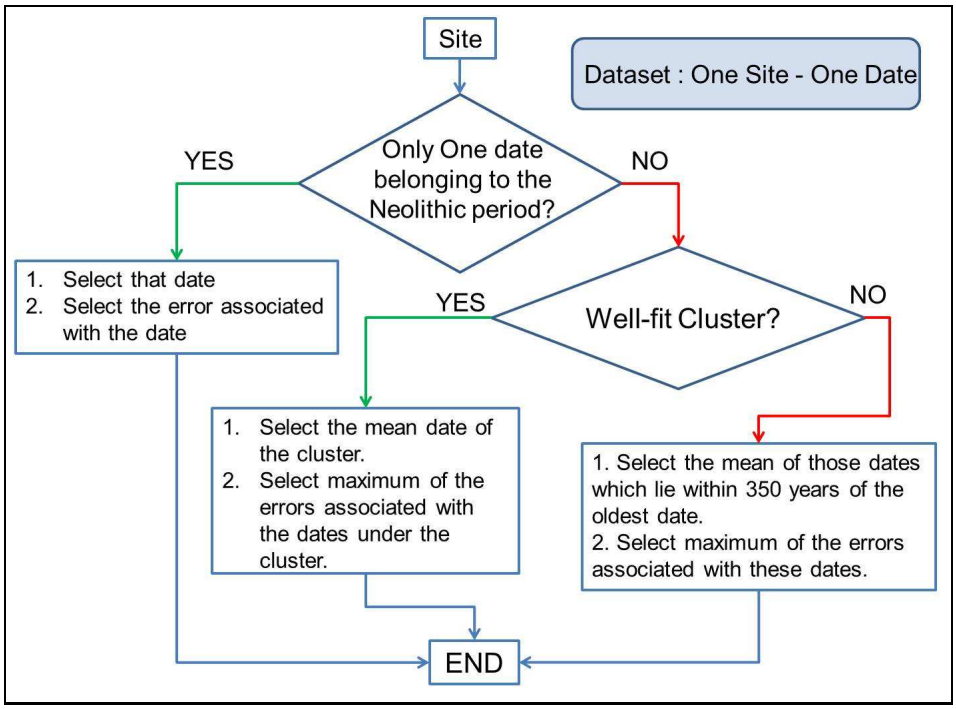


Figure 4.3: A flowchart of the selection procedure for the earliest Neolithic ^{14}C date at a site. One site, one date. See the text for further details.

earliest dates are more than 350 years apart, then only the earliest date is used.)

For criterion 2, we use the *mclust* package [63] of the R programming language, which attempts to fit the dates into up to nine separate clusters assuming a Gaussian probability distribution of the dates in each cluster. The preferred number of clusters is chosen using a Bayesian Information Criterion (BIC) which quantifies the misfit between the observed dates and the model, with a penalty for models with a larger number of parameters. For further details see Fig. 4.4.

The Gaussian mixture model is accepted for the earliest group when the following criteria are satisfied:

1. The earliest group (cluster) contains three or more dates.

2. The standard deviation of the dates in the earliest cluster, σ_1, does not exceed $\sigma_{\rm min} = 175\,{\rm yr}$, a typical accuracy of the Neolithic ^{14}C dates in our sample, as estimated below.

3. There is no overlap between the earliest and second earliest clusters within their standard deviations (half-widths). In other words, the two probability density functions do not overlap significantly within the ranges $T_1 \pm \sigma_1$ and $T_2 \pm \sigma_2$, where T_1 and T_2 are the the mean dates of the earliest and second earliest clusters, respectively, and σ_1 and σ_2 are the corresponding standard deviations.

The clustering analysis is illustrated in Fig. 4.4 for six sites selected to represent the range of situations encountered. Each dashed curve in these figures represents one cluster (i.e., one Gaussian mode). Panels (a) and (b) show well-defined clusters; the former has no significant overlap between the earliest (black dashed curve) and second earliest (red dashed curve) groups, the latter has some overlap, but the separation of the maxima is wider than the sum of the two standard deviations. For sites with a similarly strong clustering, we show the cluster mean date and standard deviation in Table A2 and label that entry with 'C' in Column 2. The dates shown in panels (c)–(f) do not have well-

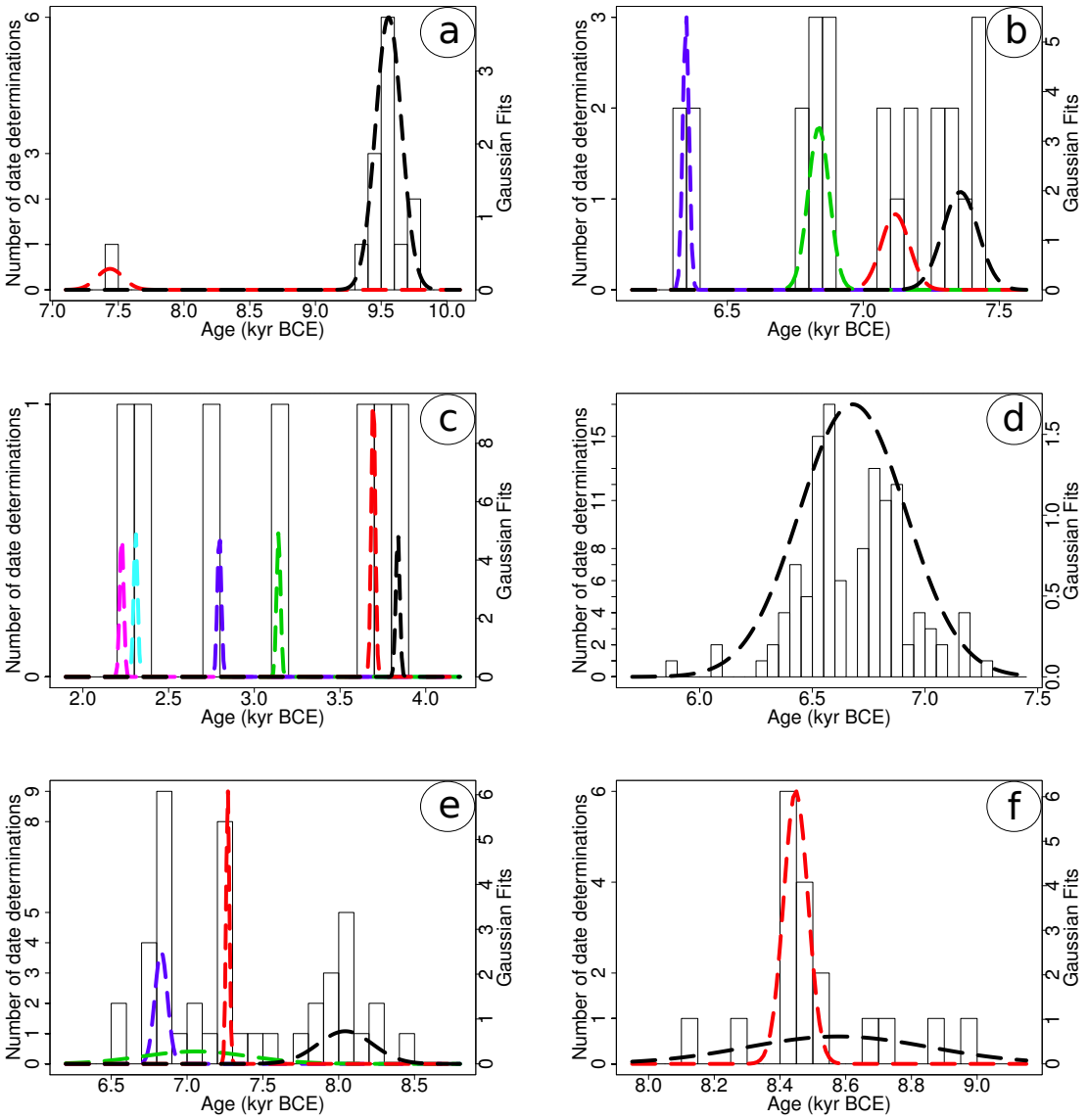


Figure 4.4: An illustration of the various situations encountered when calculating the earliest Neolithic date for sites with multiple ^{14}C date determinations, showing the histograms of the dates and the Gaussian-mixture fits to them (using the `mclust` routine of R) for each of the following sites: **(a: top left)** Körtik Tepe, **(b: top right)** Bouqras, **(c: middle left)** Sarazm, **(d: middle right)** Çatalhöyük East, **(e: bottom left)** Áin Ghazal and **(f: bottom right)** Djade. The Gaussian curves shown with a dashed black line in (a) and (b) represent well-defined earliest clusters according to criteria (1)–(3) of Section 4.1.1; dashed curves in other colors show later clusters. In Panels (c)–(f), results of the Gaussian mixture modelling are not acceptable. The histograms for the Gaussian mixture model cluster analysis for all the sites with radiocarbon dates are given in the Appendix B

defined clusters of the earliest dates according to the above criteria, and the Gaussian mixture model is not used in such cases. In Panel (c), the earliest 'cluster' contains a single ^{14}C date (fewer than the minimum of three dates). In Panels (d) and (e), the standard deviation of the dates in the cluster exceeds $\sigma_{\min} = 175\,\mathrm{yr}$. In Panel (f), the two clusters overlap.

In such cases, where a distinct earliest cluster of dates cannot be isolated, we use the average of all those dates which lie within 350 years ($= 2\sigma_{\min}$) of the earliest date at that site. The resulting average value and the standard deviation are shown in Table A2, labelled 'A'. If the gap between the two earliest ^{14}C dates exceeds 350 years, we simply take the earliest date available, labelled 'O'.

The BOUNDARY facility of OxCal might also be used to derive the probable date of the appearance of the Neolithic at a given site or at a certain distance from the source. Unfortunately, the data available in this work are too scarce to warrant this approach in any systematic manner.

To verify the stability of our results under modifications of the data set, we performed our analysis both using the full set of individual dates within the relevant clusters (see Fig. 4.2) and the single representative values for the sites with multiple ^{14}C dates (see Fig. 4.3). Another variation of the data set, used for the same purpose, was to include or exclude the dates provided in [96] where the attribution of some ^{14}C dates as Neolithic is more inclusive than that in [1]. The variation in the final fits resulting from these variations in the data sets was used to estimate the uncertainty of our results. However, our main results were obtained with the largest data set available (i.e., including the dates from [96], and without replacing groups of dates by their representative averages; Fig. 4.2).

4.1.2 Bin widths and outliers

The first step in our analysis is to group the data into bins according to their distance from a source in the Jordan Valley, chosen to give the best statistical quality of the result. The precise location of the source (Gesher in our case) is largely a technical matter (although it

is important that it is within the area of the earliest Neolithic sites in the Near East), and should not be over-interpreted as corresponding to an actual prehistoric point source. The bin width ΔD is chosen to ensure that most of the bins contain no fewer than 5 data points (see Table 4.1). The bin width is also chosen to be consistent with the accuracy of the age determinations and the expected speed of the spread. Since the *accuracy* (distinct from *precision* [123]) of the Neolithic ^{14}C dates is of order $\sigma = 100$–$200\,\mathrm{yr}$ — and archaeological dates usually have larger uncertainties — and the expected average speed of the spread is $U \sim 1\,\mathrm{km/yr}$, the width of a distance bin should be at least $\sigma U = 150\,\mathrm{km}$, comparable to the width of the propagating front. We varied the bin width around this value (by considering the range 100–300 km) to verify the stability of our results; bin widths in the range 150–250 km appear to be acceptable (Figures 4.6 to 4.13). The bin width used for the rest of this study is $\Delta D = 200\,\mathrm{km}$, and the distribution of the sites between the bins is shown in Fig. 4.1.

Table 4.1: The number of data points per bin for various bin widths

Bin width (km)	Total number of bins	Number (%) of bins with more than 5 data points
100	40	28 (70%)
150	26	21 (81%)
200	20	17 (85%)
250	16	14 (87%)
300	13	12 (92%)

Here and elsewhere in the text, BCE dates are treated as negative: for example, for $T_1 = 10,000\,\mathrm{BCE}$ and $T_2 = 8,000\,\mathrm{BCE}$, we have $T_1 < T_2$ as $-10,000 < -8,000$. (Our figures nevertheless plot BCE ages on the positive y-axis.) Since some bins contain a relatively small number of sites, it is essential to ensure that outliers do not affect the result. The outliers, shown in Fig. 4.5 (in red), were identified using the interquartile ranges [123]: with Q_1 and Q_3 being the first and third quartiles of the distribution of dates within a bin (i.e., 25% of the dates in the bin are earlier than the date Q_1 and 25% of the dates are later than Q_3), a date T is treated as an outlier if it is separated from the nearest quartile by more than three interquartile ranges; i.e., if

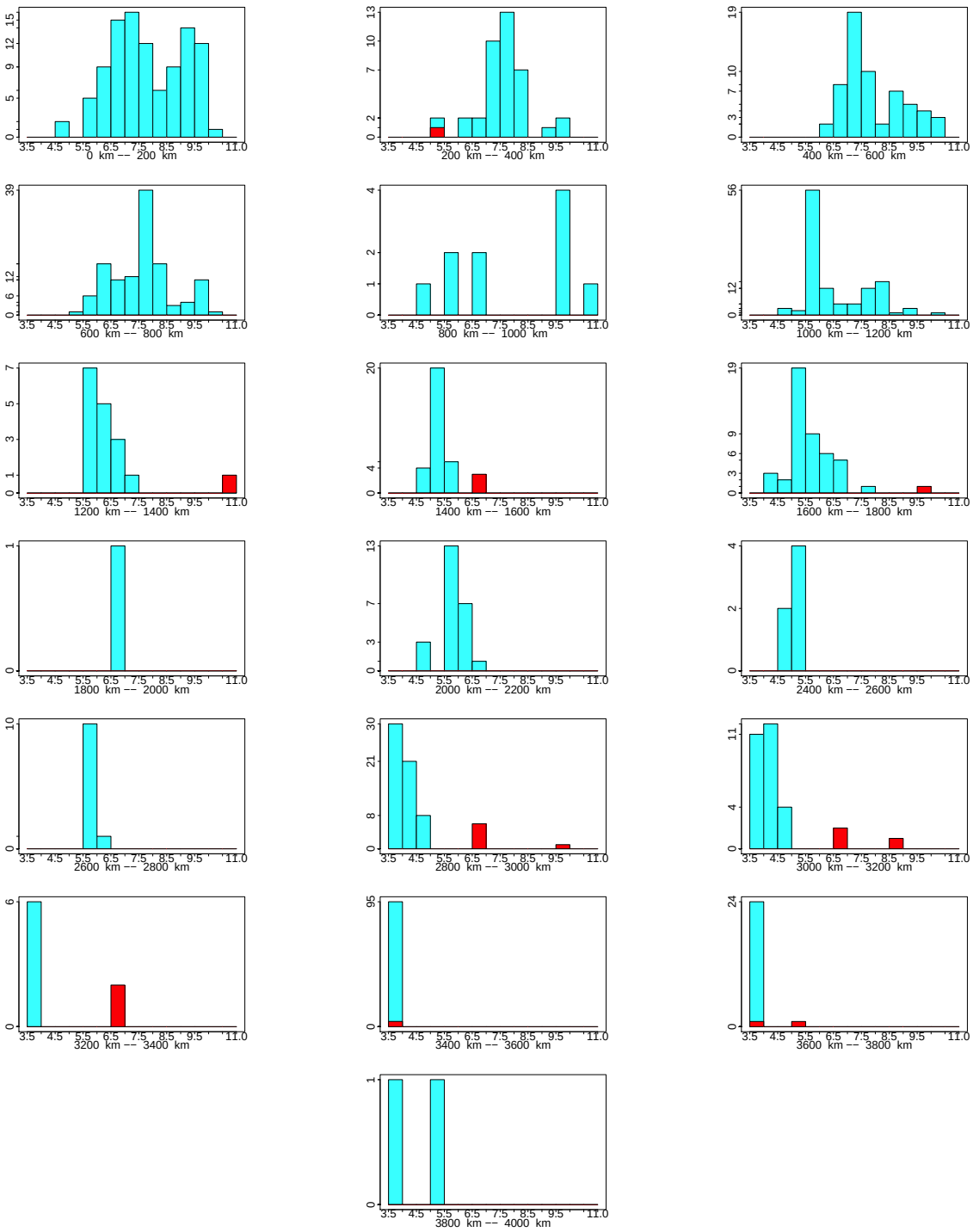


Figure 4.5: The distribution of the age determinations within each bin, of width $\Delta D = 200\,\text{km}$. Bin 12 (2,200–2,400 km) is empty. In each of these figures, the x-axis is in 'kyrBCE' and y-axis is the number of sites. The outliers are identified as described in the text, and are shaded in red.

$$T < \mathrm{Q}_1 - 3(\mathrm{Q}_3 - \mathrm{Q}_1) \quad \text{or} \quad T > \mathrm{Q}_3 + 3(\mathrm{Q}_3 - \mathrm{Q}_1).$$

Twenty five dates are classified as outliers according to this criterion. The change in the results obtained after the removal of the 25 outlier dates is not significant. Therefore, results presented in this work were obtained without discarding the outliers.

4.1.3 Data uncertainty

The published precision of the ^{14}C dates represents not their full accuracy but rather the error in the laboratory measurement of the ^{14}C content in the sample [8, 51, 123]. The calibration error alone amounts to at least 50–150 yr for the Early Neolithic period. In a specially designed experiment, Mazurkevich et al. [98] (see also [50]) obtained 35 ^{14}C dates for wooden structures belonging to four Late Neolithic dwellings in the Serteya Valley (Smolensk region, Russia) apparently constructed during a single season, so that there is every reason to expect all the dates to be contemporaneous in the sense of radiocarbon dating. The empirical standard deviation of the calibrated dates for the dwellings are 113, 83, 129 and 184 yr. Thus, the accuracy of the ^{14}C age determination is 100–200 yr in this case. Dolukhanov et al. [51] suggest that $\sigma_{\mathrm{min}} = 100$–$130$ yr for early Neolithic sites in Central and Eastern Europe. Appendix B shows the temporal cluster plots for the Neolithic sites used here. There are 35 sites for which the oldest cluster, obtained from the Gaussian mixture modelling analysis, contains 3 or more dates (26 sites have 5 or more). An average of the half widths of the normal distributions of these 35 oldest clusters is 177 yr. We use this value (rounded down) as the minimum uncertainty, $\sigma_{min} = 175$ yr. We note that underestimated errors are more problematic for statistical analyses than

overestimated ones[2].

$$\sigma_i = \max(\tilde{\sigma}_i, 175\,\mathrm{yr}),$$

where $\tilde{\sigma}_i$ includes the laboratory error (as published together with uncalibrated ^{14}C dates) and the calibration error obtained in the course of calibration with OxCal (version 4.2), combined as the quadratic sum.

For archaeological dates, the uncertainty σ_i is taken to be the time span of the relevant archaeological stage. This gives uncertainties of up to 300 yr.

For the envelope fitting method using percentile values for each bin (see percentile method in 4.2.1), the uncertainty associated with the calculation of the percentile ($\sigma_{\%}$) is typically greater than the individual dating uncertainties, and so the former value is normally used as the uncertainty σ_i for bin i. As mentioned above, the uncertainty $\sigma_{\%}$ (along with the best estimate of the percentile value, $T_{\%}$) is estimated by **bootstrapping** (resampling) [33, ch.2]. A bootstrapping procedure was implemented in which, for each bin, 10,000 synthetic data sets were created by choosing m dates at random from the full set of m dates within the bin. This is resampling with replacement, so in each synthetic set a random fraction of the original dates are replaced by duplicated original dates. (On average, this fraction will be 37% [111].) The percentile value for each synthetic set is calculated; the mean of these values is taken as $T_{\%}$, and the standard deviation as $\sigma_{\%}$. For some bins, including those where the oldest dates are multiple archaeological dates from the same stage, assigned identical dates, this procedure can give rather small values of $\sigma_{\%}$ (since the synthetic sets very often still include multiple identical earliest dates, and the relatively high percentile value becomes exactly this date). As such small uncertainties are not sensible, given that all of the individual dates going into the calculation are only known to some accuracy σ_i (for date i), the value of $\sigma_{\%}$ is taken to be a minimum of 300

[2]Overestimating errors may prevent one from drawing a right conclusion. For example take

$$X^2 = \sum_i \left(\frac{T_i - T_i^m}{\sigma_i} \right)^2 .$$

If here σ_i are overestimated, a wide range of models, T^m, would fit the data, and the amount of information we obtain would be small. If, however, σ_i are underestimated, there is a possibility that the correct model is discarded and the model that fits will have more parameters and their values will be spurious.

yr (from the maximum archaeological dating uncertainty); i.e., the uncertainty for bin i is taken as $\sigma_i = \max(\sigma_{\%}, 300\,\mathrm{yr})$.

Similar results were obtained using **alternative bootstrapping** procedures (where a fixed fraction of the dates within each bin were replaced in each synthetic data set), and using **jackknife** methods (where synthetic data sets were created by removing a fraction of the dates within each bin, without replacement).

4.2 Statistical analysis

4.2.1 The envelope of the data points in the (T, D)-plane

To the best of our knowledge, there is no *suitable* standard procedure to fit an envelope *statistically* to the type of data that we have. To ensure that our results are robust, we use two distinct approaches to find an envelope that identifies the average first arrival date of the Neolithic at a given distance from a source of the spread.

As in other analyses of this type, the precise position of the source of the spread is largely conventional [17, 40, 108], and is selected to achieve the best-quality fit to the data. We considered the six earliest Neolithic sites in the Fertile Crescent, and also all locations on a grid of $2^\circ \times 2^\circ$ encompassing this region, and identified Gesher, one of the earliest Neolithic sites in the Jordan Valley, as the best effective source.

Having chosen the bin width and the conventional source, we consider the data distribution within each bin in two different ways, to estimate the average Neolithic arrival date T_0 at a distance D from the source (see Eq. 4.1). We largely focus on the (ordinary) weighted least squares method, although the (closely related) maximum likelihood method is briefly discussed below. We also consider a refinement of this model where the propagation speed is equal to different constants in two separate ranges of D to allow, to some extent, for a change in the environment with distance from the Near East.

Weighted least squares

With the weighted least squares method, the best-fitting values of U and T_* are those that provide a minimum, $\widehat{X}^2$, with respect to U and T_*, of the sum of squares of the deviations of the model of Eq. (4.1) from the data T_i weighted with the uncertainties σ_i; i.e.,

$$X^2 = \sum_{i=1}^{n} \left[\frac{T_i - T_0(D_i)}{\sigma_i} \right]^2 , \quad \widehat{X}^2 = \min_{(U,T_*)} X^2 , \tag{4.2}$$

where n is the number of the data points. $T_0(D_i)$ is the model first arrival time calculated from Eq. (4.1) at the site i, which is at distance D_i from the source and has the observed arrival time T_i. We consider fits of the form (4.1) to either weighted data or to the percentiles of the binned data.

The statistical quality of the fit can be conveniently quantified in terms of the coefficient of determination

$$R^2 = 1 - \frac{\widehat{X}^2}{S^2} ,$$

where S^2 is a measure of the deviation of the data from their overall mean value $\overline{T}$:

$$S^2 = \sum_{i=1}^{n} \left(\frac{T_i - \overline{T}}{\sigma_i} \right)^2 ,$$

with (see, e.g., [52])

$$\overline{T} = \frac{\sum_{i=1}^{n} T_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} .$$

The coefficient of determination represents the fraction of the data variability that is accounted for by the model: $R^2 = 1$ indicates that the entirety of the scatter of the data points is explained by the model (i.e., the fit is perfect).

The weighted least-squares method implicitly assumes that the data points follow the statistical model $T_i = T_0(D_i) + \epsilon_i$, where $T_0(D) = T_* + D/U$ is our fitting function, and the errors ϵ_i are drawn from normal distributions with variance σ_i^2.

Weighted data

In this method, each date T_i in a bin is assigned a weight w_i, larger for the earlier dates within the bin, thus giving preference to the earlier local dates, with w_i defined as:

$$w_i = \frac{W_i}{w_0}\,, \quad W_i = \exp\left(-\frac{(T_i - T_{\max})^2}{\tau^2}\right), \quad w_0 = \sum_{i=1}^{m} W_i\,,$$

where $T_{\max}$ is the earliest of the m dates in the bin, τ is an adjustable parameter (a weighting scale chosen empirically as described below), and the normalization factor w_0 ensures that each bin carries the same weight independent of the number of dates in it. We have considered values of τ in the range 100–500 yr, to ensure that the results are robust with respect to this parameter; we present results using $\tau = 200\,\mathrm{yr}$ here.

In fact, for the envelope fitting method using weightings based on the relative age within each bin, the weighting associated with the individual date uncertainties is rather unimportant, compared to the imposed weighting. Calculations combining both sources of weighting have been made, but the results do not differ significantly from those using only the imposed weightings. For simplicity, in the present work we present only the results from the latter case.

Percentile method

For the percentile method, a single representative first-arrival date, $T_{\%}$, was calculated for each bin, as a certain upper percentile of the distribution of dates within the bin. Such percentiles are obtained by linear interpolation between some of the earliest dates in the bin; as such, the values are rather sensitive to the precise distribution of dates within the bin (and particularly so for bins with a small number of dates). To reduce this sensitivity, and to quantify the uncertainty in the resulting values, a bootstrapping approach was used: for each bin, the percentile value was calculated 10,000 times using sets of dates resampled randomly (with replacement)from the full set of dates in that bin; the mean of these percentile values was taken as $T_{\%}$ for that bin, and the standard deviation of the

values was taken as the associated uncertainty, $\sigma_{\%}$ (see also Section 4.1.3).

Each date $T_{\%}$ was associated with the distance D corresponding to the mid-distance of the corresponding bin from the source, and the best linear fit of the form (4.1) was obtained using $T_{\%}$ weighted by $\sigma_{\%}$, the uncertainty of this age estimate (so that less precise values of $T_{\%}$ have smaller weight). We have considered various percentile levels from 70% to 97% to ensure that the results are robust in this respect, and present those for the 95% level in Fig. 4.15.

For this method, the root mean square value of σ_i, obtained by bootstrapping, is 640 yr whereas the root mean square residual from our best constant-speed fit is 990 yr. So the uncertainties in the calculation of the percentile values do not fully explain the residuals between the data and our fit. This is not particularly surprising, or concerning, since we know that our model is highly simplified, and expect deviations arising, for example, from regional variations in the local speed of spread.

An alternative approach which explicitly includes this anticipated misfit assumes that the data points follow the statistical model $T_i = T_0(D_i) + \epsilon_i + e_i$, where ϵ_i is as above, and e_i quantifies the additional uncertainty, drawn from a normal distribution with variance σ^2; here σ is a global parameter of this model. This approach is used in the maximum likelihood fits, which seek the best fit by maximising the likelihood of the fit or equivalently, maximising the log of the likelihood, given for this model by

$$L = -\frac{1}{2}\sum_{i=1}^{n} \ln \bar{\sigma}_i^2 - \frac{1}{2}\sum_{i=1}^{n} \frac{[T_i - T_0(D_i)]^2}{\bar{\sigma}_i^2}\,, \quad \widehat{L} = \max_{(U,T_*,\sigma)} L\,,$$

where $\bar{\sigma}_i^2 = \sigma_i^2 + \sigma^2$. In the case of equal uncertainties σ_i, this approach gives the same fits as the weighted least squares formalism. Applying this method to our percentile data with the σ_i obtained from bootstrapping gives a maximum likelihood estimate of σ as 610 yr, and we note that distribution tests suggest that the residuals are well modelled as coming from a normal distribution. The values of U and T_* for this fit do not differ significantly from those for the weighted least squares method, and for simplicity we present only the

latter.

Constant-speed fits

The simplest model is that with a constant U, as presented above. The two methods of fits considered —using relative weightings based on the relative age within each bin, and using percentile values for each bin — were applied to this model as described earlier. The best fits obtained for the two methods — using bin width $\Delta D = 200\,\mathrm{km}$, $\tau = 200\,\mathrm{yr}$ for the weighting method and the 95%-ile value for the percentile method— are shown in Figures 4.14 and 4.15. Here we present the plots showing the relative insensitivity of the fits to using the two datasets: 'One Site – One Date' and 'One Site – Many Dates'. The fits also are not affected much with the choices of bin width (ΔD), weighting time scale (τ) and different percentile values. Figures 4.6 and 4.10 show that for both the datasets, 'One Site– Many Dates' and 'One Site – One Date' respectively, the speed obtained, U, varies by only 0.1 km/yr as ΔD changes from 150 to 250 km. Likewise, the results depend on τ very weakly. The speeds obtained by the percentile method for both datasets as shown in Figures 4.8 and 4.12 show that percentile levels between 85 and 97%, and bin widths in the range 150–250 km, produce a similarly small variation in the values of U. Also, as it can be seen from Figures 4.7, 4.9, 4.11 and 4.13, the effect of variations of ΔD, τ and the different percentile levels on the intercept of the fits, T_*, is similarly mild. Thus the results obtained by the percentile method are close to those obtained from the weighted data as described above, lending additional confidence in the reliability of our statistical procedures.

Variable-speed fits

There are several reasons to consider models with variable U. We are really interested here in the spread of the Neolithic beyond the Fertile Crescent, where a noticeable fraction of our ^{14}C dates belong. The Zagros Mountains separate regions with very different climate and topography, and this may have affected the Neolithic dispersal. Therefore, we also

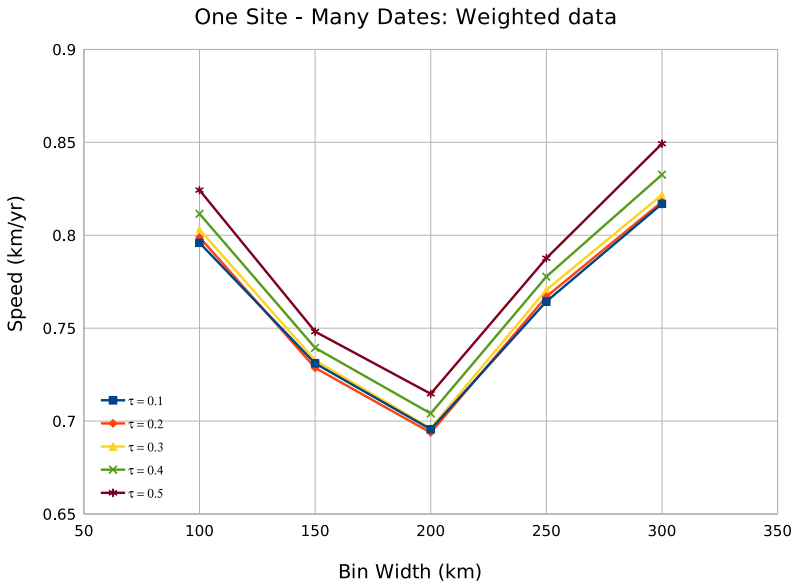


Figure 4.6: The variation in the best-fit speeds of spread U with the width of distance intervals used to bin the data for the weighted data, with different lines obtained for the different weighting time-scales τ, as specified in the key. Used here is the 'One Site – Many Dates' dataset.

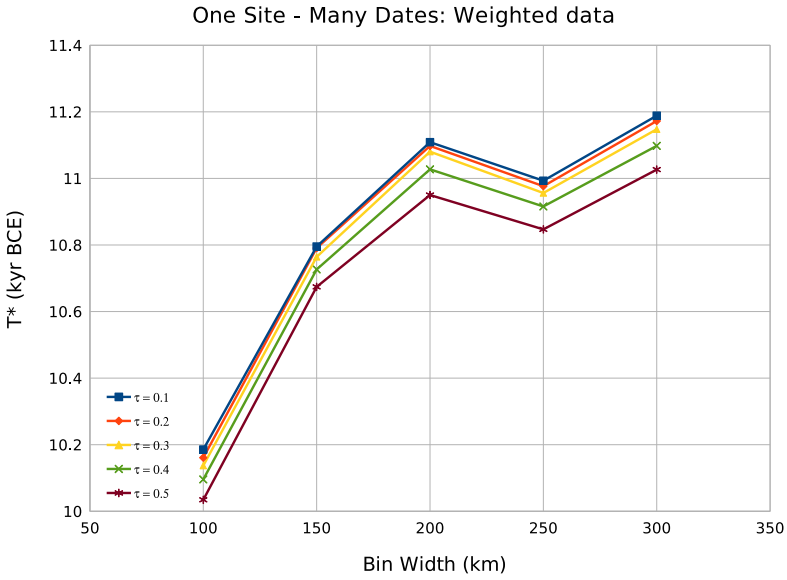


Figure 4.7: The variation in the best-fit starting times of spread T^* with the width of distance intervals used to bin the data for the weighted data, with different lines obtained for the different weighting time-scales τ, as specified in the key. Used here is the 'One Site – Many Dates' dataset.

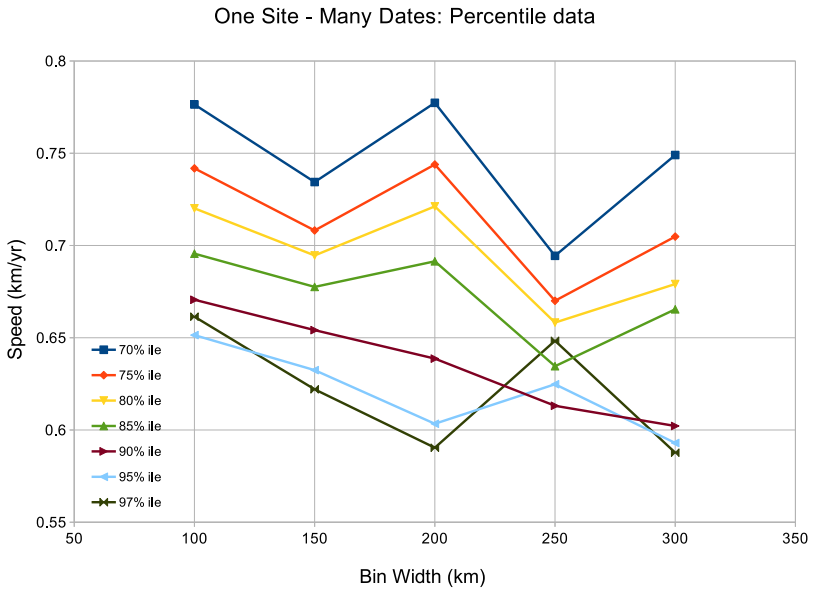


Figure 4.8: The variation in the best-fit speeds of spread U with the width of distance intervals used to bin the data for the percentile data, with different lines corresponding to the different percentile levels specified in the key. Used here is the 'One Site – Many Dates' dataset.

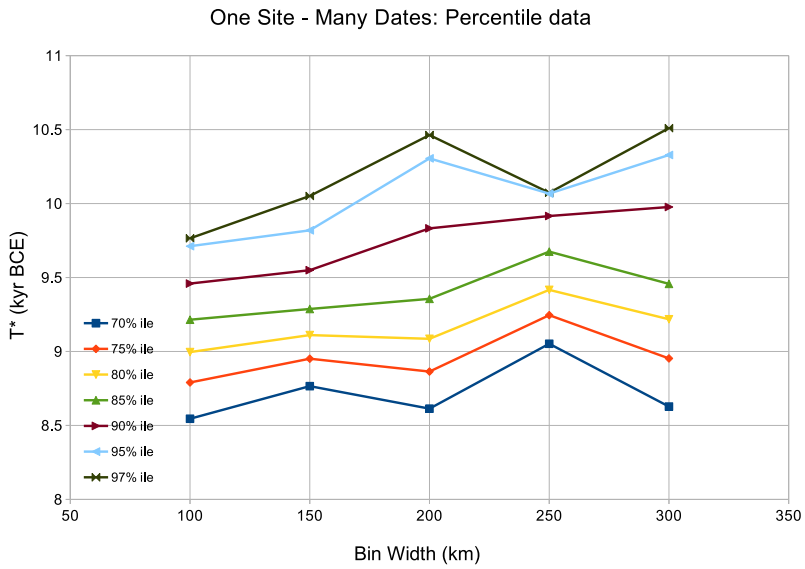


Figure 4.9: The variation in the best-fit starting times of spread T^* with the width of distance intervals used to bin the data for the percentile data, with different lines corresponding to the different percentile levels specified in the key. Used here is the 'One Site – Many Dates' dataset.

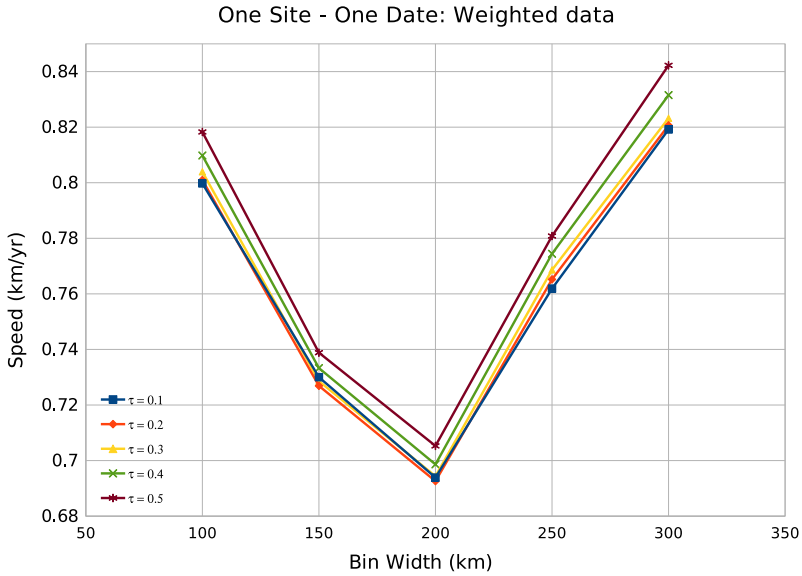


Figure 4.10: The variation in the best-fit speeds of spread U with the width of distance intervals used to bin the data for the weighted data, with different lines obtained for the different weighting time-scales τ, as specified in the key. Used here is the 'One Site – One Date' dataset.

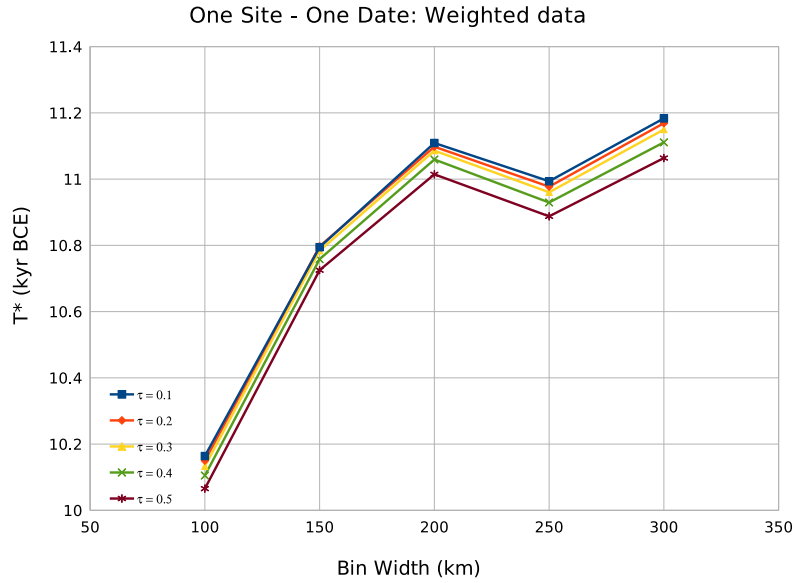


Figure 4.11: The variation in the best-fit starting times of spread T^* with the width of distance intervals used to bin the data for the weighted data, with different lines obtained for the different weighting time-scales τ, as specified in the key. Used here is the 'One Site – One Date' dataset.

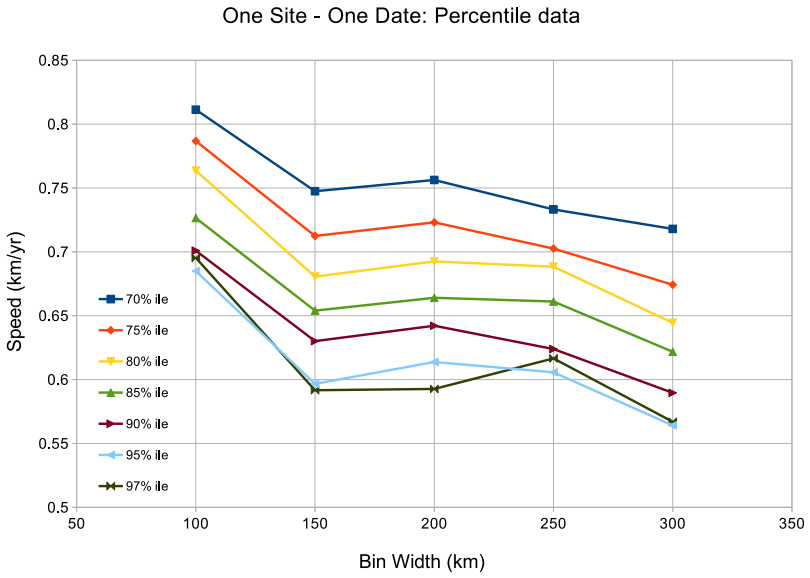


Figure 4.12: The variation in the best-fit speeds of spread U with the width of distance intervals used to bin the data for the percentile data, with different lines corresponding to the different percentile levels specified in the key. Used here is the 'One Site – One Date' dataset.

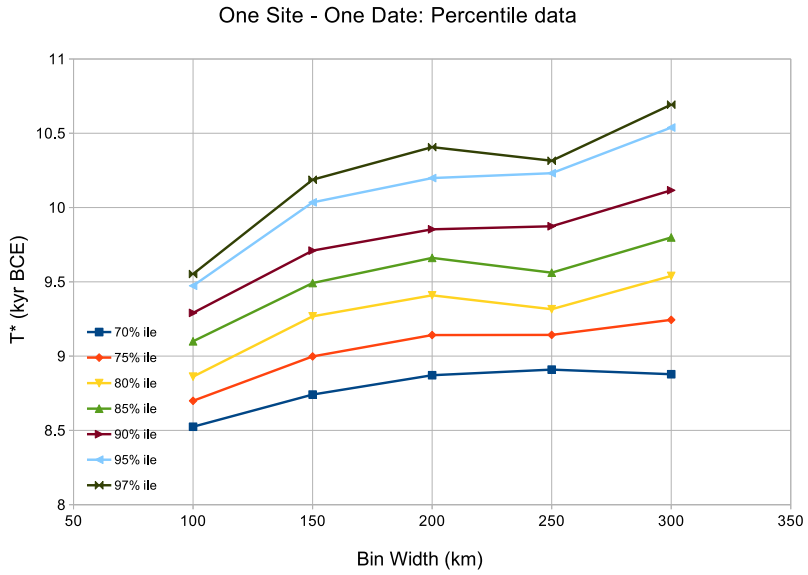


Figure 4.13: The variation in the best-fit starting times of spread T^* with the width of distance intervals used to bin the data for the percentile data, with different lines corresponding to the different percentile levels specified in the key. Used here is the 'One Site – One Date' dataset.

considered a model which allows for a change in U at a certain distance D_0 from the source. (The choice of D_0 is described below.) Given the scarcity of the data available, we only considered the simplest models of this type, where U is constant at $D < D_0$ and $D > D_0$.

Instead of Eq. (4.1), we thus consider the following model (similar to that of [94]):

$$T_0 = \begin{cases} T_{*1} + D/U_1 & \text{if } D < D_0, \\ T_{*2} + D/U_2 & \text{if } D > D_0, \end{cases} \tag{4.3}$$

with parameters T_{*1}, U_1, T_{*2}, U_2 and D_0. Here U_1 and U_2 are the speeds of dispersal west and east of the Zagros (or any other transition region), respectively. This model allows for the dependence of T_0 on D to be discontinuous or continuous at $D = D_0$. (The discontinuous fit allows for a hiatus in the spread at D_0.) In the continuous case, $T_{*2} = T_{*1} + D_0(U_1^{-1} - U_2^{-1})$ and the model has one fewer independent parameter. T_{*1} and T_{*2} are the times of the start of the spread projected to the source, $D = 0$. In the case of T_{*2}, this is outside the range of relevance of the second part of the fit, so it may be more useful to consider the time T_{*0} when the Neolithic started spreading at speed $U = U_2$ (i.e. for $D > D_0$); this is $T_{*0} = T_{*2} + D_0/U_2$. This quantity is also given in the Table 4.2.

For simplicity, we only consider these models with the data using percentile values for each bin. To identify whether or not the models with variable speed are an improvement on the simpler constant-speed fit — and also to choose the best value of D_0 for the variable-speed models — we use a form of the Information Criterion ([33, 93]): the second-order ('corrected') Akaike Information Criterion, $\mathrm{AIC_c}$. This criterion is more appropriate for small data sets than the first-order Akaike Information Criterion, or the Bayesian Information Criterion ([10, 33]); it is therefore more suitable for our fits, with a relatively small number of bins. It is given by

$$\mathrm{AIC_c} = n \ln\left(\widehat{X}^2/n\right) + \frac{2pn}{n-p-1} \ , \tag{4.4}$$

where p is the number of free parameters in the model: $p = 2$ for the constant-speed fit (T_*, U); $p = 4$ for the continuous variable-speed fit ($T_{*,1}$, U_1, U_2, D_0); $p = 5$ for the discontinuous variable-speed fit ($T_{*,1}$, $T_{*,2}$, U_1, U_2, D_0).

This definition of $\mathrm{AIC_c}$ is formulated so that a smaller value of $\mathrm{AIC_c}$ corresponds to a better model. The second term in $\mathrm{AIC_c}$ acts to penalize models with more free parameters, which should of course be better able to fit the data; a more complicated model is warranted only if it provides a smaller value for the Information Criteria (i.e., if the decrease in $\widehat{X}^2$ exceeds the increase in the penalty term). In addition to $\mathrm{AIC_c}$, we have repeated the analysis using a Bayesian Information Criteria (BIC) [33] — which differs only in the relative size of the penalty term — and obtain similar conclusions.

For both type of variable-speed model (continuous or discontinuous), we varied D_0 within the range 900, 1100, ..., 3300 km, and chose the value that minimises $\mathrm{AIC_c}$. (Within each type of model the number of free parameters does not change, so this is simply equivalent to choosing the value of D_0 that minimises $\widehat{X}^2$.) The best models of each type are shown in Table 4.2 and Fig. 4.16.

Using $\mathrm{AIC_c}$ to compare between the different types of models, the best model is the constant-speed model with $U = 0.59$ km/yr and $T_* = 10,283$ yr BCE, which has $\mathrm{AIC_c} = 24.6$. The discontinuous fit with $D_0 = 1100$ km is a close second, with $\mathrm{AIC_c} = 25.8$ (see Table 4.2). Interestingly, this fit would be preferred over the constant-speed fit on the basis of the corresponding BIC values. For this discontinuous fit, the speed U at $D < D_0$ is negative, formally corresponding to spread towards the supposed source of Gesher; this part of the fit is just an artefact of our simple model based on a point source, when an extended source in the Fertile Crescent would be more realistic. But excluding the range $D < D_0$, this fit suggests that an extended source at $D_0 \approx 1000$ km from Gesher may be a good effective source, and the fitted speed $U_2 = 0.68$ km/yr may be a good measure of the spread to the East of the Zagros mountains. A model explicitly using an extended source merits future study.

In terms of the maximum likelihood interpretation of the fitting, an appropriate definition

Table 4.2: Best-fit parameters and fit statistics for models with constant and variable speed of spread calculated using the percentile method for the 'One Site – Many Dates' database.

Fitted parameter	Dimension	Constant speed	Variable speed	
			Continuous fit	Discontinuous fit
U or U_1 (U_2)	km/yr	0.59	0.53 (0.84)	−2.5 (0.68)
D_0	km	–	2500	1100
T_* or T_{*1} (T_{*2})	kyr BCE	10.3	10.5 (8.7)	9.5 (9.6)
T_{*0}	kyr BCE	–	5.8	8.0
$\mathrm{AIC_c}$		24.6	29.2	25.8

of $\mathrm{AIC_c}$ (incorporating the second order correction, and differing from the definitions in [93, 94] to follow the sense convention adopted by [33]), is

$$\mathrm{AIC_c} = -\widehat{L} + \frac{pn}{n-p-1} \ .$$

Note that, in the context of maximum likelihood fits, σ is considered as a parameter of the model, and the $\mathrm{AIC_c}$ value for the constant-speed case would normally be calculated with $p = 3$ (U, T_*, σ). (Similarly, the continuous and discontinuous variable-speed fits would have $p = 5$ and $p = 6$. Of course, this difference in the definition of p amounts only to a constant factor, and so does not affect the comparison of $\mathrm{AIC_c}$ values between the different models considered.) Comparisons of the $\mathrm{AIC_c}$ values (and also the corresponding BIC values) for the best constant-speed and variable-speed fits obtained using the maximum likelihood fitting lead to the same conclusions as for the weighted least squares fits; the simple constant-speed fit is the best model.

We note for completeness that, within the maximum likelihood fitting, we also carried out variants of the variable-speed fits, with different variances allowed on either side of D_0 (i.e., we then had parameters σ_1, σ_2, rather than the global parameter σ). The results for the variable-speed fits were not significantly different, and the constant-speed fit was still preferred.

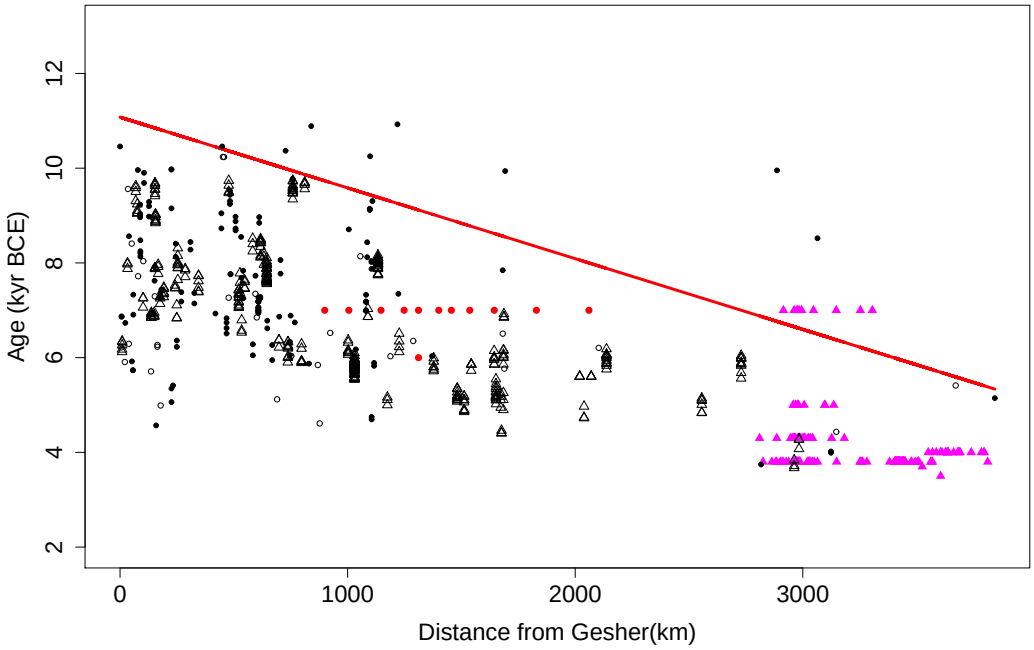


Figure 4.14: A linear envelope fit to the data using the weighted dates yields the average Neolithic dispersal speed $U = 0.67\,\mathrm{km/yr}$. The filled circles (red) and triangles (magenta) show the archaeologically dated sites from Iran and the Indus valley Civilization, respectively. Dependent on the method of choice of dates, the sites with ^{14}C dates are plotted with: open triangles (method 'C'), open circles (method 'O') and filled black circles (method 'A').

4.3 Results

When the weighted data were used, the best linear fit of the form given in Eq. (4.1) was obtained using all of the data; the binning is thus only used to calculate the weights w_i. The resulting envelope, shown in Fig. 4.14, corresponds to $U = 0.67\,\mathrm{km/yr}$ and $T_* = 11,000\,\mathrm{yr}$ BCE, with the goodness of fit quantified by the coefficient of determination $R^2 = 0.77$ (the closer R^2 is to unity, the smaller the unexplained variance of the data point deviations from the fit).

