Visualising the Neolithic transition in Europe

Thembi Russell
Department of Archaeology
University of Southampton
Southampton SO17 1BJ
England, UK
E-mail: mtr@soton.ac.uk

James Steele
Department of Archaeology
University of Southampton
Southampton SO17 1BJ
England, UK
E-mail: tjms@soton.ac.uk

In this short paper we will discuss methods of visualization and analysis of the radiocarbon record of a large-scale diffusion process. We will use the spread of the European Neolithic as a case study.

Since the 1960s, when Grahame Clark published a map of the spatial distribution of early Neolithic radiocarbon dates (Clark 1965; Figure 1), many people have analysed and interpreted this dimension of the archaeological record using map visualizations and statistical techniques. There is still substantial disagreement in the literature regarding mechanisms of diffusion of the cultural elements of early agricultural strategies (adoption-diffusion, demic diffusion with interbreeding, or population replacement). Perhaps these debates can never be resolved simply by analysing the chronological record of this diffusion process. Nevertheless, the quality and quantity of radiocarbon data available continues to make this case study an ideal test bed for techniques of data mapping and of the statistical characterization of spatial patterns.

In a recent project funded by an AHRB grant to Stephen Shennan and James Steele, a new radiocarbon database of the late Mesolithic and early Neolithic record in Europe was compiled (Gkiasta *et al.*, in press). The database contains fields identifying each site, its location and cultural affiliations, and details of published radiocarbon dates relating to late Mesolithic or early Neolithic stratigraphic units. The database is archived on the Web with the AHDS data service in York, England (http://www.ads.ahds.ac.uk). In the course of its compilation it became clear that associating dates with specific cultural elements (pottery, individual domesticates, etc.) would require extensive consultation of primary site publications and archives, since the consolidated date lists and Dating Lab records that were being used usually only gave a broad cultural affiliation. Nevertheless, it still seemed worthwhile to analyse these new data as a meaningful record of a cultural process.

In an initial attempt to visualise the new database, we replicated Clark's map in which early Neolithic sites were assigned to 1200 year 'bins' on the basis of their uncalibrated modal radiocarbon age (Figure 2). In comparison with Clark's dataset (containing 53 dates), it is clear that ours (using 508 dates) is restricted geographically to Europe and excludes sites in Turkey, the Fertile Crescent and North Africa. The spatial distribution of sites in the 4000-5200 bp range is greatly expanded, and is no longer restricted to the central European loess belt. In other respects, however, the big picture remains substantially similar to that mapped by Clark 35 years ago.

In the early 1970s, Ammerman and Cavalli-Sforza (1971) derived some rate estimates for the diffusion of agriculture (seen as a travelling wave) using regression techniques (Figure 3). Time was measured using the modal uncalibrated radiocarbon ages of 54 early Neolithic sites, and space was reduced to the single dimension of geodesic distance from Jericho (assumed to be the location of origin of the diffusion).

Repeating this major axis regression analysis with our larger early Neolithic dataset (508 uncalibrated radiocarbon dates) gives substantially similar results (Figure 4). Ammerman & Cavalli-Sforza (1971) found an average diffusion rate from an assumed origin in Jericho of about 1 km/year, and found a high value for r (0.89) in their sample of 53 Neolithic sites - suggesting that this rate was quite representative of the process generally. A similar analysis using sites indexed in the new database yielding the major axis equation:

Y (km) = 8240 - 0.89 X (yrs bp).

This suggests that the overall rate of spread is ~1.1 km/year and that the mean departure time from Jericho was ~9,250 bp (uncal.). In this case linear regression of the two variables produces a correlation coefficient, $r = 0.74$. In other words, with the larger data set now available the mean rate of spread is similar to that observed by Ammerman & Cavalli-Sforza, although the dispersion around that rate is somewhat greater (Gkiasta *et al.*, in press). We can note, at this point, one major problem with such an approach: the need to specify a single origin for the diffusion process.

Ammerman and Cavalli-Sforza (1984) subsequently mapped the diffusion in two dimensions, using spatial interpolation techniques and a larger data set of 106 early Neolithic sites (Figure 5). This technique does not require us to specify an origin to the diffusion, which can now be inferred from the orientation of the gradient in site age, apparent in the isochrons that are fitted at 500 year intervals.

We also repeated this analysis, using the interpolation (r.surf.idw, interpolation using the 12 nearest neighbours) and contour fitting (r.contour) commands in the GRASS GIS package (Figure 6). Our own visualization indicates a less simple spatial structure than that depicted by Ammerman and Cavalli-Sforza. It is likely that this is partly because in each case, all available sites were used (as opposed to those which were the earliest in each quadrat of a sampling grid) – and we have a much larger database (once again using 508 uncalibrated radiocarbon years). But we must also recognize the methodological limitations of this technique. Interpolation error will be large in regions with sparse or no sites to use as control points, and the interpolation has been continuous across sea as well as land (although the sea has subsequently been masked off). This technique therefore risks giving a misleadingly coherent impression of the spatial structure of any such dataset. As a control we repeated this exercise using our method (interpolation in GRASS GIS) with Ammerman and Cavalli-Sforza's original dataset of 106 Neolithic dates as the input data. Figure 5 (Ammerman and Cavalli-Sforza's isochron map) and Figure 7 (created using our method and Ammerman and Cavalli-Sforza's original dataset) show very similar results.

We have been experimenting with other map visualizations and statistical analyses (see also Glass *et al.* 1998), and here we report for the first time on two of them. First, we have experimented with geographically-weighted regression (Fotheringham, Brunsdon and Charlton 2000, 2001). This technique allows us to detect local variation in trends in large-scale spatial datasets, by weighting each control datum in a regression analysis inversely to its distance from a specified point location. In Figure 8 we have plotted results of a locally-weighted linear trend analysis for the ages of early Neolithic sites, for each of a regular grid of points. We have used projected co-ordinates, and we have only used the earliest sites in each 60km–by-60km quadrat of a sampling grid. The orientations of the arrows show the direction in which sites get younger. The lengths of the arrows scale to the rate of spread. The colours of the arrows indicate the strength of the locally-weighted linear trend (where red is a well-fitting model, and blue is a poorly-fitting model). Obviously we don't believe that the diffusion of agriculture was taking place out in the North Atlantic! It is merely a convention that vector field diagrams should plot values on a regular grid in this way. But what this shows us is that we can only pick up significant regional trends in regions where our dataset is full, rather than sparse. This technique therefore combines the best elements of the previous two (as used by Ammerman and Cavalli-Sforza), but avoids their weaknesses. We can observe local variation in rate and direction of diffusion, and its statistical strength. We do not need to assume any single origin, and we can see the regions where the model fits well and the regions where it fits badly.

Geographically-weighted regression also allows us to vary our characterization of 'local' (Figure 9). By increasing the bandwidth, we expand the window over which trends are detected until, in the limit (here, bandwidth=2000, Figure 9d), the weighting has no discernible effect and we simply see the major linear trend of the dataset as a whole. Unsurprisingly