When implementing the percentile method, we have considered various percentile levels from 70% to 97% to ensure that the results are robust in this respect (see Fig. 4.8), and present those for the 95% level in Fig. 4.15. This envelope has $U = 0.59\,\mathrm{km/yr}$ and

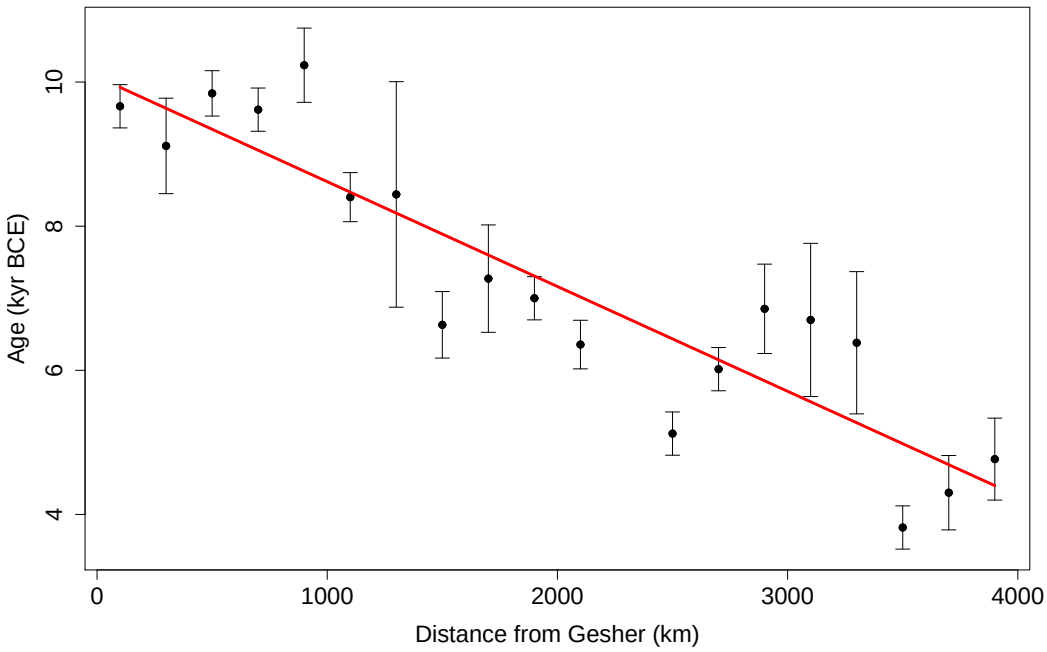


Figure 4.15: A linear envelope fit to the data using the 95-percentile points leads to a Neolithic dispersal speed $U = 0.59\,\mathrm{km/yr}$.

$T_* = 10,300\,\mathrm{yr}$ BCE, with $R^2 = 0.89$. These results are close to those obtained from the weighted data as described above, lending additional confidence in the reliability of our statistical procedures.

The Variable speed model that allows for a piece-wise constant dispersal speed (Fig. 4.16 and Table 4.2), does not produce a statistically significant improvement of the results, but provides indications that the systematic spread might be better modelled as having started from a distributed source at a distance of about 1,000 km from Gesher; i.e. from the vicinity of the Zagros mountains, which appears perfectly plausible.

The values of U and T_* given above were obtained using the largest dataset available to us. As noted above, however, we verified the robustness and accuracy of the results by also using a single representative date for each cluster. Similarly, we also used significantly modified data sets obtained by excluding the ^{14}C measurements classified as Chalcolithic in the CONTEXT database. These variations in the treatment of the ^{14}C data and in the

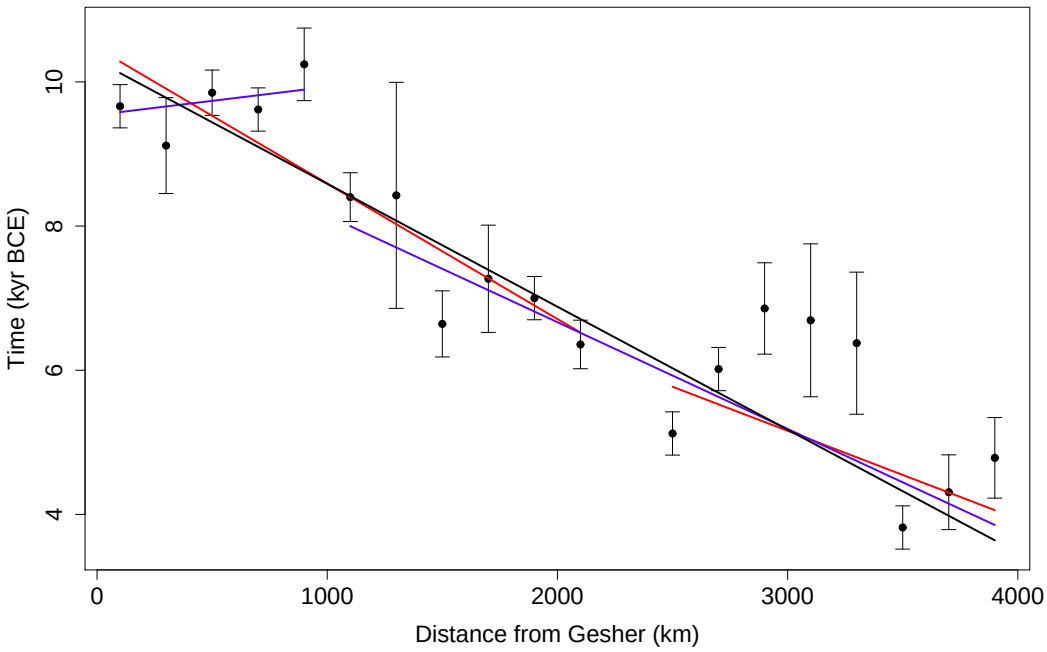


Figure 4.16: The best-fitting dependencies of the earliest Neolithic age T_0 at a distance D from Gesher: the constant-speed fit (black), the continuous variable-speed fit (red) and the discontinuous variable-speed fit (blue). The data fitted are the 95%-ile points of the binned data.

data set result in values of U ranging from 0.58 km/yr to 0.79 km/yr, and this range is a good measure of the accuracy of this estimate. The corresponding range of T_* extends from 11,000 to 8,800 yr BCE.

For comparison, the 95% confidence intervals of U (corresponding to the 2σ range), from the fitting procedures applied to our largest data set are $U = 0.63$–0.71 km/yr when using the weighted dates, and $U = 0.48$–0.76 km/yr when using $T_{\%}$ obtained as the 95%-ile. The comparable ranges for T_* are 11,200–10,800 yr BCE and 11,200–9,400 yr BCE, respectively.

These ranges of uncertainties are affected by the difference in the number of the data points used in the two fits, however: the number of dates in the weighted method by far exceeds the number of bins in the percentile method, as is evident from comparing Figures 4.14 and 4.15. Because of this, the two methods arguably under- and over-estimate the uncertainty of the fit, respectively.

A conservative summary estimates of the average speed and starting date of the Neolithic dispersal from the Near East to the Indian Subcontinent are therefore

$$U = 0.65 \pm 0.10\,\mathrm{km/yr}, \quad T_* = 10,000 \pm 1,000\,\mathrm{yr\ BCE}, \tag{4.5}$$

where the uncertainties quoted come from comparisons of the results obtained from the various analyses mentioned above, and also from variations in the bin width, weighting scale τ and the choice of the percentile-level.

4.4 Discussion

Despite their scarcity, the ^{14}C and archaeological age determinations for early Neolithic sites in Southern Asia exhibit remarkable continuity across the vast region from the Near East to the Indian Subcontinent, consistent with a systematic eastward spread at an average speed of about 0.65 km/yr. It is perhaps not surprising that the rate of spread in Asia may be lower than in Europe, 1 km/yr. Firstly, the arid climate and complicated topography of the Middle East are less favourable for agriculture. Because of this, the early Neolithic settlements in Iran apparently were relatively small and widely separated. (On the other hand, the stronger reliance of the Neolithic population on herding in arid, mountainous areas, with ensuing long-distance seasonal movements, might enhance the population mobility.) Secondly, the advancement of the Neolithic in Europe was facilitated by accelerated propagation along the major European rivers (first of all, the Danube and Rhine) and the Mediterranean coastline [46, 144]. There are no major rivers in Iran and Afghanistan that could play a similar role; and the southern coastline of Iran is more arid than the country's interior (because of the predominant northerly winds), so that the known Neolithic sites in Iran avoid the southern coastal area.

The model of the Neolithic dispersal suggested here applies at the largest, global spatial and temporal scales, as it assumes that the spread proceeded at the same speed in all directions irrespective of the local environment. Given the obvious simplicity of this model,

its success in capturing the salient features of the data is encouraging. This does not diminish the need for a more detailed analysis with allowance for the local environment and palaeoclimate; but our results provide important justification, and a basis, for more sophisticated mathematical modelling.

Dispersal concepts summarily labelled as 'wave of advance' models, similar to that considered here, are often claimed to exclude directed individual movements and to be inconsistent with a 'leap-frog' colonization such as that along major waterways. In fact, both these effects, together with many other realistic refinements, can easily be included into the models without changing their conceptual and mathematical nature. Our discussion of the Neolithic dispersal can apply to demic diffusion, cultural transmission or a combination of the two. These processes only differ in the mechanisms and efficiency (speed) of the spread, but their mathematical models and spatio-temporal manifestations are closely related and only differ in details [61].

Chapter 5

Two Routes of the Neolithic Advance

Not many studies have been carried out to investigate the possibilities of connections between the Neolithic of the Near East and that of the Indian subcontinent. Our study of the combined dataset of the archaeological and the radiometric dates from the Middle East and the Indian subcontinent states that an approximate linear relationship can be established from the conventional center of Gesher near the Mediterranean to the Indian subcontinent. Although locally the Neolithic propagation speed could vary, the average speed range for the Neolithic spread from the Near East to the Indian subcontinent is calculated to be relatively narrow 0.48 – 0.79 km/yr [67].

Iran is a country of widely varying landscapes and environments. Iran ranges from a wetter and more moist climate near the Zagros mountain ranges in the west to the semi-arid and arid plateaus with two major deserts (Dasht-e Kavir and Dasht-e Lut) in the east. It is bounded by the Caspian Sea in the north and the Persian Gulf, the Gulf of Oman and the Arabian Sea in the south. These environmental variations must have strongly affected the spread of the Neolithic through Iran. From Fig. 5.1 it can be seen that the wide spatial gap between the archaeological sites in the central and eastern Iran persists after recent archaeological surveys. Therefore, there might be an 'evidence of absence' rather than 'absence of evidence' [23, p. 146], though it has been argued that there remains

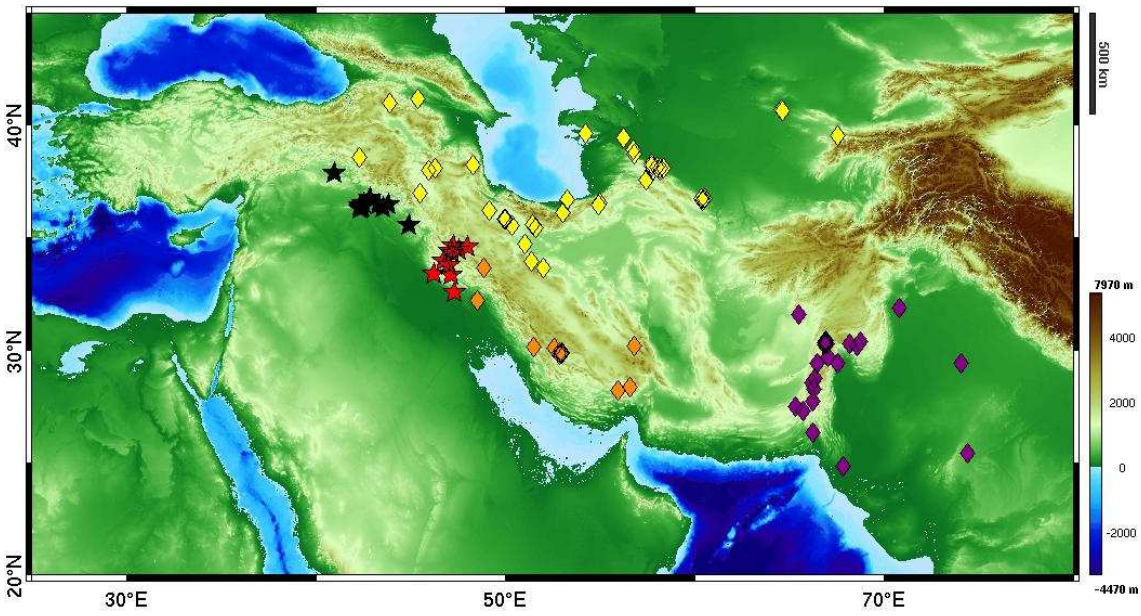


Figure 5.1: Sites shown with black stars are the source sites for the northern route. Sites shown with red stars are the source sites for the southern route. Sites in yellow and orange diamonds are the sites on the northern and the southern routes respectively. Sites in purple diamonds belong to the Indus Valley.

a possibility of finding older occupations in the southern Zagros region through intensive fieldwork [133]. Figure 5.1 also shows that, in the region east of Zagros, the sites are located mostly along the northern and southern edges of Iran, exhibiting a forked pattern. In this chapter we quantitatively explore the possibility of two separate Neolithic propagation routes from north and south of the Zagros Mountains.

5.1 Radiocarbon and Archaeological evidence

Since we are interested in studying the spread of the Neolithic from the Fertile Crescent to the Indian subcontinent, the sites in the eastern part of Fertile Crescent are more important. Hence, we discard the sites belonging to the western Fertile Crescent. Also, to analyse the spread to the Indian subcontinent, we need the earliest Neolithic arrival dates in the Indus Valley region. Hence we retain only the sites that are older than 4,000 BCE and exclude the later sites as secondary. Thus, we select a subset of the combined

database from [67] and the additional Neolithic sites from the Kopet Dag region.

The data is separated in two categories: sites in the source regions for the northern and southern Neolithic dispersals, and sites on the northern and southern routes (see Fig. 5.1). The source sites for the northern and the southern routes are listed in Tables 5.2 and 5.3 respectively (black and red stars in Fig. 5.1). Similarly, the sites on the northern and southern routes are listed in Tables 5.4 and 5.5 respectively (yellow and orange diamonds in Fig. 5.1). The Indus Valley Civilisation (IVC) sites used in the analysis are given in a separate table (Table 5.6, purple diamonds in Fig. 5.1).

5.2 Time to Move

Along with the topography, the precipitation (and as a consequence the vegetation) also varies widely in West Asia. With the notable exception of the Fertile Crescent, the area is mostly arid or semi arid. Due to the presence of rich soil and a good amount of rainfall, the Fertile Crescent is the most agriculturally productive area in West Asia.

Figures 5.2 and 5.3 show selected cumulative spatio-temporal snapshots of the Neolithic sites from 10,900 BCE to 4,000 BCE on the topographic background. To make these plots, we used one Neolithic date per site using the selection procedure from Chapter 4.

The two frames of Figures 5.2a and 5.2b, with seven sites older than 10,000 BCE (M'lefaat, Cayönü, Gesher, Mureybet, Chogha Bonut, Chogha sefid, Tbeik) show that these sites highlight the crescent shape of the region. From the later snapshots it is seen that with the exception of the two sites (Aq Kupruk in present-day Afghanistan and Mehrgarh in present-day Pakistan), the Neolithic sites are confined within the Fertile Crescent till about 7300 BCE. The spread is outside this region is only noticeable later than 7300 BCE. It is in this frame (Fig. 5.3a) that we see the first Neolithic sites appearing in the Baluchistan region of the Indian subcontinent. The sites that originated in 7300–7000 BCE in the north and east of the Fertile Crescent are mostly dated with the comparative dating technique. Between 7,000 BCE and 6,100 BCE, the plots show that the spread is

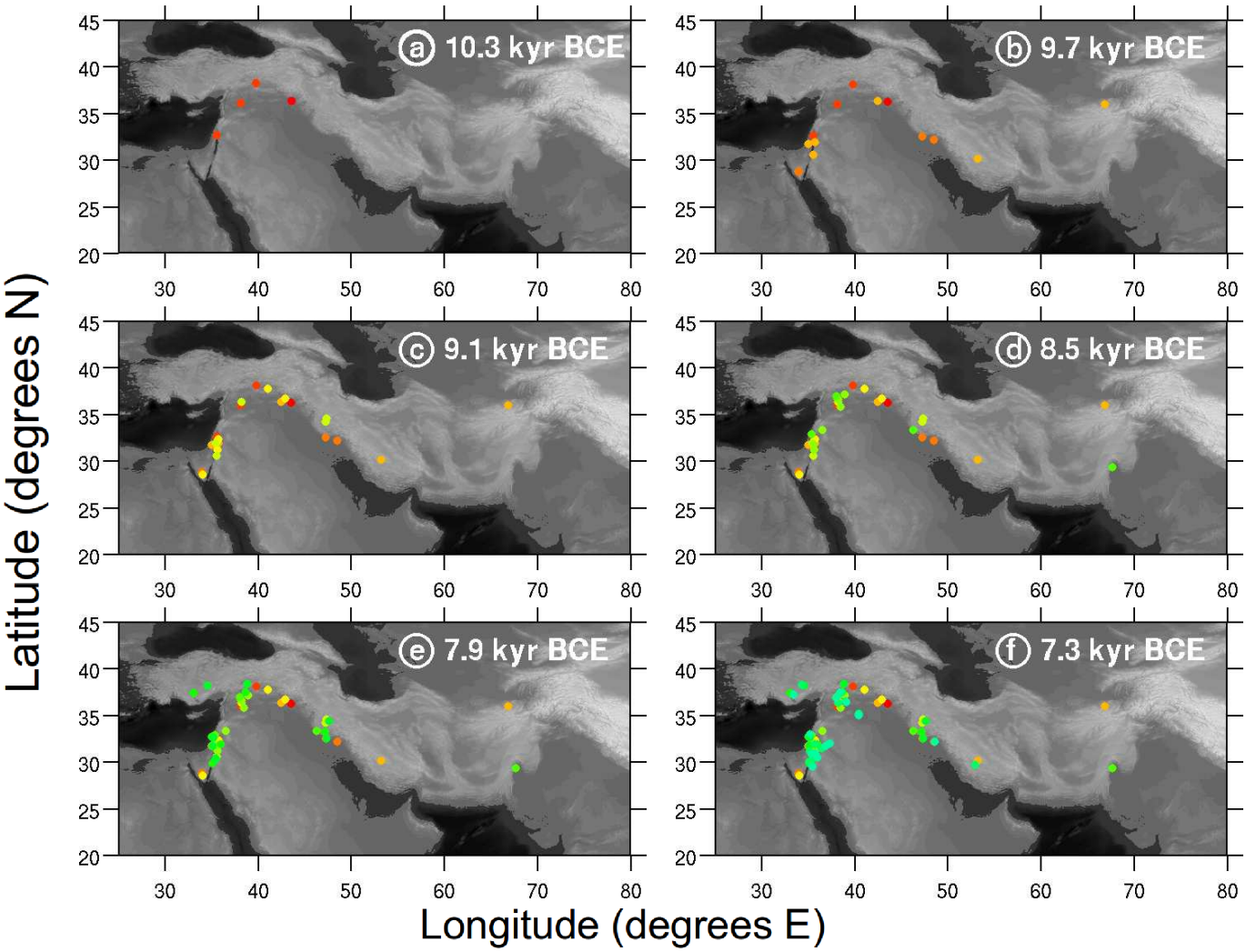


Figure 5.2: The spatio-temporal maps show the Neolithic sites on a grey topographic background. The sites with ^{14}C dates are shown in filled circles. The colour scheme for sites goes from warm to cool colours. The oldest sites are shown in red filled circles and progressively the colour changes to blue. The maps are 'cumulative', i.e. they include all the sites from 10,900 yr BCE till the age shown on the top of the plot. The six frames show all the sites in our database till 7,000 BCE. See the text for a detailed description of the plots.

restricted to the modern-day countries of Turkey, northern Syria, northern Iraq. In the next frame (Fig. 5.3c, 5800 BCE) an eastward spread can be seen in the north with the site of Ayakagytma in Uzbekistan. The later frames show newer sites near the already existing sites in the northern Iran (near the Caspian Sea) and a progression of sites is seen in the Fars and Kerman provinces of southern Iran. By this time the IVC in the Indian subcontinent is already thriving and in the last frame (Fig. 5.3f) a proliferation of sites is

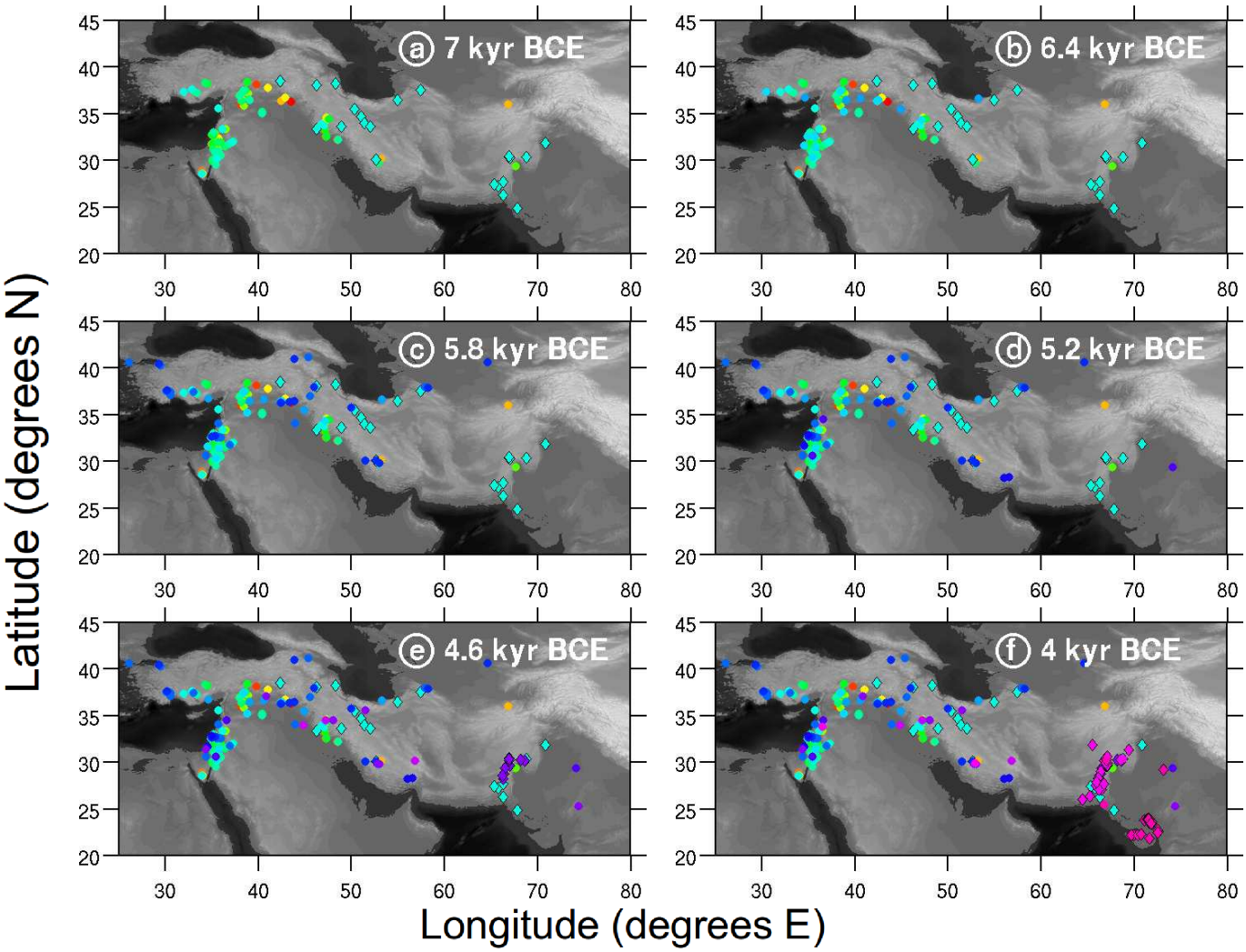


Figure 5.3: The plots show the 'cumulative' site distributions till 4,000 yr BCE. The sites dated with comparative dating are shown in diamonds. See Fig. 5.2 and also the text for a detailed description.

seen in the Indian subcontinent.

As a summary, the Neolithic occupation is restricted to the Fertile Crescent for the first 4000 year period starting from 10,900 yr BCE. It is around 7000 BCE that Neolithic sites are seen east of the Fertile Crescent. In addition to population pressure and resource depletion, a climate deterioration event noted around 6200 BCE [139] might have been a prominent cause for search of a habitable zone in the less hospitable areas outside the Fertile Crescent.

Table 5.1: The sites in addition to datasets in [67]. The date selection keys: IR – Archaeological dates from Iran and regions close to Iran, RC_O– Only a single ^{14}C date available for the site.

Site Name	Latitude (° N)	Longitude (° E)	Neolithic Date (kyr BCE)	Sigma (kyr)	Date Selection Key	Reference
Bacha well	39.43	56.23	6.10	0.30	IR	[73]
Bami	38.72	56.82	6.10	0.30	IR	[73]
Chagylly	36.70	60.47	6.10	0.30	IR	[73]
Chakmakli	36.73	60.55	6.10	0.30	IR	[73]
Chopan	38.07	58.20	6.10	0.30	IR	[73]
Gadymi	36.63	60.43	6.10	0.30	IR	[73]
Gievdzhik	38.17	57.72	6.10	0.30	IR	[73]
Jebel	39.63	54.25	5.81	0.18	RC_O	[49, p. 249]
Kelyata	38.27	57.72	6.10	0.30	IR	[73]
Monjukli	36.75	60.35	6.10	0.30	IR	[73]
Naiza	38.92	56.75	6.10	0.30	IR	[73]
Sang-i Chakhmak	36.47	54.93	6.73	0.38	RC_O	[73]

Table 5.2: Sites within the source region for the Northern route. The mean Neolithic start date for the source region is 7.63 kyr BCE. The date selection keys: RC_O – Only a single ^{14}C date available for the site, RC_A – date selected using the 'Average' method, RC_C– date selected using the 'Clustering' method.

Name	Latitude (° N)	Longitude (° E)	Neolithic Date (kyr BCE)	Sigma (kyr)	Date Selection Key	Reference
Körtik Tepe	37.83	40.97	9.56	0.10	RC_C	[26]
Qermez Dere	36.38	42.45	9.76	0.18	RC_A	[1]
Nemrik 9	36.73	42.88	9.65	0.06	RC_C	[1]
M'lefaat	36.31	43.54	10.89	0.26	RC_A	[1]
Ali Agha	36.45	43.82	5.84	0.18	RC_O	[1]
Gawra	36.43	43.34	5.87	0.18	RC_A	[1]
Thalathat II	36.48	42.50	6.74	0.28	RC_A	[1]
Jarmo	35.56	44.92	6.52	0.18	RC_O	[1]
Sotto	36.28	42.37	6.19	0.20	RC_A	[1]
Magzaliyah	36.39	42.33	6.89	0.18	RC_A	[1]
Yarim Tepe I	36.31	42.43	6.03	0.19	RC_O	[1]

Table 5.3: Sites within the source region for the Southern route. The mean Neolithic start date for the source region is 7.78 kyr BCE. The date selection keys: IR – Archaeological dates from Iran and regions close to Iran,RC_O – Only a single ^{14}C date available for the site, RC_A– date selected using the 'Average' method, RC_C – date selected using the 'Clustering' method.

Name	Latitude (° N)	Longitude (° E)	Neolithic Date (kyr BCE)	Sigma (kyr)	Date Selection Key	Reference
Ganj Dareh	34.42	47.57	8.02	0.37	RC_C	[1]
Sheikh-e Abad	34.61	47.27	9.31	0.18	RC_A	[45]
Ali Kosh	32.56	47.32	7.98	0.18	RC_A	[1]
Asiab	34.30	47.19	9.13	0.18	RC_A	[1, 96]
Sarab	34.38	47.09	6.92	0.18	RC_A	[1, 96]
Chia Sabz	33.34	47.14	8.27	0.23	RC_A	[118]
Tepe Guran	33.73	47.07	7.13	0.18	RC_A	[96]
Jani	33.95	46.78	7.11	0.24	RC_O	[44]
Čogā Golān	33.38	46.27	8.69	0.18	RC_A	[118]
Seh Gabi	34.58	48.00	5.14	0.18	RC_C	[96]

5.3 Analysis and Discussion

From the cumulative plots in Figures 5.2 and 5.3, it is seen that the Neolithic dispersal east of the Zagros Mountain ranges primarily occurred near the northern and the southern edges of Iran giving the appearance of two branches of a fork. In the following text we quantitatively establish the existence of the two separate Neolithic propagation routes.

In the earlier studies of the Neolithic dispersals, a single site was identified and used as a conventional source for the Neolithic spread (e.g. Jericho in [17] and Gesher in [67]). It is corroborated by archaeological studies that the path to neolithisation was traversed locally in small groups [97]. The central Zagros is one such 'formative zone' [102, p. 19],[97], where the Neolithic development is said to have taken place without any influence of outside regions.

In the following analysis we introduce these 'spatial source regions' rather than using a single archaeological site as the origin. By inspection of the Neolithic data we see that two areas, one in the northern and the other in the eastern Fertile Crescent (central Zagros)

Table 5.4: Sites on the northern route of Neolithic dispersal. The date selection keys: IR – Archaeological dates from Iran and regions close to Iran, RC_O– Only a single ^{14}C date available for the site, RC_A – date selected using the 'Average' method, RC_C – date selected using the 'Clustering' method.

Name	Latitude (° N)	Longitude (° E)	Neolithic Date (kyr BCE)	Sigma (kyr)	Date Selection Key	Reference
Aq Tappe	37.57	57.42	7.00	0.30	IR	[21]
Arisman	33.67	52.00	7.00	0.30	IR	[21]
Arukhlo 1	41.00	43.90	6.03	0.18	RC_O	[1]
Ayakagytma	40.70	64.60	5.88	0.23	RC_C	[131]
Bacha well	39.43	56.23	6.10	0.30	IR	[73]
Bami	38.72	56.82	6.10	0.30	IR	[73]
Belt	36.70	53.30	6.51	0.97	RC_O	[96]
Chagylly	36.70	60.47	6.10	0.30	IR	[73]
Chakmakli	36.73	60.55	6.10	0.30	IR	[73]
Cheshmeh Ali	35.60	51.45	5.14	0.18	RC_C	[57]
Chopan	38.07	58.20	6.10	0.30	IR	[73]
Gŭsa Tappe (Ardabil)	38.25	48.28	7.00	0.30	IR	[21]
Gadymi	36.63	60.43	6.10	0.30	IR	[73]
Gievdzhik	38.17	57.72	6.10	0.30	IR	[73]
Golestān Park	38.08	46.28	7.00	0.30	IR	[21]
Hajji Firuz	37.00	45.50	6.17	0.19	RC_A	[1]
Jebel	39.63	54.25	5.81	0.18	RC_	[49, p. 249]
Jeitun	38.00	58.20	5.97	0.40	RC_C	[96]
Kelyata	38.27	57.72	6.10	0.30	IR	[73]
Monjukli	36.75	60.35	6.10	0.30	IR	[73]
Naiza	38.92	56.75	6.10	0.30	IR	[73]
Qomrud	34.73	51.07	7.00	0.30	IR	[21]
Sang-i Chakhmak	36.47	54.93	6.73	0.38	RC_	[73]
Shōmu	41.16	45.38	6.35	0.18	RC_O	[1]
Tapeh Sialk	33.97	51.40	7.00	0.30	IR	[96]
Tappe J̌olbar	38.56	42.32	7.00	0.30	IR	[21]
Tappeh Deh Keir	36.53	54.98	7.00	0.30	IR	[21]
Tappeh ozbaki	35.54	50.34	7.00	0.30	IR	[21]
Tepe Chahar Boneh	35.81	50.03	5.83	0.18	RC_C	[109]
Tepe Ebrahim Abad	36.12	53.05	5.24	0.18	RC_C	[109]
Tepe Pardis	35.45	51.60	5.18	0.18	RC_C	[96]
Tepeh Khaleseh	36.19	49.17	6.00	0.30	IR	[11, 134]
Togolok	38.00	57.80	6.21	0.19	RC_O	[1]
Yanik	37.99	45.95	5.86	0.18	RC_A	[1]
Zagheh	35.82	49.95	6.03	0.19	RC_A	[96]

Table 5.5: Sites on the southern route of Neolithic dispersal.The date selection keys: IR – Archaeological dates from Iran and regions close to Iran, RC_O – Only a single ^{14}C date available for the site, RC_A – date selected using the 'Average' method, RC_C – date selected using the 'Clustering' method.

Name	Latitude (° N)	Longitude (° E)	Neolithic Date (kyr BCE)	Sigma (kyr)	Date Selection Key	Reference
Abadah	33.97	44.83	4.61	0.11	RC_O	[1]
Bakun B	29.85	52.83	5.12	0.11	RC_C	[96]
Chogha Mish	32.21	48.55	7.35	0.18	RC_A	[96]
Čoḡā Mīš	32.21	48.55	7.35	0.18	RC_A	[13]
Deh Hajj	33.68	48.88	7.00	0.30	IR	[21]
Iblis	30.20	56.80	4.81	0.18	RC_A	[96]
Mushki	29.78	52.90	7.85	0.35	RC_A	[1]
Sawwan I	34.12	43.93	6.24	0.03	RC_C	[1]
Siahbid	34.50	47.25	4.72	0.04	RC_A	[96]
Tall-e Jari B	29.85	52.96	6.04	0.18	RC_A	[96]
Tall-e Bakun A	29.90	52.90	4.43	0.18	RC_A	[14]
Tall-e Jari A	29.90	53.00	5.09	0.21	RC_C	[14]
Tepe Gaz Tavila	28.34	56.58	5.60	0.18	RC_A	[7]
Tepe Rahmatabad	30.11	53.06	6.92	0.01	RC_C	[27]
Tepe Yahya	28.20	55.98	5.60	0.18	RC_A	[96]
Tol-e Basi	30.08	52.59	5.97	0.18	RC_A	[96]
Tol-e Nurabad	30.12	51.52	5.83	0.18	RC_A	[2]
Tang-e Bolaghi	30.15	53.13	9.94	0.18	RC_A	[96]

Table 5.6: Neolithic sites from the Indus Valley [110].The date selection keys: IN – Archaeological dates from Indus Valley region, RC_O – Only a single ^{14}C date available for the site, RC_A – date selected using the 'Average' method.

Name	Latitude	Longitude	Neolithic Date		Date Selection
	(° N)	(° E)	(yr BCE)		Key
Anjira	28.28	66.32	5.00	0.30	IN
Bagor	25.40	74.40	5.15	0.43	RC_A
Baleli	30.33	66.88	7.00	0.30	IN
Duki Mound	30.17	68.57	5.00	0.30	IN
Gumla	31.88	70.83	7.00	0.30	IN
Isplinji Two	29.69	67.04	5.00	0.30	IN
Jebri Damb Two	27.29	65.75	7.00	0.30	IN
Kalibangan	29.40	74.10	5.41	0.19	RC_O
Kargushki Damb	27.48	65.32	7.00	0.30	IN
Kasiano Dozakh	30.45	66.93	7.00	0.30	IN
Khakhar Buthi	26.32	66.27	7.00	0.30	IN
Kili Ghul Mohammad	30.28	66.97	7.00	0.30	IN
Kuki Damb	28.75	66.35	5.00	0.30	IN
L-2	30.30	68.17	5.00	0.30	IN
L-3	30.30	68.20	5.00	0.30	IN
Mehrgarh	29.42	67.58	8.52	0.43	RC_A
Nal	27.73	66.27	7.00	0.30	IN
Neghar Damb	28.27	66.30	5.00	0.30	IN
Q-17	30.23	66.90	5.00	0.30	IN
Q-25	30.35	66.93	5.00	0.30	IN
Rana Ghundai	30.40	68.75	7.00	0.30	IN
Saiyid Maurez Damb	29.43	66.45	5.00	0.30	IN
Siah Damb,Surab	28.57	66.18	5.00	0.30	IN
Tharro Hill	24.83	67.82	7.00	0.30	IN
Thok Valley One	28.73	66.35	5.00	0.30	IN

region (sites shown in black and red stars in Fig. 5.1) are likely to have been the source regions for the northern and the southern branches of the Neolithic fork. Tables 5.2 and 5.3 list the sites in the source regions for the northern and the southern dispersal respectively. Tables 5.4, 5.5 and 5.6 list the sites on the northern route, the southern route and the sites belonging to the Neolithic of the Indus Valley region, as used in the current analysis. Using the procedures from Chapter 4, we select two subsets of sites for each of the two routes, one with the mean date of the oldest cluster of Neolithic dates (One Site – One Date), the other with all the dates belonging to the oldest cluster of the Neolithic dates (One Site – Many Dates), along with the respective uncertainties. Further analyses use these subsets with inclusion and exclusion of the Indus data. The distances of the northern and the southern route sites are the minimum geodesic distances calculated from the respective northern and southern site groups in the source region.

We group the data into intervals or bins according to the distances of the sites from the respective source regions. In each case, we find the earliest Neolithic arrival times at every point, by fitting an upper envelope to the data using the percentile and weighting procedures from Chapter 4. However, we retain only the percentile plots as the results obtained from them are robust.

The number of data points, especially in the subsets of the One Site – One Date dataset, is small. Similar to the treatment followed in Chapter 4, to estimate the optimum bin width, we calculate the variations in the Neolithic propagation speeds and the time of the Neolithic commencement by varying the distance bin widths from 100 to 400 km in steps of 50 km for subsets of both datasets (Fig. 5.8 and Fig. 5.9).

We consider first the results for the subsets of the One Site – One Date dataset (Fig. 5.8). For these datasets, the number of empty bins is large in the case of smaller distance intervals (100, 150 and 200 km). For the dataset "South with Indus (One date)" for the bin width of 250 km, for all the percentiles, the time values obtained for each bin by bootstrapping procedure as in Chapter 4 are such that the time-distance fits are nearly horizontal giving very large (positive / negative) values for the speeds and the starting

times, and thus the results cannot be accepted. The smallest variation in the propagation speeds obtained using different bin widths occurs for the bin widths of 300 and 350 km. The same is true for the Neolithic arrival times. Hence, the best results are obtained for the intervals of 300 and 350 km for the One Site – One Date datasets. We give here the results for the bin width of 350 km.

A similar analysis for the One Site – Many Dates datasets shows that in case of "South with Indus (all dates)", "North no Indus (all dates)" and "North with Indus (all dates)" (see Fig. 5.9) some bin widths display similar behaviour as mentioned earlier, i.e. the time values obtained after bootstrapping are such that the time-distance plots are almost horizontal giving very large positive or negative values for speed and starting times, and cannot be relied upon. Thus, the best results are obtained for bin widths of 200 and 250 km, and we use a bin width of 200 km in this case.

Table 5.7 summarises the speeds and start times obtained from the percentile-based envelope procedure for the northern and southern routes with and without the Indus sites. Figures 5.4 and 5.5 show the percentile plots for the northern and the southern routes.

5.3.1 The two routes of the Neolithic dispersal

The average Neolithic ages in the two source regions for the northern and the southern routes are calculated to be 7.78 and 7.63 kyr BCE respectively (Tables 5.2 and 5.3). So from both the northern and the southern source regions, the Neolithic could have started spreading outside the Fertile Crescent at around the same time. The upper envelope fits for the sites on the southern route (sites from Table 5.5) give the average Neolithic commencement time in the source region to be $T_S^* = 7.2$ kyr BCE with an average propagation speed $U_S = 0.66$ km/yr. This speed is consistent with the average spread speed of 0.65 km/yr calculated in Chapter 4 with the Neolithic spread starting from the conventional source of the site of Gesher at about 10,000 BCE. The northern route tells a different story. The upper envelope fits to the northern route data (sites from Table 5.4) give an average Neolithic commencement time of $T_N^* = 7.1$ kyr BCE with an average speed

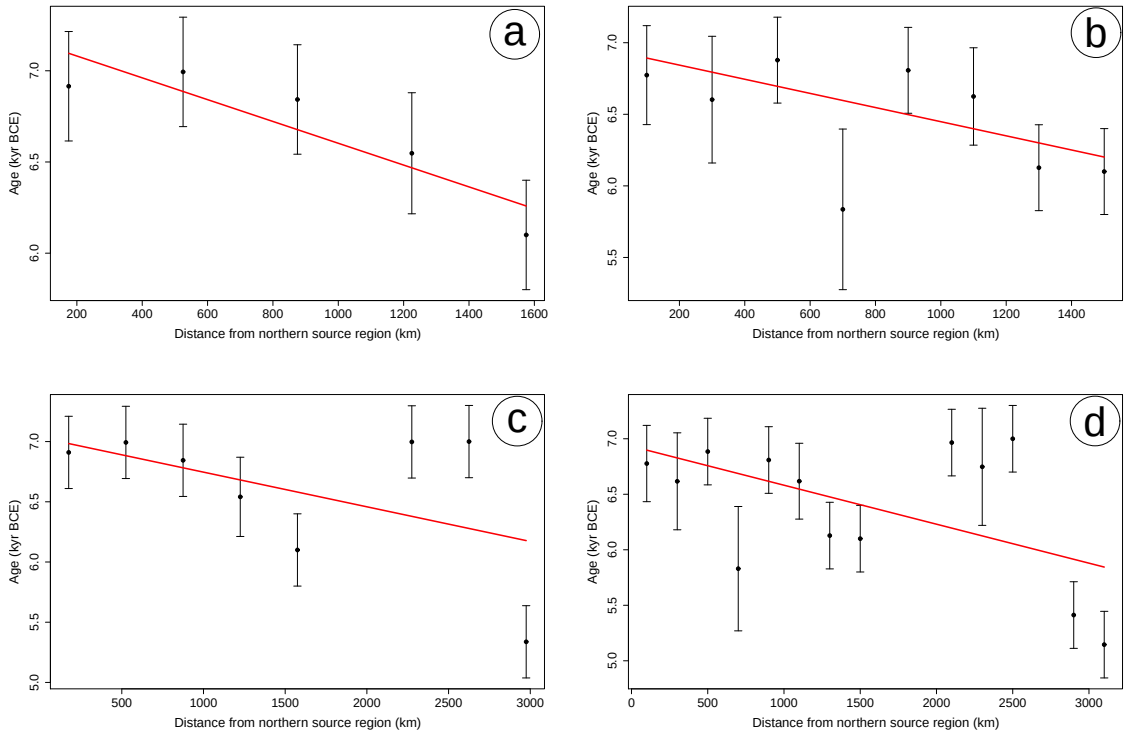


Figure 5.4: Northern route: the envelope plots generated using the percentile method from [67]. Panel a is the plot for the 'One Site – One Date' dataset and panel b is the plot for the 'One Site – Many Dates' dataset, excluding the Indus dates. Panel c is the plot for the 'One Site – One Date' dataset and panel d is the plot for the 'One Site – Many Dates' dataset,including the Indus dates. The suggested route is shown by the yellow arrow in Fig. 5.7

.

$U_N = 1.84$ km/yr. This speed is more than twice the propagation speed for the southern route. There are significant differences in the climates of northern and southern Iran. Northern Iran, near the Caspian Sea, exhibits Mediterranean climate which is warmer and wetter. In contrast, southern Iran is more hot, dry and arid, a semi-desert and desert area that is not as hospitable. This difference in the climates can explain the difference in the Neolithic spread speeds in the two routes. The analysis thus shows that the Neolithic started spreading outside the Fertile Crescent at about 7,000 BCE and indeed there were two separate routes for the Neolithic dispersal, one from the northern Iran and south of Caspian Sea (a part of the future Silk Road) and the other through the semi-desert regions of Iran in the Fars and Mamasani districts (red arrow in Fig. 5.7). The propagation via

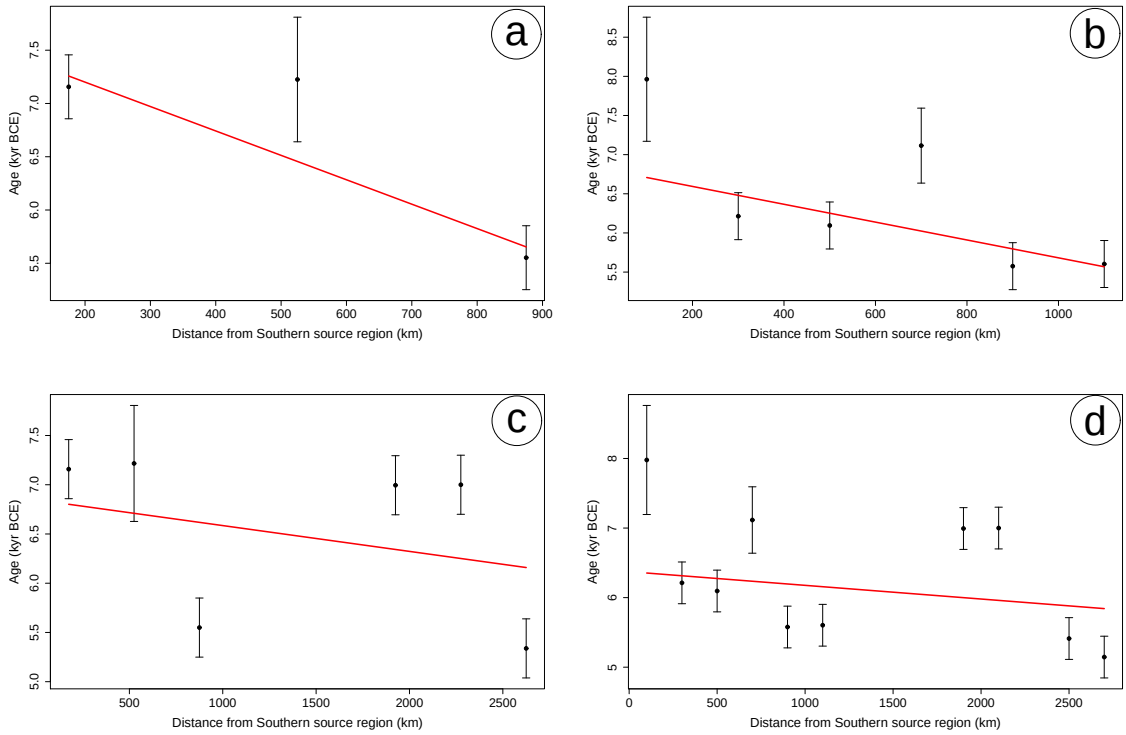


Figure 5.5: Southern route: the envelope plots are generated using the percentile method from [67]. Panel a is the plot for the 'One Site – One Date' dataset and panel b is the plot for the 'One Site – Many Dates' dataset, excluding the Indus dates. Panel c is the plot for the 'One Site – One Date' dataset and panel d is the plot for the 'One Site – Many Dates' dataset, including the Indus dates. The suggested route is shown by the red arrow in Fig. 5.7.

the northern route was twice as fast as that via the southern route.

The extension to the northern route

The site of Ayakagytma with the mean Neolithic date 5,900 BCE is in present-day Uzbekistan. This is the only site in the region that is older than 4,000 BCE. The site of Sarazm in Tajikistan is younger with a mean date of 3,700 BCE. There is a wide spatial gap between the northern route sites and the two mentioned sites, and hence are not included in the northern route sites. We separately analyse the northern route sites with and without the inclusion of Ayakagytma and Sarazm. The upper envelope fits to the data (Figures 5.6a and 5.6b including the two extra sites give an average speed of 1.75 km/yr. This speed is

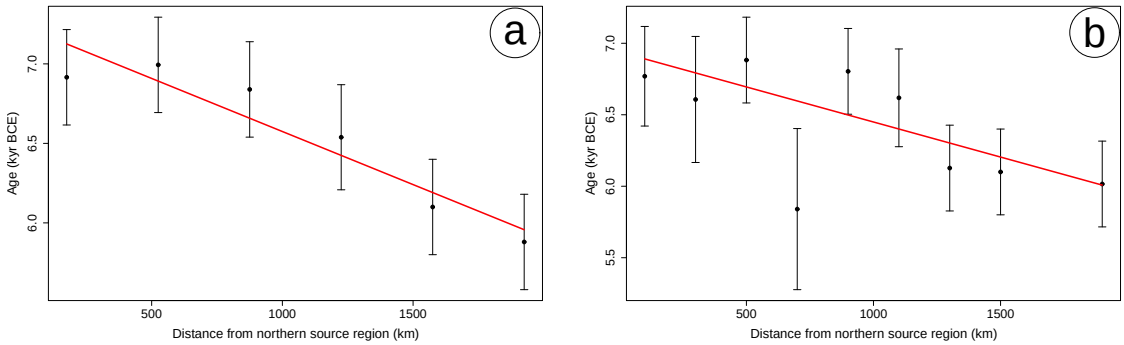


Figure 5.6: Plots for the dataset that includes the northern route Neolithic sites from table 5.4 with the sites of Ayakagytma and Sarazm in the south-eastern Kazyl-Kum plains of present-day Uzbekistan. The speeds obtained are consistent with the speeds obtained by analysing only the north route sites in table 5.4. Thus the analysis suggests that the spread continued further north-east with a speed of 1.75 km/yr. The suggested route is shown by the dashed white arrow in Fig. 5.7.

consistent with the speed obtained by the upper envelope fits to only the northern route sites without them. We therefore suggest one more branch of the Neolithic spread as an extension of the northern route from the south-east of the Caspian Sea to Central Asia. This route is shown by the small dashed white arrow in Fig. 5.7.

5.3.2 Indus Valley, a result of the Neolithic spread from the northern route?

The influence of the Neolithic of the Near East is evident in the Neolithic of the Indus Valley sites. Also, an approximate linear relationship is apparent between the geodesic distance of the Indus Valley sites and their distance from the conventional source of the site of Gesher in the Near East, with an average dispersal speed of about 0.65 km/yr (see Chapter 4). Since there are two separate routes of the eastward Neolithic propagation from the Fertile Crescent, it is of interest to see if the Neolithic reached the Indus Valley along the northern or the southern route. First we analyse the southern route sites combined with the Neolithic Indus sites (sites from Tables 5.5 and 5.6). The upper envelope fits give an average value of Neolithic start time of $T^*_{SI} = 6.60$ kyr BCE with an average speed $U_{SI} = 3.58$ km/yr. The Neolithic start time obtained from the analysis is slightly later than that of the southern source region with the IVC sites excluded. Also the

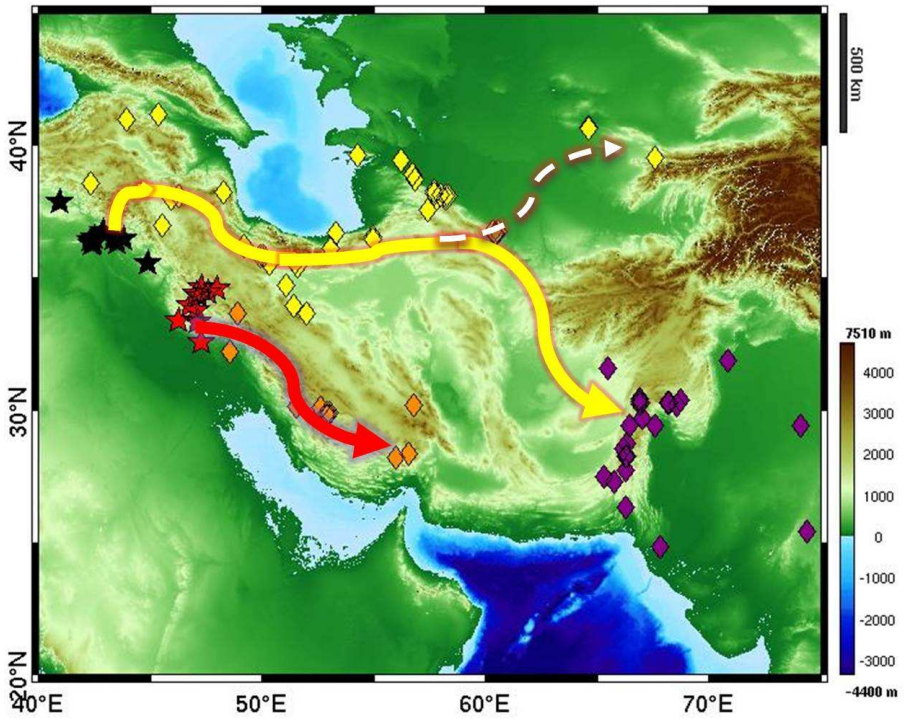


Figure 5.7: The two main paths of the Neolithic dispersal are shown in this figure. The red arrow shows the Neolithic dispersal from the southern Zagros. The yellow arrow shows the movement of the Neolithic from the north of Zagros Mountains (north-central Fertile Crescent) to the Indian subcontinent, and the small dashed white arrow indicates dispersal further north-east.

speed obtained with inclusion of IVC sites is more than four times faster than 0.66 km/yr obtained for the southern route in Iran. Similar analysis after combining the northern route sites with the Indus sites (Tables 5.4 and 5.6) gives an average starting time $T^*_{NI} = 6.93$ kyr BCE with an average speed $U_{SI} = 2.92$ km/yr. This speed is about 1.5 times faster than the speed obtained for the northern route without the Indus Valley dates. The R^2 values are measures of how good the linear fit is to the data. The closer the R^2 value to unity, the better is the fit. Table 5.7 shows the R^2 values obtained for each of the fits. For the southern route sites with Indus, the goodness of fit values obtained are about 0.10 and that for the northern route sites combined with Indus sites are about 0.30. Thus the Neolithic of the Indus Valley is more likely an outcome of the spread from the northern route through Iran. This route is shown with the long yellow arrow in Fig. 5.7.

	One-Site-One Date				One Site – Many Dates		
Data Subset	Number of sites in a subset	Speed (U) (km/yr)	Start Time (T^*) (kyr BCE)	R^2	Speed (U) (km/yr)	Start Time (T^*) (kyr BCE)	R^2
S	18	0.44	7.66	0.88	0.88	6.82	0.41
SI	43	3.19	6.83	0.13	3.97	6.36	0.10
SInM	42	3.81	6.85	0.09	5.08	6.37	0.06
N	34	1.67	7.20	0.81	2.02	6.94	0.86
NI	59	2.95	7.05	0.32	2.89	6.93	0.27
NInM	58	2.84	6.93	0.30	2.85	6.93	0.30
NAS	36	1.51	7.24	0.78	0.91	7.38	0.56
NA	35	1.50	7.24	0.89	2.04	6.94	0.58

Table 5.7: Spread speeds and Starting times calculated by the percentile method, both the 'One Site – One Date' and the 'One Site – Many Dates' datasets. The abbreviations are as follows: **S**: Southern route sites excluding the Indus dates; **SI**: Southern route sites including Indus dates; **SInM**: Southern route sites **with** Indus but **without** Mehrgarh; **N**: Northern route sites excluding the Indus dates; **NI**: Northern route sites including Indus dates; **NInM**: Northern route sites **with** Indus but **without** Mehrgarh; **NAS**: Northern route sites excluding the Indus dates, **with Ayakagytma and Sarazm BOTH included**; **NA**: Northern route sites excluding the Indus dates, **with Ayakagytma included but NOT Sarazm**. The values are calculated using a bin width of 350 km for the One Site – One Date datasets and a bin width of 200 km for the One-Side-All-Dates datasets.

5.4 Conclusion

The terrain of West and South Asia is geographically diverse. On the one hand, the region contains the very productive Fertile Crescent, and on the other it has two large deserts, Dasht-e Kavir and Dasht-e Lut. In such an environment it is impossible that the Neolithic spread uniformly. Our analysis shows two distinct routes of the Neolithic propagation across southern Asia. The first via northern Iran, and possibly continuing towards the Indus Valley; and the second through southern Iran, possibly terminating in the Kerman province. According to our results, the Neolithic dispersal via the northern route progressed with an average speed of 1.85 km/yr, and that via southern route had an average speed of 0.66 km/yr. These two routes are shown by the long yellow arrow and the short red arrow respectively in Fig. 5.7. The red arrow in Fig. 5.7 ends short

of the Indus Valley, and makes it necessary to find if this wide spatial gap really exists or is due to lack of archaeological investigation. The white arrow in the figure shows another possible path of Neolithic dispersal north-east of Iran, pointing towards one more interesting region in the present day Tajikistan, Kyrgyzstan, Uzbekistan and south east Kazakhstan. Exploration of this region might be of interest as this might shed more light on the Neolithic propagation north of the Himalayas.

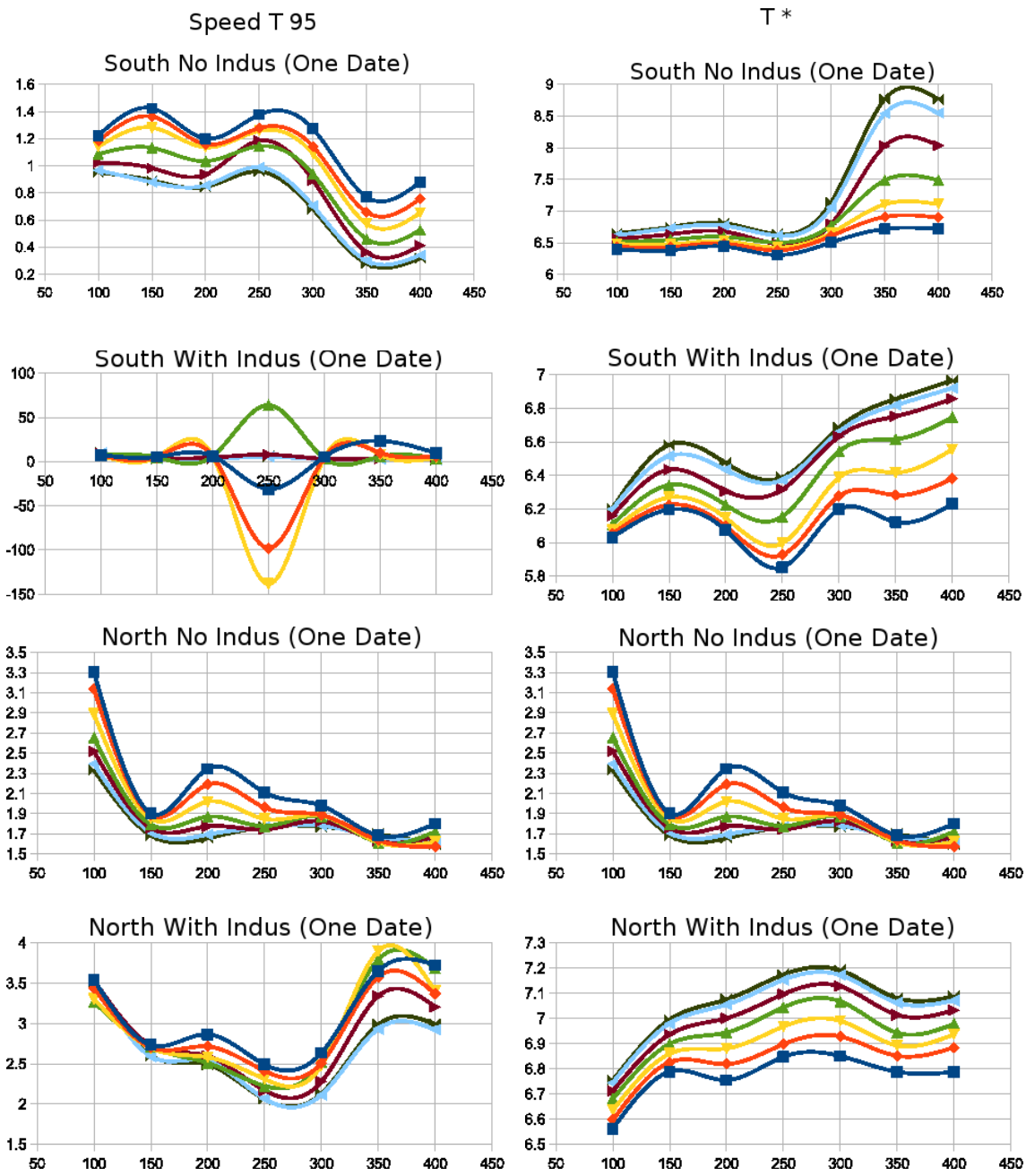


Figure 5.8: The plots of change in speed (left column) and starting times (right column) with change in bin widths (from 100 km to 400 km in steps of 50km) for One Site – One Date datasets for South (S), South With Indus (SI), North (N), and North With Indus(NI), respectively from top to bottom.

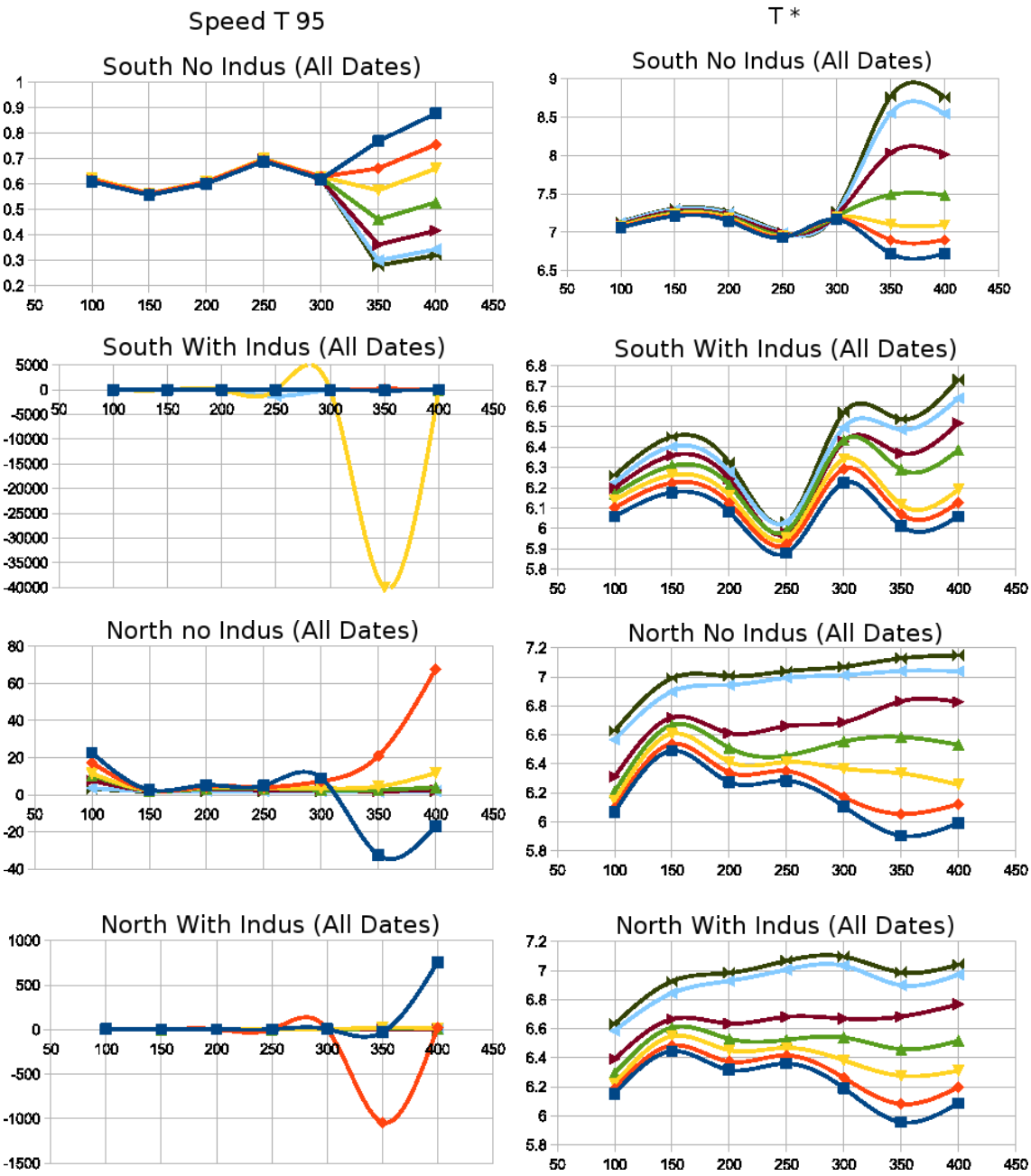


Figure 5.9: The plots of change in speed (left column) and starting times (T*, right column) with change in bin widths (from 100 km to 400 km in steps of 50km) for One Site – Many Dates datasets for South (S), South With Indus (SI), North (N), and North With Indus(NI), respectively from top to bottom.

Chapter 6

Differential Equation Models and Preliminary Results

In this chapter we explore two mathematical models that are based on continuous differential equation models. Both models are modifications of the Fisher-Kolmogorov-Petrowskii-Piscounov (FKPP) equation. The first model (see Section 6.1) uses the FKPP equation with constant and variable diffusivities. The variability of the diffusivity is achieved by introducing it as a function of the palaeovegetation data. The second model (see Section 6.2) is based on the idea of spread of a crop that is of significance to a population and thus represents the spread of that population. The model uses modern altitude and precipitation data. Both the models are currently in the testing stage and only the preliminary results obtained are given in this chapter. The models still require work which is described in the 'Future Work' for each section.

6.1 FKPP Model and Palaeovegetation

The geographically diverse region from the Fertile Crescent to the Indus Valley contains nine different palaeoecological zones. Not all of these zones are favourable for living.

Motivated by this, we introduce a mathematical model for the spread of the Neolithic, which makes use of the palaeovegetation data. Chapter 3 describes the traditional and non-traditional mathematical methods used in studies of population dynamics. Here, the deterministic FKPP model is used to study this Neolithic spread. As given in Section 3.2.2, the Neolithic spread speed is $U = 2\sqrt{\gamma\,\nu}$ with the net growth rate γ and the diffusivity ν. The larger the growth rate and diffusivity, the faster is the spread.

Here, we study the spread of the Neolithic across the region in three different ways:

1. The FKPP model is applied to the whole region from the Mediterranean to the Indus Valley. Firstly, the growth rate, carrying capacity and diffusion are all assumed to be constants. Secondly, we introduce a heterogeneous diffusion based on the palaeovegetation data for the region.

2. The FKPP model is applied to the southern Neolithic route with a source in the Zagros region.

3. In the third and last part, the FKPP model is applied to the northern Neolithic route. As with the southern region, we study this domain with and without the Indus Valley sites.

6.1.1 The FKPP model

We recall that the FKPP equation for the growth and diffusion of population is

$$\frac{\partial N}{\partial t} \;=\; \gamma\, N \left(1 - \frac{N}{K}\right) \;+\; \nabla \,\cdot\, (\nu\, \nabla N)\,, \tag{6.1}$$

Where N is population density, K is carrying capacity, ν is the variable diffusion coefficient and $\partial N/\partial t$ is the rate of change of population density with time. The logistic growth term, the first term on the RHS of Eq. (6.1), represents population growth with the net rate γ. The population growth saturates over time as $N \to K$. The second term in the equation models population movement. For a detailed description, see Section 3.2.2.

6.1.2 Palaeovegetation

The vegetation of the world can be subdivided into different types:

- **Mountain tundra** is an ecosystem dominated by low temperatures and smaller duration for growing seasons, which hinder tree growth. Tundra vegetation is mainly composed of shrubs, grasses, lichen and mosses.
- **Savannah** is a grassland ecosystem with high tree density, with the trees being sufficiently widely spaced to allow for sunlight to reach the land, supporting a grass layer.
- **Steppe and prairie** lands are regions with a large temperature difference between summer and winters. The vegetation mainly consists of grass, herbs and shrubs more than trees.
- **Montane forests (including broad-leaf forests) and bushes** are a type of a mountain system. The climate of the zone is strongly altitude dependent and the characteristic flora and fauna change with elevation. Regions of this type are resource rich, and it is possible to work the land for agriculture.
- **Dry steppes and Semi desert** ecosystems fall in between desert and humid climate ecosystems. The ecology supports scrubby vegetation, with semi-arid areas usually dominated by either grasses or shrubs.
- **Desert** regions contain barren lands with very little or no precipitation, and thus are hostile for plant and animal life.

The forests readily provide food (fruits, nuts, and also game that can be hunted for meat), wood is easily available for fuel and also for building houses. Thus, the forests would have been the most desirable choice of lands to build settlements. Steppes and prairies have large open land areas, in comparison to the forest lands they are more difficult to protect, but nevertheless fertile steppe lands are good for agriculture and hence would have been

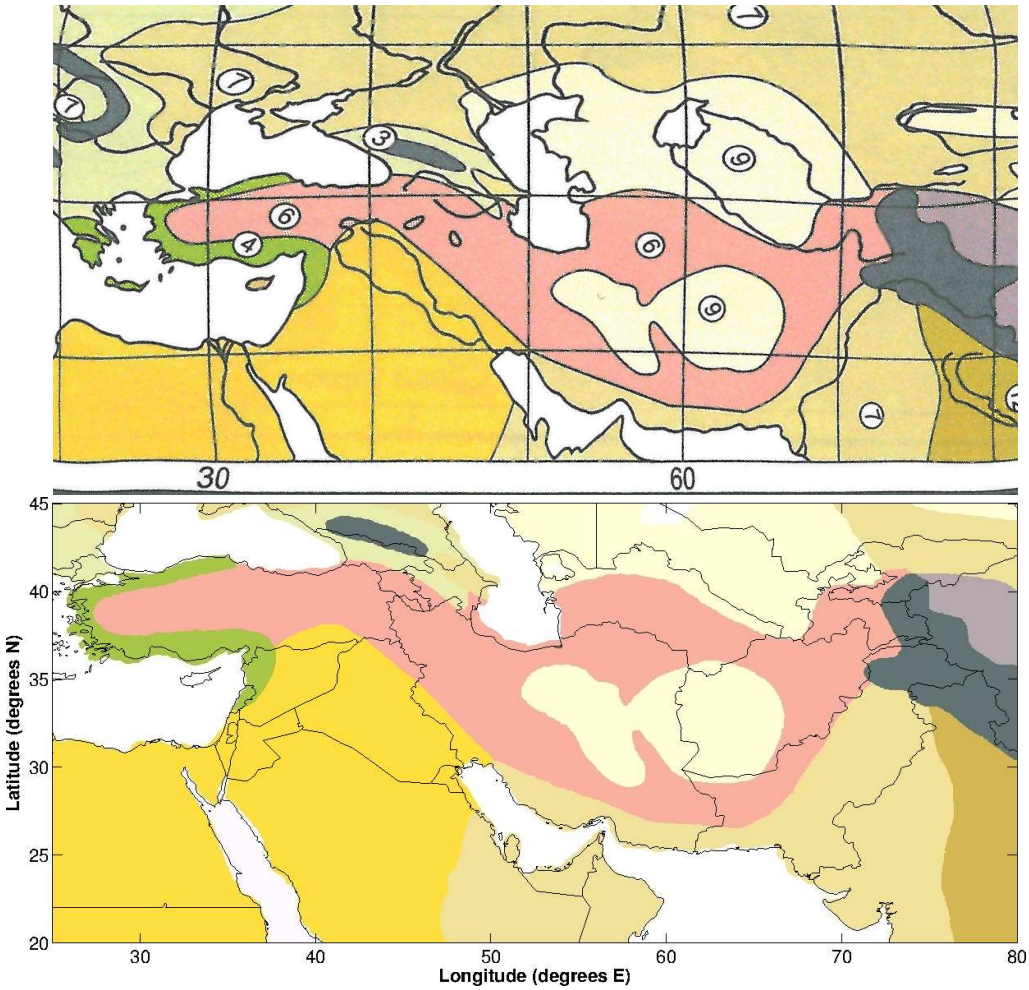


Figure 6.1: The vegetation map (top) [56, p. XX] is converted from planar projection to cylindrical projection using ArcGIS software, and the same map(bottom), after removal of the grid and with the addition of the extrapolated region, with current political boundaries plotted to ease the interpretation. Numbered colours represent different vegetation zones; 3: mixed and broadleaf forests, 4: forests with evergreen elements, 6: mountain forests and bushes, 7: steppe and prairies, 8: dry steppes and semideserts, 9: deserts, 10: high mountain deserts, 11: mountain tundra, 12: savannahs.

the next habitat choice. Savannah lands are not that easy to turn into agricultural fields, but are still habitable. Though the harsh climates of dry steppes and semi-deserts make human survival difficult, some prehistoric settlements occur in these ecosystems. Deserts are the most hostile lands for the survival of humans.

Studies of the climatic change and terrestrial landscape evolution for the extratropical northern hemisphere for the last 130 kyr have been summarised in [56]. The palaeovegetation data of the Holocene period for the region is illustrated in Fig. 6.1 (top). The original

map is in planar projection [56, p. XX]. The map was converted into cylindrical projection using the ArcGIS software. The new map in cylindrical projection was further processed to remove the grid lines and numbers from the original map, and was extrapolated to fit the region in study (20–45°N and 25–80° E) (see Fig. 6.1 (bottom)). Current political boundaries are added for ease of visualisation.

From Fig. 6.1 (top) it is seen that there are nine palaeovegetation zones in the region. Starting with the West Asian countries, Sudan, Egypt, Jordan, Iraq, Syria, Saudi Arabia mostly had *dry steppe and semi-deserts* type vegetation (No. 8 in Fig. 6.1). Turkey had mostly *mountain forests* (No. 6) with *Mediterranean forests* (No. 4) on the periphery. The southern coast of Iran had a narrow *dry steppe and semi-deserts* (No. 8) area with a fringe of *steppes and prairies* (No. 7) in the south-east, but was mainly comprised of *mountain forests and bushes* (No. 6) with *deserts* (No. 9) in the central and eastern parts. Turkmenistan again was mainly comprised of *mountain forests and bushes* (No. 6) type vegetation with the *Karakum desert* at the northern boundary. The vegetation of Uzbekistan and a part of Kazakhstan was mostly of *desert* (No. 9) type (the Karakum and Kyzylkum deserts). South-western Afghanistan also had *desert* (No. 9) vegetation whereas the north-eastern part was *mountain forests and bushes* (No. 6). Pakistan was mainly occupied with *steppe and prairies* (No. 7) with a stout fringe of *mountain forests and bushes* (No. 6) in the south-west. North-eastern Pakistan contained *mountain tundra* (No. 11) type vegetation. North-west India showed *steppe and prairie* (No. 7) type vegetation and the eastern part showed *savannah* (No. 12) type vegetation; extreme northern India along with north-eastern Pakistan and some part of China exhibited *mountain tundra* (No. 11) type vegetation along with *high mountain deserts* (No. 10).

Figure 6.2 shows the Neolithic sites with altitude contours, with the palaeovegetation map background. The map shows that, with an exception of four sites (Ayakagytma, Iblis, Mundigak and Sarazm) situated in desert areas, all other Neolithic sites occur in the *mountain forests and bushes*, *steppe and prairies*, *dry steppe and semi-deserts* zones. We assign trial diffusion coefficients to the ecological zones according to their habitability.

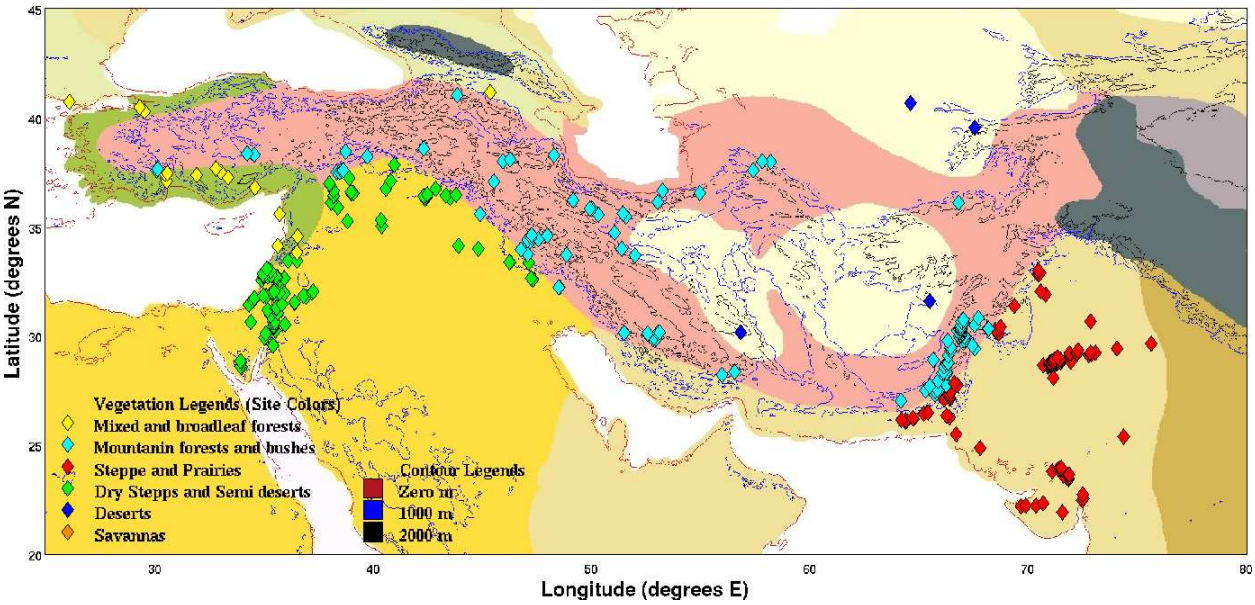


Figure 6.2: The archaeological sites on the palaeovegetation background are shown here. The altitude contours at 0, 1000 and 2000 meters are displayed on the map. The sites are colour coded as given in the legend.

The preferred ecological zones (the mountain forests and the Mediterranean forests) are assigned the highest coefficient. The less hospitable the zone, the smaller is the relative diffusion coefficient (Table 6.1). The desert areas are the harshest environments for human survival, and yet sporadic oasis settlements can exist there. Hence, instead of completely ruling out the possibility of a settlement formation in desert, we assign a very small diffusion coefficient $1/d$, where d is the distance of a point in desert from the nearest fertile land. The local diffusion coefficient is obtained as

$$\nu_i = \frac{\mu_i}{\mu_{max}}\,\nu_0, \tag{6.2}$$

where, μ_i is the palaeovegetation-dependent factor, $\mu_{max} = \max(\mu_i) = 6$ and ν_0 is the background constant diffusion coefficient.

Table 6.1: The table shows the vegetation zones arranged in decreasing order of habitability. A relative trial diffusion coefficient defined in Eq. (6.2), along with the normalised value is also given. Deserts are possibly the worst zones for human habitability, yet sporadic prehistoric settlements are seen there, hence, the relative diffusivity there, though non-zero, would have been really small as compared to other vegetation zones. As a trial we take this relative diffusion coefficient (μ_i) in the deserts to be an average of the inverse of the distances (d) of the oasis settlements from the nearest fertile lands: $1/d$.

Vegetation zone	Relative diffusion coefficient (μ_i)	Normalised diffusion coefficient (μ_i/μ_{max})
Mountain forests and bushes	6	1.00
Mediterranean forests	6	1.00
Steppes and prairies	4	0.70
Savannah	3	0.50
Dry Steppes and semi-deserts	2	0.30
Deserts	0.14	0.02

6.1.3 FKPP model: results and discussion of the preliminary runs

The model is implemented using the finite differences in spherical polar coordinates in the latitude range: . Equation (6.1) is solved on a grid with resolution ($1^\circ/15 \times 1^\circ/15$) containing 825×375 mesh points.

The FKPP models, for which some trial results are shown below, are implemented for different diffusivities and different domains, as given in Table 6.2, using the finite differences in spherical polar coordinates. The primary ('Whole') domain is an area spanned by latitude: 20–45°N and longitudes 25–80°E. The models use growth rate $\gamma = 0.02\,\mathrm{yr}^{-1}$ that corresponds to a population doubling in 30 years [46]. The models are implemented making use of the source sites given in tables 5.2 and 5.3 and the site of Gesher. The model is initiated at 10,000 BCE when Gesher is used as the conventional source (see Eq. (4.5)) and for the Neolithic dispersal from the extended sources in northern and southern part of the Zagros mountains the model start times are taken to be around 7,000 BCE (see Table 5.7).

To compare the model with the data, we calculate the residuals (the difference between the actual Neolithic time at site and that obtained from the FKPP model) at an interval $\sigma = 300$ years making the $\pm 3\sigma$ interval $\pm 900\,\mathrm{yr}$, a round figure close to the ball park

Table 6.2: Summary of preliminary results from the FKPP model from the domain specified (either the whole domain from the near East to the Indus Valley or its part) with a constant (C) or vegetation-dependent (V) local diffusion. The third column gives the starting location of the spread: Gesher (source is at the site of Gesher), South (southern source sites given in Table 5.3), North (northern source sites given in Table 5.2). The average spread speed is given in column 5 and column 6 gives the figures illustrating the solutions. The results tabulated below are for a growth rate $\gamma = 0.02\,\mathrm{yr}^{-1}$.

Domain	Diffusion type (C or V)	Source	Background diffusivity $\nu_0(\mathrm{km}^2/\mathrm{yr})$	Average speed km/yr	Summary figures
Whole domain	C	Gesher	5.45	0.65	6.3, 6.4
Whole domain	V	Gesher	9.00	0.65	6.5, 6.6
Southern Iran	V	South	13.75	0.66	6.7, 6.8
Southern Iran and Indus	V	South	275.0	3.58	6.9, 6.10
Northern Iran	V	North	77.0	1.85	6.11, 6.12
Northern Iran and Indus	V	North	231.0	2.92	6.13, 6.14

estimate of $\pm 1000\,\mathrm{yr}$ (see Eq. (4.5)). The histograms given in figures 6.3 to 6.13 show the residuals in the arrival times. For the sites with negative residuals, the model arrives there earlier than the actual time and for sites with positive residuals, the model arrives at the sites later than the actual time. In the maps below the histograms we have plotted the sites that lie between ± 900 years of the zero residual. In each of the maps, from Fig. 6.4 to Fig. 6.14, the green stars are the source sites. Model times for the sites shown by black filled circles lie within 1σ interval (i.e. $\pm 300\,\mathrm{yr}$) of the actual time. The sites where the model reaches earlier (where the residuals are more negative than $-300\,\mathrm{yr}$) are shown with filled blue circles and sites where the model reaches late (where the residuals are more positive than $+300\,\mathrm{yr}$) are shown in red filled circles. Sites with residuals outside the $\pm 900\,\mathrm{yr}$ range are not shown in these maps.

Table 6.2 presents a summary of a preliminary run of the FKPP model. The first column of the table contains the domain modelled. The second column contains the type of the diffusivity, either constant (C) or variable (V). The source of the spread is given in the third column. The fourth column gives the background diffusivity coefficient used in each case to get the constant average speed given in the next column. The last column points to the summary figures presenting each case.

The analysis from Chapter 4 shows that the Neolithic spread eastwards from Gesher took place with a speed $0.65 \pm 0.1\,\text{km/yr}$. When solved on a homogeneous domain with a constant diffusivity $5.45\,\text{km}^2/\text{yr}$ the FKPP model, with the Neolithic spread starting at Gesher at 10,000 BCE, agrees reasonably with the Neolithic data. As can be seen from Fig. 6.3, the deviations in the site occupation times obtained from the FKPP model lie within ± 900 years of the observed dates for more than 230 sites ($> 60\%$). Introducing the palaeovegetation-dependent diffusion, as given in Table 6.1, improves the results as now the histogram of Fig. 6.5 shows an even more prominent peak with more than 270 ($> 70\%$) sites lying within the ± 900 years ($\pm 3\sigma$ interval) of the observed dates. The introduction of the palaeovegetation-based diffusivity instead of a constant diffusivity thus improves the agreement of the FKPP model with the Neolithic data. For both, homogeneous domain and domain with palaeovegetation-dependent diffusion, the sites within ± 900 years are plotted in Fig. 6.4 and Fig. 6.6 respectively.

The FKPP model with palaeovegetation-based diffusion is now implemented on two sub-domains within the main domain, with the source in the southern Zagros (the southern route), with and without the Indus sites and with the source in the northern Zagros (the northern route), with and without the Indus sites.

When the model is applied to the southern route without the Indus Valley data, 12 sites ($> 70\%$) lie within the $\pm 3\sigma$ (± 900 years) of the observed Neolithic times. Even though it is a high percentage of sites, the histogram in Fig. 6.7 is not peaked around a value and is flat and hence it would not be wise to conclude either in favour of or against the FKPP model in this region at this preliminary stage. Figure 6.8 shows these southern route sites (without the Indus Valley sites) with residuals lying within ± 900 years of the observed Neolithic times.

When the Indus Valley sites are included in these southern route sites, only 23 sites ($\sim 50\%$) lie within the $\pm 3\sigma$ range i.e. ± 900 years of the observed Neolithic times. Again the histogram shown in Fig. 6.9 is flat and not peaked around one value. Hence it cannot be concluded whether the FKPP model is a good model for explaining the southern sites

as a result of the spread of the Neolithic from the source in southern Zagros mountains. Figure 6.10 maps the southern route sites (with the Indus Valley sites) with residuals within ±900 years.

Thus our preliminary executions of the FKPP model, when the southern route is considered with a source in the southern Zagros mountains, seem to reiterate our conclusions from Chapter 5 that the Neolithic propagation most probably did not occur via the route from southern Zagros.

With the northern route excluding the Indus Valley sites the FKPP model looks more promising as 21 ($\sim 64\%$) sites lie within ±300 years (1σ interval) of the observed values and about 91% lie within 3σ i.e. ±900 years. As opposite to the histogram in Fig. 6.7 for the southern route, the histogram of the residuals for the northern route shown in Fig. 6.11 peaks around zero. The FKPP model thus fits better to the northern route than the southern route when Indus Valley sites are not included. The map from Fig. 6.12 shows the northern route sites with residuals lying within the ±900 years (3σ interval).

When the Indus Valley sites are included in the northern route 44 sites ($> 70\%$) have residuals within the 3σ range as displayed by the histogram in Fig. 6.13. Compared to the southern route, this percentage is significantly higher. The histogram looks a little flatter than the one for the northern route in the Fig. 6.11, yet it is a much better fit than the one in Fig. 6.9 for the southern route. Fig. 6.14 shows the northern route sites including the Indus Valley sites with residuals within ±900 years on a current political map.

The FKPP model with the palaeovegetation-based diffusivity thus fits the data better for the northern rather than the southern route, for the datasets with and without the Indus Valley sites. The values of the background diffusivities in case of both, the northern and the southern routes when the Indus Valley sites are included, are very high as compared to their values when the Indus Valley sites are excluded. Hence, even though the relative coefficients for the different ecoregions remain the same, the background diffusivities used in each case are different and hence the effective diffusion coefficients in each case are different for the same vegetation zone.

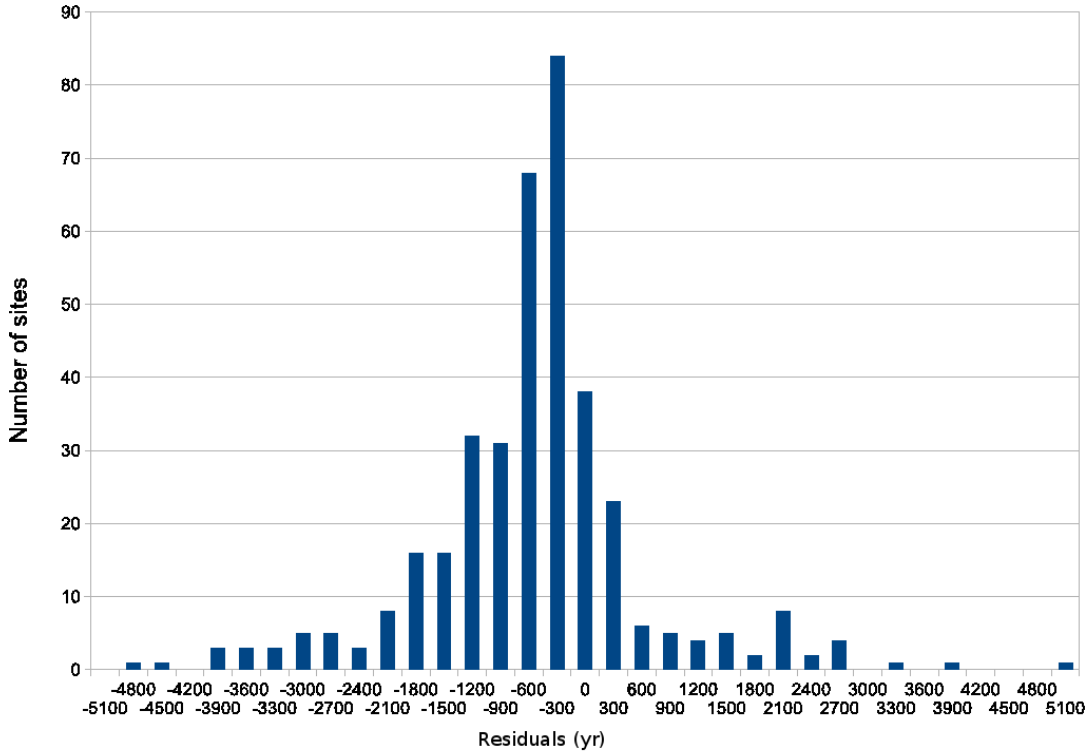


Figure 6.3: Residuals between observed and model Neolithic arrival times in a homogeneous domain. The residuals are given by: $\Delta T = T_{\mathrm{o}} - T_{\mathrm{m}}$,where T_{o} is the observed time and T_{m} is the model time.

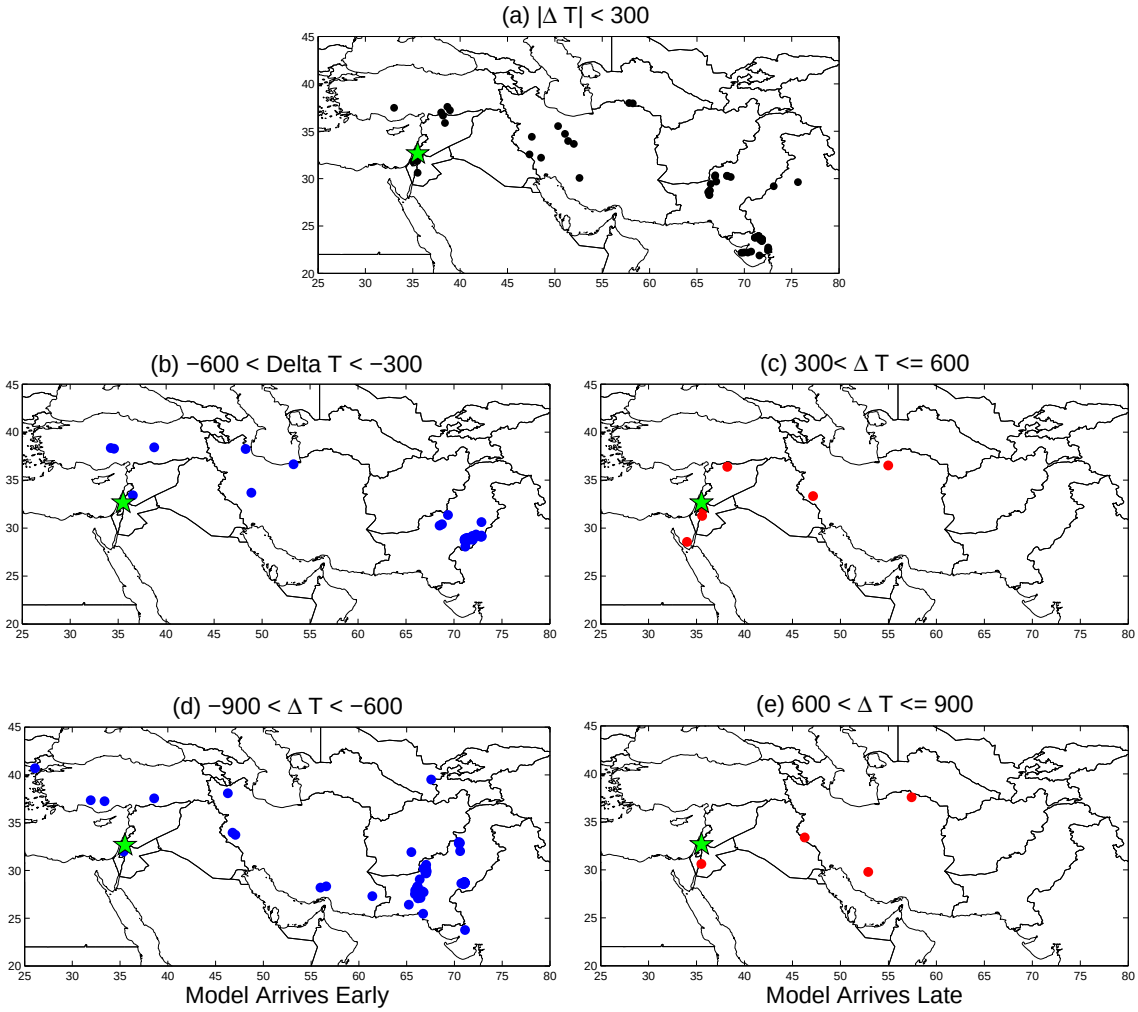


Figure 6.4: Sites plotted according to the residuals between the observed and the model Neolithic arrival times for Neolithic spread in a homogeneous domain with $\gamma = 0.02\text{yr}^{-1}$ and $\nu = 5.45\text{km}^2\text{yr}^{-1}$ and average Neolithic spread speed is $U = 0.66\text{km/yr}$. Negative residuals occur when model arrives earlier than the observed time and positive residuals occur when the model arrives later than the observed time. The source site, Gesher, is shown as a green star. 129 (34%) sites with residuals that lie outside the 3σ band ($\pm$900 yr) are not shown on these maps.

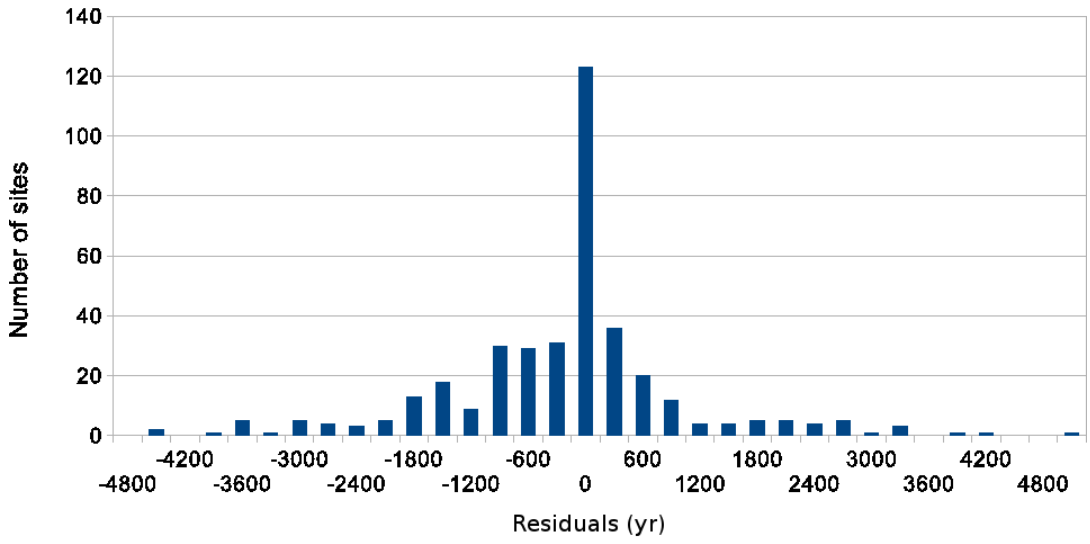


Figure 6.5: Residuals between observed and model Neolithic arrival times. The model uses palaeovegetation-based diffusion coefficients for a spread starting from Gesher.

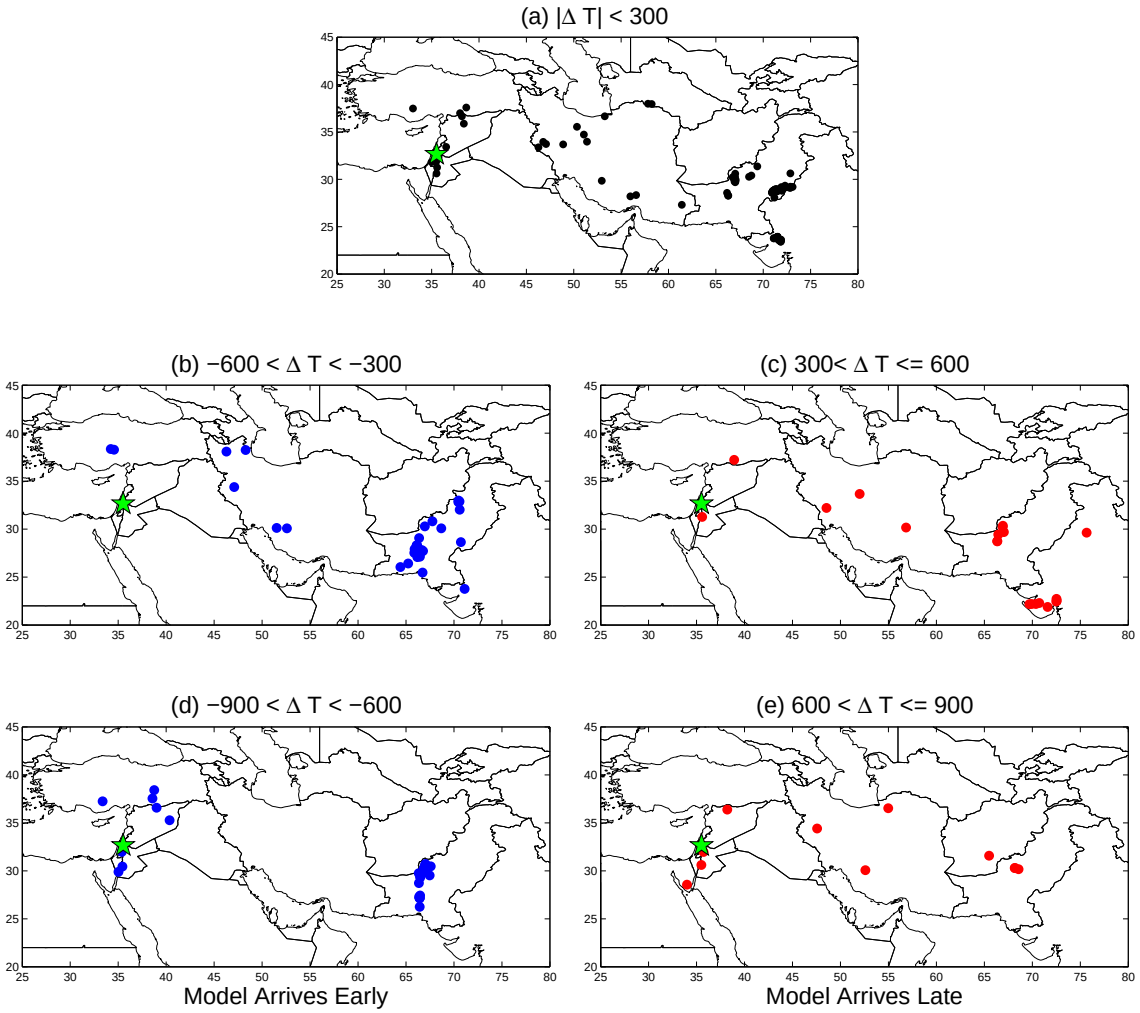


Figure 6.6: Sites plotted according to the residuals between observed and model Neolithic arrival times for the Neolithic spread starting at Gesher, and involving palaeovegetation-based diffusion. 103 (27%) sites with residuals that lie outside the 3σ band ($\pm$900yr) are not shown on these maps. Other details as in Fig. 6.4.

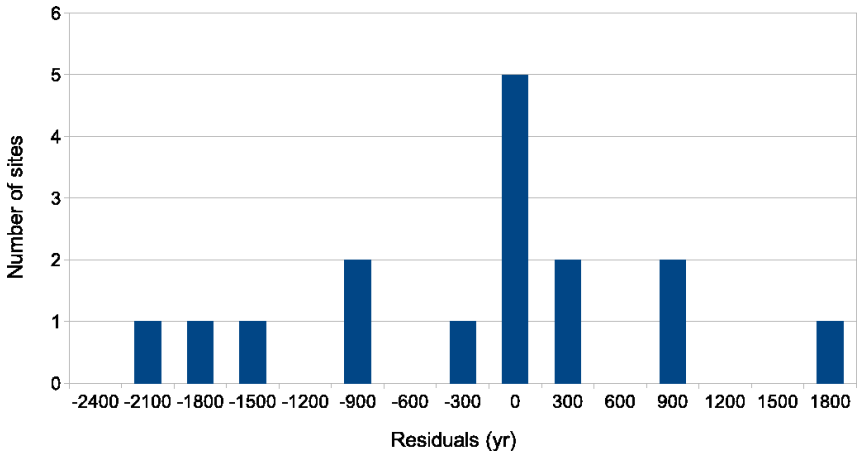


Figure 6.7: Residuals between observed and model Neolithic arrival times, for the southern route without Indus valley sites. The model uses palaeovegetation-based diffusion coefficients.

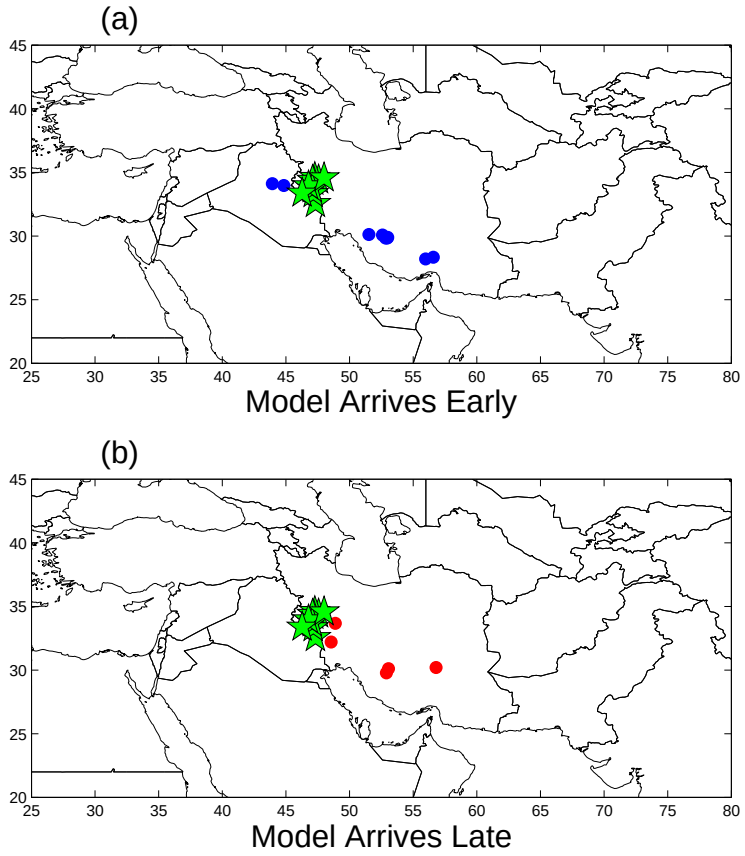


Figure 6.8: Sites plotted according to the residuals between observed and model Neolithic arrival times for the FKPP model with palaeovegetation-based diffusion for the southern route without the Indus valley sites. 6 (37%) sites with residuals that lie outside the 3σ band ($\pm$900yr) are not shown on these maps. Other details as in Fig. 6.4

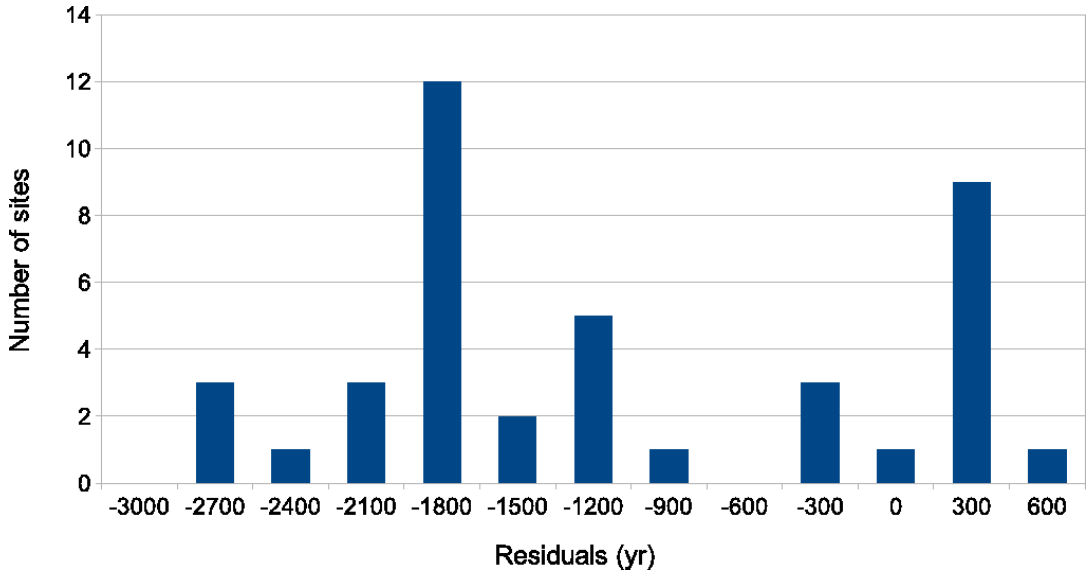


Figure 6.9: Residuals between observed and model Neolithic arrival times for the FKPP model for southern route including the Indus Valley sites.

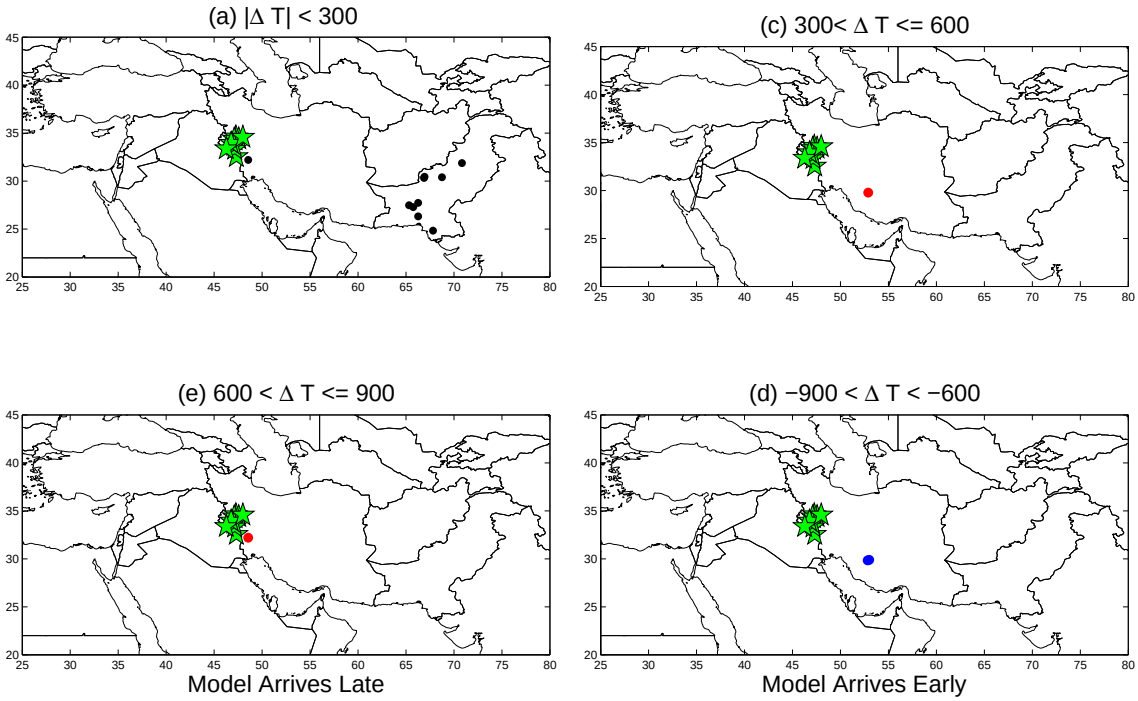


Figure 6.10: Sites plotted according to the residuals between observed and model Neolithic arrival times for the FKPP model with Palaeovegetation-based diffusion for the southern route including Indus valley sites. 28 (68%) sites with residuals that lie outside the 3σ band ($\pm$900yr) are not shown on these maps. Other details as in Fig. 6.4.

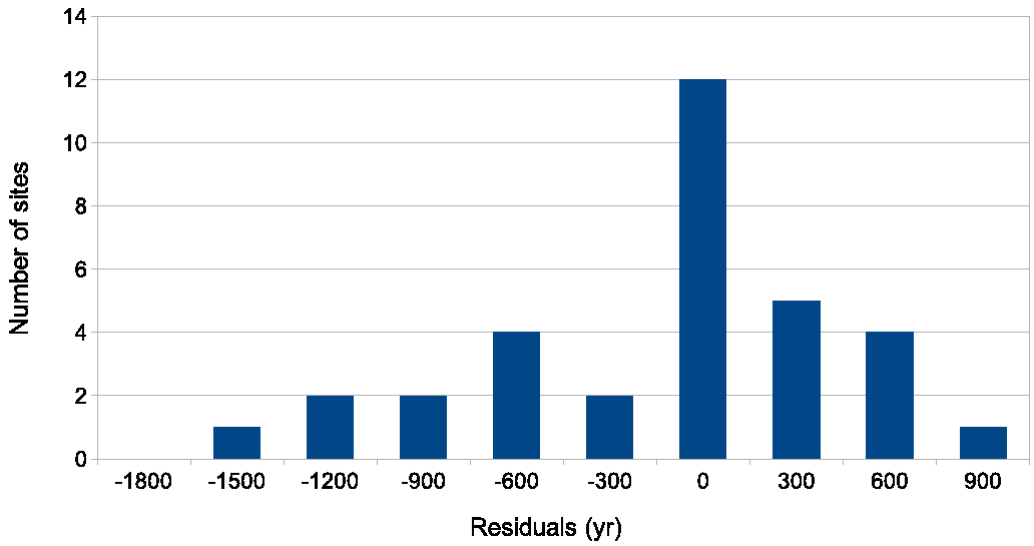


Figure 6.11: Residuals between observed and model Neolithic arrival times for the FKPP model for the Neolithic spread via northern route excluding the Indus sites. The model uses palaeovegetation-based diffusion coefficients.

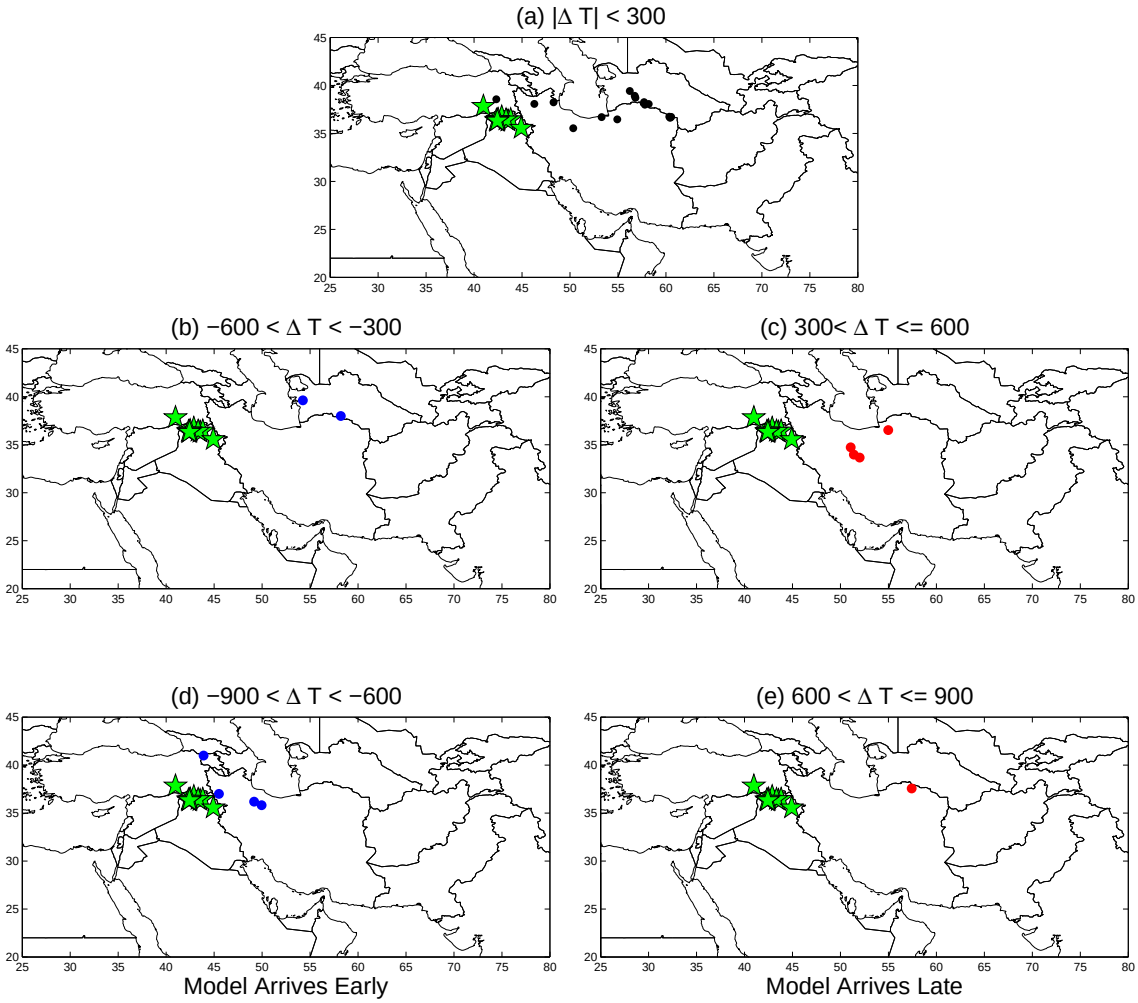


Figure 6.12: Sites plotted according to the residuals between observed and model Neolithic arrival times for the Neolithic spread through the northern route without the Indus valley sites and involving palaeovegetation-based diffusion. 4 (12%) sites with residuals that lie outside the 3σ band ($\pm$900yr) are not shown on these maps. Other details as in Fig. 6.4.

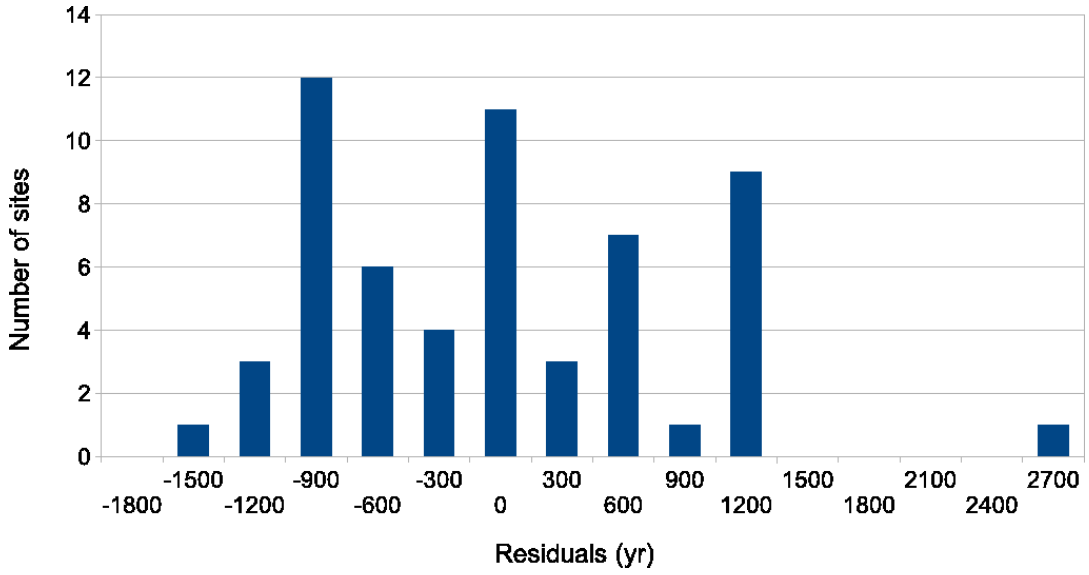


Figure 6.13: Residuals between observed and model Neolithic arrival times for the FKPP model for the Neolithic spread via northern route including the Indus sites. The model uses palaeovegetation-based diffusion coefficients.

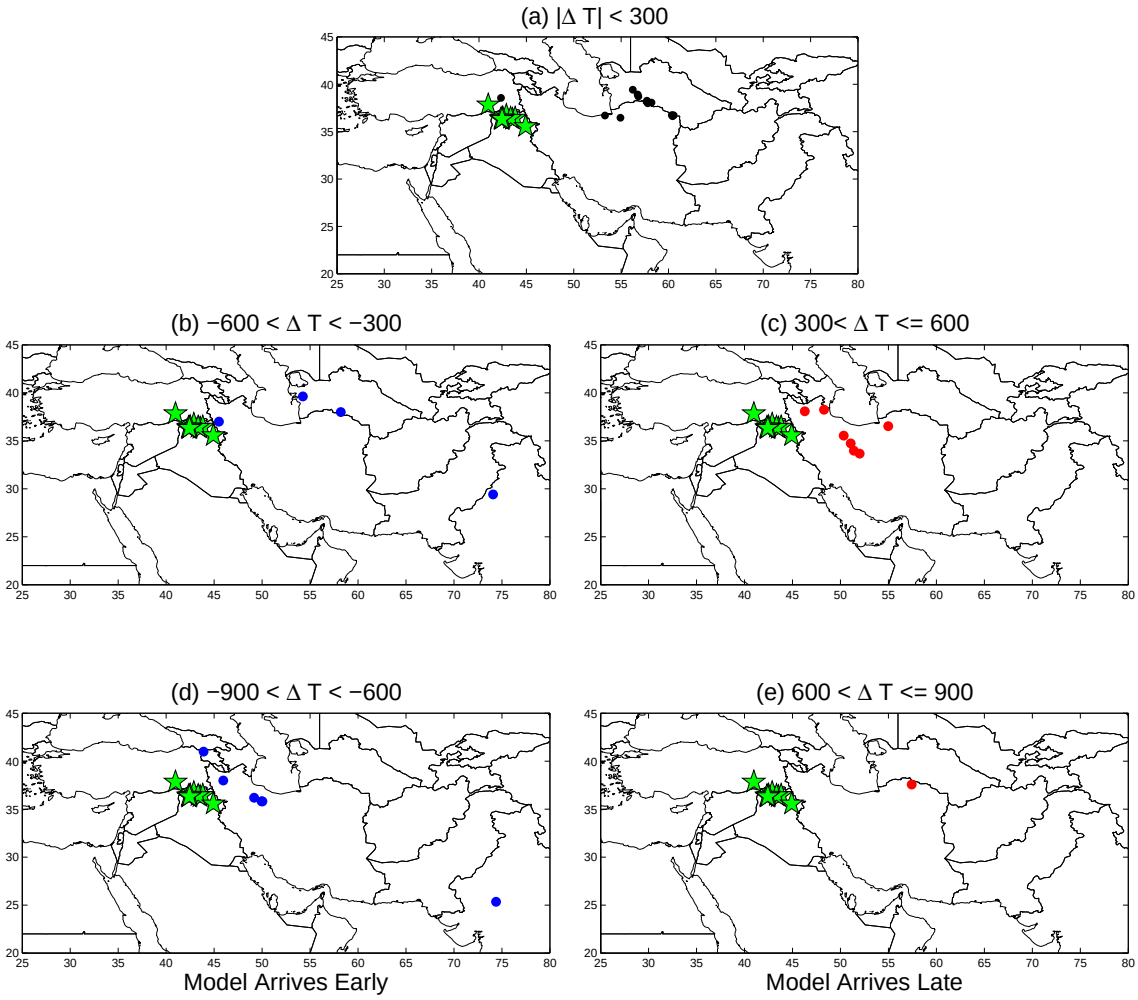


Figure 6.14: Sites plotted according to the residuals between observed and model Neolithic arrival times for the Neolithic spread through the northern route including the Indus valley sites involving palaeovegetation-based diffusion. 15 (26%) sites with residuals that lie outside the 3σ band ($\pm$900yr) are not shown on these maps. Other details as in Fig. 6.4.

6.1.4 Future Work

In the current version of the model, the same carrying capacity ($K = 3.5\text{persons/km}^2$) is assigned to all vegetation zones. A variable carrying capacity needs to be included in the model, but this would not affect the propagation speed in the standard FKPP equation.

Other than the vegetation, various other factors (e.g. topography, precipitation, soil type, land fertility, temperature) affect the growth and spread of human populations. Figure 6.15 shows the Neolithic sites plotted on the backgrounds representing these major environmental factors. The environment datasets used here are present-day conditions. Clear preference among the three environmental factors is evident from these three plots. The histograms of Fig. 6.16 quantify these preferences. Thus it might be necessary to include precipitation, topography and temperature in order to better understand the Neolithic settlements and make the deterministic model more realistic and accurate. However, palaeovegetation may remain the best summary environmental characteristic to use since it is sensitive to all other environmental variables.

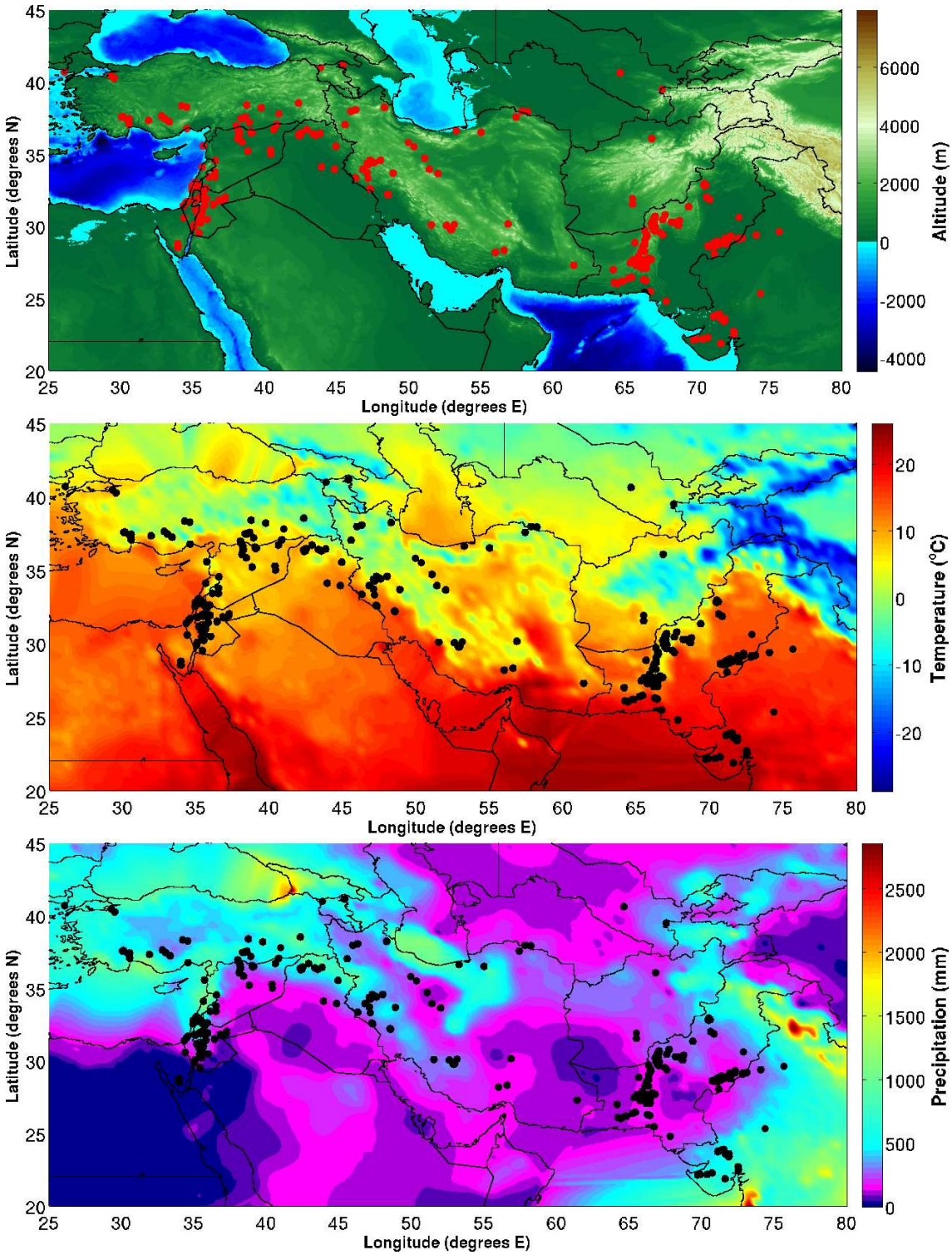


Figure 6.15: The Neolithic sites plotted on maps of current topography (upper) [100], annual mean temperature (middle) [37] and annual precipitation (bottom) [37] data for the study zone, with current political boundaries shown for ease of visualisation.

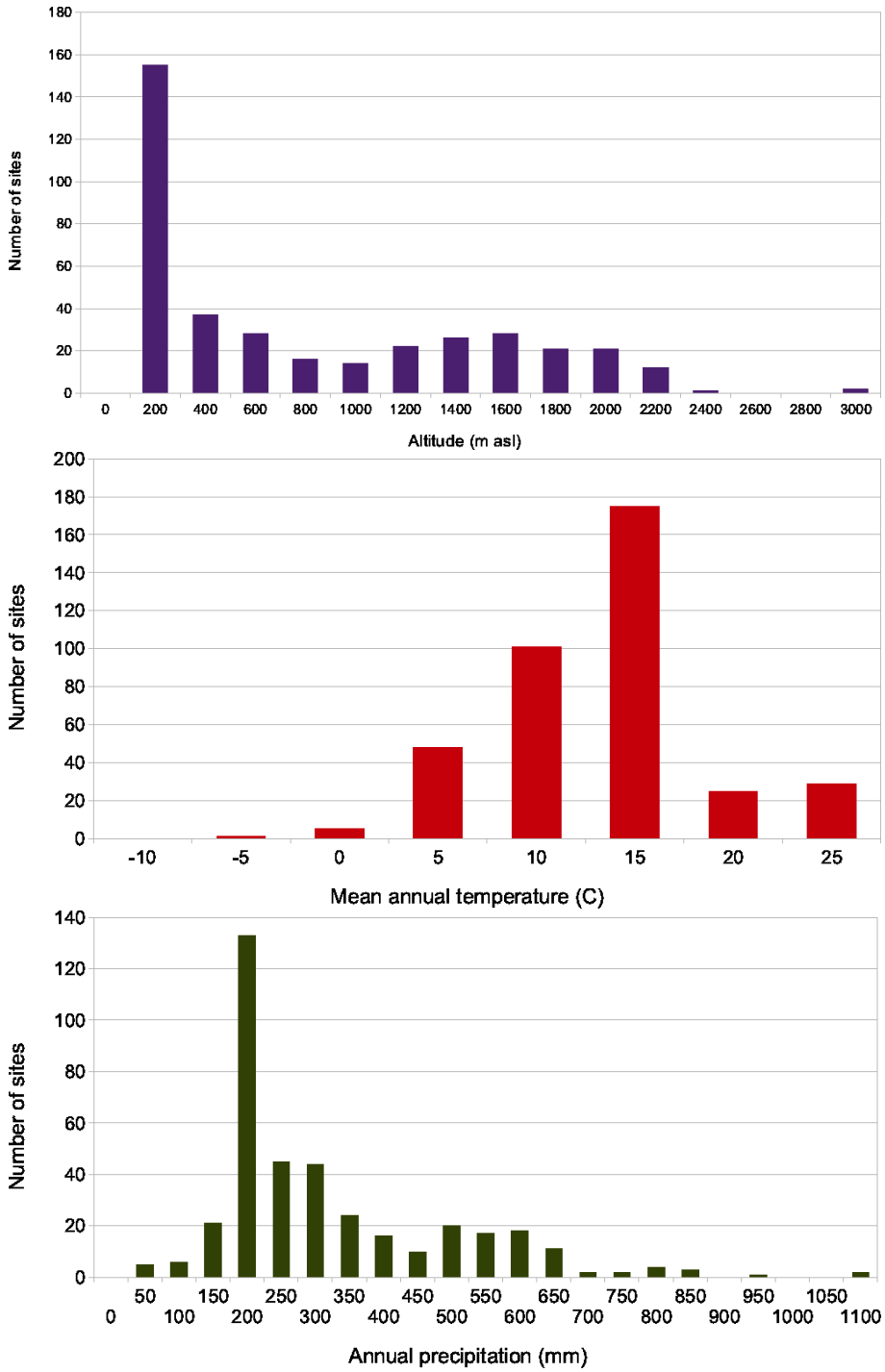


Figure 6.16: Histograms of site distributions according to their altitude, annual mean temperature and annual precipitation. The values are interpolated from standard datasets.

6.2 Crops, Competition and Languages

The Indian subcontinent was an independent origin for Neolithic rice farming in the Gangetic plains. Taking this center to be Lahuradeva in the Gangetic plain, we analyse the Neolithic rice farming data [65] as before to see whether the rice farming actually spread from Lahuradeva to western India and to find the spread speed. Figure 6.17 shows plots for the rice data obtained by the percentile method from Chapter 4. The speed obtained is 0.68 km/yr with the goodness of fit $R^2 = 0.75$.

6.2.1 The competition model

Depending upon the type of crops used, populations in different geographical conditions have different dietary habits. Usually, there is at least one staple crop that determines the economy of the population. We use the concept of the signature crop of a population to de-

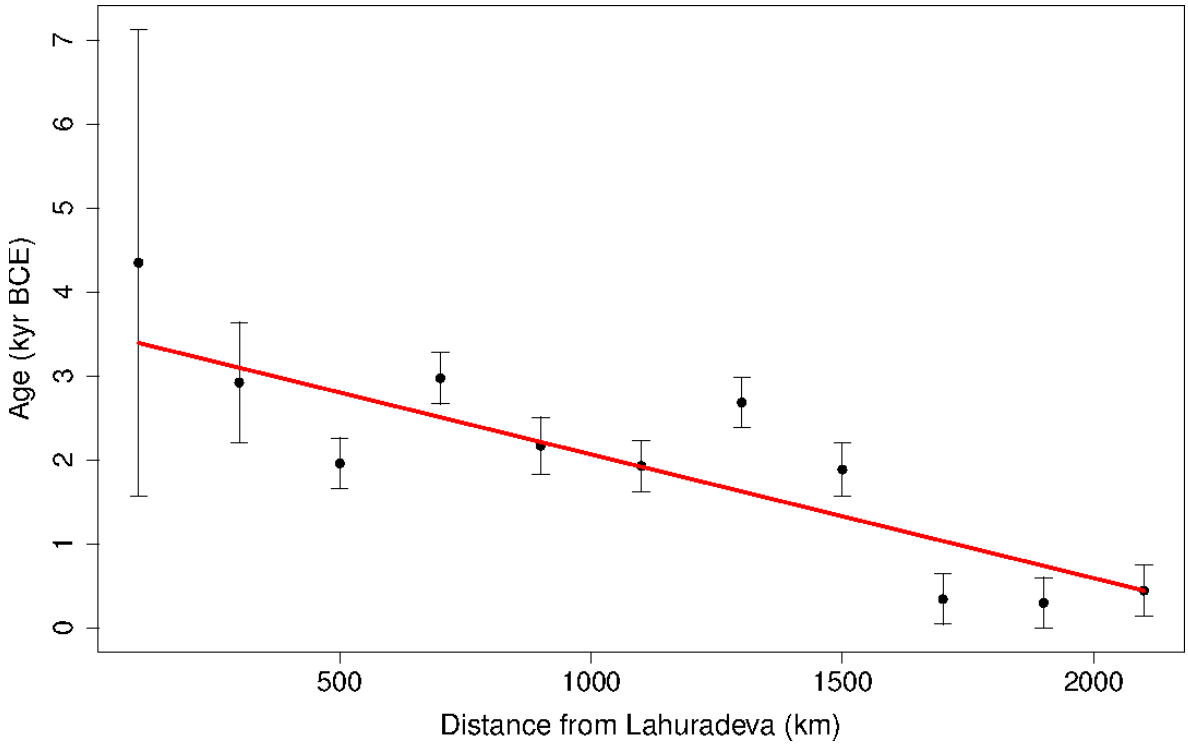


Figure 6.17: Envelope fit for the rice data [65] in Indian subcontinent by the percentile method gives a spread speed 0.68 km/yr.

fine our model. The spread of important staple crops (the signature crops) is synonymous with the spread of the farming population constrained by the environmental requirements of the crop and the geographical conditions. We model the spread of the signature crop numerically on a spatial grid. We use the modified Lotka-Volterra model [99], based on the principle of competitive exclusion. The basic model is modified to include the logistic growth and diffusion terms. It accounts for the competitive interactions of more than one populations through the system of coupled differential equations (Equations (6.4) and (6.5)). Although there are various important factors (e.g. soil type, water availability, soil fertility, altitude) that affect the growth of the crops, we take only two of these major factors into account, namely, altitude [100] and water availability (in the form of current rainfall data [141]) in defining the fertility parameter for the crops.

As mentioned in section 1.2.2 various studies have been performed.

Our model is particularly very close to two studies that model the process of conversion of the Mesolithic hunter-gatherers to Neolithic farmers [4, 106]. Even though this conversion process could be important, evidence for a complete conversion is very rare [24]. Hence we focus on Neolithic dispersal rather than the conversion process and even though the equations defining the model are similar, the underlying concept is entirely different. We apply this model to the northern region of the Indian subcontinent.

In the prehistoric Indian subcontinent, there were at least three population groups. A record of more than two thousand archaeological sites (the present-day India, Pakistan and Afghanistan) shows the existence of two distinct prehistoric population groups [68], one in the north-western part (in the Punjab and Indus river valleys) and the other in the north-eastern India (Gangetic valley). The third population group, the Dravidian group, belonged to the southern Indian peninsula. Our model is applied to the northern region.

In this region, at least two sources for Early Farming are identified [95], Mehrgarh (longitude: $67^{\circ}35^{'}$E, latitude: $29^{\circ}25^{'}$N) in the western Indian subcontinent and Lahuradeva (longitude $82^{\circ}57^{'}$E, latitude: $26^{\circ}46^{'}$N) in the east.

We propose a model accounting for competition for resources between the two crop com-

plexes, the rice complex originating in the east and the wheat complex in the west, based on the principle of competitive exclusion [99]. The model takes into account topography [100] and rainfall [141] for the specified region.

The signature crops for these two regions were wheat and rice respectively. The environmental requirements of these two crops are very different. Wheat grows in colder and dryer climates and can be grown in places with moderate rainfall, whereas rice is practically grown in water and needs copious amounts throughout its lifetime. We enforce this condition in the fertility parameter σ (see Eqs. 6.6, 6.7). We define this parameter to include the altitude dependence and rainfall required by the crop. We consider N_r and N_w to be the densities of rice and wheat producers respectively. K_r and K_w are their saturation carrying capacities respectively. These variables are changed into dimensionless variables as,

$$n_r = \frac{N_r}{K_r}, \quad n_w = \frac{N_w}{K_w} \tag{6.3}$$

The model is then defined mathematically with the following coupled differential equations:

$$\frac{dn_r}{dt} = \nabla \cdot [\, \nu(h) \nabla n_r \,] + r_r n_r (1 - n_r) - c_{rw} n_r n_w \tag{6.4}$$

$$\frac{dn_w}{dt} = \nabla \cdot [\, \nu(h) \nabla n_w \,] + r_w n_w (1 - n_w) - c_{wr} n_w n_r \tag{6.5}$$

In both Eqs. (6.4) and (6.5), the first term on the RHS is the diffusion. The second term gives the logistic growth. The last term on RHS is the competition term which gives the effect of competition interaction. The subscript 'r' corresponds to the eastern Indian farmers whose signature crop was rice and subscript 'w' corresponds to the farmers in western part of the subcontinent whose signature crop was wheat. n_r and n_w are the density fractions for rice and wheat respectively. ν is the spatially varying diffusivity, which is taken to be a simple linear function of the altitude h at a point and is constant in time, $\nu = \nu_0 \left(1 - \frac{h}{h_0}\right)$. The cut-off is taken to be $h_0 = 1,000$ meters above sea level

beyond which ν is defined to be zero. r_r and r_w are the logistic growth rates. c_r and c_w are the competition parameters determining the incursion abilities of the two crops.

Our model uses the basic framework defined by Ammerman and Cavalli-Sforza [17] and Ackland et al. [4]. The diffusion and the logistic terms are formulated in the same manner as Ackland et.al. [4] and their values of a few parameters are used.

The fertility parameters σ_w and σ_r for wheat and rice respectively are given by the trial functions,

$$\sigma_w = f(h)\, g_r(p), \tag{6.6}$$

$$\sigma_r = f(h)\, g_w(p). \tag{6.7}$$

Here

$$f(h) = \left[1 - \tanh\left(\frac{h - h_0}{dh}\right)\right], \tag{6.8}$$

$$g_w(p) = \left[\tanh\left(\frac{p - p_{uw}}{dpu}\right) - \tanh\left(\frac{p - p_{lw}}{dpl}\right)\right], \tag{6.9}$$

$$g_r(p) = \left[1 + \tanh\left(\frac{p - p_{lr}}{dp}\right)\right]. \tag{6.10}$$

Where h is the altitude at a point and p denotes the precipitation at that point. h_0 is the cut-off altitude. p_{uw} is the upper limit and p_{lw} is the lower limit of the rainfall requirement. Similarly, the lower limit of rainfall for rice is given by p_{lr}; there is no upper limit. The water requirement of wheat is taken to be between 300 and 650 mm per year and that of rice is taken to be a minimum of 600 mm per year The three functional forms are shown in the three frames in Fig. 6.18.

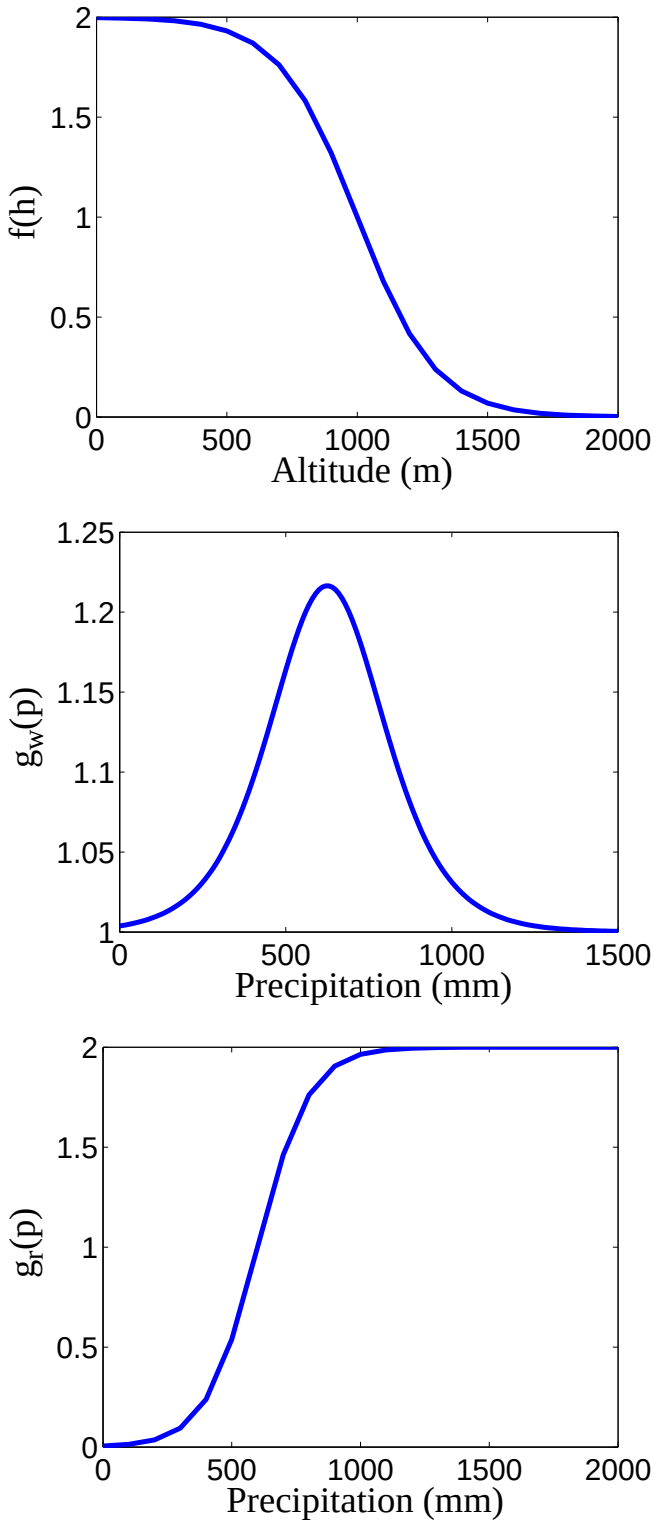


Figure 6.18: Uppermost frame displays the functional form of Eq. (6.8) using the altitude data. Middle frame is the plot of the water required for wheat growth given by Eq. (6.10) and the frame at the bottom is the functional form of Eq. (6.10), the water requirement of rice.

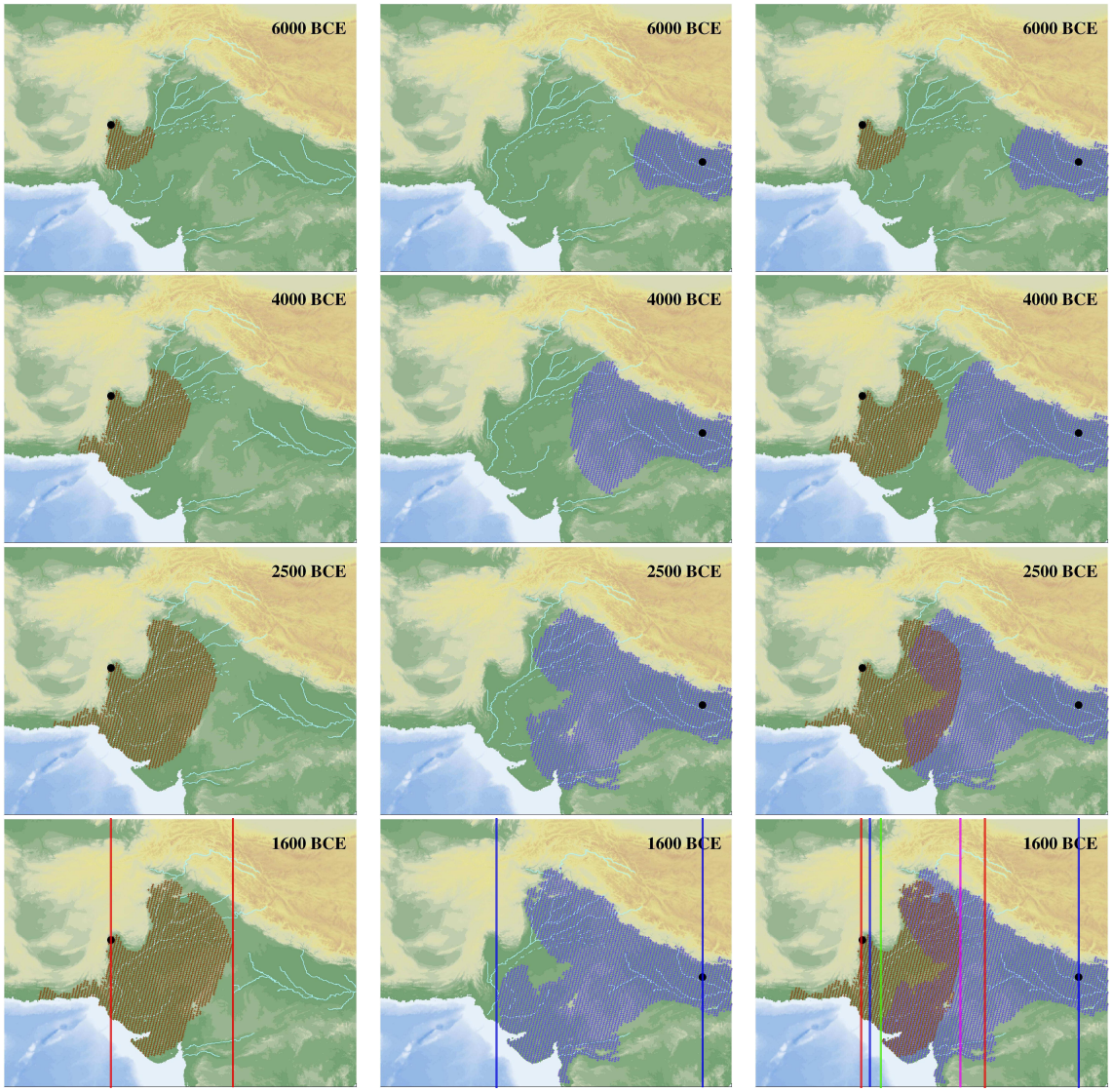


Figure 6.19: Leftmost column shows unhindered spread of wheat without the presence of rice to compete with, the second column shows unhindered spread of rice without the presence of wheat to compete with and the third column shows the result from the model including competitive farming. In the last row, the blue lines show the extent of spread of rice when wheat is absent, and the red lines show the extent of spread of wheat when rice is absent. The green line shows the boundary of rice and the magenta line shows extent of wheat, when both rice and wheat are spreading simultaneously. The change in the resultant spread boundaries, which is an effect of competition, is evident from a comparison of these three snapshots. This competition model is implemented using a birth rate $b_0 = 0.06\,\mathrm{yr}^{-1}$, a death rate $d_0 = 0.02\,\mathrm{yr}^{-1}$, and an altitude cut-off at $h_0 = 1000\,\mathrm{m}$. Water requirement for wheat is taken between 300 mm and 650 mm and that for rice is taken to be 600 mm.

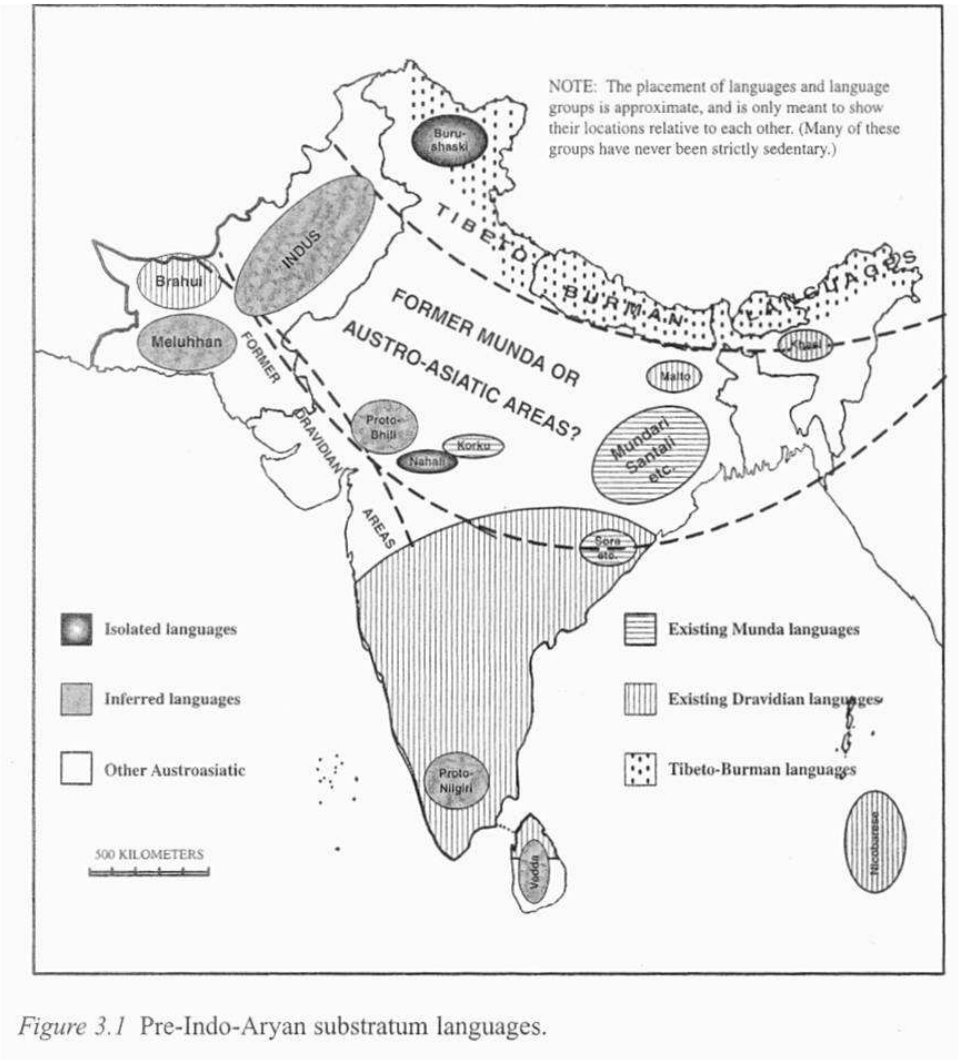


Figure 3.1 Pre-Indo-Aryan substratum languages.

Figure 6.20: The map reproduced from [129], displays the pre-IndoAryan language distribution in South Asia.

The results of the preliminary simulation runs are shown in Fig. 6.19. Column three in Fig. 6.19 shows the inception of the wheat wave (shown in shades of brown) in the north-west, and rice wave (shown in blue colour) in the east, originating from Mehrgarh and Lahuradeva respectively. In the same column the next snapshots show the gradual progression of the two waves (rice from east-to-west and wheat from west-to-east) and their interaction.

6.2.2 Linguistic analysis and prehistoric languages of the Indian subcontinent

A language is an attribute of a population and requires people to carry it. Hence, on a broad scale, study of linguistic spread is study of population dynamics. Distinct models based on similar ideas (e.g. initial colonization, subsistence / demography model, elite dominance etc.) have been proposed for linguistic expansion through language change and replacement [34, 117]. Renfrew and Bellwood note that, "*The homeland zones of agriculture and the homeland zones of the major agriculturist language families are correlated geographically to a high degree*" [24]. This has lead to the proposal of the farming-language dispersal (FLD) model. The farming-language dispersal hypothesis (FLDH) can be stated as *"The expansion of farming practices replaced or assimilated those of adjacent hunter-gatherer populations, leading to language spreads.... The present-day distributions of many of the world's languages and language families can be traced back to the early developments and dispersals of farming from several nuclear areas where animal and plant domestication emerged"* [128]. The FLD and FLDH models are theoretical frameworks for interpretation and *not* mathematical models. The few exceptions noted for this hypothesis, as in the case of non-expansionist agricultural families [34], suggest that the FLD model might not be the universal model for explaining the linguistic spread (about which Harris [74] argues that only the population groups that had pulses and cereals in their staple diet, and probably had domesticated animals – so that it would make them self-sufficient and thus independent of foraging and hunting – were capable of expansion). Despite these exceptions it is a commonly accepted fact that demic diffusion processes did take place when Neolithic farmers inhabited Europe [36, 145] and the FLD model best explains the spread of the Indo-European language family across Europe.

The Indian subcontinent is linguistically diverse. The four major language families spanning the subcontinent (Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman) belong to the agriculturalist spread zone category [85, 129]. Hence, we apply the FLD model, though with a slight modification of the FLDH, to the northern part of the Indian subcontinent. We thus try to provide an explanation for the linguistic scenario in

pre-historic times.

6.2.3 Discussion and inferences

Neolithic farming originated independently in different parts of the world. A few Neolithic source regions have been identified all over the world. The inception of early agriculture around 10,000 BCE in the Fertile Crescent in the Near East was responsible for the farming spread across Europe. It involved a single crop complex comprising mainly of wheat and barley. The wave of advance model, used by Ammerman and Cavalli-Sforza [17], quantitatively described this spread of incipient farming by replacement or conversion of Mesolithic hunter-gatherers to Neolithic farmers. The same model was later elaborated by Renfrew to provide the first explanation of the origin and spread of Indo-European languages. Although some more complex wave of advance models have been proposed [4, 41], they essentially model the same basic process stated in the FLDH, i.e. the spread of agriculture by conversion or replacement of hunter-gatherers by farmers. In Europe, farming spread occurred from a single source in the Fertile crescent and the Indo-European language family dominated Europe's linguistic distribution. The aforementioned models were successful in explaining this unhindered spread of the Indo-European family. The dominance of the Indo-European language family throughout Europe is then adequately explained with the help of FLDH. The same model when applied to the Indian subcontinent [4] depicts the farming spread starting from a single source region in Baluchistan and eventually covering the entire subcontinent. If the FLDH is applied to this situation, then a single language family should have spanned the whole subcontinent, which is not the scenario in India.

The prehistoric reconstruction of South Asia's linguistic past shows the presence of more than one language families [129]. Figure 6.20 shows the placement of languages in pre-IndoAryan times [129]. Along with currently existing language families like Austro-Asiatic, Dravidian, Tibeto-Burman, it also shows the relative positions of isolated (geographically) and inferred (whose presence can be determined by traces found in existing languages

[129]) languages in pre-IndoAryan times. The region of interest for the discussion here is the belt bounded by two dashed lines on north and south, from north-west to north-east India, shown in the map as belonging to "Former Munda" or "Austro-Asiatic" areas. The inferred language 'Indus' belongs to this belt.

Our mathematical model is implemented for the same region. The model involves competitive interaction for farming resources between two crop complexes. The crucial idea involved in the model is the fact that certain form of vegetation grows best in certain environmental conditions and cannot be sustained without those. Altitude is one of the key components in deciding the presence of a crop at a location. Another influential factor is the water requirement of crops. The water requirement for rice and wheat, which were the main dietary crops of Lahuradeva (north-east) and Mehrgarh (north-west) respectively, is entirely different. Rice requires ample water supply throughout its gestation period as well as maturity. It is practically grown in water. Thus, the existence of a crop complex in a region is constrained by the rainfall received by and the altitude of the region. We incorporate these requirements in our model. Due to unavailability of palaeo data, we use the current topographic and precipitation measurements.

According to our model, in the absence of rice farming to compete with, wheat farming would have spread as shown in column 1 from Fig. 6.19 and in the absence of wheat farming, rice farming would have spread as shown in column 2 of Fig. 6.19. From a comparison of the three columns in Fig. 6.19, it is clear that the rice and wheat farming spreads, when competing, have distributions quite different than unhindered rice or wheat farming.

The rice farming wave advanced east-to-west and the wheat wave advanced west-to-east. An independent cogent evidence is provided through archaeological records which asserts that indeed rice spread in India from east to west [105]. The two waves interact and interfere thus forming an intermediate region, where mixed farming could exist. In this region the competition for resources would be significant and the two waves would impede each other, eventually forming stationary boundaries of the intermixed region. Thus,

eventually rice would stop spreading westwards and wheat would stop spreading eastwards. The intermixed region is approximately the same as the 'Indus' region shown in Fig. 6.20.

Based on Fig. 6.20 from [129] and our simulation results, we propose that, rice farmers from the east would have belonged to the Munda (or more broadly the Austro-Asiatic) speaking category and the wheat farmers from the west were Indo-European and/or Brahui (Dravidian) speakers. These rice and wheat farming populations interacted with each other producing the intermediate zone where both types of crops would have existed. The linguistic scenario also would then have been different in this intermediate zone (the region of the "Indus" language). All the three families would then have influenced the language spoken in the intermediate zone. The inferred Indus language thus would be either a mixture of the three families or at the least would have been strongly influenced by the three distinct families.

6.2.4 Future work

The model needs to be tested and calibrated using the rice data from the Indian subcontinent.

A similar model, with different values of parameters, could be applied to other Neolithic origins to study the spread of major crops (e.g. Africa, Europe, China, Americas). The model could be useful in studying the spread of Bantu languages in Africa.

Prehistoric China shows presence of more than one signature crops in the north eastern and the south eastern area, and the competition model can be useful in understanding the distribution of these crops.

Chapter 7

Summary and Conclusions

The Neolithic period at the end of the Stone Age, about 12,000 yrs ago, is of great importance and interest as it involves a gradual change in the human lifestyle (from hunting-gathering to sedentism). The Neolithic is better explored in Europe in terms of the ^{14}C dates as compared to the rest of the world, especially the Near East, the Middle East and the Indian subcontinent. With a large number of radiometric records it has been shown that the Neolithic in Europe is a consequence of the spread from the Near East at a rate $\sim 1\,\mathrm{km/yr}$. In comparison, the region of our study from the Near East to the Indian subcontinent shown in Fig. 1.7 has been explored with only a handful of Neolithic ^{14}C dates. The region, however, is dated with comparative dating techniques.

The nature of the comparative and ^{14}C dates and their uncertainties is different. Due to this inconsistent nature any quantitative studies of the Neolithic in this region have been hindered. In this thesis we made use of both of these databases. We assigned different uncertainties to the comparative dates and compiled a comprehensive database for the Neolithic dates in South Asia, which included both the comparative as well as the ^{14}C dates.

In Chapters 4 and 5, we studied the possibility of an eastward spread of the Neolithic from the origins in the Fertile Crescent and Zagros Mountains. For this study only the

first arrival dates at each of the sites are of significance. The first arrival date(s) are chosen in two ways and two different databases, 'One Site – One Date' and 'One Site – Many Dates', are created with procedures shown in the flowcharts in Figures 4.2 and 4.3.

In Chapter 4 we analysed the possibility of an eastward Neolithic spread from the Fertile Crescent. We first considered a model of the *constant speed fit* from the Fertile Crescent to the Indian subcontinent with 'Gesher', one of the oldest Neolithic sites from the Fertile Crescent (in present-day Israel), as the conventional source of the Neolithic spread. The two datasets were processed with two new statistical procedures, weighted and percentile methods, for finding the 'upper envelopes' to the data (see section 4.2.1 for details). With the weighted data method, for a bin width of 200 km, the constant speed fit model gives a Neolithic spread speed of $U = 0.67\,\mathrm{km/yr}$ with the spread starting at $T^* = 11,000\,\mathrm{yr\,BCE}$. The same model using the percentile data method gives the Neolithic spread speed $U = 0.59\,\mathrm{km/yr}$ with the spread starting at $T^* = 10,300\,\mathrm{yr\,BCE}$. Figures 4.6 to 4.13 show that the values of the spread speeds and the Neolithic start times obtained from these two methods are insensitive to use of different percentile values (in the range 70–97 percentile), bin widths (ΔD), weighting time-scales (τ) and choice of databases (whether one date or many dates per site). Taking into account all of the aforementioned variations, we give a range for the constant spread speed and the Neolithic starting times to be $U = 0.65 \pm 0.10\,\mathrm{km/yr}$, and $T_* = 10,000 \pm 1,000\,\mathrm{yr\,BCE}$.

The region from the Fertile Crescent to the Indian subcontinent spans a very large area with strongly varying geography and climate. The Zagros Mountains separate the rich fertile lands on the west and the arid and semi-arid lands on the east. Hence there is a strong possibility that the Neolithic spread speeds in the regions west and east of the Zagros Mountains were different. For this reason we also considered a *variable speed fit* model that allows for a change in the spread speed U at a certain distance D_0 from the source. Given the scarcity of the data available, we only considered the simplest models of this type, where U is constant at distances $D < D_0$ and $D > D_0$. The process was repeated with D_0 varying in the range 900, 1100, ..., 3300 km. The model allows for

the dependence of T_0 on distance D to be discontinuous or continuous at $D = D_0$ (see section 4.2.1: Variable-speed fits for details).

Out of the three, the constant speed fit, the variable-speed fit (continuous) and the variable-speed fit (discontinuous) models, the best model is chosen on the basis of an information criterion: the second-order ('corrected') Akaike Information Criterion $\mathrm{AIC_c}$, as it is more appropriate for small datasets as ours than the Bayesian Information Criterion (BIC). Based on the value of $\mathrm{AIC_c}$ the best-fit model is the constant-speed model with $U = 0.59$ km/yr and $T_* = 10,283$ yr BCE, which has $\mathrm{AIC_c} = 24.6$. The discontinuous fit with $D_0 = 1100\,\mathrm{km}$ is a close second, with $\mathrm{AIC_c} = 25.8$ (see Table 4.2 for details). Our model assumes that the Neolithic spread proceeded with the same speed in all directions and neglects any effects of the local environment. Hence it applies at the global spatial and temporal scales. Despite its simplicity the model captures the salient features of the data quite well.

The Neolithic spread speed obtained by our constant-speed fit model is slightly lower ($\sim 0.65\,\mathrm{km/yr}$) than the one obtained for the spread of the Neolithic in Europe ($\sim 1\,\mathrm{km/yr}$). The difference in the Neolithic spread speeds from the two regions can be understood in terms of their diverse climates and topographies. Although the Fertile Crescent is the most productive land in the Near and the Middle East, the arid climate of the Middle East is not favourable for agriculture and thus may have hindered the Neolithic spread beyond the Fertile Crescent. Also a dependence on herding instead of farming in these arid lands may have been in favour of seasonal long-distance movements, thus increasing population mobility instead of promoting sedentism. A reason for the accelerated spread of the Neolithic across Europe was the presence of major rivers like the Danube and Rhine. The spread was also accelerated along the Mediterranean coastline. There are no large rivers in Iran and Afghanistan that could help accelerate the Neolithic spread east of the Zagros Mountains. The Iranian coastline is also primarily arid and could not have supported sedentism. Hence, as can be seen from Fig. 1.7, there are no Neolithic sites near the southern Iranian coast. Ignoring the local topographic and climatic variations

we determined the Neolithic spread speed across the Near and the Middle East to be 0.65 km/yr.

The Zagros Mountains separate the Fertile Crescent from the arid and semi-arid lands in the east. Iran has a diverse climate, ranging from a wetter climate near the Zagros mountains in the west, to the two major deserts, Dasht-e Kavir and Dasht-e Lut in the east. From Fig. 5.1 it can be seen that there is a wide gap between the archaeological sites in central and eastern Iran. The the gap persists even after recent archaeological surveys. The Neolithic sites are mostly located along the northern and southern edges of Iran, and resemble a forked pattern. With the help of the methods outlined in Chapter 4, we explored this interesting feature in Chapter 5.

We investigated the forked nature of the Neolithic in Iran by considering the possibility of separate Neolithic advancements via the routes north and south of the Zagros Mountains. We also determine whether the Indus Valley Neolithic was influenced by one or both of these propagations. From Figures 5.2 and 5.3 we found that commencing at about 10,900 yr BCE, the Neolithic occupation was restricted to the Fertile Crescent for about 4,000 years. Around 7,000 yr BCE the Neolithic sites are seen outside the Fertile Crescent. A climate deterioration event at about 6,200 yr BCE may have been one of the major causes of a search for a habitable zone outside the Fertile Crescent, along with the usual motivations such as population pressure and resource depletion.

The fact that the Neolithic was restricted to the Fertile Crescent for about 4,000 years and then started spreading out, has encouraged us to consider different source regions in the Fertile Crescent, for the northern and the southern routes, as effective sources. The sites on the northern and southern edges of Iran are chosen for the sites on the two branches of the fork (see Fig. 5.1). The sites belonging to these source regions and those on the north and south routes and the Indus Valley are listed in Tables 5.2 to 5.6.

As before, we used both versions of the datasets, 'One Site – One Date' and One Site – Many Dates' and analysed them with the procedures developed in section 4.2.1. Figures 5.4 to 5.6 and Table 5.7 show the results of the upper envelope fits for southern and northern

routes with and without the Indus Valley Neolithic data. The Neolithic sites Sarazm and Aykagytma are in the south-eastern Kazyl-Kum plains of present-day Uzbekistan. There is large spatial gap between these sites and other Neolithic sites. Mehrgarh is a site for which the ^{14}C dates show the Neolithic presence much earlier (around 8,500 BCE). With and without these three sites we checked if the analyses were consistent.

As can be seen from Table 5.7, the upper envelope fits for the southern route give an average starting time from the source region as $T_S^* = 7.2$ kyr BCE and an average propagation speed $U_S = 0.66$ km/yr. This speed is consistent with the one obtained in Chapter 4 i.e. 0.65 km/yr. For the northern route however, the average speed is $U_N = 1.84$ km/yr which is twice the southern route propagation speed. Northern Iran has a warm and wet Mediterranean climate whereas southern Iran exhibits a hot, dry and arid, semi-desert and desert type climate. This difference in the climates can explain the difference in the Neolithic spread speeds of the two routes. Our analysis showed that there were two distinct routes of the Neolithic propagation, one from the northern Iran, south of the Caspian Sea which may have been a part of the Silk Road and the other through southern Iran from the Fars and Mamasani districts (red arrow in Fig. 5.7).

The northern route sites were separately analysed with and without the inclusion of Ayakagytma and Sarazm. The upper envelope plots to this data give an average speed of 1.75 km/yr, consistent with the speed obtained for the envelope fitted to only the northern route sites. We therefore suggest one more branch of the Neolithic spread, an extension of the northern route from the south-east of the Caspian Sea to Central Asia, shown by the white dashed arrow in Fig. 5.7.

We analysed the northern and southern route datasets by adding the Indus Valley sites. For the southern route, the average speed is $U_{SI} = 3.58$ km/yr, more than four times faster than than the average speed of 0.66 km/yr obtained for the southern sites without the Indus Valley sites. With the northern route, the average speed obtained was 2.92 km/yr, a speed 1.5 times faster than for the northern route without the Indus Valley sites. The speeds for the northern route, with and without the Indus Valley sites, were

found to be similar compared to the southern route where there appears an inconsistent four fold increase. Thus we conclude that it is more likely that the Indus Valley Neolithic was influenced by the spread from the northern route through Iran, shown with the long yellow arrow in Fig. 5.7.

Our first models, although they ignore the local environmental effects, were encouraging and form a basis for more complex mathematical models. In Chapter 6 we investigate two such models and discussed the preliminary results. Both models are based on the FKPP equation and require more work.

In the first model we assumed the diffusivity is a function of the palaeovegetation data. The model was executed on three different domains: the whole domain from the Near East to the Indian subcontinent, the southern Iranian domain and the northern domain. The background diffusivity, used in order to obtain a desired spread speed in each domain, is different in each case and in some cases turns out to be high (see Table 6.2). The model uncertainties are shown in histograms and the sites between $\pm 3\sigma = \pm 900$ years are plotted on maps, in Figures 6.3 to 6.14. The results obtained from this model corroborate our conclusions from Chapters 4 and 5. The palaeovegetation-based diffusivity model fits the data better, as is seen from a comparison of the histograms in Figures 6.3 and 6.5 for the whole domain. The results of the model for the northern and southern routes, with and without the Indus data, favour an influence of the Neolithic, on the Indian subcontinent, from the northern route rather than the southern route. The model can be made more complex with the inclusion of other important factors such as rainfall, temperature, altitude (see Figures 6.15 and 6.16) and we leave this as future work.

The second model is based on the idea of the spread of a signature crop for a population which would in turn represent the spread of the population. The competitive interactions between the crops would then represent the interactions between the population groups. This model is executed with the incorporation of the altitude and rainfall data and its domain of application is the Indian subcontinent. The preliminary analysis indicates the effect of competitive interaction between the two crops, rice from the east and wheat

from the west (see Fig. 6.19). We also compare these preliminary model results with the prehistoric linguistic scenario of the Indian subcontinent displayed in Fig. 6.20. Even though the model requires more work, based on this scenario and the simulation results we can propose that the rice farmers in the east may have belonged to the Munda or Austro-Asiatic speaking category and the wheat farmers in the west could have been Indo-European or Brahui (Dravidian) speakers. The two populations would have interacted with each other producing an intermediate zone (the zone where both rice and wheat would exist). The 'Indus' language of this zone would then either be a mixture of all three families or at least would be strongly influenced by them.

Appendix A

The radiometric and archaeological dates

Table A1: The ^{14}C dates used in the analysis, arranged in the alphabetical order according to the site name. Column 2 shows the laboratory number for those dates. For some sites, we have omitted ^{14}C dates for reasons given in Table A3

No.	**Lab number**	**Site Name**	**Lat. °N**	**Long. °E**	**Age (yr cal BCE)**	$\tilde{\sigma}$ **(yr)**	**Distance from Gesher (km)**	**Ref.**
1	BM-1823	Abadah	33.97	44.83	4612	109	877	[1]
2	RT 2453	Abu Gosh	31.80	35.11	8037	213	102	[55]
3	OxA-1228	Abu Hureyra	35.87	38.40	9048	241	445	[1]
4	BM-1122	Abu Hureyra	35.87	38.40	8729	379	445	[1]
5	Pta-2699	Abu Madi I	28.56	34.00	9739	382	477	[1]
6	Pta-4568	Abu Madi I	28.56	34.00	9636	384	477	[1]
7	Pta-4552	Abu Madi I	28.56	34.00	9505	252	477	[1]
8	Pta-4577	Abu Madi I	28.56	34.00	9478	324	477	[1]
9	A-11802	Àin Abu Nukhayla	29.55	35.41	7735	206	345	[55, 77]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
10	A-11806	Àin Abu Nukhayla	29.55	35.41	7726	229	345	[55, 77]
11	A-11804	Àin Abu Nukhayla	29.55	35.41	7618	97	345	[55, 77]
12	A-11803	Àin Abu Nukhayla	29.55	35.41	7480	304	345	[55, 77]
13	A-11807	Àin Abu Nukhayla	29.55	35.41	7395	199	345	[55, 77]
14	GrN-12966	Àin Ghazal	31.93	35.94	8478	237	89	[1]
15	GrN-12965	Àin Ghazal	31.93	35.94	8248	281	89	[1]
16	GrN-12960	Àin Ghazal	31.93	35.94	8207	247	89	[1]
17	GrN-12959	Àin Ghazal	31.93	35.94	8134	303	89	[1]
18	Beta-138548	Akarçay Tepe	36.92	38.02	7784	170	527	[1]
19	Beta-174035	Akarçay Tepe	36.92	38.02	7587	59	527	[1]
20	Beta-138583	Akarçay Tepe	36.92	38.02	7370	228	527	[1]
21	P-1499	Ali Agha	36.45	43.82	5844	137	869	[1]
22	SI-160	Ali Kosh	32.56	47.32	8019	274	1105	[1]
23	Sh-1174	Ali Kosh	32.56	47.32	8038	501	1105	[96]
24	Beta-118721	Ali Kosh	32.56	47.32	7893	309	1105	[96]
25	Beta-536	Aqab	37.08	40.82	5120	181	690	[1]
26	OxA-574	Arjoune	34.56	36.55	5414	191	233	[1]
27	TB-300	Arukhlo 1	41.00	43.87	6029	180	1188	[1]
28	OxA-7882*	Ashkelon	31.61	34.50	6954	340	150	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
29	OxA-7883*	Ashkelon	31.61	34.50	6893	248	150	[1]
30	OxA-7995*	Ashkelon	31.61	34.50	6881	180	150	[1]
31	OxA-7916*	Ashkelon	31.61	34.50	6860	179	150	[1]
32	GrN-6434	Asiab	34.30	47.19	9117	282	1097	[1]
33	GrN-6413	Asiab	34.30	47.19	9144	305	1097	[96]
34	GrN-19116	Asikli Höyük	38.35	34.23	8106	165	644	[1]
35	Hd-19640	Asikli Höyük	38.35	34.23	8036	195	644	[1]
36	GrN-19865	Asikli Höyük	38.35	34.23	8003	244	644	[1]
37	GrN-20349	Asikli Höyük	38.35	34.23	7983	227	644	[1]
38	GrN-19120	Asikli Höyük	38.35	34.23	7950	272	644	[1]
39	GrN-19858	Asikli Höyük	38.35	34.23	7903	302	644	[1]
40	GrN-19115	Asikli Höyük	38.35	34.23	7891	311	644	[1]
41	GrN-19118	Asikli Höyük	38.35	34.23	7885	277	644	[1]
42	GrN-19869	Asikli Höyük	38.35	34.23	7883	289	644	[1]
43	GrN-20353	Asikli Höyük	38.35	34.23	7882	285	644	[1]
44	GrN-19870	Asikli Höyük	38.35	34.23	7881	295	644	[1]
45	GrN-20684	Asikli Höyük	38.35	34.23	7878	289	644	[1]
46	GrN-20354	Asikli Höyük	38.35	34.23	7876	290	644	[1]
47	P-1239	Asikli Höyük	38.35	34.23	7821	350	644	[1]
48	GrN-19119	Asikli Höyük	38.35	34.23	7787	173	644	[1]
49	GrN-18617	Asikli Höyük	38.35	34.23	7773	167	644	[1]
50	GrN-18618	Asikli Höyük	38.35	34.23	7772	169	644	[1]
51	GrN-18620	Asikli Höyük	38.35	34.23	7771	172	644	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
52	GrN-19860	Asikli Höyük	38.35	34.23	7770	169	644	[1]
53	GrN-20352	Asikli Höyük	38.35	34.23	7770	169	644	[1]
54	GrN-19861	Asikli Höyük	38.35	34.23	7760	177	644	[1]
55	GrN-19121	Asikli Höyük	38.35	34.23	7712	223	644	[1]
56	GrN-20351	Asikli Höyük	38.35	34.23	7687	96	644	[1]
57	GrN-19360	Asikli Höyük	38.35	34.23	7677	75	644	[1]
58	GrN-19363	Asikli Höyük	38.35	34.23	7667	67	644	[1]
59	GrN-19867	Asikli Höyük	38.35	34.23	7665	90	644	[1]
60	GrN-19362	Asikli Höyük	38.35	34.23	7651	65	644	[1]
61	GrN-19863	Asikli Höyük	38.35	34.23	7651	61	644	[1]
62	GrN-18619	Asikli Höyük	38.35	34.23	7646	100	644	[1]
63	GrN-19361	Asikli Höyük	38.35	34.23	7640	108	644	[1]
64	GrN-19868	Asikli Höyük	38.35	34.23	7633	306	644	[1]
65	GrN-19364	Asikli Höyük	38.35	34.23	7627	89	644	[1]
66	GrN-19862	Asikli Höyük	38.35	34.23	7625	92	644	[1]
67	GrN-19359	Asikli Höyük	38.35	34.23	7623	126	644	[1]
68	GrN-20356	Asikli Höyük	38.35	34.23	7620	103	644	[1]
69	GrN-19358	Asikli Höyük	38.35	34.23	7610	123	644	[1]
70	MC-865	Assouad	36.58	39.00	7836	363	541	[1]
71	LY-11385	Aswad	33.42	36.53	9289	459	127	
72	Gif-2633	Aswad	33.42	36.53	9198	447	127	[1]
73	Gif-2372	Aswad	33.42	36.53	8979	322	127	[1]
74	Pta-3950	Atlit-Yam	32.55	34.91	6907	258	58	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
75	Gif/LSM-11205	Ayakagytma	40.65	64.62	6045	33	2729	[131]
76	Gif-11206 / GifA-99182	Ayakagytma	40.65	64.62	5997	226	2729	[131]
77	Gif-11207 / GifA-99183	Ayakagytma	40.65	64.62	5990	218	2729	[131]
78	Gif/LSM-11103	Ayakagytma	40.65	64.62	5987	61	2729	[131]
79	Gif/LSM-11104	Ayakagytma	40.65	64.62	5984	74	2729	[131]
80	Gif-11208 / GifA-99181	Ayakagytma	40.65	64.62	5895	161	2729	[131]
81	Gif-11101 / GifA-99148	Ayakagytma	40.65	64.62	5869	142	2729	[131]
82	Gif-11099 / GifA-99131	Ayakagytma	40.65	64.62	5862	144	2729	[131]
83	Gif/LSM-11102	Ayakagytma	40.65	64.62	5835	60	2729	[131]
84	Gif-10660	Ayakagytma	40.65	64.62	5681	166	2729	[131]
85	Gif-11105	Ayakagytma	40.65	64.62	5563	80	2729	[131]
86	OxA-870	Azraq 31	31.83	36.82	7334	251	153	[1]
87	Bln-5123	Ba´ja	30.41	35.46	7110	68	249	[1]
88	Bln-5036	Ba´ja	30.41	35.46	6840	190	249	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
89	Bln-5035	Ba´ja	30.41	35.46	6834	195	249	[1]
90	KIA-11579	Baaz	33.81	36.51	4569	111	158	[1]
91	Hd-22340	Bademağaci	37.40	30.48	6865	167	699	[1]
92	TF-786	Bagor	25.35	74.38	5146	425	3844	[110]
93	P-931	Bakun B	29.85	52.83	5246	216	1673	[1, 96]
94	P-438	Bakun B	29.85	52.83	4948	257	1673	[1, 96]
95	UCLA-1923A	Balakot	25.48	66.73	4019	318	3124	[110]
96	UCLA-1923B	Balakot	25.48	66.73	3992	221	3124	[110]
97	GrN-14537	Basta	30.23	35.53	7384	208	269	[1]
98	GrN-14538	Basta	30.23	35.53	7185	128	269	[1]
99	P-1380	Beidha	30.37	35.45	8313	321	254	[1]
100	GrN-5062	Beidha	30.37	35.45	8155	166	254	[1]
101	GrN-5063	Beidha	30.37	35.45	8009	215	254	[1]
102	P-1382	Beidha	30.37	35.45	7976	315	254	[1]
103	GrN-5136	Beidha	30.37	35.45	7964	242	254	[1]
104	K-1410	Beidha	30.37	35.45	7955	297	254	[1]
105	P-1381	Beidha	30.37	35.45	7902	304	254	[1]
106	K-1082	Beidha	30.37	35.45	7892	317	254	[1]
107	P-1379	Beidha	30.37	35.45	7644	293	254	[1]
108	P-19c	Belt	36.65	53.28	6506	970	1682	[96]
109	RT-1394	Betzet 1	33.07	35.15	7329	244	58	[1]
110	GrN-8262	Bouqras	35.03	40.39	7442	101	522	[1]
111	GrN-13080	Bouqras	35.03	40.39	7427	118	522	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
112	GrN-13102	Bouqras	35.03	40.39	7426	91	522	[1]
113	GrN-8263	Bouqras	35.03	40.39	7355	212	522	[1]
114	GrN-13104	Bouqras	35.03	40.39	7338	155	522	[1]
115	GrN-13103	Bouqras	35.03	40.39	7334	158	522	[1]
116	GrN-4852	Bouqras	35.03	40.39	7286	226	522	[1]
117	GrN-13101	Bouqras	35.03	40.39	7265	182	522	[1]
118	GrN-8261	Bouqras	35.03	40.39	7183	125	522	[1]
119	GrN-8258	Bouqras	35.03	40.39	7168	128	522	[1]
120	OxA-2770	Burqu 35	31.95	37.20	7289	208	176	[1]
121	OxA-2769	Burqu 35	31.95	37.20	7255	213	176	[1]
122	OxA-2768	Burqu 35	31.95	37.20	7140	314	176	[1]
123	GrN-1544	Byblos	34.12	35.65	6231	159	164	[1]
124	Ly-4437	Cafer	38.42	38.75	8061	234	705	[1]
125	Ly-2523	Cafer	38.42	38.75	7773	407	705	[1]
126	HU-11	Canhasan III	37.25	33.37	7635	115	548	[1]
127	HU-12	Canhasan III	37.25	33.37	7601	115	548	[1]
128	BM-1667R	Canhasan III	37.25	33.37	7470	281	548	[1]
129	BM-1662R	Canhasan III	37.25	33.37	7455	268	548	[1]
130	OxA-9778	Catalhöyük East	37.65	32.82	7269	187	608	[1]
131	OxA-9777	Catalhöyük East	37.65	32.82	7186	128	608	[1]
132	OxA-9893	Catalhöyük East	37.65	32.82	7185	128	608	[1]
133	OxA-9892	Catalhöyük East	37.65	32.82	7183	128	608	[1]
134	P-779	Catalhöyük East	37.65	32.82	7178	336	608	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
135	OxA-9775	Catalhöyük East	37.65	32.82	7063	236	608	[1]
136	OxA-9948	Catalhöyük East	37.65	32.82	7062	234	608	[1]
137	AA-47057	Catalhöyük East	37.65	32.82	7043	265	608	[1]
138	P-782	Catalhöyük East	37.65	32.82	7020	323	608	[1]
139	PL-972425A	Catalhöyük East	37.65	32.82	7003	303	608	[1]
140	AA-18104	Catalhöyük East	37.65	32.82	6998	181	608	[1]
141	P-1370	Catalhöyük East	37.65	32.82	6977	324	608	[1]
142	P-775	Catalhöyük East	37.65	32.82	6976	319	608	[1]
143	OxA-9949	Catalhöyük East	37.65	32.82	6959	183	608	[1]
144	PL-9800565A	Catalhöyük East	37.65	32.82	6939	243	608	[1]
145	GrN-8103	Cayönü	38.22	39.73	10368	250	727	[1]
146	UCLA-305	Chagha Sefid	32.63	47.26	10252	428	1099	[1]
147	OxA-9994	Cheshmeh Ali	27.31	61.41	5144	147	2556	[57]
148	OxA-9995	Cheshmeh Ali	27.31	61.41	5109	109	2556	[57]
149	OxA-9996	Cheshmeh Ali	27.31	61.41	5092	129	2556	[57]
150	OxA-9855	Cheshmeh Ali	27.31	61.41	5010	202	2556	[57]
151	OxA-9956	Cheshmeh Ali	27.31	61.41	4841	116	2556	[57]
152	OxA-9937	Cheshmeh Ali	27.31	61.41	4834	129	2556	[57]
153	Erl-14835	Chia Sabz	33.34	47.14	8436	124	1086	[118]
154	Erl-14836	Chia Sabz	33.34	47.14	8107	141	1086	[118]
155	KIA43836	Čoḡā Golān	33.38	46.27	8694	123	1005	[118]
156	Beta-104553	Čoḡā Bonut	32.22	48.51	10125	446	1219	[13]
157	Beta-106167	Čoḡā Mīš	32.21	48.55	7348	172	1223	[13]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
158	UtC-1094/-1095/-1096	Damishliyah	36.49	39.03	6794	287	534	[1]
159	UtC-1124	Damishliyah	36.49	39.03	6574	181	534	[1]
160	UtC-1097/-1098/-1099	Damishliyah	36.49	39.03	6533	100	534	[1]
161	AA-38142	Dhra	31.27	35.58	9688	332	154	[1]
162	AA-38141	Dhra	31.27	35.58	9659	345	154	[1]
163	ISGS-2898	Dhra	31.27	35.58	9630	378	154	[1]
164	AA-38144	Dhra	31.27	35.58	9557	257	154	[1]
165	AA-38143	Dhra	31.27	35.58	9550	252	154	[1]
166	ISGS-A0246	Dhra	31.27	35.58	9462	196	154	[1]
167	ISGS-A0248	Dhra	31.27	35.58	9413	232	154	[1]
168	OxA-1637	Dhuweila	32.03	37.25	7362	220	177	[1]
169	BM-2349	Dhuweila	32.03	37.25	7253	195	177	[1]
170	Ly-12110	Djade	36.65	38.19	8975	202	507	[1]
171	Ly-11329	Djade	36.65	38.19	8877	244	507	[1]
172	Ly-11330	Djade	36.65	38.19	8727	380	507	[1]
173	Ly-8842	Djade	36.65	38.19	8686	125	507	[1]
174	Ly-4400	El Kowm 2	35.22	38.83	6932	249	419	[1]
175	Bln 4973	Es Sifiya	31.44	35.82	6951	175	137	[72]
176	Bln 4969	Es Sifiya	31.44	35.82	6925	149	137	[72]
177	Bln 4971	Es Sifiya	31.44	35.82	6886	176	137	[72]
178	Bln 4972	Es Sifiya	31.44	35.82	6881	178	137	[72]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
179	Bln 4968	Es Sifiya	31.44	35.82	6876	176	137	[72]
180	ISGS-3279	Es Sifiya	31.44	35.82	6848	195	137	[72]
181	Bln 4970	Es Sifiya	31.44	35.82	6843	188	137	[72]
182	HD-12335	Feinan	30.62	35.43	5348	122	226	[1]
183	HD-12338	Feinan	30.62	35.43	5062	225	226	[1]
184	OxA-2100	Ganj Dareh	34.42	47.57	8164	370	1133	[1]
185	B-108245	Ganj Dareh	34.42	47.57	8116	154	1133	[1]
186	B-108242	Ganj Dareh	34.42	47.57	8116	154	1133	[1]
187	B-108243	Ganj Dareh	34.42	47.57	8106	165	1133	[1]
188	SI-4741	Ganj Dareh	34.42	47.57	8068	224	1133	[1]
189	B-108239	Ganj Dareh	34.42	47.57	8063	217	1133	[1]
190	P-1484	Ganj Dareh	34.42	47.57	8048	296	1133	[1]
191	B-108248	Ganj Dareh	34.42	47.57	8045	204	1133	[1]
192	B-108246	Ganj Dareh	34.42	47.57	8024	207	1133	[1]
193	P-1486	Ganj Dareh	34.42	47.57	8004	279	1133	[1]
194	B-108244	Ganj Dareh	34.42	47.57	7983	227	1133	[1]
195	B-108249	Ganj Dareh	34.42	47.57	7983	227	1133	[1]
196	B-108247	Ganj Dareh	34.42	47.57	7979	230	1133	[1]
197	OxA-2101	Ganj Dareh	34.42	47.57	7955	297	1133	[1]
198	OxA-2099	Ganj Dareh	34.42	47.57	7932	318	1133	[1]
199	B-108238	Ganj Dareh	34.42	47.57	7896	284	1133	[1]
200	B-108240	Ganj Dareh	34.42	47.57	7896	284	1133	[1]
201	OxA-2102	Ganj Dareh	34.42	47.57	7875	327	1133	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
202	P-1494	Gawra	36.43	43.34	5874	146	830	[1]
203	RT-814a	Gesher	32.65	35.52	10459	348	0	[1]
204	ISGS 4366	Ghuwayr 1	30.62	35.53	9151	506	226	[125]
205	RT-777A	Gilgal	32.00	35.44	9628	477	69	[1]
206	RT-777B	Gilgal	32.00	35.44	9585	836	69	[1]
207	Pta-4588	Gilgal	32.00	35.44	9502	244	69	[1]
208	Pta-4583	Gilgal	32.00	35.44	9309	348	69	[1]
209	Hd-20036	Göbekli Tepe	37.22	38.92	8967	208	595	[1]
210	Hd-20025	Göbekli Tepe	37.22	38.92	8845	286	595	[1]
211	Beta-13216	Gritille	37.55	38.57	7779	427	611	[1]
212	AA-41602	Hacilar	37.57	30.07	6334	96	738	[1]
213	AA-41603	Hacilar	37.57	30.07	6328	96	738	[1]
214	AA-41604	Hacilar	37.57	30.07	6256	159	738	[1]
215	P-313A	Hacilar	37.57	30.07	6228	174	738	[1]
216	P-314	Hacilar	37.57	30.07	6216	185	738	[1]
217	P-313	Hacilar	37.57	30.07	6022	209	738	[1]
218	P-455	Hajji Firuz	37.04	45.54	6175	185	1036	[1]
219	Beta-58928	Halula	36.42	38.19	7766	173	485	[1]
220	Beta-50856	Halula	36.42	38.19	7758	187	485	[1]
221	GifA-91139	Hatoula	31.82	34.98	9903	528	105	[1]
222	GifA-91360	Hatoula	31.82	34.98	9689	428	105	[1]
223	GrN-12510	Hayaz Höyük	37.48	38.33	7348	172	595	[1]
224	Pta-3625	Hemar	31.17	35.18	7965	285	168	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
225	OxA-1016	Hemar	31.17	35.18	7920	312	168	[1]
226	OxA-1014	Hemar	31.17	35.18	7773	407	168	[1]
227	Bln-4609	Hoca Çeşme	40.70	26.09	6510	81	1227	[1]
228	Hd-16725/119145	Hoca Çeşme	40.70	26.09	6347	115	1227	[1]
229	GrN-19779	Hoca Çeşme	40.70	26.09	6227	135	1227	[1]
230	Hd-16724/17186	Hoca Çeşme	40.70	26.09	6122	90	1227	[1]
231	Hd-16727/17038	Hoca Çeşme	40.70	26.09	5902	109	1227	[1]
232	Hd-16726/17084	Hoca Çeşme	40.70	26.09	5896	90	1227	[1]
233	RT-1397	Horvat Galil	32.96	35.32	8564	216	39	[1]
234	HD-14219/14007	Höyücek	37.45	30.57	6372	108	698	[1]
235	HD-14218/14002	Höyücek	37.45	30.57	6370	108	698	[1]
236	HD-14217/13822	Höyücek	37.45	30.57	6218	138	698	[1]
237	Gx-864§	Iblis	30.16	56.84	4970	351	2039	[96]
238	P-925§	Iblis	30.16	56.84	4737	192	2039	[96]
239	P-926§	Iblis	30.16	56.84	4724	184	2039	[96]
240	GrN-19351	Ilipinar	40.46	29.30	6099	108	1030	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
241	GrN-22046	Ilipinar	40.46	29.30	6037	190	1030	[1]
242	GrN-19354	Ilipinar	40.46	29.30	6035	51	1030	[1]
243	GrN-17046	Ilipinar	40.46	29.30	5976	71	1030	[1]
244	GrN-19795	Ilipinar	40.46	29.30	5976	77	1030	[1]
245	GrN-15085	Ilipinar	40.46	29.30	5975	90	1030	[1]
246	GrN-19352	Ilipinar	40.46	29.30	5950	67	1030	[1]
247	GrN-22787	Ilipinar	40.46	29.30	5950	103	1030	[1]
248	GrN-15087	Ilipinar	40.46	29.30	5938	95	1030	[1]
249	GrN-24613	Ilipinar	40.46	29.30	5931	123	1030	[1]
250	GrN-17045	Ilipinar	40.46	29.30	5916	71	1030	[1]
251	GrN-19793	Ilipinar	40.46	29.30	5903	95	1030	[1]
252	GrN-15077	Ilipinar	40.46	29.30	5895	112	1030	[1]
253	GrN-17048	Ilipinar	40.46	29.30	5895	168	1030	[1]
254	GrN-24615	Ilipinar	40.46	29.30	5892	98	1030	[1]
255	GrN-19794	Ilipinar	40.46	29.30	5886	113	1030	[1]
256	GrN-17054	Ilipinar	40.46	29.30	5885	98	1030	[1]
257	GrN-22788	Ilipinar	40.46	29.30	5880	104	1030	[1]
258	GrN-19792	Ilipinar	40.46	29.30	5880	133	1030	[1]
259	GrN-24614	Ilipinar	40.46	29.30	5877	108	1030	[1]
260	GrN-17052	Ilipinar	40.46	29.30	5876	110	1030	[1]
261	GrN-18478	Ilipinar	40.46	29.30	5868	123	1030	[1]
262	GrN-17055	Ilipinar	40.46	29.30	5866	116	1030	[1]
263	GrN-17051	Ilipinar	40.46	29.30	5858	120	1030	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
264	GrN-17056	Ilipinar	40.46	29.30	5854	121	1030	[1]
265	GrN-24616	Ilipinar	40.46	29.30	5850	132	1030	[1]
266	GrN-17047	Ilipinar	40.46	29.30	5828	154	1030	[1]
267	GrN-15078	Ilipinar	40.46	29.30	5827	154	1030	[1]
268	GrN-16144	Ilipinar	40.46	29.30	5814	82	1030	[1]
269	GrN-16149	Ilipinar	40.46	29.30	5808	174	1030	[1]
270	GrN-19791	Ilipinar	40.46	29.30	5785	112	1030	[1]
271	GrN-22786	Ilipinar	40.46	29.30	5780	107	1030	[1]
272	GrN-19353	Ilipinar	40.46	29.30	5777	65	1030	[1]
273	GrN-19350	Ilipinar	40.46	29.30	5777	106	1030	[1]
274	GrN-21215	Ilipinar	40.46	29.30	5762	131	1030	[1]
275	GrN-22785	Ilipinar	40.46	29.30	5738	99	1030	[1]
276	GrN-22043	Ilipinar	40.46	29.30	5734	101	1030	[1]
277	GrN-16145	Ilipinar	40.46	29.30	5720	170	1030	[1]
278	GrN-19384	Ilipinar	40.46	29.30	5690	73	1030	[1]
279	GrN-18480	Ilipinar	40.46	29.30	5679	127	1030	[1]
280	GrN-18484	Ilipinar	40.46	29.30	5671	108	1030	[1]
281	GrN-18483	Ilipinar	40.46	29.30	5661	102	1030	[1]
282	GrN-24611	Ilipinar	40.46	29.30	5652	84	1030	[1]
283	GrN-17053	Ilipinar	40.46	29.30	5644	108	1030	[1]
284	GrN-22041	Ilipinar	40.46	29.30	5637	72	1030	[1]
285	OxA-2567	Iraq ed-Dubb	32.39	35.73	9561	312	35	[1]
286	KN-I 336	Iraq el-Barud	32.73	34.98	8408	199	51	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
287	RT-1607	Issaron	29.90	35.03	8442	159	309	[1]
288	RT-1510	Issaron	29.90	35.03	8281	286	309	[1]
289	Beta-258649	Jani	33.95	46.78	7106	239	1056	[44]
290	GrN-6353	Jarmo	35.56	44.92	6523	120	923	[1]
291	OxA-2914	Jeitun	37.95	58.23	6182	197	2137	[1]
292	OxA-2915	Jeitun	37.95	58.23	6069	178	2137	[1]
293	OxA-2916	Jeitun	37.95	58.23	6064	177	2137	[1]
294	OxA-2913	Jeitun	37.95	58.23	6059	178	2137	[1]
295	OxA-2912	Jeitun	37.95	58.23	5990	218	2137	[1]
296	OxA-4916	Jeitun	37.95	58.23	6037	395	2137	[96]
297	OxA-4694	Jeitun	37.95	58.23	6025	181	2137	[96]
298	OxA-4915	Jeitun	37.95	58.23	5940	130	2137	[96]
299	OxA-4690	Jeitun	37.95	58.23	5893	131	2137	[96]
300	OxA-4692	Jeitun	37.95	58.23	5884	135	2137	[96]
301	OxA-4693	Jeitun	37.95	58.23	5872	131	2137	[96]
302	OxA-4914	Jeitun	37.95	58.23	5834	173	2137	[96]
303	OxA-4691	Jeitun	37.95	58.23	5756	123	2137	[96]
304	Ly-10651	Jerf el Ahmar	36.39	38.20	9523	229	483	[1]
305	Ly-10648	Jerf el Ahmar	36.39	38.20	9437	218	483	[1]
306	Beta-71870	Jerf el Ahmar	36.39	38.20	9306	130	483	[1]
307	Ly-275	Jerf el Ahmar	36.39	38.20	9244	398	483	[1]
308	P-378	Jericho	31.86	35.47	9231	422	88	[1]
309	P-379	Jericho	31.86	35.47	9035	233	88	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
310	P-377	Jericho	31.86	35.47	8988	258	88	[1]
311	BM-1327	Jericho	31.86	35.47	8964	228	88	[1]
312	OxA-2969	Jilat 26	31.50	36.42	7898	309	153	[1]
313	OxA-2407	Jilat 26	31.50	36.42	7893	309	153	[1]
314	OxA-1802	Jilat 26	31.50	36.42	7875	327	153	[1]
315	OxA-526	Jilat 7	31.52	36.42	7919	308	152	[1]
316	OxA-527	Jilat 7	31.52	36.42	7624	313	152	[1]
317	GifA-100396	Kaletepe	38.28	34.57	8207	247	632	[1]
318	GifA-100631	Kaletepe	38.28	34.57	8134	303	632	[1]
319	GifA-99090	Kaletepe	38.28	34.57	7965	285	632	[1]
320	TF-439	Kalibangan	29.42	74.08	5412	188	3672	[110]
321	HD-10818/10747	Karain B	37.07	30.55	6267	167	669	[1]
322	Hd-10817/10764	Karain B	37.07	30.55	5949	148	669	[1]
323	TK-859	Kashkashok II	36.71	40.59	6781	275	647	[1]
324	TK-803	Kashkashok II	36.71	40.59	6619	199	647	[1]
325	GrN-26147	Khirbet Hammam	31.02	35.65	7437	94	182	[1]
326	P-524	Kili Ghul Mohammad I	30.28	66.97	4277	188	2983	[110]
327	L-180A	Kili Ghul Mohammad I	30.28	66.97	4271	1165	2983	[110]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
328	UW-61	Kili Ghul Mohammad I	30.28	66.97	4074	246	2983	[110]
329	ETH-39511	Körtik Tepe	37.83	40.97	9739	292	759	[26]
330	ETH-38851	Körtik Tepe	37.83	40.97	9726	281	759	[26]
331	ETH-38849	Körtik Tepe	37.83	40.97	9639	234	759	[26]
332	ETH-38855	Körtik Tepe	37.83	40.97	9600	209	759	[26]
333	ETH-38850	Körtik Tepe	37.83	40.97	9597	206	759	[26]
334	ETH-38853	Körtik Tepe	37.83	40.97	9568	230	759	[26]
335	ETH-38854	Körtik Tepe	37.83	40.97	9540	213	759	[26]
336	ETH-38848	Körtik Tepe	37.83	40.97	9529	212	759	[26]
337	ETH-39509	Körtik Tepe	37.83	40.97	9521	231	759	[26]
338	ETH-38852	Körtik Tepe	37.83	40.97	9483	182	759	[26]
339	ETH-39512	Körtik Tepe	37.83	40.97	9479	182	759	[26]
340	ETH-39510	Körtik Tepe	37.83	40.97	9468	186	759	[26]
341	Beta-178242	Körtik Tepe	37.83	40.97	9347	89	759	[103, 104]
342	GrN-12652	Kumartepe	37.48	38.48	6849	207	601	[1]
343	HD-12915/12673	Kuruçay Höyük	37.62	30.15	6196	165	737	[1]
344	HD-12916/12674	Kuruçay Höyük	37.62	30.15	5997	71	737	[1]
345	HD-12917/12830	Kuruçay Höyük	37.62	30.15	5902	174	737	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
346	IGAN-772	Magzaliyah	36.39	42.33	6888	181	749	[1]
347	Pta-3652	Megadim	32.72	34.97	5921	140	52	[1]
348	BETA-1721	Mehrgarh	29.42	67.58	8520	430	3064	[110]
349	GrN-25819	Menteşe	40.27	29.52	6368	112	1002	[1]
350	GrN-25821	Menteşe	40.27	29.52	6321	75	1002	[1]
351	GrN-25822	Menteşe	40.27	29.52	6154	83	1002	[1]
352	GrN-25823	Menteşe	40.27	29.52	6138	75	1002	[1]
353	GrN-25824	Menteşe	40.27	29.52	6117	96	1002	[1]
354	GrN-24463	Menteşe	40.27	29.52	6102	116	1002	[1]
355	GrN-24461	Menteşe	40.27	29.52	6066	147	1002	[1]
356	B-2737	Mezad Mazal	30.93	35.32	7481	125	192	[1]
357	Hv-9108	Mezad Mazal	30.93	35.32	7458	143	192	[1]
358	KN-2444	Mezad Mazal	30.93	35.32	7376	192	192	[1]
359	Hv-9107	Mezad Mazal	30.93	35.32	7346	201	192	[1]
360	Hv-9106	Mezad Mazal	30.93	35.32	7280	219	192	[1]
361	AA-49102	Mezraa Teleilat	36.98	38.00	8549	200	532	[1]
362	Gd-6150	M´lefaat	36.31	43.54	10888	258	840	[1]
363	TF-1129	Mundigak	31.58	65.50	3744	226	2817	[110]
364	Ly-4927	Munhata	32.60	35.55	6216	157	6	[1]
365	Lv-607	Mureybet	36.08	38.10	10459	348	449	[1]
366	TK-34	Mushki	29.78	52.90	7845	354	1681	[1]
367	OxA-375	Naja	31.77	36.92	6268	187	164	[1]
368	Gd-2970	Nemrik 9	36.73	42.88	9699	399	811	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
369	Gd-4209	Nemrik 9	36.73	42.88	9689	407	811	[1]
370	Gd-5257	Nemrik 9	36.73	42.88	9653	353	811	[1]
371	Gd-5595	Nemrik 9	36.73	42.88	9561	312	811	[1]
372	Pta-4557	Netiv HaGdud	31.98	35.38	9240	405	76	[1]
373	Pta-4555	Netiv HaGdud	31.98	35.38	9105	282	76	[1]
374	Pta-4590	Netiv HaGdud	31.98	35.38	9061	235	76	[1]
375	Pta-4556	Netiv HaGdud	31.98	35.38	9042	219	76	[1]
376	OxA-8303	Nevali Çori	37.58	38.65	8509	188	618	[1]
377	KIA-14756	Nevali Çori	37.58	38.65	8482	136	618	[1]
378	Hd-16782-351	Nevali Çori	37.58	38.65	8460	151	618	[1]
379	Hd-16783-769	Nevali Çori	37.58	38.65	8453	166	618	[1]
380	OxA-8302	Nevali Çori	37.58	38.65	8428	134	618	[1]
381	KIA-14762	Nevali Çori	37.58	38.65	8424	124	618	[1]
382	OxA-8235	Nevali Çori	37.58	38.65	8418	133	618	[1]
383	KIA-14760	Nevali Çori	37.58	38.65	8344	102	618	[1]
384	Hv-8509	Nizzanim	31.72	34.58	5710	170	136	[1]
385	OxA-5500	Pinarbasi Site A	37.48	33.03	8516	210	583	[1]
386	OxA-5501	Pinarbasi Site A	37.48	33.03	8401	165	583	[1]
387	OxA-5499	Pinarbasi Site A	37.48	33.03	8248	281	583	[1]
388	OxA-5504	Pinarbasi Site B	37.49	33.04	6283	170	584	[1]
389	OxA-5503	Pinarbasi Site B	37.49	33.04	6050	162	584	[1]
390	SMU-662	Qadesh Barnea 3	30.62	34.40	6361	237	249	[1]
391	Pta-3662	Qadesh Barnea 3	30.62	34.40	6228	168	249	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
392	RT-1544	Qanah	32.75	35.33	5908	150	21	[1]
393	Pta-2968	Qatif Y-3	31.41	34.32	4990	219	178	[1]
394	OxA-3752	Qermez Dere	36.38	42.45	9775	375	758	[1]
395	OxA-3756	Qermez Dere	36.38	42.45	9743	371	758	[1]
396	GrN-14539	Rahub	32.62	35.93	6292	192	39	[1]
397	GrN-4426	Ramad	33.42	36.10	7259	188	101	[1]
398	GrN-4428	Ramad	33.42	36.10	7256	203	101	[1]
399	GrN-4821	Ramad	33.42	36.10	7062	234	101	[1]
400	P-2148	Rana Ghundai	30.40	68.75	4433	100	3147	[110]
401	P-460	Ras Shamra	35.58	35.73	7365	222	326	[1]
402	P-459	Ras Shamra	35.58	35.73	7144	326	326	[1]
403	Pta-3137	Reu´el	30.10	35.10	7878	289	286	[1]
404	Pta-3202	Reu´el	30.10	35.10	7845	318	286	[1]
405	Pta-2848	Reu´el	30.10	35.10	7705	120	286	[1]
406	GrN-21319	Sabi Abyad II	36.53	39.10	7569	87	542	[1]
407	GrN-22273	Sabi Abyad II	36.53	39.10	7253	195	542	[1]
408	Pta-3821	Samir	32.82	34.95	5735	107	57	[1]
409	Pta-3820	Samir	32.82	34.95	5733	163	57	[1]
410	P-466	Sarab	34.38	47.09	6863	264	1089	[1]
411	Beta-159550	Sarab	34.38	47.09	7032	260	1089	[96]
412	Beta-159548	Sarab	34.38	47.09	6865	183	1089	[96]
413	LE2172	Sarazm	39.52	67.57	3837	128	2962	[81]
414	LE2174	Sarazm	39.52	67.57	3716	63	2962	[81]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
415	LE2173	Sarazm	39.52	67.57	3673	37	2962	[81]
416	P-855	Sawwan I	34.12	43.93	6286	173	798	[1]
417	P-856	Sawwan I	34.12	43.93	6193	176	798	[1]
418	SI-2668§	Seh Gabi	34.58	48.00	5163	199	1175	[96]
419	SI-2669§	Seh Gabi	34.58	48.00	5110	262	1175	[96]
420	SI-2670§	Seh Gabi	34.58	48.00	5000	213	1175	[96]
421	OxA-7919*	Sha´ar Hagolan	32.68	35.61	6345	98	9	[1]
422	OxA-7918*	Sha´ar Hagolan	32.68	35.61	6333	95	9	[1]
423	OxA-7917*	Sha´ar Hagolan	32.68	35.61	6263	154	9	[1]
424	OxA-7885*	Sha´ar Hagolan	32.68	35.61	6175	180	9	[1]
425	OxA-7920*	Sha´ar Hagolan	32.68	35.61	6122	100	9	[1]
426	Beta-258647	Sheikh-e Abad	34.61	47.27	9306	130	1109	[45]
427	Wk-15160	Shkarat Msaied	30.45	35.44	8405	135	245	[1]
428	Wk-15159	Shkarat Msaied	30.45	35.44	8128	164	245	[1]
429	LE-631	Shōmu	41.16	45.38	6352	118	1288	[1]
430	Bta-35081	Shu´eib	31.95	35.69	7719	245	79	[1]
431	QU-1035§	Siahbid	34.50	47.25	4748	290	1105	[96]
432	P-442§	Siahbid	34.50	47.25	4695	202	1105	[96]
433	GrN-9833	Sinn	35.28	40.36	7685	104	534	[1]
434	Hv 1355	Snake Cave (Ghar-i-Mar, Aq Kupruk I)	36.08	66.83	9954	657	2887	[110]
435	IGAN-769	Sotto	36.28	42.37	6335	103	746	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
436	IGAN-774	Sotto	36.28	42.37	6054	149	746	[1]
437	P-1391	Suberde	37.35	31.93	7281	215	616	[1]
438	P-1388	Suberde	37.35	31.93	7252	214	616	[1]
439	TO-1407	Tabaqat al-Buma	32.53	35.72	6734	268	23	[1]
440	AA-56411	Tal-e Jari B	29.85	52.96	6163	172	1685	[96]
441	AA-56410	Tal-e Jari B	29.85	52.96	6065	154	1685	[96]
442	AA-56415	Tal-e Jari B	29.85	52.96	6026	181	1685	[96]
443	AA-56412	Tal-e Jari B	29.85	52.96	5858	133	1685	[96]
444	AA65264	Tal-e Jari B	29.85	52.96	6150	88	1685	[14]
445	Beta-207565	Tal-e Jari B	29.85	52.96	5998	78	1685	[14]
446	AA63491	Tall-e Bakun A	29.91	52.89	4462	123	1676	[14]
447	Beta-210983	Tall-e Bakun A	29.91	52.89	4415	73	1676	[14]
448	Beta-207562	Tall-e Bakun A	29.91	52.89	4400	62	1676	[14]
449	AA63489	Tall-e Bakun B	29.85	52.83	5179	178	1673	[14]
450	Beta-210985	Tall-e Bakun B	29.85	52.83	5109	109	1673	[14]
451	AA63492	Tall-e Jari A	29.86	52.96	5259	207	1684	[14]
452	Beta-207564	Tall-e Jari A	29.86	52.96	5110	111	1684	[14]
453	Beta-210982	Tall-e Jari A	29.86	52.96	4898	102	1684	[14]
454	NUTA2-12459	Tang-e Bolaghi	30.15	53.13	9939	177	1692	[96]
455	Pta-2700	Tbeik	28.82	33.94	10236	361	452	[1]
456	OxA-7886*	Tel Ali	32.70	35.56	6872	191	7	[1]
457	OxA-7921*	Tel Ali	32.70	35.56	6863	178	7	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name		Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
458	OxA-17739	Tepe Boneh	Chahar	35.80	50.03	5751	85	1378	[109]
459	OxA-17740	Tepe Boneh	Chahar	35.80	50.03	5805	76	1378	[109]
460	OxA-17741	Tepe Boneh	Chahar	35.80	50.03	5800	76	1378	[109]
461	OxA-17742	Tepe Boneh	Chahar	35.80	50.03	5991	73	1378	[109]
462	OxA-17743	Tepe Boneh	Chahar	35.80	50.03	5921	77	1378	[109]
463	OxA-17744	Tepe Boneh	Chahar	35.80	50.03	5717	74	1378	[109]
464	OxA-17704	Tepe Boneh	Chahar	35.80	50.03	5176	120	1378	[109]
465	OxA-17745	Tepe Boneh	Chahar	35.80	50.03	5344	120	1378	[109]
466	OxA-17746	Tepe Boneh	Chahar	35.80	50.03	5190	118	1378	[109]
467	OxA-17747	Tepe Boneh	Chahar	35.80	50.03	5201	121	1378	[109]
468	OxA-17748	Tepe Boneh	Chahar	35.80	50.03	5290	72	1378	[109]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
469	OxA-17749	Tepe Chahar Boneh	35.80	50.03	5287	70	1378	[109]
470	OxA-17750	Tepe Chahar Boneh	35.80	50.03	5348	120	1378	[109]
471	OxA-17751	Tepe Chahar Boneh	35.80	50.03	5116	105	1378	[109]
472	OxA-17752	Tepe Chahar Boneh	35.80	50.03	5281	70	1378	[109]
473	OxA-17585	Tepe Ebrahim Abad	36.12	53.05	5144	148	1652	[109]
474	OxA-17597	Tepe Ebrahim Abad	36.12	53.05	5385	84	1652	[109]
475	OxA-17598	Tepe Ebrahim Abad	36.12	53.05	5286	69	1652	[109]
476	OxA-17599	Tepe Ebrahim Abad	36.12	53.05	5104	106	1652	[109]
477	OxA-17600	Tepe Ebrahim Abad	36.12	53.05	5021	173	1652	[109]
478	OxA-17601	Tepe Ebrahim Abad	36.12	53.05	5181	119	1652	[109]
479	OxA-17602	Tepe Ebrahim Abad	36.12	53.05	5200	120	1652	[109]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
480	OxA-17603	Tepe Ebrahim Abad	36.12	53.05	5446	73	1652	[109]
481	OxA-17604	Tepe Ebrahim Abad	36.12	53.05	5201	121	1652	[109]
482	OxA-17605	Tepe Ebrahim Abad	36.12	53.05	5269	57	1652	[109]
483	OxA-17606	Tepe Ebrahim Abad	36.12	53.05	5341	121	1652	[109]
484	OxA-17607	Tepe Ebrahim Abad	36.12	53.05	5546	67	1652	[109]
485	OxA-17736	Tepe Ebrahim Abad	36.12	53.05	5116	105	1652	[109]
486	OxA-17737	Tepe Ebrahim Abad	36.12	53.05	5161	129	1652	[109]
487	OxA-17738	Tepe Ebrahim Abad	36.12	53.05	5171	120	1652	[109]
488	PRL-749	Tepe Gaz tavila	28.34	56.58	5604	374	2070	[7]
489	PRL-744	Tepe Gaz tavila	28.34	56.58	5603	276	2070	[7]
490	PRL-748	Tepe Gaz tavila	28.34	56.58	5599	372	2070	[7]
491	Beta-177177	Tepe Guran	33.73	47.07	7327	143	1081	[96]
492	Beta-147122	Tepe Guran	33.73	47.07	7187	122	1081	[96]
493	Beta-177116	Tepe Guran	33.73	47.07	7173	126	1081	[96]
494	Beta-147118	Tepe Guran	33.73	47.07	7002	173	1081	[96]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
495	Beta-147120	Tepe Guran	33.73	47.07	6983	158	1081	[96]
496	OxA-14739	Tepe Pardis	35.45	51.60	4768	77	1512	[96]
497	OxA-14738	Tepe Pardis	35.45	51.60	3927	116	1512	[96]
498	OxA-14737	Tepe Pardis	35.45	51.60	3862	96	1512	[96]
499	OxA-14740	Tepe Pardis	35.45	51.60	4895	98	1512	[96]
500	OxA-14741	Tepe Pardis	35.45	51.60	4809	92	1512	[96]
501	OxA-14742	Tepe Pardis	35.45	51.60	4879	105	1512	[96]
502	OxA-14743	Tepe Pardis	35.45	51.60	4856	107	1512	[96]
503	OxA-14744	Tepe Pardis	35.45	51.60	4893	97	1512	[96]
504	OxA-14745	Tepe Pardis	35.45	51.60	5060	148	1512	[96]
505	OxA-14746	Tepe Pardis	35.45	51.60	5183	121	1512	[96]
506	OxA-14747	Tepe Pardis	35.45	51.60	5183	126	1512	[96]
507	OxA-14749	Tepe Pardis	35.45	51.60	5106	111	1512	[96]
508	OxA-14750	Tepe Pardis	35.45	51.60	5107	107	1512	[96]
509	KIA33174	Tepe Rahmatabad	30.11	53.06	6866	160	1687	[27]
510	KIA33173	Tepe Rahmatabad	30.11	53.06	6924	150	1687	[27]
511	UZ5331/ETH31882	Tepe Rahmatabad	30.11	53.06	6846	200	1687	[27]
512	OxA-22347	Tepe Sialk	33.97	51.40	5200	120	1481	[96]
513	OxA-22505	Tepe Sialk	33.97	51.40	5108	104	1481	[96]
514	OxA-22504	Tepe Sialk	33.97	51.40	5154	128	1481	[96]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
515	OxA-22503	Tepe Sialk	33.97	51.40	5195	119	1481	[96]
516	OxA-22502	Tepe Sialk	33.97	51.40	5351	119	1481	[96]
517	OxA-22501	Tepe Sialk	33.97	51.40	5105	105	1481	[96]
518	OxA-22500	Tepe Sialk	33.97	51.40	5113	102	1481	[96]
519	OxA-22499	Tepe Sialk	33.97	51.40	5122	100	1481	[96]
520	OxA-22498	Tepe Sialk	33.97	51.40	5157	129	1481	[96]
521	OxA-22497	Tepe Sialk	33.97	51.40	5171	119	1481	[96]
522	OxA-22496	Tepe Sialk	33.97	51.40	5161	126	1481	[96]
523	OxA-22495	Tepe Sialk	33.97	51.40	5076	131	1481	[96]
524	OxA-22494	Tepe Sialk	33.97	51.40	4875	89	1481	[96]
525	OxA-22508	Tepe Sialk	33.97	51.40	5341	122	1481	[96]
526	OxA-22507	Tepe Sialk	33.97	51.40	5344	121	1481	[96]
527	OxA-22506	Tepe Sialk	33.97	51.40	5271	59	1481	[96]
528	PRL-749	Tepe Yahya	28.20	55.98	5604	374	2020	[96]
529	PRL-744	Tepe Yahya	28.20	55.98	5603	276	2020	[96]
530	PRL-748	Tepe Yahya	28.20	55.98	5599	372	2020	[96]
531	TK-198	Thalathat II	36.48	42.50	6744	280	768	[1]
532	Bln-719	Togolok	37.98	57.83	6206	191	2103	[1]
533	AA56351	Tol-e Basi	30.08	52.59	6146	85	1644	[96]
534	AA56355	Tol-e Basi	30.08	52.59	6009	88	1644	[96]
535	AA56354	Tol-e Basi	30.08	52.59	5995	78	1644	[96]
536	AA-56340	Tol-e Basi	30.08	52.59	5986	87	1644	[96]
537	AA-56353	Tol-e Basi	30.08	52.59	5959	73	1644	[96]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
538	AA-56343	Tol-e Basi	30.08	52.59	5864	117	1644	[96]
539	AA56339	Tol-e Basi	30.08	52.59	5854	122	1644	[96]
540	WK13990	Tol-e Nurabad	30.12	51.52	5863	122	1543	[2]
541	WK13991	Tol-e Nurabad	30.12	51.52	5854	128	1543	[2]
542	OZI128	Tol-e Nurabad	30.12	51.52	5854	123	1543	[2]
543	WK13992	Tol-e Nurabad	30.12	51.52	5850	127	1543	[2]
544	WK13993	Tol-e Nurabad	30.12	51.52	5728	106	1543	[2]
545	Pta-2703	Ujrat el-Mehed 1	28.58	33.93	7265	202	477	[1]
546	NITA2-11408	Wadi Abu Tulayha	30.51	35.97	7525	67	242	[55, 64]
547	NITA2-11409	Wadi Abu Tulayha	30.51	35.97	7478	109	242	[55, 64]
548	NITA2-11406	Wadi Abu Tulayha	30.51	35.97	7466	108	242	[55, 64]
549	Beta-120210	Wadi Faynan 16	30.62	35.50	9973	281	226	[1]
550	Beta-135111	Wadi Faynan 16	30.62	35.50	9973	281	226	[1]
551	Beta-35080/WS-1	Wadi Shu´eib	31.97	35.73	9961	675	78	[55, 126]
552	Pta 2700	Wadi Tbeik	28.77	33.95	10236	361	457	[55]
553	P-1244	Yanik	37.99	45.95	5889	137	1116	[1]
554	P-1243	Yanik	37.99	45.95	5827	156	1116	[1]
555	LE-1086	Yarim Tepe I	36.31	42.43	6034	188	753	[1]
556	Pta-4242	Yiftahel	32.72	35.18	7999	274	33	[1]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
557	RT-736b	Yiftahel	32.72	35.18	7973	318	33	[1]
558	Pta-4245	Yiftahel	32.72	35.18	7878	289	33	[1]
559	Rome-467	Yümüktepe	36.79	34.60	6832	226	468	[1]
560	Rome-734	Yümüktepe	36.79	34.60	6740	282	468	[1]
561	R-1344	Yümüktepe	36.79	34.60	6624	187	468	[1]
562	R-1343	Yümüktepe	36.79	34.60	6509	139	468	[1]
563	TUNC-12	Zagheh	35.82	49.95	6033	188	1371	[96]
564	WK 9633	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	9032	210	156	[55]
565	Wk 9568	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	9012	246	156	[55]
566	Wk 9447	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	9011	210	156	[55]
567	Wk 9445	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	8965	219	156	[55]
568	Wk 9570	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	8907	250	156	[55]
569	OZE-605	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	8881	244	156	[55]
570	OZE-607	ZahratAdh Dhra´ 2 (ZAD 2)	31.25	35.57	8873	247	156	[55]

Continued on next page

Table A1 – continued from previous page

No.	Lab number	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)	Ref.
571	OZE-606	ZahratAdh Dhra´2 (ZAD 2)	31.25	35.57	8843	272	156	[55]

Notes

[§]: In [96] the date is classified as Neolithic, but is classified as belonging to the chalcolithic period in the CONTEXT database [1].

Table A2: In this table we give one date per site for the first Neolithic arrival. The method of selection is mentioned in Column 2 of the table. (See the notes at the end of the table for a description of methods of selection.) Figure 4.3 also shows the various steps in this selection procedure. For the sites with archeological time periods, we use the start of the period and thus question of selection of one date does not arise there. These times are given in the column 'Age' of tables A5 and A6.

No.	**Method of selection**[†]	**Site Name**	**Lat. °N**	**Long. °E**	**Age (yr cal BCE)**	$\tilde{\sigma}$ **(yr)**	**Distance from Gesher (km)**
1	O	Abadah	33.97	44.83	4612	175	877
2	O	Abu Gosh	31.80	35.11	8037	213	102
3	A	Abu Hureyra	35.87	38.40	8888	226	445
4	A	Abu Madi I	28.56	34.00	9589	175	477
5	A	Àin Abu Nukhayla	29.55	35.41	7591	175	345
6	A	Àin Ghazal	31.93	35.94	8267	175	89
7	C	Akarçay Tepe	36.92	38.02	7580	228	527
8	O	Ali Agha	36.45	43.82	5844	175	869
9	A	Ali Kosh	32.56	47.32	7983	175	1105
10	O	Aqab	37.08	40.82	5120	181	690
11	A	Arjoune	34.56	36.55	5414	191	233
12	O	Arukhlo 1	41.00	43.87	6029	180	1188
13	A	Ashkelon	31.61	34.50	6897	175	150
14	A	Asiab	34.30	47.19	9130	175	1097
15	C	Asikli Höyük	38.35	34.23	7897	350	644
16	A	Assouad	36.58	39.00	7836	363	541
17	A	Aswad	33.42	36.53	9088	175	127
18	A	Atlit-Yam	32.55	34.91	6907	258	58
19	C	Ayakagytma	40.65	64.62	5882	226	2729

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
20	O	Azraq 31	31.83	36.82	7334	251	153
21	A	Ba´ja	30.41	35.46	6928	175	249
22	A	Baaz	33.81	36.51	4569	175	158
23	A	Bademağaci	37.40	30.48	6865	175	699
24	A	Bagor	25.35	74.38	5146	425	3844
25	C	Bakun B	29.85	52.83	5120	258	1673
26	A	Balakot	25.48	66.73	4005	175	3124
27	A	Basta	30.23	35.53	7284	175	269
28	C	Beidha	30.37	35.45	7979	321	254
29	O	Belt	36.65	53.28	6506	970	1682
30	A	Betzet 1	33.07	35.15	7329	244	58
31	C	Bouqras	35.03	40.39	7357	259	522
32	A	Burqu 27	31.95	37.20	7228	175	176
33	O	Byblos	34.12	35.65	6231	175	164
34	A	Cafer	38.42	38.75	7917	203	705
35	A	Canhasan III	37.25	33.37	7540	175	548
36	A	Catalhöyük East	37.65	32.82	7069	175	608
37	A	Cayönü	38.22	39.73	10368	250	727
38	A	Chagha Sefid	32.63	47.26	10252	428	1099
39	A	Cheshmeh Ali	27.31	61.41	5005	175	2556
40	A	Chia Sabz	33.34	47.14	8272	233	1086
41	A	Čoḡā Golān	33.38	46.27	8694	175	1005
42	A	Čoḡā Bonut	32.22	48.51	10125	446	1219

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
43	A	Čoḡā Mīš	32.21	48.55	7348	175	1223
44	A	Damishliyah	36.49	39.03	6633	175	534
45	C	Dhra	31.27	35.58	9565	378	154
46	A	Dhuweila	32.03	37.25	7307	175	177
47	A	Djade	36.65	38.19	8816	175	507
48	A	El Kowm 1	35.24	38.82	6932	249	419
49	A	Es Sifiya	31.44	35.82	6887	175	137
50	A	Feinan	30.62	35.43	5205	202	226
51	C	Ganj Dareh	34.42	47.57	8021	370	1133
52	A	Gawra	36.43	43.34	5874	175	830
53	A	Gesher	32.65	35.52	10458	348	0
54	A	Ghuwayr 1	30.62	35.51	9151	506	226
55	A	Gilgal	32.03	35.47	9506	175	69
56	A	Gritille	37.55	38.57	7779	427	611
57	A	Göbekli Tepe	37.22	38.92	8906	175	595
58	A	Hacilar	37.57	30.07	6230	175	738
59	A	Hajji Firuz	37.04	45.54	6175	185	1036
60	A	Halula	36.42	38.19	7762	175	485
61	A	Hatoula	31.82	34.98	9796	175	105
62	O	Hayaz Höyük	37.48	38.33	7348	175	595
63	A	Hemar	31.17	35.18	7886	175	168
64	C	Hoca Çeşme	40.70	26.09	6301	175	1227

Continued on next page

Table A2 – continued from previous page

No.	**Method of selection**[†]	**Site Name**	**Lat. °N**	**Long. °E**	**Age (yr cal BCE)**	$\tilde{\sigma}$ **(yr)**	**Distance from Gesher (km)**
65	A	HorseCaveAqKupruk II	36.08	66.83	9954	657	2887
66	A	Horvat Galil	32.96	35.32	8564	216	39
67	A	Höyücek	37.45	30.57	6320	175	698
68	A	Iblis	30.16	56.84	4810	175	2039
69	C	Ilipinar	40.46	29.30	5880	190	1030
70	O	Iraq ed-Dubb	32.39	35.73	9561	312	35
71	O	Iraq el-Barud	32.73	34.98	8408	199	51
72	A	Issaron	29.90	35.03	8361	175	309
73	O	Jani	33.95	46.78	7106	239	1056
74	O	Jarmo	35.56	44.92	6523	175	923
75	C	Jeitun	37.95	58.23	5969	395	2137
76	A	Jerf el Ahmar	36.39	38.20	9377	175	483
77	A	Jericho	31.86	35.47	9054	175	88
78	A	Jilat 13	31.50	36.42	7888	175	153
79	A	Jilat 7	31.52	36.42	7771	208	152
80	A	Kaletepe	38.28	34.57	8102	175	632
81	O	Kalibangan	29.42	74.08	5412	188	3672
82	A	Karain B	37.07	30.55	6108	225	669
83	A	Kashkashok II	36.71	40.59	6700	175	647
84	A	Khirbet Hammam	31.02	35.65	7437	175	182
85	A	Kili Ghul Mohammad I	30.28	66.97	4207	175	2983

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
86	C	Körtik Tepe	37.83	40.97	9556	292	759
87	O	Kumartepe	37.48	38.48	6849	207	601
88	A	Kuruçay Höyük	37.62	30.15	6032	175	737
89	A	M´lefaat	36.31	43.54	10888	258	840
90	A	Magzaliyah	36.39	42.33	6888	181	749
91	A	Megadim	32.72	34.97	5921	175	52
92	A	Mehrgarh	29.42	67.58	8520	430	3064
93	A	Menteşe	40.27	29.52	6181	175	1002
94	A	Mezad Mazal	30.93	35.32	7388	175	192
95	A	Mezraa Teleilat	36.98	38.00	8549	200	532
96	A	Mundigak	31.58	65.50	3744	226	2817
97	O	Munhata	32.60	35.55	6216	175	6
98	A	Mureybet	36.08	38.10	10458	348	449
99	A	Mushki	29.78	52.90	7845	354	1681
100	O	Naja	31.77	36.92	6268	187	164
101	C	Nemrik 9	36.73	42.88	9650	480	811
102	A	Netiv HaGdud	31.98	35.38	9112	175	76
103	C	Nevali Çori	37.58	38.65	8444	311	618
104	O	Nizzanim	31.72	34.58	5710	175	136
105	C	Pinarbasi Site A	37.48	33.03	8388	281	583
106	A	Pinarbasi Site B	37.49	33.04	6166	175	584
107	A	Qadesh Barnea 3	30.62	34.40	6294	175	249
108	O	Qanah	32.75	35.33	5908	175	21

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
109	O	Qatif Y-3	31.41	34.32	4990	219	178
110	A	Qermez Dere	36.38	42.45	9759	175	758
111	O	Rahub	32.62	35.93	6292	192	39
112	A	Ramad	33.42	36.10	7192	175	101
113	O	Rana Ghundai	30.40	68.75	4433	175	3147
114	A	Ras Shamra	35.58	35.73	7254	175	326
115	A	Reu´el	30.10	35.10	7809	175	286
116	A	Sabi Abyad	36.53	39.10	7411	224	542
117	A	Samir	32.82	34.95	5734	175	57
118	A	Sarab	34.38	47.09	6920	175	1089
119	A	Sarazm	39.52	67.57	3742	175	2962
120	A	Sawwan I	34.12	43.93	6239	175	798
121	A	Seh Gabi	34.58	48.00	5091	175	1175
122	A	Sha´ar Hagolan	32.68	35.61	6247	175	9
123	A	Sheikh-e Abad	34.61	47.27	9306	175	1109
124	A	Shkarat Msaied	30.45	35.44	8266	196	245
125	O	Shōmu	41.16	45.38	6352	175	1288
126	O	Shu´eib	31.95	35.69	7719	245	79
127	A	Siahbid	34.50	47.25	4722	175	1105
128	A	Sinn	35.28	40.36	7685	175	534
129	A	Sotto	36.28	42.37	6194	199	746
130	A	Suberde	37.35	31.93	7267	175	616
131	A	Tabaqat al-Buma	32.53	35.72	6734	268	23

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
132	A	Tal-e Jari B	29.85	52.96	6043	175	1685
133	A	Tall-e Bakun A	29.91	52.89	4426	175	1676
134	C	Tall-e Jari A	29.86	52.96	5089	207	1684
135	A	Tang-e Bolaghi	30.15	53.13	9939	177	1692
136	O	Tbeik	28.82	33.94	10236	361	452
137	A	Tel Ali	32.70	35.56	6867	175	7
138	C	Tepe Chahar Boneh	35.81	50.03	6	175	1378
139	C	Tepe Ebrahim Abad	36.12	53.05	5	175	1652
140	A	Tepe Gaz Tavila	28.34	56.58	5602	175	2070
141	A	Tepe Guran	33.73	47.07	7134	175	1081
142	C	Tepe Pardis	35.45	51.60	5	175	1512
143	C	Tepe Rahmatabad	30.11	53.06	7	175	1687
144	C	Tepe Sialk	33.97	51.40	5	175	1481
145	A	Tepe Yahya	28.20	55.98	5602	175	2020
146	A	Thalathat II	36.48	42.50	6744	280	768
147	O	Togolok	37.98	57.83	6206	191	2103
148	A	Tol-e Basi	30.08	52.59	5973	175	1644
149	A	Tol-e Nurabad	30.12	51.52	5830	175	1543
150	O	Ujrat el-Mehed 1	28.58	33.93	7265	202	477
151	A	Wadi Abu Tulayha	30.51	35.97	7490	175	242
152	A	Wadi Faynan 16	30.62	35.50	9973	175	226
153	A	Wadi Shu´eib	31.97	35.73	9961	675	78
154	O	Wadi Tbeik	28.77	33.95	10236	361	457

Continued on next page

Table A2 – continued from previous page

No.	Method of selection†	Site Name	Lat. °N	Long. °E	Age (yr cal BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
155	A	Yanik	37.99	45.95	5858	175	1116
156	O	Yarim Tepe I	36.31	42.43	6034	188	753
157	A	Yiftahel	32.72	35.18	7950	175	33
158	A	Yümüktepe	36.79	34.60	6676	175	468
159	A	Zagheh	35.82	49.95	6033	188	1371
160	A	ZahratAdh Dhra' 2 (ZAD 2)	31.25	35.57	8941	175	156

Notes

† **C: Cluster**: The oldest Neolithic cluster at the site is used in the analysis; standard deviation of the dates in the cluster is given in σ column.

A: Average: Where the cluster method is unapplicable, we use the dates within 350 years of the oldest date. The mean of these dates within 350 years of the oldest date is given in the Age column. The associated value of σ in Column 7 is the maximum of the errors associated with each of these dates.

O: One Date: There is only a single ^{14}C date belonging to the Neolithic period at the site.

Table A3: Discarded ^{14}C dates (The reason for exclusion of the date is given in more details in the reference provided.)

No.	**Lab number**	**Site Name**	**Age (yr cal BCE)**	$\tilde{\sigma}$ **(yr)**	**Reason for discarding**	**Ref.**
3	BM-1720	Abu Hureyra	24731	826	An anomalously early date for the Neolithic, with the later dates BM-1718 and BM-1719 classified as Mesolithic. The date is also noted as suspicious.	[1, 32]
6	AA-25038	Ain Ghazal	22672	4245	An anomalously early date for the Neolithic	[119]
17	MC-607	Assouad	12662	5725	Failed attempt at ^{14}C dating (note the value of$\tilde{\sigma}$)	[35]
38	UCLA-300	Chagha Sefid	11178	214	Disturbed stratigraphy of the site, the date may have been wrongly recorded.	[1, 79]
90		M´lefaat	8970	331	Classified as unacceptable	[89]
90		M´lefaat	14392	713	Classified as unacceptable	[89]
102	Gd-5440	Nemrik 9	19888	5015	An anomalously early date for the Neolithic; classified as unacceptable	[88, 89]
102	Gd-4208	Nemrik 9	9741	450	Classified as unacceptable	[89]
102	Gd-2777	Nemrik 9	9907	528	Classified as unacceptable	[89]
102	Gd-5451	Nemrik 9	10623	349	Classified as unacceptable	[89]
102	Gd-2714	Nemrik 9	10892	259	Classified as unacceptable	[89]
102	Gd-5249	Nemrik 9	11087	243	Classified as unacceptable	[89]
111	OxA-3753	Qermez Dere	11906	255	Two samples from the same material gave divergent dates; OxA-3753 is inconsistent with the age of the site implied by other dates.	[76, 138]
118	Pta-3385	Salibiya 9	20031	421	Possibly recorded under wrong lab number.	[1]

A4

Table A4: The archaeological phases for the Indus Valley region [68, 110]

Sr. No.	**Phase**	**Start time**	**End time**
1	Kili Ghul Mohammad	7000	5000
2	Early Kulli	7000	3500
3	Burj Basket-marke	5000	4300
4	Togau	4300	3800
5	Anarta blade making	4000	3500
6	Anarta with Microlith	4000	3500
7	Anarta (pre-Harappan)	4000	3500
8	Microlith blade makin	4000	3500
9	Microliths	4000	3500
10	Hakra ware	3800	3200
11	Kechi Beg	3800	3200

Table A5: The archaeological dates for the Indus Valley region [110]

No.	**Site Name**	**Lat.**	**Long.**	**Age**	$\tilde{\sigma}$	**Distance from**
		°N	**°E**	**(yr BCE)**	(yr)	**Gesher (km)**
1	Abduwali	28.76	71.34	3800	200	3436
2	Adhi One	28.77	71.08	3800	200	3411
3	Ahmed Khanzai North	30.18	66.97	3800	200	2986
4	Ahmed Khanzai South	30.15	66.95	3800	200	2985
5	Akkanwali Their	28.83	71.41	3800	200	3440
6	Ambrawali	28.79	71.97	3800	200	3494
7	Anjira	28.28	66.32	5000	300	2981
8	Ashal	26.06	64.42	3800	200	2885
9	Awaran Niabat	26.42	65.23	4300	300	2946
10	Azimwala Two	28.79	71.19	3800	200	3421
11	Azimwali C	28.78	71.21	3800	200	3423
12	Badalwala Five	28.69	71.08	3800	200	3414
13	Badalwala Four	28.69	71.09	3800	200	3414
14	Badrang Damb	27.67	65.52	3800	200	2927
15	Bagaya-no Timbo	23.43	71.83	4000	200	3688
16	Baggewali	28.83	71.15	3800	200	3416
17	Bahilawala B	28.85	71.48	3800	200	3446
18	Bahilawala C	28.87	71.47	3800	200	3444
19	Bajaniya-no Thumdo	23.83	71.49	4000	200	3638
20	Baleli	30.33	66.88	7000	300	2974
21	Bandwali	28.87	71.43	3800	200	3441
22	Belar Damb	27.12	66.45	4300	300	3033
23	Bhootanwala C	28.78	71.05	3800	200	3408
24	Bhootanwali Two	28.77	71.04	3800	200	3407

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
25	Binjor Three	29.20	73.10	4000	200	3587
26	Chak 353 West	29.19	72.27	3800	200	3509
27	Chambrawala Ther	29.33	72.30	3800	200	3508
28	Chandnewala Two	28.74	71.21	3800	200	3424
29	Changalawala C	28.85	71.38	3800	200	3437
30	Changda	22.53	72.55	4000	200	3797
31	Channanwala Ther	29.13	72.90	3800	200	3570
32	Chaudhryanwala	28.79	71.27	3800	200	3428
33	Chikrala	28.75	71.20	3800	200	3423
34	Chimri	27.82	66.63	4300	300	3025
35	Chore	28.76	71.16	3800	200	3419
36	Choteria Timbo	23.60	71.85	4000	200	3682
37	Dabar Kot	30.08	68.68	3800	200	3149
38	Dabli East	28.90	71.47	3800	200	3443
39	Dabli West	28.91	71.47	3800	200	3443
40	Damb Sadaat	30.05	66.95	3800	200	2988
41	Darkhanwala Ther	28.72	71.23	3800	200	3427
42	Datrana Eight	23.77	71.11	4000	200	3606
43	Datrana Four	23.77	71.11	3500	200	3606
44	Datrana Seven	23.78	71.12	4000	200	3606
45	Dhuni South	28.58	70.94	3800	200	3404
46	Dhuni, Hakra	28.59	70.93	3800	200	3403
47	Dosia Khal Damb	27.30	66.37	3800	200	3019
48	Drakalo Damb	27.15	66.42	3800	200	3030
49	Duki Mound	30.17	68.57	5000	300	3136

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
50	Faiz Mohammad	29.95	67.10	4300	300	3004
51	Gajjuwala Two	28.84	71.12	3800	200	3413
52	Ganario-no Thumdo	23.94	71.52	4000	200	3636
53	Gate Dap	26.12	64.22	3800	200	2864
54	Ghuram Damb	28.70	66.28	3800	200	2964
55	Godavari One	22.20	69.92	4000	200	3570
56	Gokhijadio-no Timbo	23.62	71.88	4000	200	3684
57	Gudel	22.73	72.52	4000	200	3785
58	Gumla	31.88	70.83	7000	300	3306
59	Gwani Kalat	27.48	65.92	3800	200	2971
60	Harappa	30.63	72.87	3700	200	3525
61	Harhari-no Thumdo	23.88	71.39	4000	200	3627
62	Hathala	32.02	70.60	3800	200	3281
63	Hotewala Two	28.92	71.23	3800	200	3420
64	Islam Chowki	32.98	70.48	3800	200	3252
65	Isplinji One	29.69	67.05	4300	300	3007
66	Isplinji Two	29.69	67.04	5000	300	3006
67	Jafawala Three	28.71	71.14	3800	200	3418
68	Jafawala Two	28.70	71.13	3800	200	3418
69	Jalwali A	28.86	71.38	3800	200	3436
70	Jalwali B	28.86	71.38	3800	200	3436
71	Jangipar	28.68	71.08	3800	200	3414
72	Janoya-no Timbo	23.42	71.86	4000	200	3691
73	Jaren	26.22	64.75	3800	200	2909
74	Jawaiwala Two	28.73	71.07	3800	200	3411

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
75	Jawarji Kalat	27.52	65.87	4300	300	2965
76	Jebri Damb Two	27.29	65.75	7000	300	2962
77	Jhalar	28.71	71.12	3800	200	3417
78	Jhandewala Two	28.72	70.98	3800	200	3403
79	Kalharwala B	28.87	71.25	3800	200	3424
80	Kanewal, Sai No Tekro	22.45	72.50	4000	200	3796
81	Karezgai	30.81	67.75	3800	200	3043
82	Kargushki Damb	27.48	65.32	7000	300	2915
83	Kasiano Dozakh	30.45	66.93	7000	300	2976
84	Kechi Beg	30.12	66.95	3800	200	2986
85	Khakhar Buthi	26.32	66.27	7000	300	3047
86	KhanKandewala D	28.84	71.41	3800	200	3440
87	Khanpuri Two	28.75	71.27	3800	200	3429
88	Khiplewali	28.73	71.02	3800	200	3407
89	Khiplewali Three	28.72	71.03	3800	200	3408
90	Khiplewali Two	28.72	71.02	3800	200	3407
91	KI	29.96	66.85	3800	200	2981
92	Kikrl Two	28.72	71.33	3800	200	3436
93	Kili Ghul Mohammad	30.28	66.97	7000	300	2984
94	Killianwali	28.89	71.43	3800	200	3440
95	Killianwali D	28.88	71.45	3800	200	3442
96	Kirta	29.53	67.47	3800	200	3051
97	Kota	22.17	69.70	4000	200	3551
98	Kotada, Jamnagar	22.20	70.37	4000	200	3611
99	Kowas	30.47	67.58	3800	200	3036

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
100	Kuchanwala	29.11	71.91	3800	200	3478
101	Kuchnai Ghundai	30.72	67.04	3800	200	2979
102	Kuki Damb	28.75	66.35	5000	300	2969
103	Kullu Kalat	29.07	66.37	4300	300	2961
104	Kunal	29.63	75.66	3800	200	3812
105	L-2	30.30	68.17	5000	300	3095
106	L-3	30.30	68.20	5000	300	3098
107	Lak Largai	32.82	70.52	3800	200	3259
108	Lakhman	28.72	71.17	3800	200	3421
109	Lathwala Two	28.83	71.20	3800	200	3420
110	Lewan	32.88	70.58	3800	200	3263
111	Litanwala	28.78	71.38	3800	200	3439
112	Loharki Theri	29.17	72.25	3800	200	3508
113	Loteshwar	23.60	71.84	4000	200	3681
114	Lundewali Four	28.89	71.41	3800	200	3438
115	Lundewali Three	28.89	71.41	3800	200	3438
116	Luppewala	28.82	71.21	3800	200	3422
117	Luppewala Three	28.83	71.21	3800	200	3421
118	Marki Mas	27.17	66.42	3800	200	3029
119	Mehwali Two	28.66	71.03	3800	200	3410
120	Merechi Kanda	28.82	71.24	3800	200	3424
121	Merechi Kanda Two	28.83	71.24	3800	200	3424
122	Moniwala	28.64	70.72	3800	200	3381
123	Mundigak	31.92	65.50	4300	300	2810
124	Musafarwali	28.78	71.14	3800	200	3416

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
125	Musafarwali Two	28.77	71.15	3800	200	3418
126	Naharnwala	28.84	71.50	3800	200	3448
127	Naharwali	28.84	71.39	3800	200	3438
128	Naharwali B	28.83	71.39	3800	200	3438
129	Nahrenwala	28.84	71.50	3800	200	3448
130	Nal	27.73	66.27	7000	300	2995
131	Nani Chandur	23.58	71.63	4000	200	3663
132	Neghar Damb	28.27	66.30	5000	300	2979
133	Niai Buthi	26.25	66.43	3800	200	3065
134	Niwaniwala Ther West	28.79	71.17	3800	200	3419
135	Niwaniwala Three	28.79	71.17	3800	200	3419
136	Nundara	26.47	65.42	3800	200	2962
137	Oinwala Ther	28.84	71.38	3800	200	3437
138	Old Balor	26.05	64.42	4300	300	2886
139	Oriyo Timbo	21.89	71.60	4000	200	3740
140	Pal	22.30	70.72	4000	200	3639
141	Panju Damb	27.32	66.42	4300	300	3023
142	Parhara	28.75	71.19	3800	200	3422
143	Parharewala B	28.07	71.18	3800	200	3443
144	Payunewala Bhit Three	28.81	71.37	3800	200	3437
145	Payunewala Bhit Two	28.98	71.37	3800	200	3432
146	Periano Ghundai	31.37	69.38	4300	300	3182
147	Phusi Damb	27.08	66.18	4300	300	3010
148	Pir Haidar Shahr	28.27	66.10	4300	300	2961
149	Q-06	29.77	66.97	4300	300	2997

Continued on next page

Table A5 – continued from previous page

No.	**Site Name**	**Lat.**	**Long.**	**Age**	$\tilde{\sigma}$	**Distance from**
		°N	**°E**	**(yr BCE)**	(yr)	**Gesher (km)**
150	Q-17	30.23	66.90	5000	300	2978
151	Q-18	30.18	66.88	4300	300	2978
152	Q-23	30.27	66.98	4300	300	2985
153	Q-25	30.35	66.93	5000	300	2978
154	Q-26	30.32	66.87	3800	200	2973
155	Q-28	30.32	66.87	3800	200	2973
156	Q-30	30.27	66.97	4300	300	2984
157	Q-32	30.30	66.95	4300	300	2981
158	Q-33	29.78	67.07	4300	300	3006
159	Q-35	30.22	66.78	4300	300	2967
160	Q-36	29.97	66.95	4300	300	2990
161	Qadir Bux Their	28.78	71.40	3800	200	3441
162	Quetta Miri	30.25	66.98	4300	300	2985
163	R.D. 66	29.22	72.87	3800	200	3565
164	Rahmanwali	28.65	71.22	3800	200	3428
165	Rais Sher Mohammad	28.32	66.13	4300	300	2962
166	Rana Ghundai	30.40	68.75	7000	300	3147
167	Rodkan	26.10	64.40	3800	200	2882
168	Sadwala Kanda	28.81	71.11	3800	200	3413
169	Safuwala Ther	28.64	70.98	3800	200	3406
170	Safuwala Three	28.65	71.00	3800	200	3407
171	Safuwala Two	28.64	70.98	3800	200	3406
172	Sahib Khan	30.60	67.05	4300	300	2983
173	Saiyid Maurez Damb	29.43	66.45	5000	300	2958
174	Sala Khan	29.30	66.48	3800	200	2965

Continued on next page

Table A5 – continued from previous page

No.	Site Name	Lat. °N	Long. °E	Age (yr BCE)	$\tilde{\sigma}$ (yr)	Distance from Gesher (km)
175	Santhli Five	23.90	71.50	4000	200	3636
176	Santhli Four	23.90	71.48	4000	200	3634
177	Santhli One	23.90	71.50	4000	200	3636
178	Santhli Six	23.90	71.51	4000	200	3637
179	Santhli Three	23.90	71.48	4000	200	3634
180	Santtili Two	23.90	71.49	4000	200	3635
181	Sanukewala Two	28.86	71.17	3800	200	3417
182	Shahr Sardar	29.45	66.48	3800	200	2960
183	Sheri Khan Tarakai	32.82	70.45	3800	200	3252
184	Sheruwala Three	28.73	71.22	3800	200	3425
185	Sheruwala Two	28.73	71.24	3800	200	3427
186	Shidiwala A	28.78	71.23	3800	200	3425
187	Siah Damb, Surab	28.57	66.18	5000	300	2959
188	Site Near Kuki Damb	28.73	66.35	3800	200	2970
189	Sohniwali	28.75	71.02	3800	200	3406
190	Sohniwali Two	28.75	71.03	3800	200	3407
191	Sorak Damb	27.43	66.47	3800	200	3024
192	Sraduk	27.02	64.18	3800	200	2825
193	SraKala	30.63	66.98	3800	200	2976
194	Sumer Damb	27.16	66.43	3800	200	3030
195	Suneri Damb	27.45	65.75	3800	200	2956
196	SurJangal	30.27	68.50	4300	300	3127
197	Surkh Damb	28.30	66.27	3800	200	2976
198	Tegak	28.32	66.15	4300	300	2964
199	Tharro Hill	24.83	67.82	7000	300	3253

Continued on next page

Table A5 – continued from previous page

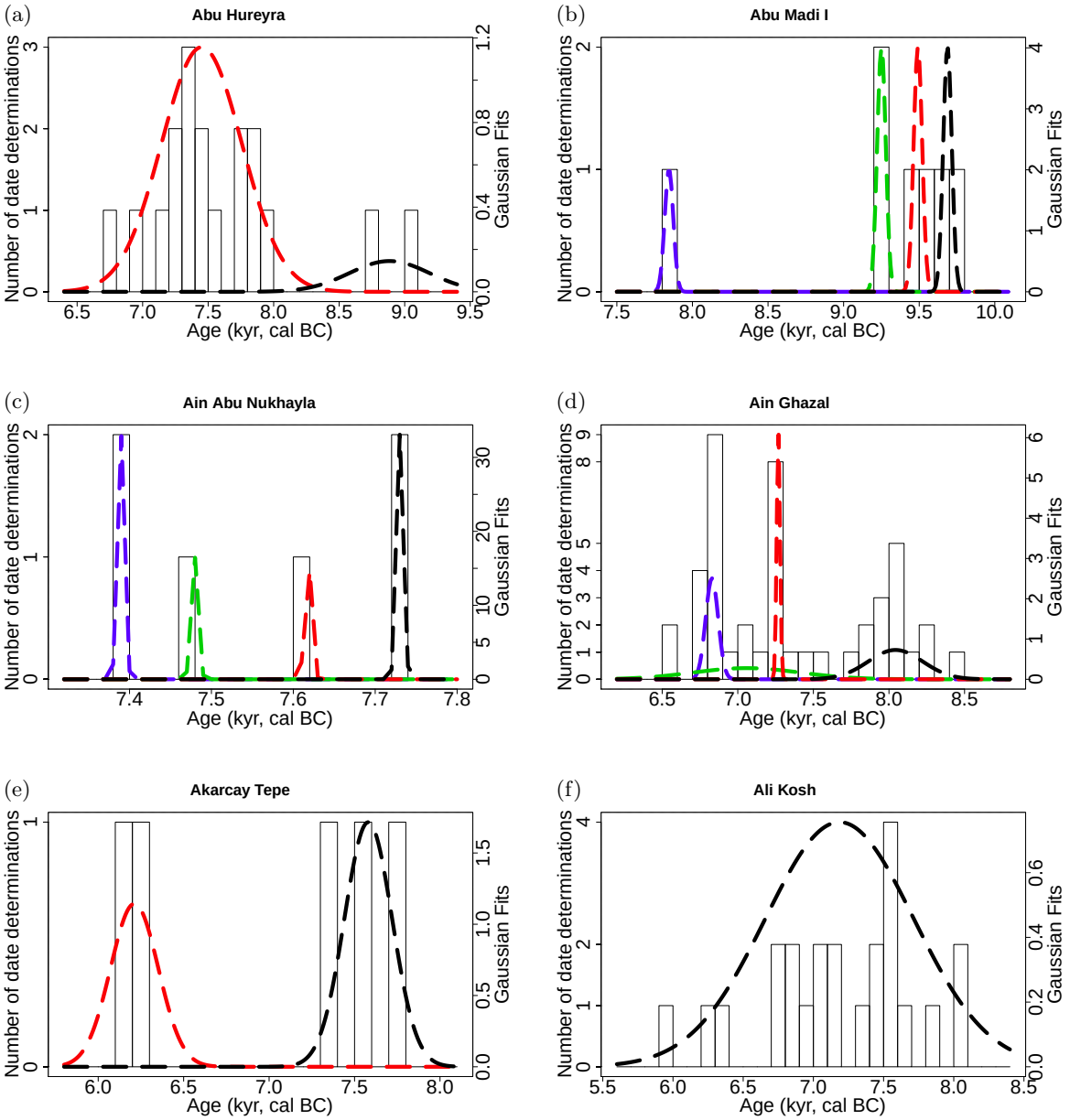
No.	**Site Name**	**Lat.**	**Long.**	**Age**	$\tilde{\sigma}$	**Distance from**
		°N	**°E**	**(yr BCE)**	(yr)	**Gesher (km)**
200	Theriwala	29.10	72.81	3800	200	3563
201	Thok Valley One	28.73	66.35	5000	300	2970
202	Thoom Thali	28.77	71.36	3800	200	3437
203	Thoriwala	28.60	71.03	3800	200	3412
204	Toji Damb	28.88	65.67	3800	200	2901
205	Tokaria Timbo	23.47	71.83	4000	200	3686
206	Tor Ghundai	29.75	66.33	3800	200	2938
207	Trillar	29.18	72.21	3800	200	3504
208	Turawewala B	28.78	71.50	3800	200	3450
209	Turawewala C	28.78	71.51	3800	200	3451
210	Turawewali Their	28.78	71.50	3800	200	3450
211	Valwala Two	28.62	70.98	3800	200	3406
212	Valwali	28.63	70.98	3800	200	3406
213	Waddenwali	28.87	71.44	3800	200	3442
214	Wariyal C	29.18	71.91	3800	200	3476
215	Zayak North	27.92	65.90	4300	300	2954
216	Zidi	27.72	66.78	4300	300	3043
217	Zik	26.20	64.78	3800	200	2913

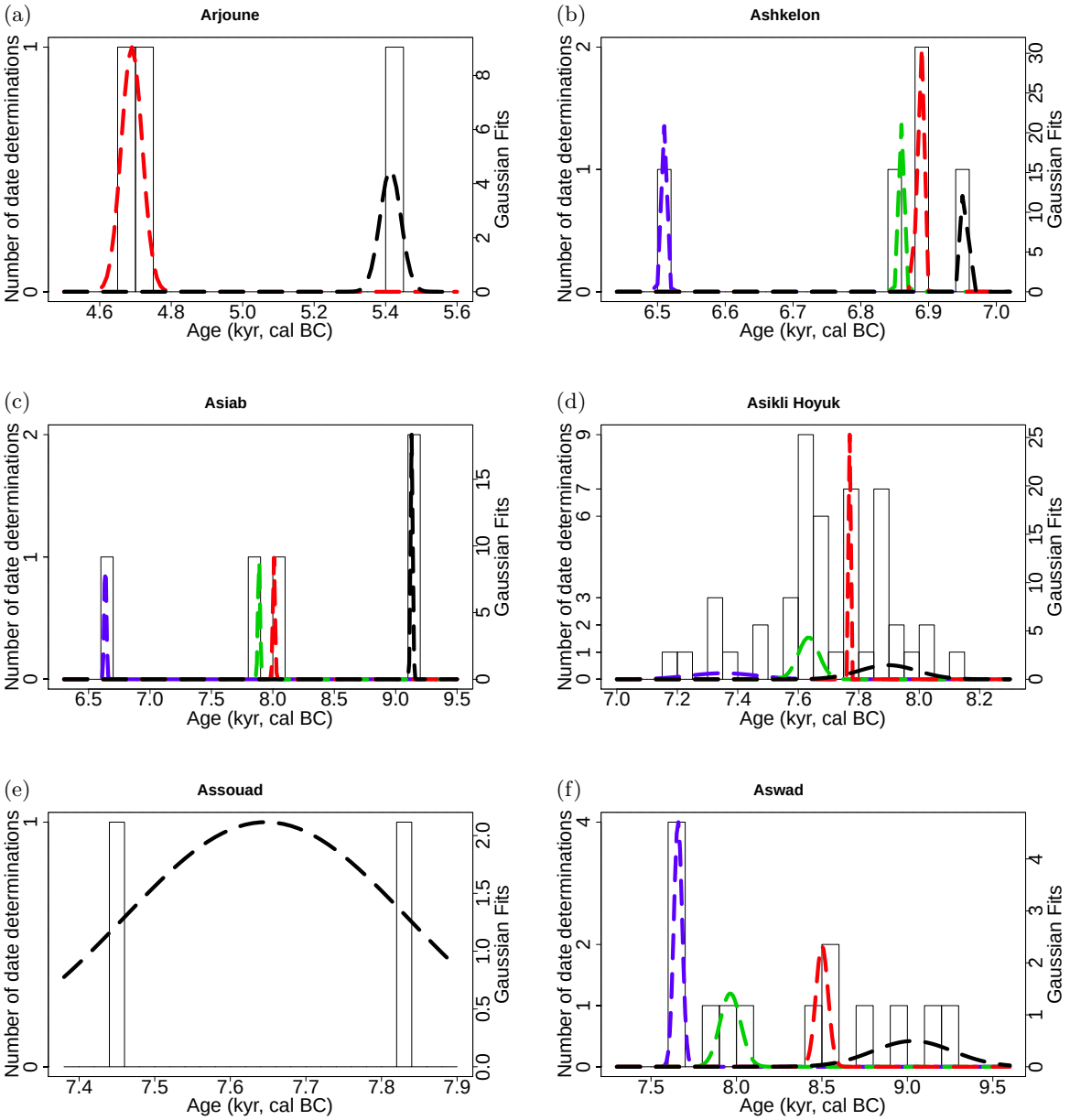
Table A6: The archaeological dates for Iran and Afghanistan [21], and , Kopet Dag region [73] (Tepeh Khaleseh is from [11, 134]).

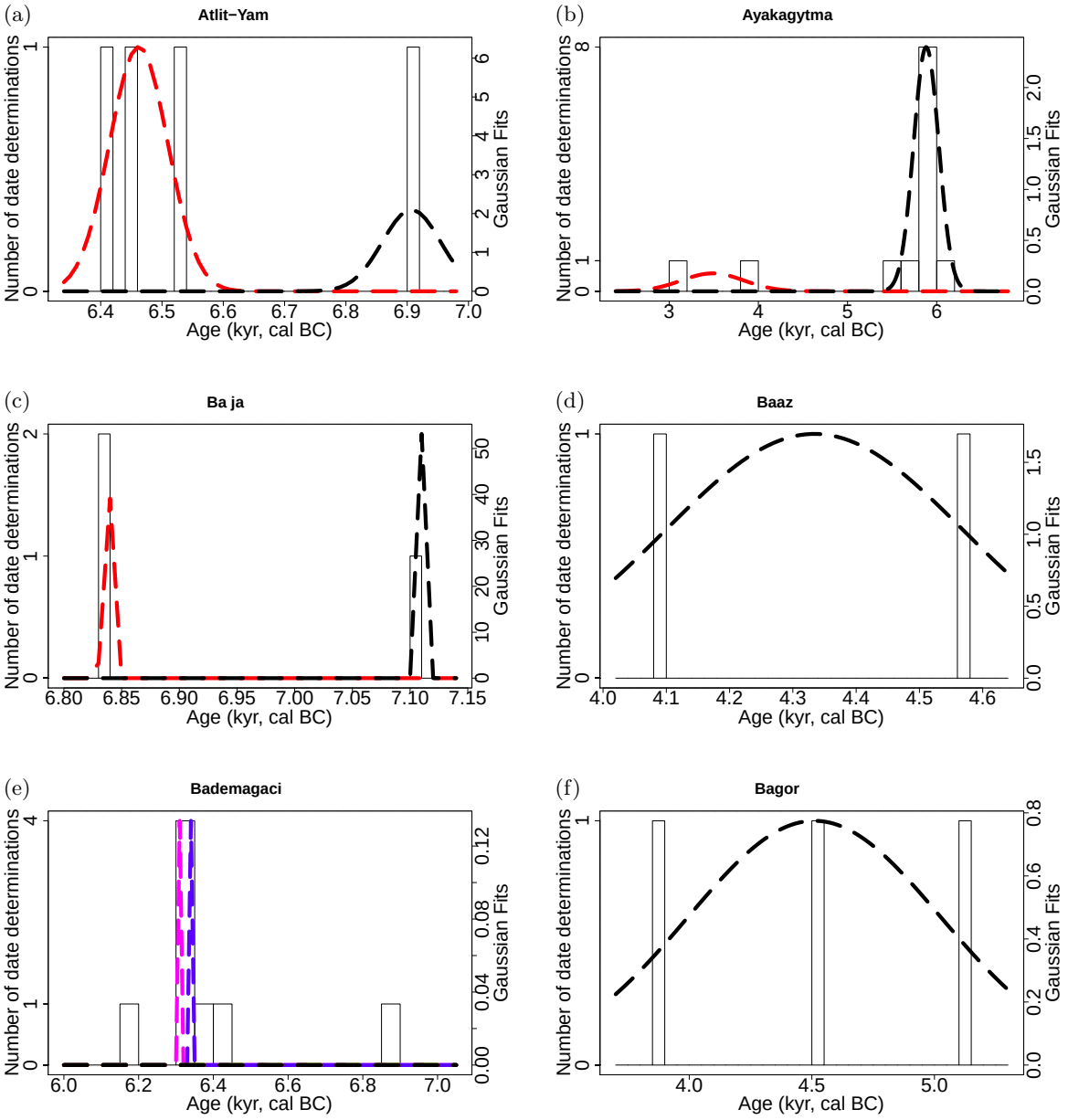
No.	**Site Name**	**Lat. °N**	**Long. °E**	**Age (yr BCE)**	$\tilde{\sigma}$ (yr)	**Distance from Gesher (km)**
1	Aq Tappe	37.57	57.42	7000	300	2060
2	Arismān	33.67	52.00	7000	300	1537
3	Čoḡā Āhuwān / Golān	33.38	46.27	7000	300	1005
4	Deh Hajj	33.68	48.88	7000	300	1248
5	Golestān Park	38.08	46.28	7000	300	1147
6	Gŭsa Tappe(Ardabil)	38.25	48.28	7000	300	1311
7	Qomrud	34.73	51.07	7000	300	1456
8	Tepeh Khaleseh	36.19	49.17	6000	300	1312
9	Tappe J̌olbar	38.56	42.32	7000	300	899
10	Tappeh Deh Keir	36.53	54.98	7000	300	1829
11	Tappeh Ozbaki	35.54	50.34	7000	300	1400
12	Tol-e Baši	30.08	52.59	7000	300	1644
13	Chopan	38.07	58.2	6100	300	1847
14	Togolok	38.15	58	6000	300	2136
15	Sang-i Chakhmak	36.47	54.93	6730	380	2121
16	Ayakagytma	40.65	64.62	6000	300	1824
17	Jeitun	38.1	58.33	6040	50	2729
18	Gadymi	36.63	60.43	6100	300	2149
19	Chagylly	36.7	60.47	6100	300	2315
20	Chakmakli	36.73	60.55	6100	300	2319
21	Monjukli	36.75	60.35	6100	300	2326
22	Bacha well	39.43	56.23	6100	300	2308
23	Naiza	38.92	56.75	6100	300	2003
24	Bami	38.72	56.82	6100	300	2032
25	Kelyata	38.27	57.72	6100	300	2032
26	Gievdzhik	38.17	57.72	6100	300	2099
27	Jebel	39.63	54.25	5810	180	2097

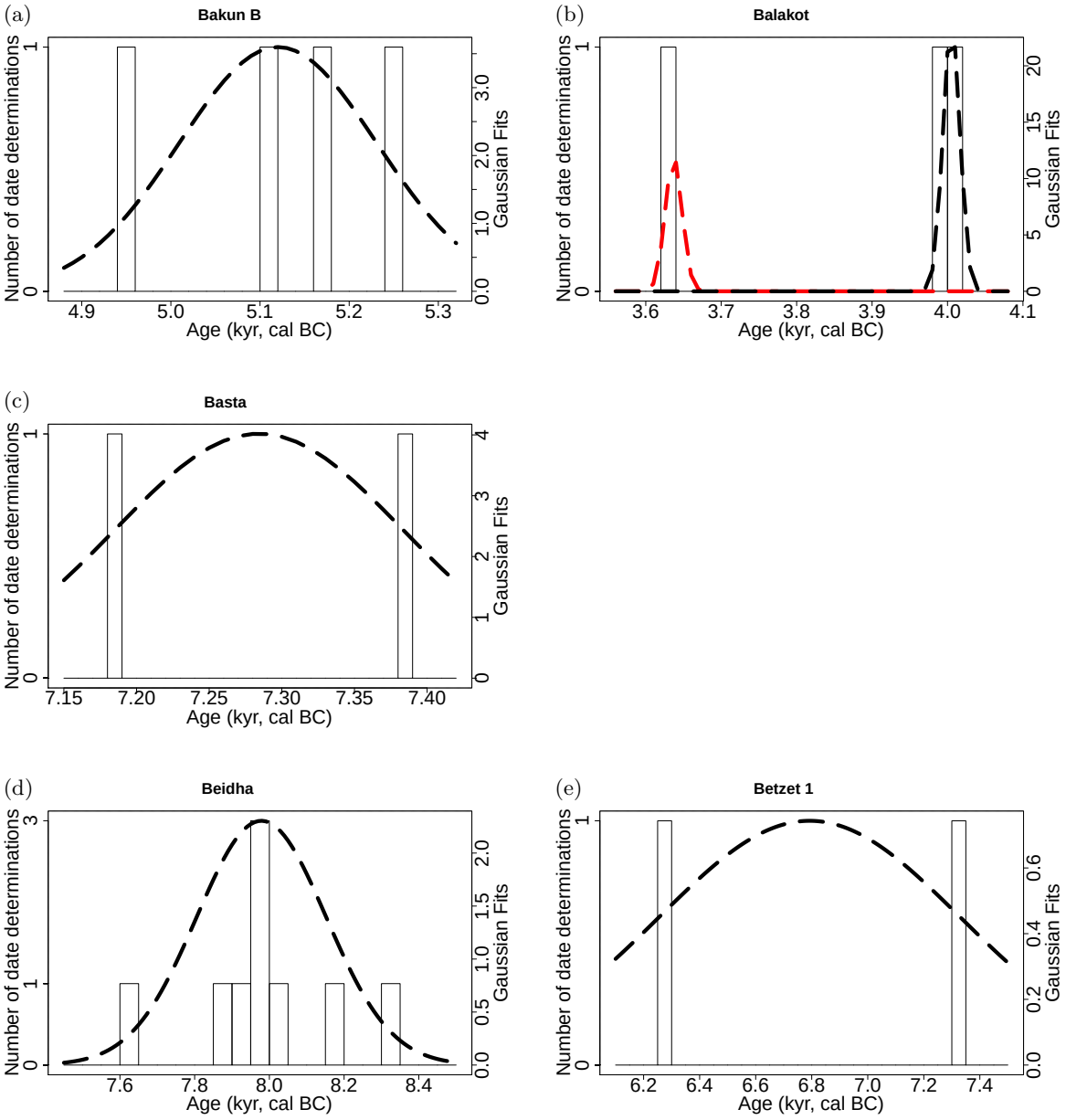
Appendix B

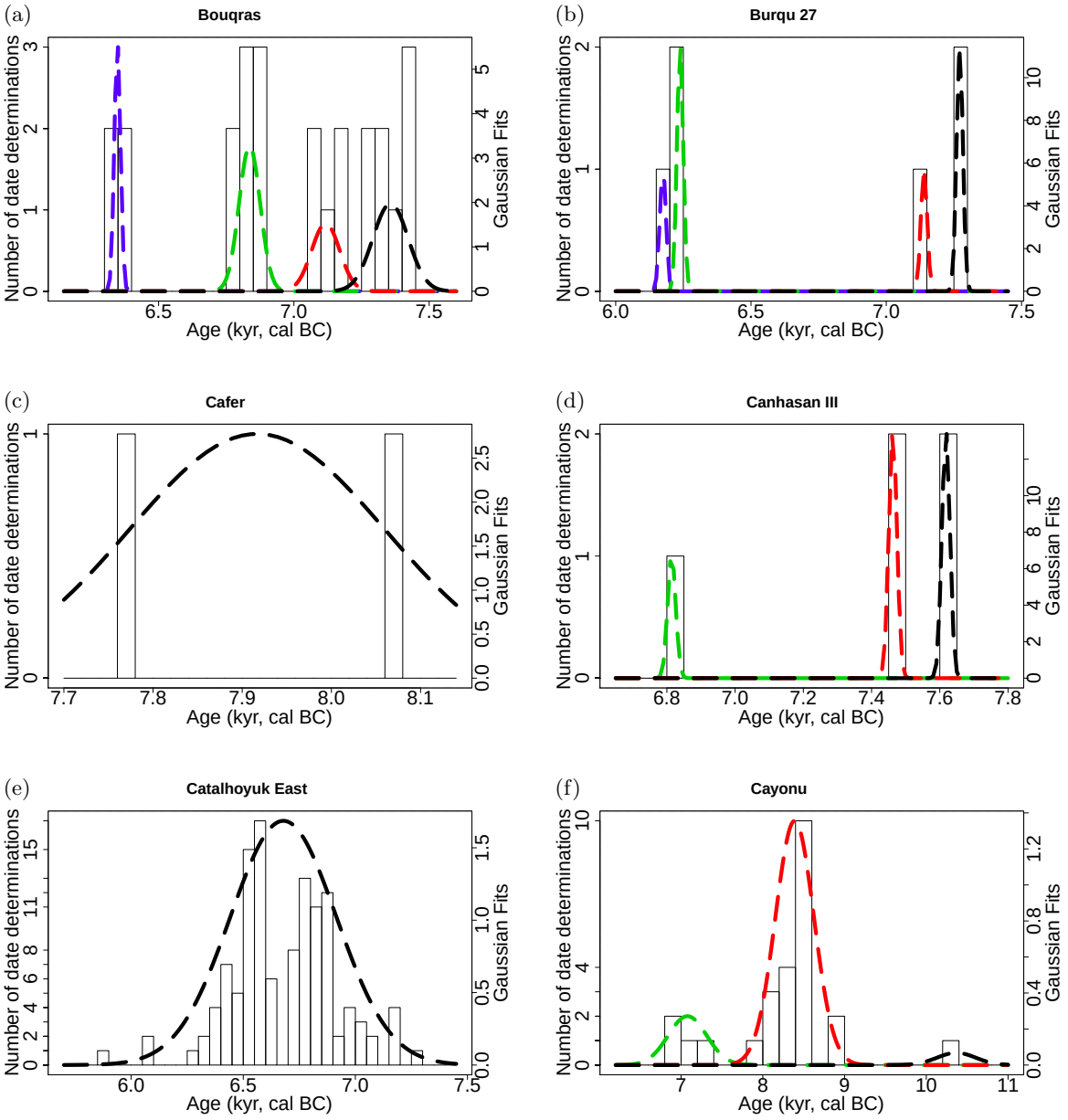
Histograms of the Gaussian mixture modelling clusters for the sites with ^{14}C dates

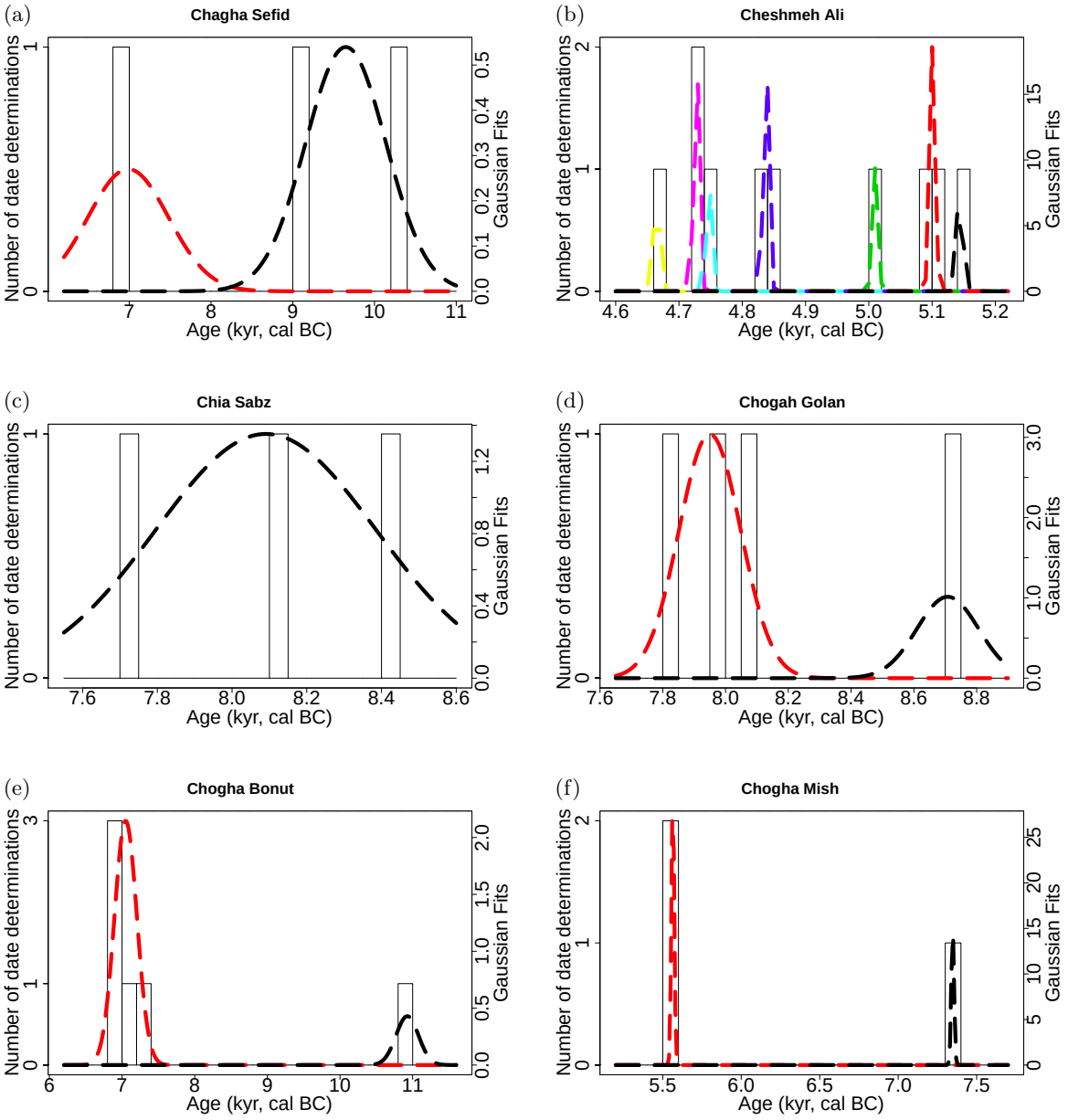

(a)
Abu Hureyra
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Abu Madi I
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Ain Abu Nukhayla
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Ain Ghazal
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Akarcay Tepe
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Ali Kosh
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

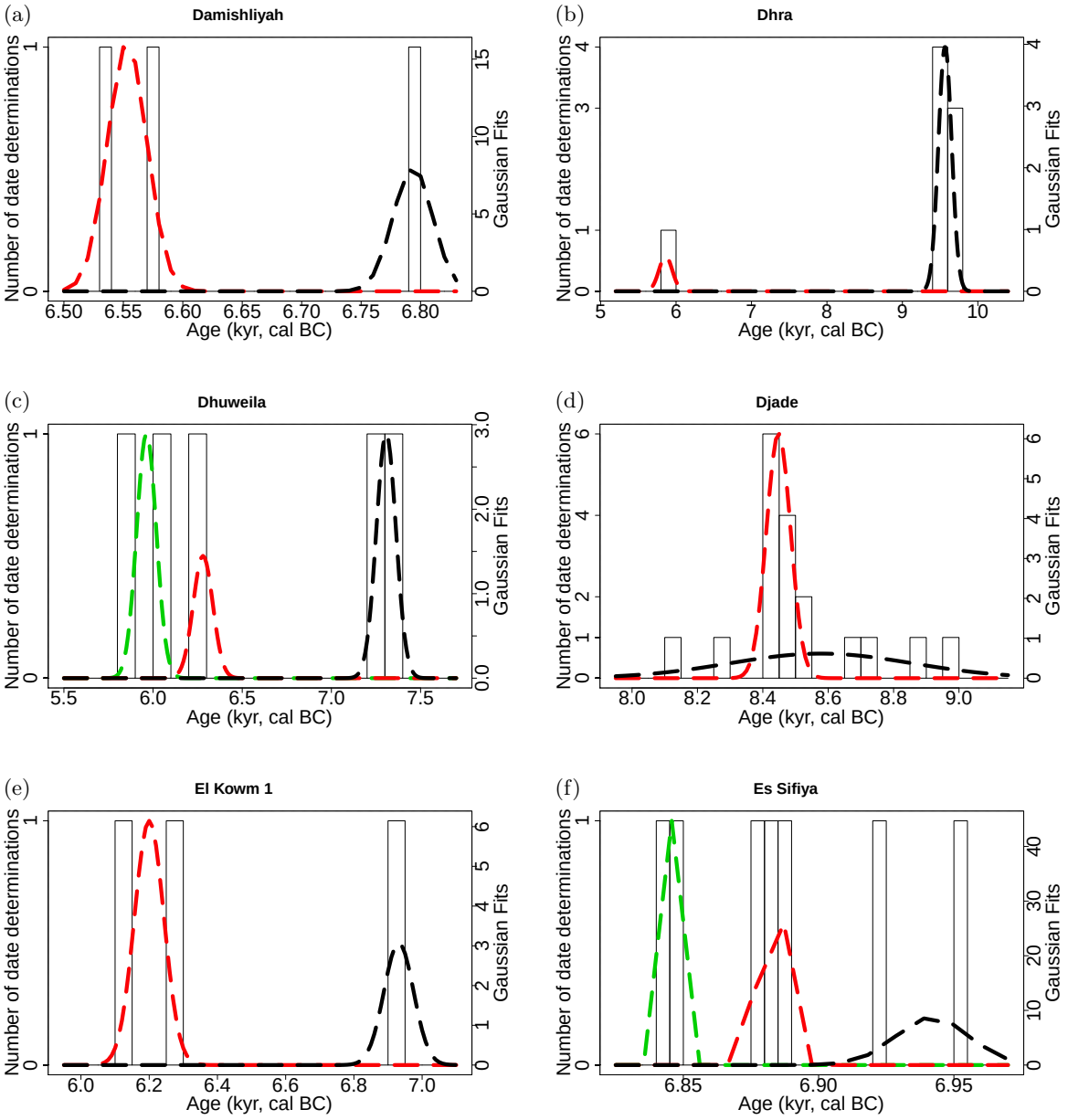

(a)
Arjoune
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Ashkelon
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Asiab
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Asikli Hoyuk
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Assouad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Aswad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

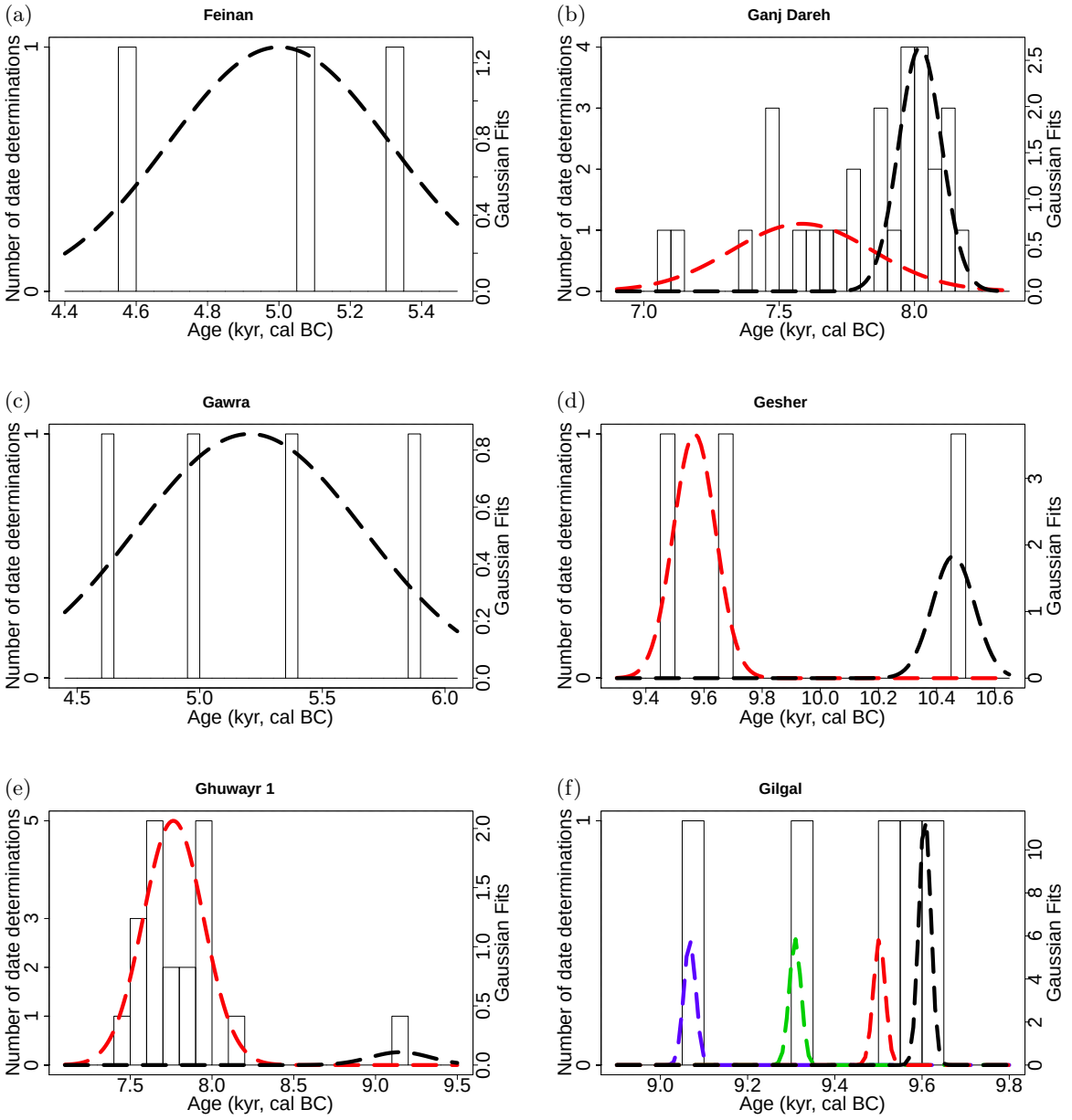

(a)
Atlit−Yam
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Ayakagytma
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Ba ja
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Baaz
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Bademagaci
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Bagor
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

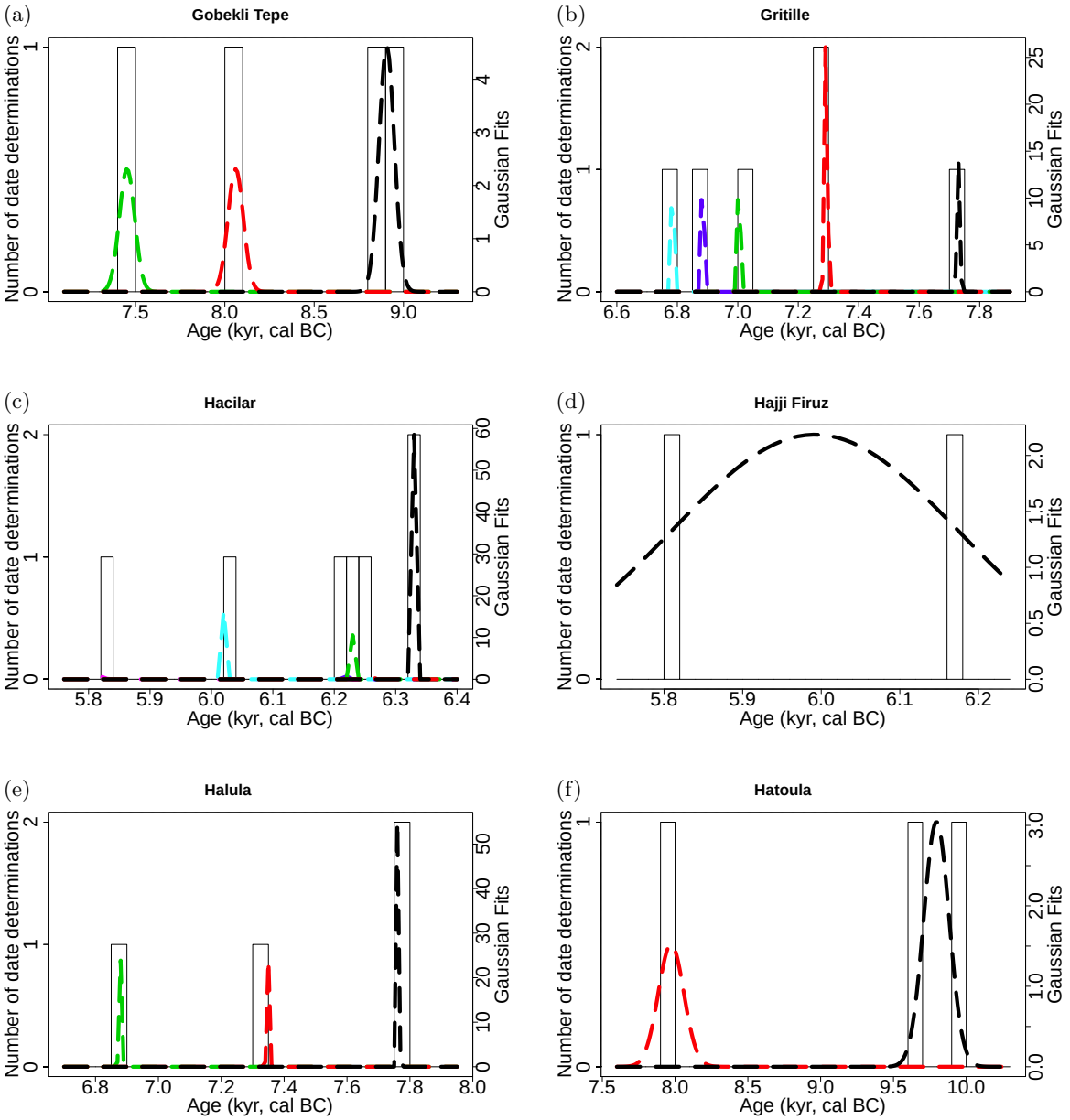

(a)
Bakun B
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Balakot
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Basta
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Beidha
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Betzet 1
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

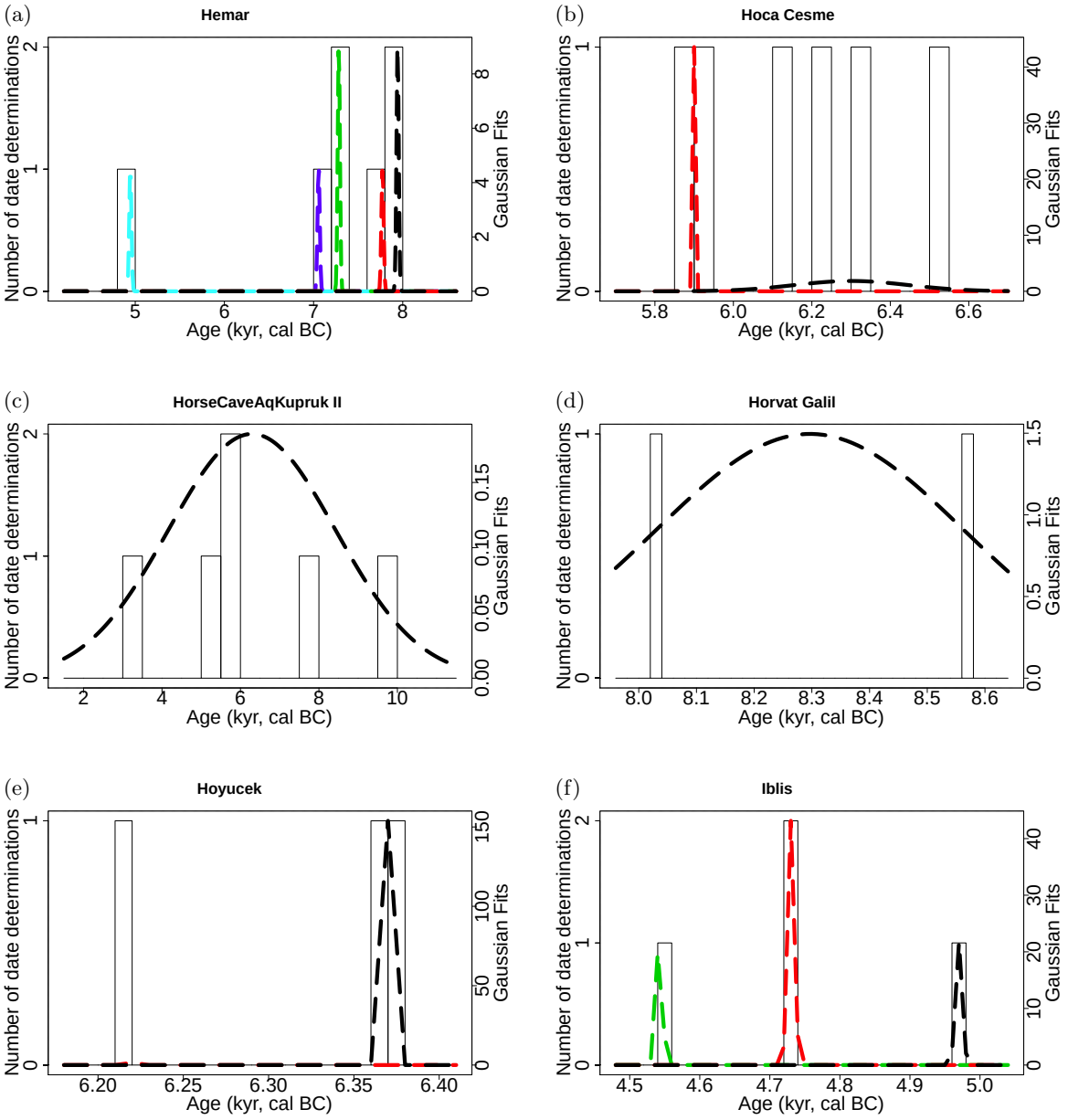

(a)
Bouqras
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Burqu 27
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Cafer
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Canhasan III
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Catalhoyuk East
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Cayonu
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

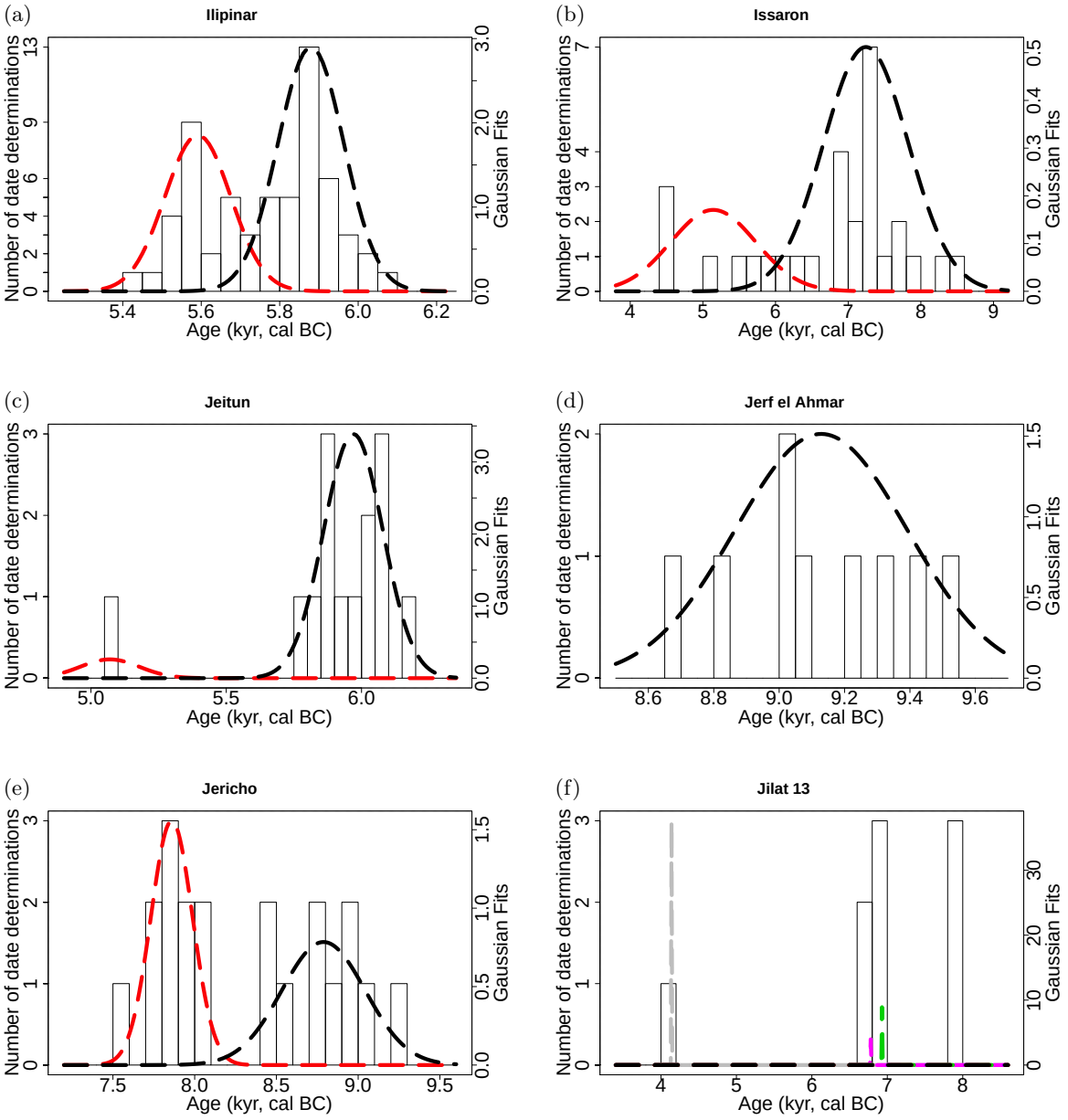

(a)
Chagha Sefid
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Cheshmeh Ali
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Chia Sabz
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Chogah Golan
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Chogha Bonut
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Chogha Mish
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

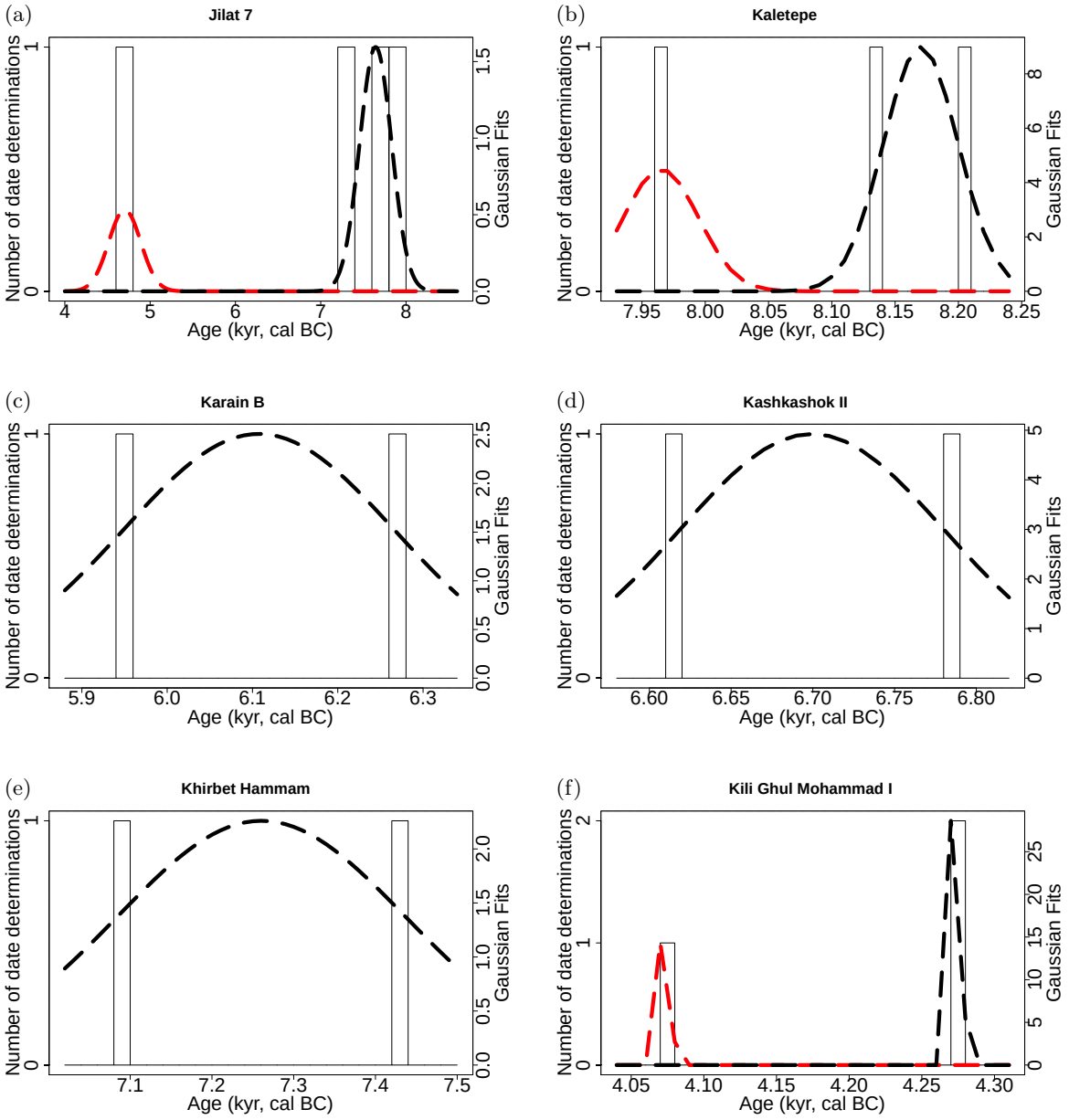

(a)
Damishliyah
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Dhra
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Dhuweila
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Djade
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
El Kowm 1
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Es Sifiya
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

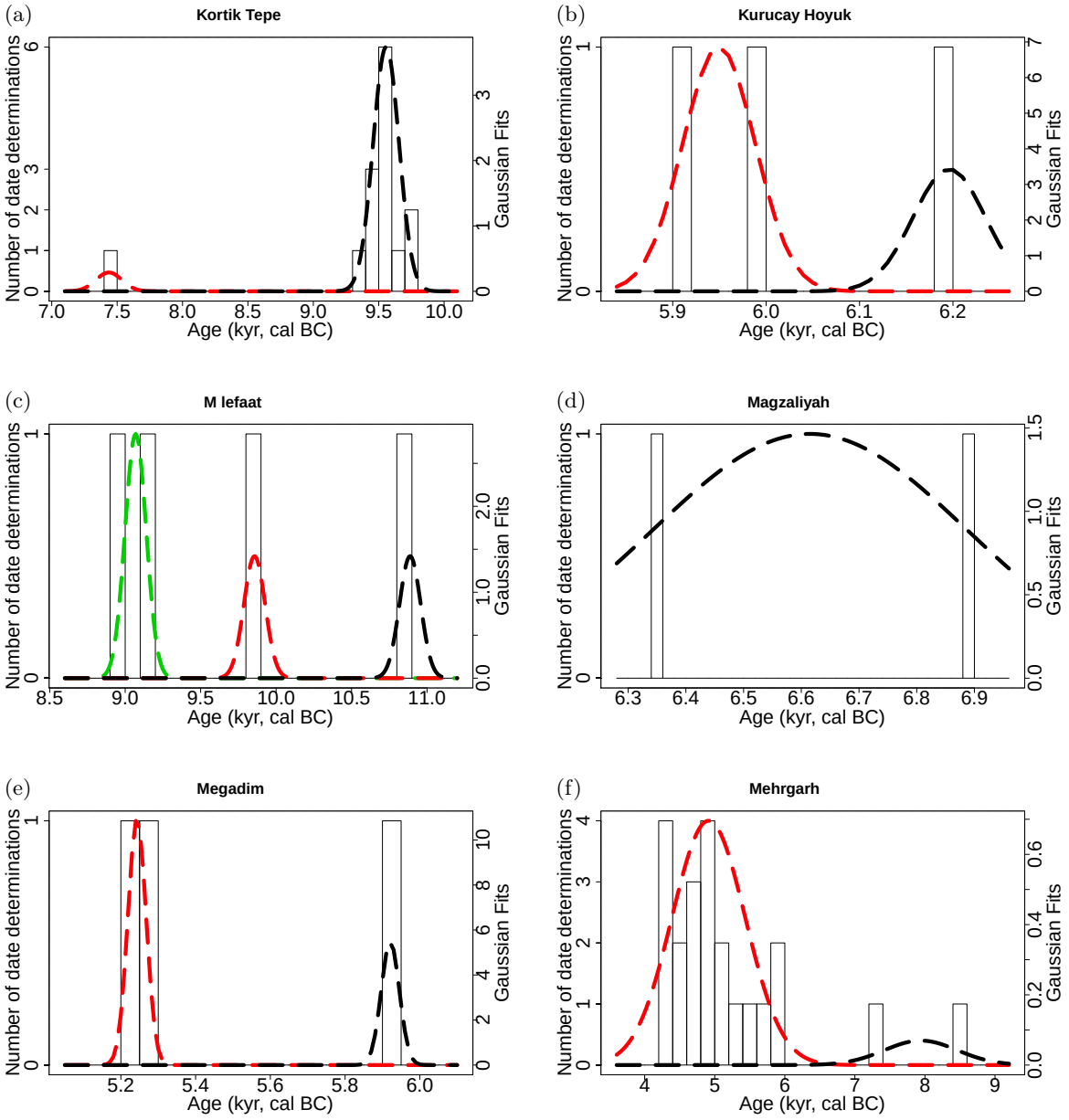

(a)
Feinan
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Ganj Dareh
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Gawra
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Gesher
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Ghuwayr 1
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Gilgal
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

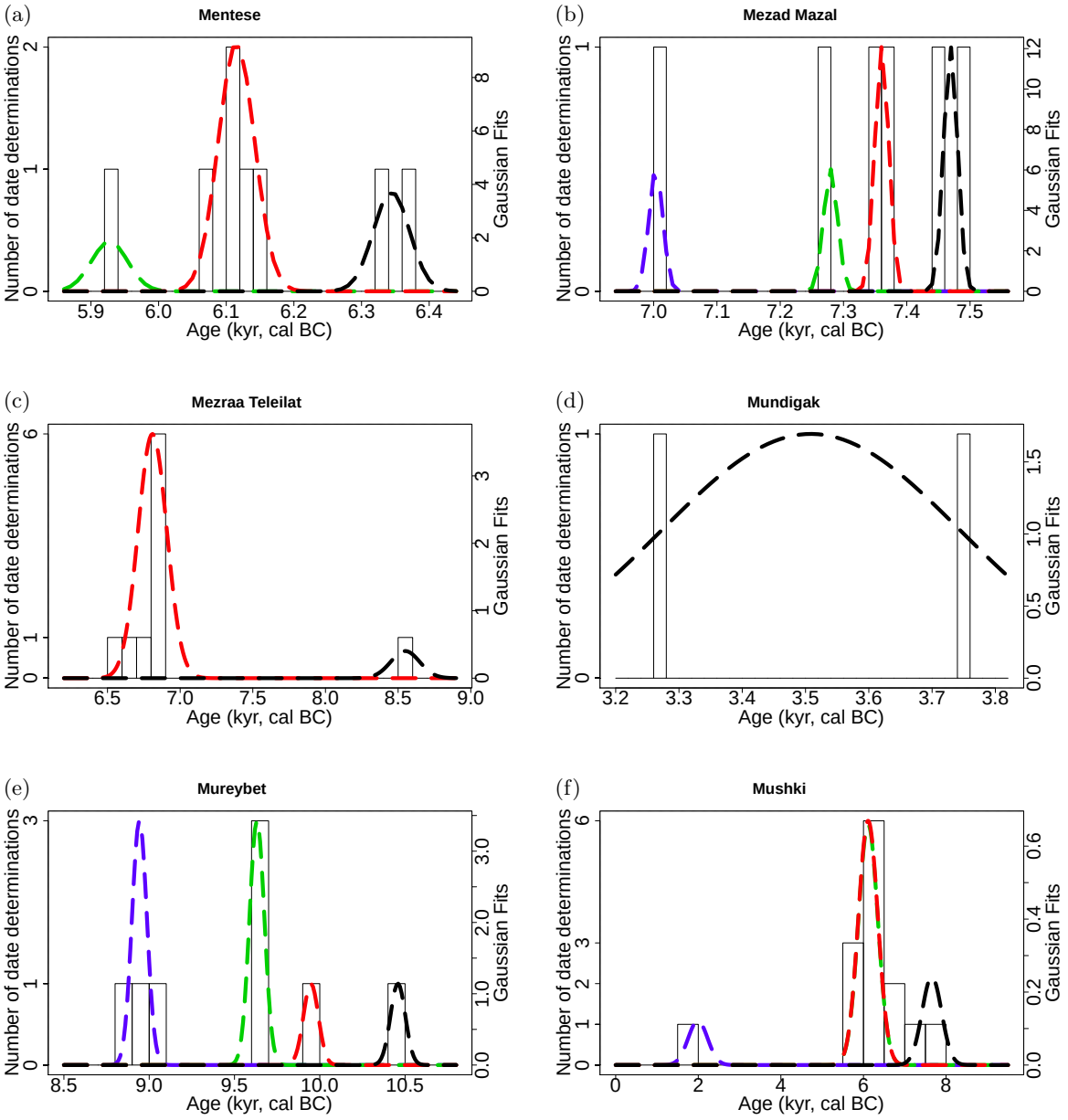

(a)
Gobekli Tepe
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Gritille
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Hacilar
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Hajji Firuz
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Halula
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Hatoula
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

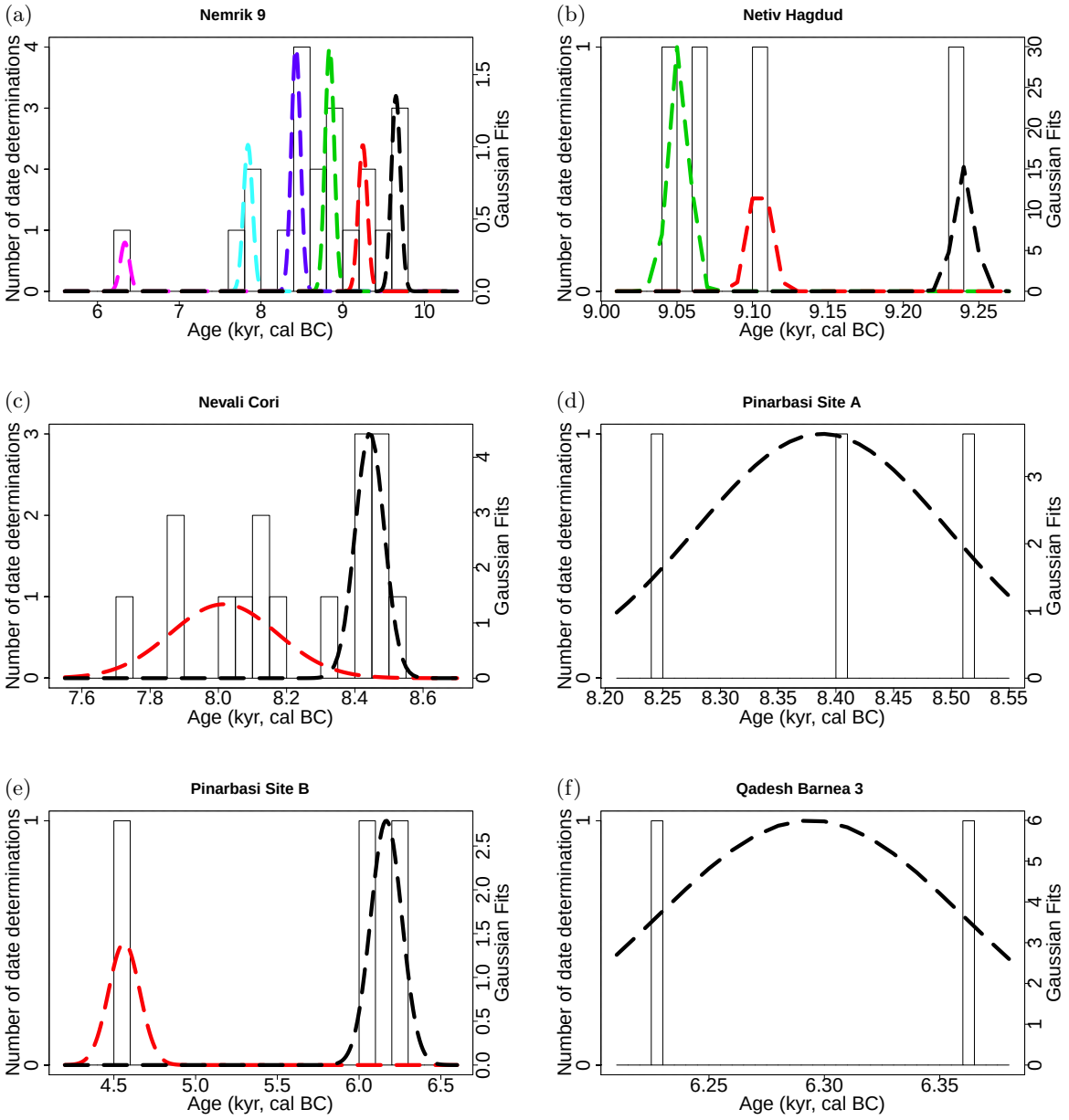

(a)
Hemar
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Hoca Cesme
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
HorseCaveAqKupruk II
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Horvat Galil
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Hoyucek
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Iblis
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

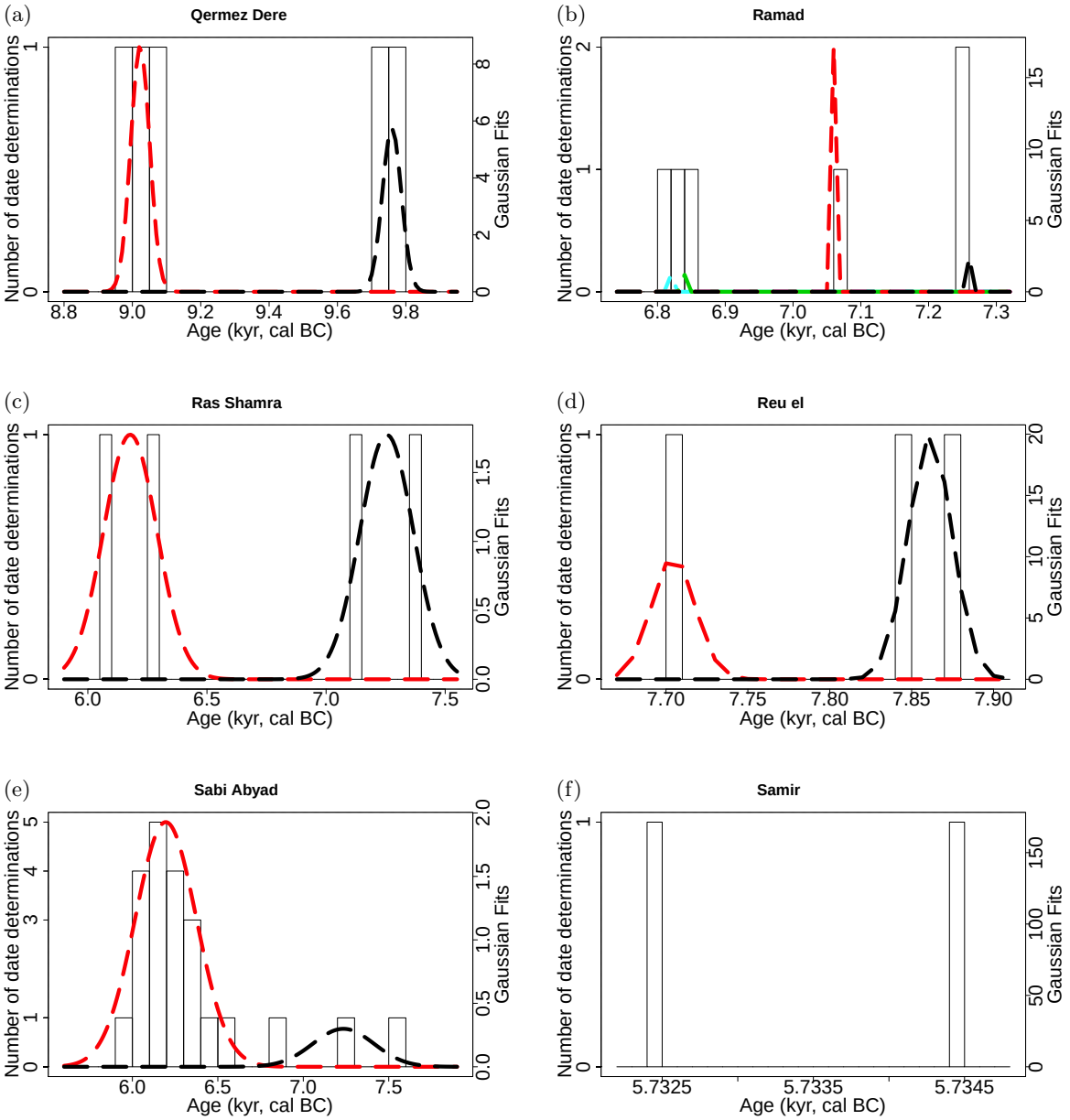

(a)
Ilipinar
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Issaron
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Jeitun
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Jerf el Ahmar
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Jericho
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Jilat 13
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

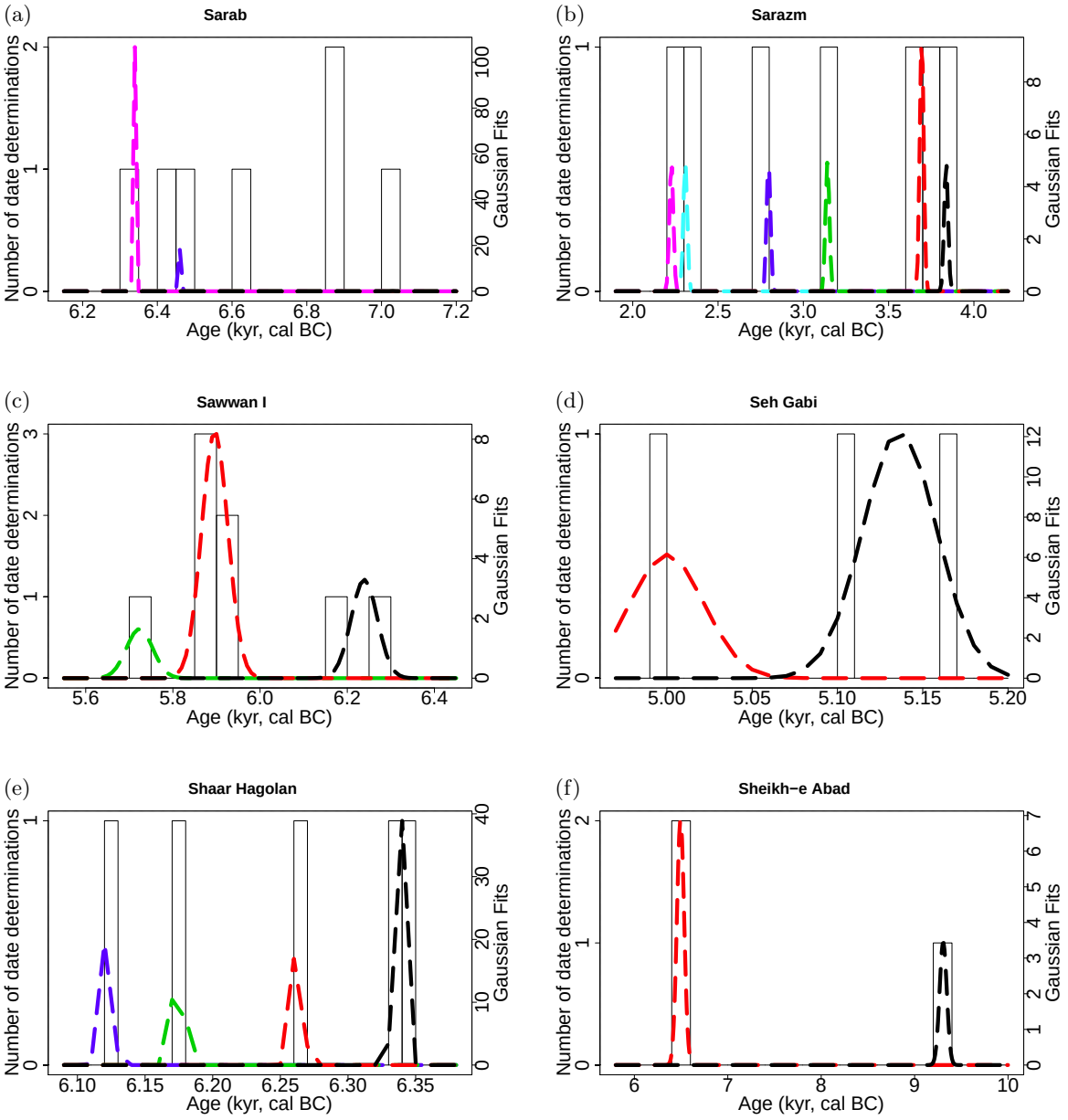

(a)
Jilat 7
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Kaletepe
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Karain B
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Kashkashok II
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Khirbet Hammam
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Kili Ghul Mohammad I
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

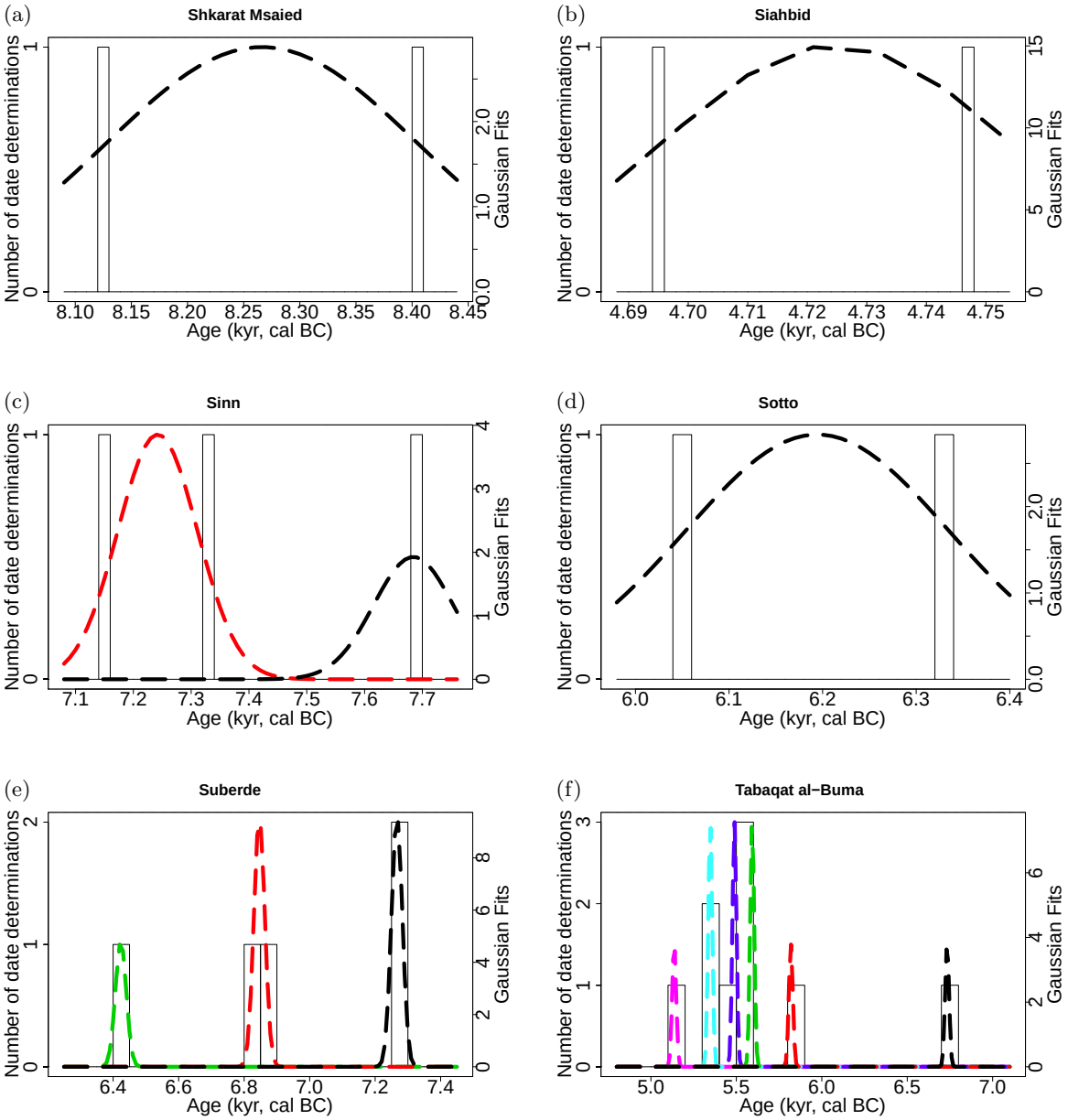

(a)
Kortik Tepe
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Kurucay Hoyuk
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
M Iefaat
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Magzaliyah
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Megadim
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Mehrgarh
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

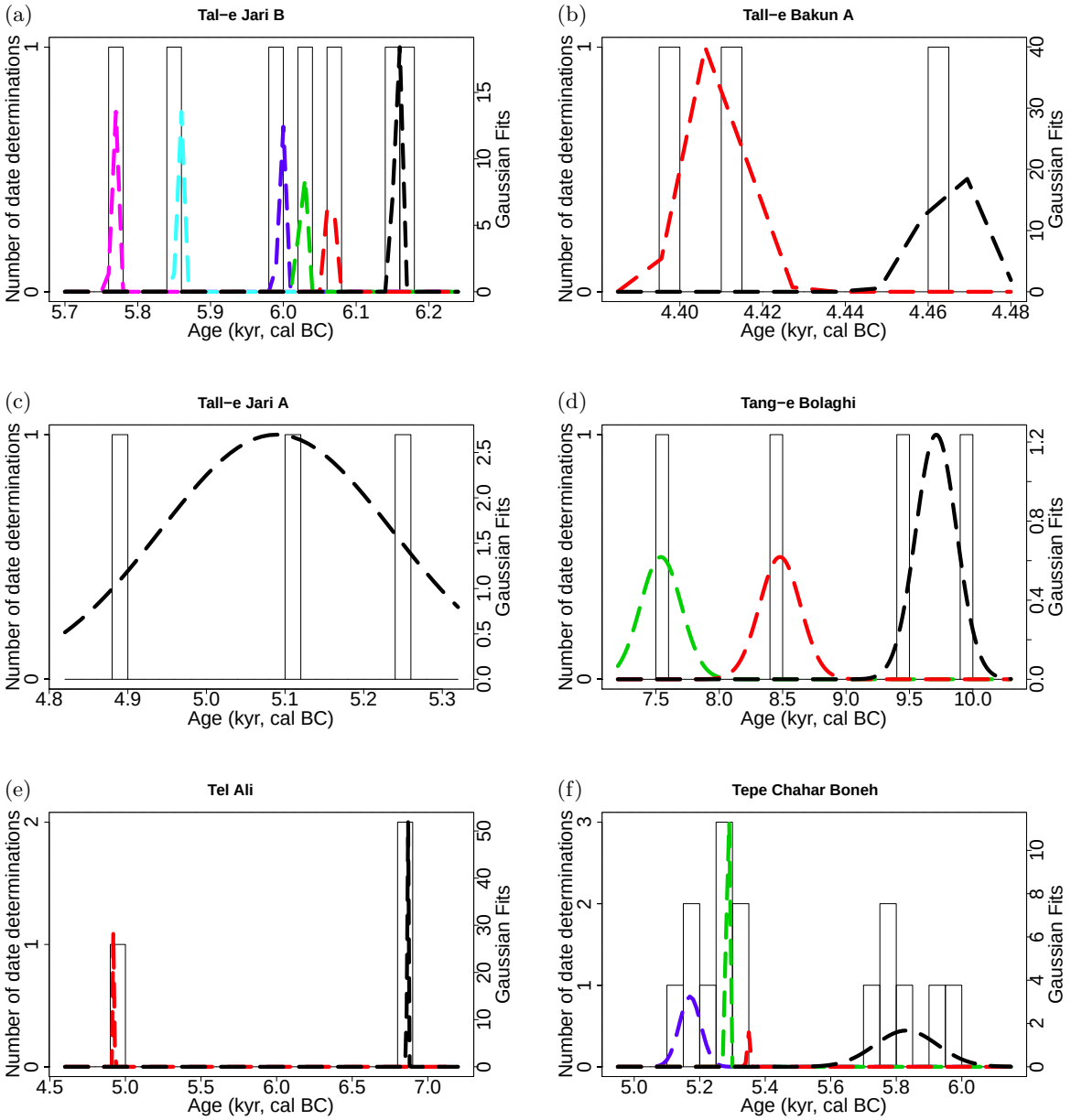

(a)
Mentese
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Mezad Mazal
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Mezraa Teleilat
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Mundigak
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Mureybet
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Mushki
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

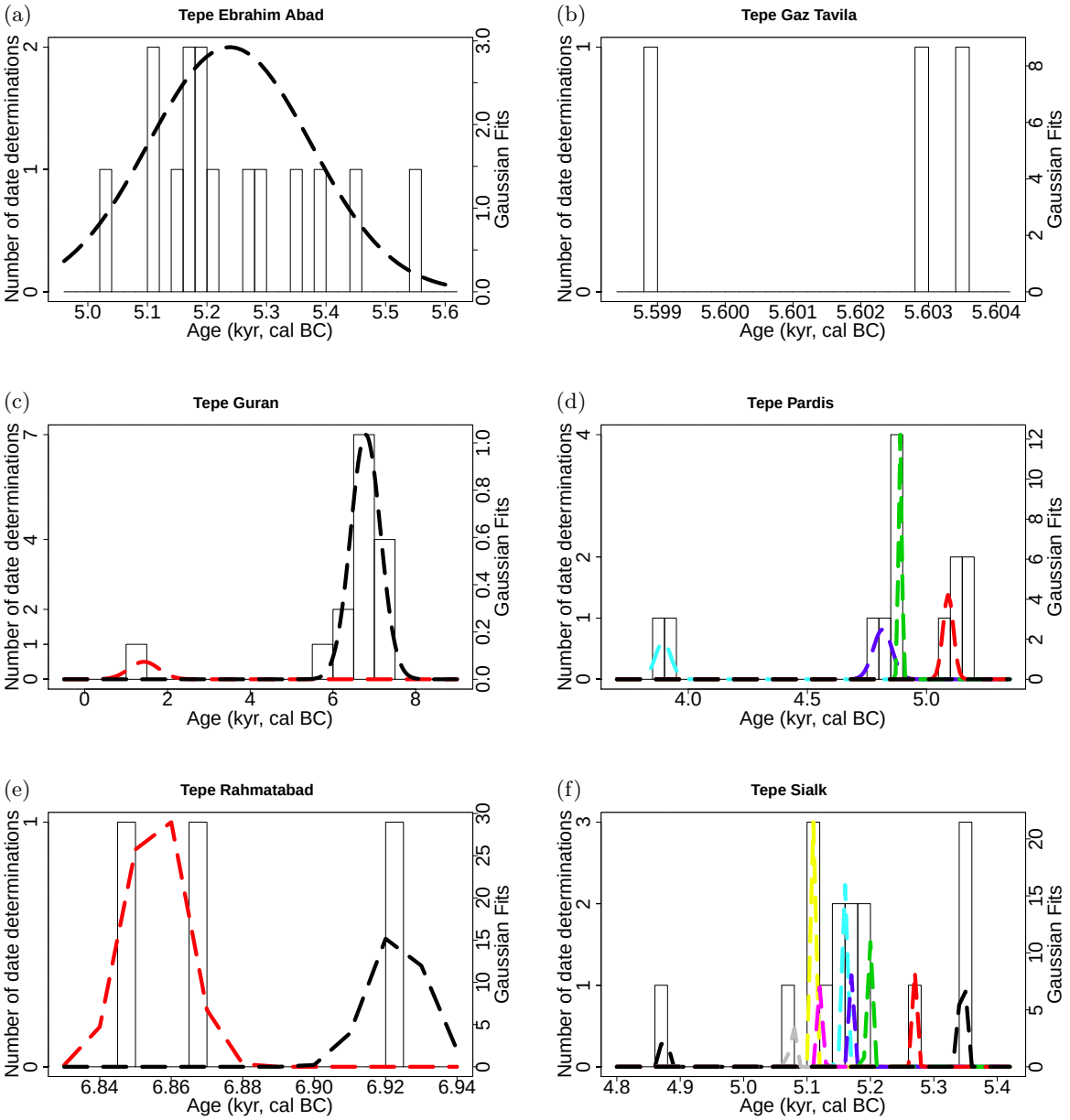

(a)
Nemrik 9
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Netiv Hagdud
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Nevali Cori
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Pinarbasi Site A
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Pinarbasi Site B
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Qadesh Barnea 3
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

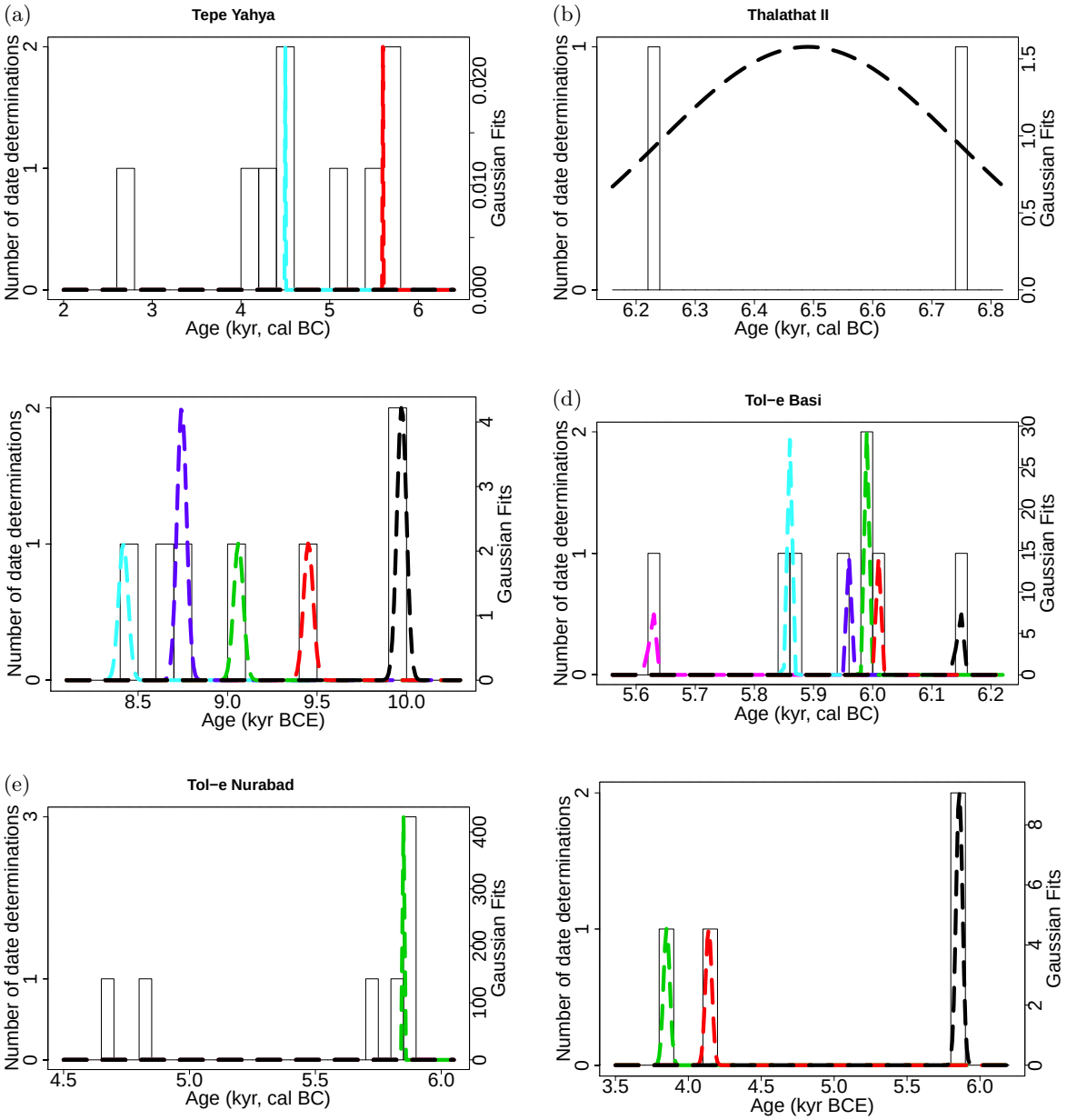

(a)
Qermez Dere
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Ramad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Ras Shamra
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Reu el
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Sabi Abyad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Samir
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

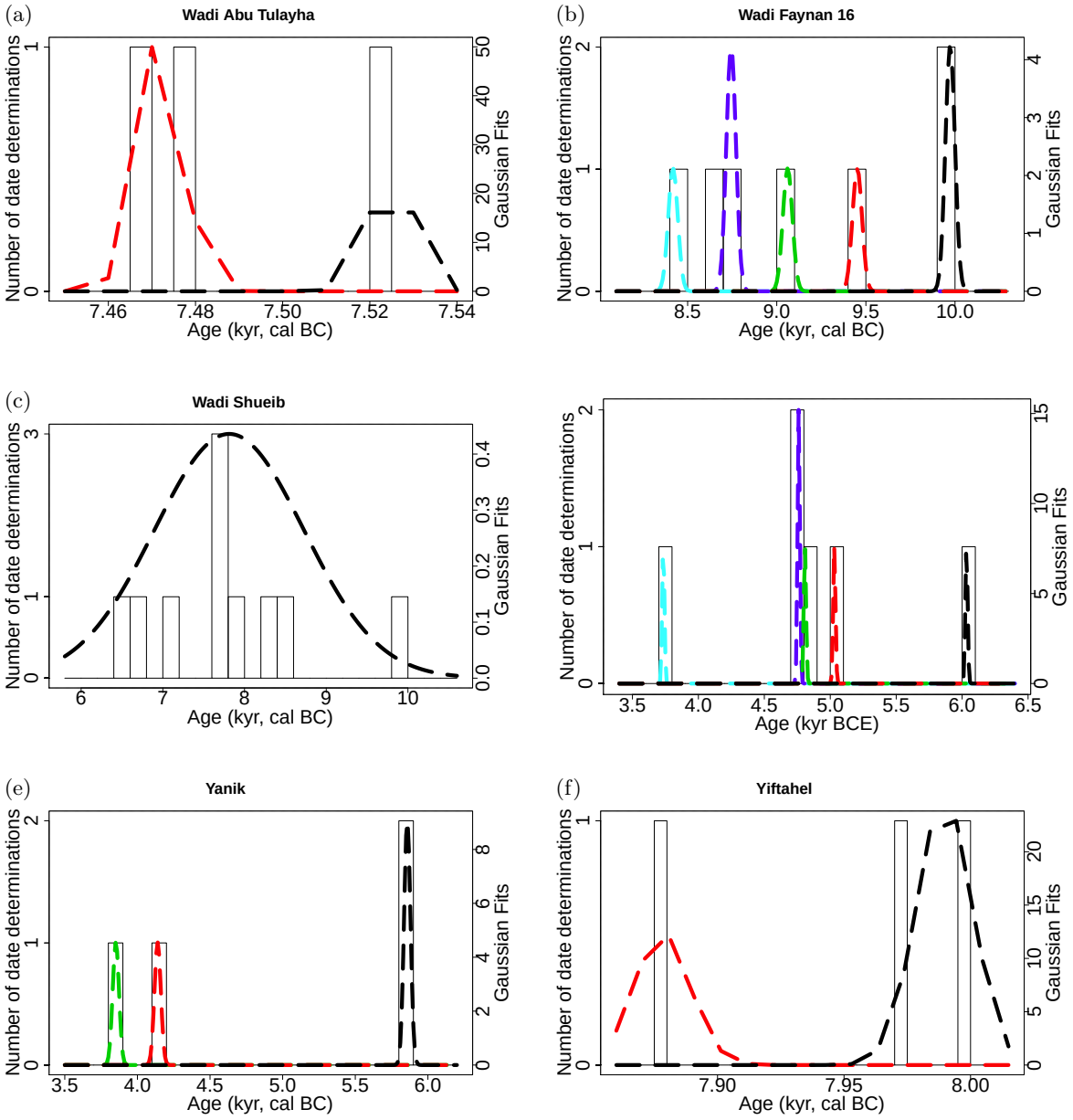

(a)
Sarab
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Sarazm
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Sawwan I
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Seh Gabi
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Shaar Hagolan
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Sheikh−e Abad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

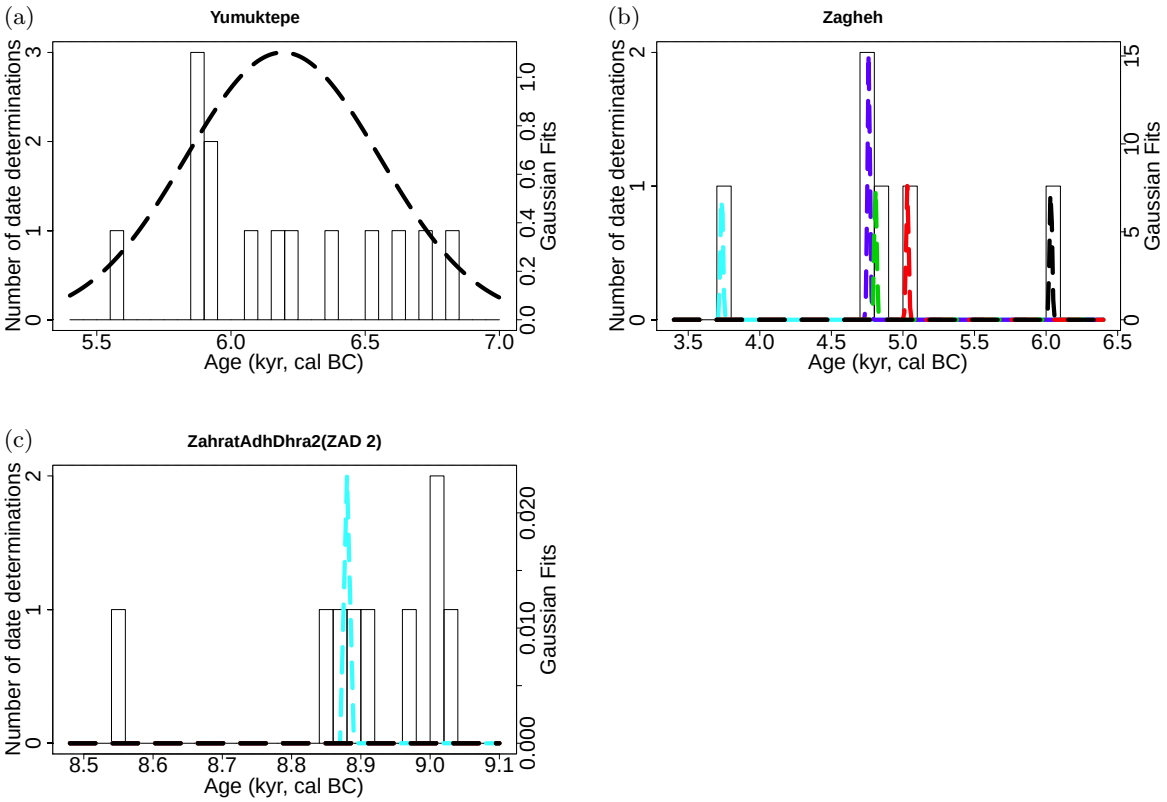

(a)
Shkarat Msaied
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Siahbid
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Sinn
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Sotto
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Suberde
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Tabaqat al−Buma
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)


(a)
Tal−e Jari B
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Tall−e Bakun A
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Tall−e Jari A
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Tang−e Bolaghi
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Tel Ali
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Tepe Chahar Boneh
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)


(a)
Tepe Ebrahim Abad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Tepe Gaz Tavila
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Tepe Guran
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(d)
Tepe Pardis
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Tepe Rahmatabad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Tepe Sialk
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)


(a)
Tepe Yahya
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Thalathat II
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
Number of date determinations
Gaussian Fits
Age (kyr BCE)
(d)
Tol−e Basi
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(e)
Tol−e Nurabad
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
Number of date determinations
Gaussian Fits
Age (kyr BCE)


(a)
Wadi Abu Tulayha
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Wadi Faynan 16
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
Wadi Shueib
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
Number of date determinations
Gaussian Fits
Age (kyr BCE)
(e)
Yanik
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(f)
Yiftahel
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)


(a)
Yumuktepe
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(b)
Zagheh
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)
(c)
ZahratAdhDhra2(ZAD 2)
Number of date determinations
Gaussian Fits
Age (kyr, cal BC)

Bibliography

[1] CONTEXT Radiocarbon Database. University of Köln. (Accessed in 2013).

[2] In Potts DT, Roustaei K, Petrie CA, Weeks LR, editor, *The Mamasani Archaeologica Project Stage One, A report on first two seasons of the ICAR- University of Sydney expedition to the Mamasani District, Fars Province, Iran*. Archaeopress, Oxford, 2009.

[3] Abdou M, Hamill L, Gilbert N. Designing and Building an Agent-Based Model. In Heppenstall AJ et al., editor, *Agent-Based Models of Geographical Systems*, pages 141–166. Springer Science+Business Media B.V., 2012.

[4] Ackland GJ, Signitzer M, Stratford K, Cohen MH. Cultural hitchhiking on the wave of advance of beneficial technologies. *Proc Natl Acad Sci USA*, 104(21):8714–8719, May 2007.

[5] Adamatzky A. Life's Still Lifes. In McIntosh HV, editor, *Game of Life Cellular Automata*. Springer Publications, 2010.

[6] Aerospace Systems Design Laboratory. Agent Based Modeling. http://www.asdl.gatech.edu/INIT:AGENT.html.

[7] Agrawal DP, Krishnamurthy RV, Kusum S. Physical Research Laboratory Radiocarbon Date List V. *Radiocarbon*, 27(1):95–110, 1985.

[8] Aitken M. *Science-based Dating in Archaeology*. . Longman, London, edition, 1990.

[9] Aitken MJ. *Science-based Dating in Archaeology.* Longman, London and New York, 1990.

[10] Akaike H. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mechanics, A*, 30:9–14, 1978.

[11] Alibaigi S, Khosravi S. Tepeh Khaleseh: a new Neolithic and Palaeolithic site in the Abharrud basin in north-western Iran. *Antiquity*, 83(319):, 2009.

[12] Alizadeh A. Excavations at the prehistoric mound of Chogha Bonut, Khuzestan, Iran. Technical report, University of Chicago, Illinois, 2003.

[13] Alizadeh A with contributions by Miller NF, Rosen AM, Redding RW. *Excavations at the Prehistoric Mound of Chogha Bonut; Khuzestan; Iran; Seasons 1976/77; 1977/78; and 1996.* Oriental Institute of the University of Chicago, Illinois, 2003.

[14] Alizadeh A with Miller NF, Kimiaie M, Mashkour M. *The Origins of state organizations in prehistoric highland FARS Southern Iran; Excavations at Tall-e Bakun; Appendix A : Tables 9;10;11*, volume 128. Oriental Institute Publications, 2006.

[15] Allianz.com. http://knowledge.allianz.com/environment/climate_change/?619/earths-climate-history-written-in-ice-wood-stone. Earth's climate history written in ice wood and stone.

[16] Ammerman AJ, Cavalli-Sforza LL. Measuring the Rate of Spread of Early Farming in Europe. *Man New Series*, 6(4):674–688, December 1971.

[17] Ammerman AJ, Cavalli-Sforza LL. A population model for the diffusion of early farming in Europe. In Renfrew C, editor, *The Explanation of Culture Change: Models in Prehistory*, pages 343–357. Duckworth, London, 1973.

[18] Aoki K, Shida M, Shigesada N. Travelling Wave Solutions for the Spread of Farmers into a Region Occupied by Hunter Gatherers. *Theoretical Population Biology*, 50:1–17, 1996.

[19] Axtell R, Epstein JM. Agent-Based modelling: Understanding our creations. *The Bulletin of the Santa Fe Institute*, pages 28–32, Winter 1994.

[20] Axtell RL, Epstein JM, Dean JS, Gumerman GJ, Swedlund AC, Harburger J, Chakravarty S, Hammond R, Parker J, Parker M. Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proc Natl Acad Sci USA*, 99(3):7275–7279, March 14 2002.

[21] Azarnoush M, Helwing B. Recent archaeological research in Iran Prehistory to Iron Age. *Archaeologische Mitteilungen aus Iran und Turan*, 37:189–246, 2005.

[22] Baggaley AW, Sarson GR, Shukurov A, Boys RJ, Golightly A. Bayesian inference for a wavefront model of the Neolthisation of Europe. *Phys Rev E*, 86:016105, 2012.

[23] Barker G. *The Agricultural Revolution in Prehistory: Why did foragers become farmers?* Oxford University Press, 2006.

[24] Bellwood P. Farmers; foragers; languages; genes: the genesis of agricultural societies. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.

[25] Bellwood PS. *First Farmers: The Origins of Agricultural Societies*. Blackwell, London, 2005.

[26] Benz M, Coskun A, Weninger B, Alt KW, Ozkaya V. Stratigraphy and Radiocarbon dates of the PPNA site of Kortik tepe Diyarbakir. *Arkeometri Sonuçlari Toplantisi*, 26:81–100, 2010.

[27] Bernback R, Pollock S, Nashli HF. Rahmatabad : Dating the aceramic Neolithic in Fars province. *Neo-Lithics*, 1(08):37–39, 2008.

[28] Bocquet-Appel J-P, Naji S, Vander Linden M, Kozlowski JK. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *J Archaeol Sci*, 36:807–820, March 2009.

[29] Bradley SP, Hax AC, Magnanti TL. *Applied Mathematical Programming*. Addison-Wesley Publishing Company, Reading, MA, 1977.

[30] Bronk Ramsey C. Bayesian analysis of radiocarbon dates. . *Radiocarbon*, 51:337–360, 2009.

[31] Buck CE, Litton CD, Smith AFM. Calibration of radiocarbon results pertaining to related archaeological events. *J Archaeol Sci*, 19:497–512, 1992.

[32] Burleigh R, Ambers J, Matthews K. British Museum natural radiocarbon measurements XV. *Radiocarbon*, 24(3):262–290, 1982.

[33] Burnham KP, Anderson DR. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer, NY, 2 edition, 2010.

[34] Campbell L. What drives linguistic diversification and language spread. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.

[35] Cauvin J. Les Debuts de la ceramique sur le Moyen-Euphrate : nouveaux documents. *Paleorient*, 2(1):199–205, 1974.

[36] Cavalli-Sforza LL. Demic diffusion as the basic process of human expansions. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.

[37] Center for climatic research, Department of Geography, University of Delaware. Precipitation and temperature data. Willmott, Matsuura and Collaborators' Global Climate Resource Pages. (Accessed in 2014).

[38] Childe VG. *The Dawn of European Civilization*. K. Paul, Trench, Trubner & Co., London, 1925.

[39] Clark JGD. Radiocarbon dating and the expansion of farming culture from the Near East over Europe. *The Prehistoric Society*, 1965.

[40] Clark JGD. Radiocarbon dating and the spread of farming economy. *Antiquity*, 39:45–48, 1965.

[41] Cohen MH. Nonlinearity, disorder, the spread of Neolithic farming, and the origin of Indo-European languages. In Abdullaev F, Bishop AR, Pnevmatikos S, editor, *Nonlinearity with Disorder*, volume 67 of *Springer proceedings in Physics*, pages 161–170. Springer Verlag, Berlin, 1992.

[42] Costantini L. The first farmers in Western Pakistan: the evidence of the Neolithic agro-pastoral settlement of Mehrgarh. *Pragdhara*, 18:167–178, 2008.

[43] Crooks AT, Heppenstall AJ. Introduction to Agent-Based Modelling. In Heppenstall AJ et al., editor, *Agent-Based Models of Geographical Systems*, pages 85–105. Springer Science+Business Media B.V., 2012.

[44] CZAP. http://www.czap.org/jani.

[45] CZAP : The Central Zagros Archaeological Project. http://www.czap.org/sheikh-e-abad, Accessed in 2013.

[46] Davison K, Dolukhanov P, Sarson GR, Shukurov A. The role of waterways in the spread of the Neolithic. *J Archaeol Sci*, 33:641–652, 2006.

[47] Davison K, Dolukhanov PM, Sarson GR, Shukurov A, Zaitseva GI. Multiple sources of the European Neolithic: mathematical modelling constrained by radiocarbon dates. *Quaternary International*, 203:10–18, 2009.

[48] Dolukhanov P. *Environment and Ethnicity in the Ancient Middle East*. Ashgate, Aldershot, 1994.

[49] Dolukhanov P. *Environment and Ethnicity in the Middle East*, volume 7 of *World-wide Archaeology Series*. Avebury, 1994.

[50] Dolukhanov P, Shukurov A. Modelling the Neolithic dispersal in Northern Eurasia. *Documenta Praehistorica*, 31:35–47, 2004.

[51] Dolukhanov P, Shukurov A, Gronenborn D, Sokoloff D, Timofeev V, Zaitseva G. The chronology of Neolithic dispersal in Central and Eastern Europe. *J Archeo Sci*, 32:1441–1458, 2005.

[52] Dolukhanov P, Sokoloff D, Shukurov A. Radiocarbon chronology of Upper Palaeolithic sites in Eastern Europe at improved resolution. *J Archaeol Sci*, 28:699–712, 2001.

[53] Drechsler P. *The dispersal of the Neolithic over the Arabian Peninsula*. British Archaeological Reports International Series S1969, Archaeopress, Oxford, 2009.

[54] Edmonson MS. Neolithic diffusion rates. *Current Anthropology*, 2:71–102, 1961.

[55] Ex Oriente. PPND: the Platform for publication of Neolithic Radiocarbon Dates, 2013. (Accessed in 2013).

[56] Novenko EYu. 9B: Vegetation. In Velichko AA, editor, *Paleoclimates and paleoenvironments of extra-tropical area of the northern hemisphere: Late Pleistocene-Holocene, Atlas monograph*. Russian Academy of Sciences, Institute of geography, Moscow, GEOS, 2009.

[57] Fazeli H, Conningham RAE, Batt CM. Chashmeh-Ali Revisited: towards an absolute dating of the Late Neolithic and Chalcolithic of Iran's Tehran Plain. *IRAN*, 42:13–23, 2004.

[58] Feugier FG, Sarson GR, Shukurov A, Dolukhanov PM. Population spread along self-organized paths. In Dolukhanov PM, Sarson GR, Shukurov AM, editor, *The East European Plain on the Eve of Agriculture*, pages 217–228. British Archaeological Reports International Series S1964, Archaeopress, Oxford, 2009.

[59] Fisher RA. The wave of advance of advanageous genes. *Annals of Eugenics*, 7:355–369, 1937.

[60] Fort J. Mathematical models of the Neolithic transition: a review for non-mathematicians. In Dolukhanov PM, Sarson GR, Shukurov AM, editor, *The East*

European Plain on the Eve of Agriculture, pages 211–216. British Archaeological Reports International Ser. S1964, Archaeopress, Oxford, 2009.

[61] Fort J. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. *Proc Natl Acad Sci USA*, 109:18669–18673, 2012.

[62] Fort J, Pujol T. Progress in front propagation research. *Rep Prog Phys*, 71: 086001:41 pp., 2008.

[63] Fraley C, Raftery AE. MCLUST version 3 for R: normal mixture modeling and model-based clustering. *Technical Report No. 504, Department of Statistics, Univ. Washington*, 2006.

[64] Fujii S. Wadi Abu Tulayha and Wadi Ruweishid ash-Sharqi; An investigation of PPNB barrage systems in the Jafr Basin. *Neo-Lithics A newsletter of Southwest Asian Lithics Research*, 2:6–17, 2007.

[65] Fuller DQ. Private communication.

[66] Fuller DQ. Agricultural origins and frontiers in South Asia: a working synthesis. *J World Prehistory*, 20:1–86, 2006.

[67] Gangal K, Sarson GR and Shukurov A. The Near-Eastern roots of the Neolithic in South Asia. *PLoS ONE*, 9(5), May 7 2014.

[68] Gangal K, Vahia MN, Adhikari R. Spatio-temporal analysis of the Indus urbanization. *Curr Sci*, 98(6):846–852, March 2010.

[69] Gkiasta M, Russell T, Shennan S, Steele J. Neolithic transition in Europe: the radiocarbon record revisited. *Antiquity*, 77:45–62, 2003.

[70] Golub GH, Van Loan CF. An analysis of the total least squares problem. *Numer. Anal.*, 17:883–893, 1980.

[71] Goring-Morris AN, Belfer-Cohen A. Neolithization processes in the Levant: the outer envelope. *Curr Anthropol*, 52(S4):S195–S208, 2011.

[72] Görsdorf J. 14C-Datings of the Es-Sifiya Settlement (Area C). In Bienert HD, Müller-Neuhof B, Wagner-Lux U, Liedgens I, editor, *At the Crossroads. Essays on the Archaeology, History and Current Affairs of the Middle East*, pages 15–19. German Protestant Institute of Archaeology in Amman, 2000.

[73] Harris D, Coolidge J. *The Mesolithic and Neolithic periods: sites, sequences and subsistence*. University of Pennsylvania Press, Pennsylvania, 2010.

[74] Harris DR. The expansion capacity of early agricultural systems: a comparative perspective on the spread of agriculture. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.

[75] Harris DR. *Origins of Agriculture in Western Central Asia: An Environmental-Archaeological Study*. Univ Pennsylvania Press, Philadelphia, 2010.

[76] Hedges REM, Housley RA, Pettitt PB,Ramsey C,Bronk CR,Van Klinken GJ. Radiocarbon Dates from the Oxford AMS system: Archaeometry datelist 21. *Archaeometry*, 38(1):181–207, 1996.

[77] Henry DO, Cordova C, White JJ, Dean RM, Beaver JE, et al. The Early Neolithic Site of Ayn Abu Nukhayla, Southern Jordan. *Bulletin of the American Schools of Oriental Research (BASOR)*, 330:1–30, May 2003.

[78] Hiebert FT, Dyson RH. Prehistoric Nishapur and frontier between Central Asia and Iran. *Iranica Antiqua*, XXXVII:113–149, 2002.

[79] Hole F. Chronologies in the Iranian Neolithic, Chronologies in the Near East : relative chronologies and absolute chronology 16.000-4.000 B.P. : C.N.R.S. International symposium, Lyon (France). In Aurenche O. et al., editor, *BAR International Series 379(i)*, pages 353–379. 24–28 November 1986.

[80] How stuff works? http://science.howstuffworks.com/environmental/earth/geology/carbon-141.htm.

[81] Isakov A, Kohl PL, Lamberg-Karlovsky CC, Maddin R. Metallurgical Analysis From Sarazm; Tadjikistan SSR. *Archaeometry*, 29(1):90–102, 1987.

[82] Jarrige C. The figurines of the first farmers at Mehrgarh and their offshoots. *Pragdhara*, 18:155–166, 2008.

[83] Jarrige JF. Mehrgarh Neolithic. *Pragdhara*, 18:136–154, 2008.

[84] John Rafferty (Ed.). *Geochronology, Dating and Precambrian time: The beginning of the world as we know it.* Encyclopedia Britannica, Britannica educational publishing, New York, 2011.

[85] Kivisild T, Rootsi S, Metspalu M, Metspalu E, Parik J, et al. The genetics of language and farming spread in India. In Bellwood P and Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis.* McDonald Institute for Archaeological Research, Cambridge, 2002.

[86] Kohler TA, Gumerman GJ, Reynolds RG. Simulating Ancient Societies. *Scientific American*, pages 77–84, July 2005.

[87] Kolmogorov AN, Petrovskii IG, Piskunov NS. A study of the diffusion equation-with increase in the quantity of matter, and its application to a biological problem. *Bulletin of Moskow University, Mathematic Series A*, 1:1–25, 1937.

[88] Kozlowski SK. Nemrik 9; a PPN Neolithic site in Northern Iraq. *Paleorient*, 15(1):25–31, 1989.

[89] Kozlowski SK. Radiocarbon dates from aceramic Iraq. In Bar-Yosuf O, Kra RS, editor, *Late Quaternary Chronology and Paleoclimates of the Eastern Mediterranean*, pages 255–264. Tucson: Radiocarbon, 1994.

[90] Kozlowski SK. *Eastern Wing of the Fertile Crescent: Late Prehistory of Greater Mesopotamian Lithic Industries (British Archaeological Reports (BAR) International).* Number 760 in British Archaeological Reports. Oxford, 1999.

[91] Kreft JU, Booth G, Wimpenny JW. BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology*, 144(12):3275–3287, December 1998.

[92] Kuzmina EE, Mair VH. *The Prehistory of the Silk Road*. Univ. Pennsylvania Press, Philadelphia, 2008.

[93] Leonard T, Hsu JSJ. *Bayesian methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1999.

[94] Main IG, Leonard T, Papasouliotis O, Hatton CG, Meredith PG. One slope or two? Detecting statistically significant breaks of slope in geophysical data, with application to fracture scaling relationships. *Geophys Res Lett*, 26(18):2801–2804, 1999.

[95] Mani BR. Kashmir Neolithic and Early Harappan: a linkage. *Pragdhara*, 18, 2008.

[96] Marshall JL. *Missing Links: Demic Diffusion and the Development of Agriculture on the Central Iranian Plateau*. PhD thesis, 2012.

[97] Matthews R, Fazeli NH. The Neolithisation of Iran: Patterns of change and continuity. In Matthews R, Fazeli NH, editor, *The Neolithisation of Iran, The formation of new societies*, pages 1–13. Oxbow Books, Oxford, UK, 2013.

[98] Mazurkevich A, Dolukhanov P, Shukurov A, Zaitseva G. Late Stone – early Bronze age cites in the Western Dvina–Lovat Area . In Dolukhanov PM, Sarson GR, Shukurov AM, editor, *The East European Plain on the Eve of Agriculture*, pages 145–153. British Archaeological Reports International Series 1964, Archaeopress, Oxford, 2009.

[99] Murray JD. *Mathematical Biology I: An Introduction*. Springer-Verlag, New York, third edition, 1989,1993.

[100] NGDC. Topography data. National Geophysical Data Center. (Accessed in 2013).

[101] North MJ, Macal CM, Vos JR. Terrorist organization modelling. In *North American Association for Computational, Social and Organizational Science Conference 2004, Pittsburgh*, 2004.

[102] Özdoğan M. The expansion of the Neolithic way of life: what we know and what we do not know . In Lichter C, editor, *How Did Farming Reach Europe? Anatolian-European Relations from the Second Half of the 7^{th} Through the First Half of the 6^{th} Millennium Cal BC (Byzas)*, pages 13–27. Istanbul: Ege, 2005.

[103] Ozkaya V, Coskun A. Kŏrtik Tepe Kazilari: Erken Neolithik Dŏnemde Bŏlgesel Kŭlturel Ilişkiler Ŭzerine Bazi Gŏzlemler. *Işik D. Arkeologji Yazilari;Atatürk Üniversitesi 50. Kuruluş Yildönümü Arkeoloji Bölümü Armağani;Istanbul*, pages 85–98, 2007.

[104] Ozkaya V, San O. Körtik Tepe: Bulgular Işiğinda Külturel Doku Üzerine-Ilk Gözlemler. *Anadolu'da Uygarliğin Doğuşu ve Avrupa'ya Yayilimi: Türkije'de Neolithik Dönem, Yeni Kazilar, Yeni Bulgular*, pages 21–36, 2007.

[105] Pant PC. First farmers in global perspective: some considerations. *Pragdhara*, 18, 2008.

[106] Patterson MA, Sarson GR, Sarson HC, Shukurov A. Modelling the Neolithic transition in a heterogeneous environment. *J Archeol Sci*, 37:2929–2937, 2010.

[107] Petrie CA, Thomas KD. The topographic and environmental context of the earliest village sites in western South Asia. *Antiquity*, 86:1055–1067, 2012.

[108] Pinhasi R, Fort J, Ammerman AJ. Tracing the origin and spread of agriculture in Europe. *PLoS Biol*, 3:2220–2228, 2005.

[109] Pollard AM, Davoudi H, Mostafapour I, Valipour HR, Nashli HF. A New radiocarbon chronology for the late Neolithic to Iron age in the Qazvin plain, Iran. *Intl J Humanities*, 19(3):110–151, 2012.

[110] Possehl GL. *Indus Age: The Beginnings*. Univ. Pennsylvania Press, Philadelphia, 1999.

[111] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, , 3 edition, 2007.

[112] Price TD. *Europe's First Farmers*. Cambridge Univ. Press, Cambridge, 2000.

[113] Quintana-Murci L, Chaix R, Spencer Wells R, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Mehdi SQ, Torroni A, McElreavey K. Where West meets East: the complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet*, 74:827–845, 2004.

[114] Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K. Y-chromosome lineages trace diffusion of people and languages in Southwestern Asia. *Am J Hum Genet*, 68:537–542, 2001.

[115] Railsback SF,Grimm V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton University Press, 2012.

[116] Reimer PJ, Baillie MGL, Bard E et al. INTCAL09 and MARINE09 radiocarbon age calibration curves, 050,000 years cal BP. *Radiocarbon*, 51(4):1111–1150, 2009.

[117] Renfrew C. The emerging synthesis: the archaeogenetics of farming/language dispersals and other spread zones. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.

[118] Riehl S , Benz M , Conard NJ , Darabi H , Deckers K , Fazeli Nashli H , Zeidi-Kulehparcheh M. Plant use in three Pre-Pottery Neolithic sites of the northernand eastern Fertile Crescent: a preliminary report. *Vegetation History and Archaeobotany*, 21(2):95–106, 2011.

[119] Rollefson G. Expanded Radiocarbon Chronology from Ain Ghazal. *Neo-Lithics A newsletter of Southwest Asian Lithics Research*, 2:8–10, 1998.

[120] Rowley-Conwy P. Westward Ho! The spread of agriculture from Central Europe to the Atlantic. *Curr Anthropol*, 52(Supplement 4):S431–S451, 2011.

[121] Salem HJ. Cultural transmission and change in raditional Palestinian pottery production. *Leiden journal of pottery studies*, 22, 2006.

[122] Sauer CO. *Agricultural origins and dispersals*. series two. American Geographical Society, New York, 1952.

[123] Scott EM, Cook GT, Naysmith P. Error and uncertainty in radiocarbon measurements. *Radiocarbon*, 49(2):427–440, 2007.

[124] Shiffman D. *The Nature of Code*. The Nature of Code, 2011.

[125] Simmons AH, Najjar M. Ghwair I: A Neolithic community in Southern Jordan. *J Field Archaeol*, 31:77–95, 2006.

[126] Simmons AH, Rollefson GO, Kafafi Z, Mandel R, Al-Nahar M, Cooper J, Koehler-Rollefson I. Wadi Shu'eib; a large Neolithic community in central Jordan: Final report of test investigations. *Bulletin of the American Schools of Oriental Research*, 321:1–39, 2001.

[127] Smaldino PE. Cooperation in harsh environments and the emergence of spatial patterns. *Chaos, Solitons & Fractals*, 56:6–12, November 2013.

[128] Smith DG. Examining the farming/language dispersal hypothesis. *The American Society of Human Genetics*, 76(1):190–192, January 2005.

[129] Southworth FC. *Linguistic archaeology of South Asia*. Routledge Curzon, 2005.

[130] Steele J, Adamsb J, Sluckinc T. Modelling Paleoindian dispersals. *World Archaeology*, 30(2):286–305, 1998.

[131] Szymczak K, Khudzhanazarov M. *Exploring the Neolithic of the Kyzyk-Kums*. Institute of Archaeology, Warsaw University, 2006.

[132] Thiele JC, Kurth W, Grimm V. Agent-and Individual-based Modelling with NetLogo: Introduction and New NetLogo Extensions. In Römisch K, Nothdurft A, Wunn U, editor, *Die Grüne Reihe 22. Tagung der Sektion Forstliche Biometrie und Informatik des Deutschen Verbandes Forstlicher Forschungsanstalten und der Arbeitsgemeinschaft Ökologie und Umwelt der Internationalen Biometrischen Gesellschaft-Deutsche Region*, pages 68–101. Göttingen (Germany), September 2010.

[133] Tsuneki A. Proto-Neolithic caves and neolithisation in the southern Zagros. In Matthews R, Fazeli NH, editor, *The Neolithisation of Iran, The formation of new societies*, pages 84–96. Oxbow Books, Oxford, UK, 2013.

[134] Valipour HR, Davoudi H, Sadati JH, Nashli HF. Tepe Khaleseh: archaeological evaluation of a Late Neolithic site in north-western Iran. *Antiquity*, 86(331):, 2012.

[135] Vavilov NI. The problem of the origin of the world's agriculture in the light of the latest investigations. In *Science at the Crossroads: Papers Presented to the International Congress of the History of Science and technology*. Kniga, London, 1931.

[136] Walker M. *Geochronology, Dating and Precambrian time: The beginning of the world as we know it*. John Wiley and Sons Ltd., West Sussex, England, 2005.

[137] Wall JV. Practical statistics for astronomers II. Correlation data modelling and sample comparison. *Q J R astr soc*, 37:519–563, Winter 1996.

[138] Watkins T, Betts A, Dobney K, Nesbitt M. *Qermez Dere; Tel Afar: Interim Report 3*. Department of Archaeology; University of Edinburgh, 1995.

[139] Weeks L. The Neolithisation of Fars, Iran. In Matthews R, Fazeli NH, editor, *The Neolithisation of Iran, The formation of new societies*, pages 97–107. Oxbow Books, Oxford, UK, 2013.

[140] Whittle A. *Europe in the Neolithic: The Creation of New Worlds*. Cambridge World Archaeology. Cambridge University Press, United Kingdom, 1996.

[141] Willmott CJ, Matsuura K. with support from NASAs Seasonal to Interannual ESIP, 2000.

[142] Wolfram Demonstrations Project. 2D Cellular Automaton Animations. http://demonstrations.wolfram.com/2DCellularAutomatonAnimations/.

[143] Yang Y, Atkinson PM. Computers in Urban Planning and Urban Management (CUPUM). In Batty S, editor, *An integrated ABM and GIS model of infectious disease transmission*. London, 2005.

[144] Zilhao J. Radiocarbon evidence for maritime pioneer colonization and the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci USA*, 98:14181–14185, 2001.

[145] Zvelebil M. Demography and dispersal of early farming populations at the Mesolithic-Neolithic transition: linguistic and genetic implications. In Bellwood P, Renfrew C, editor, *Examining the Farming/Language Dispersal Hypothesis*. McDonald Institute for Archaeological Research, Cambridge, 2002.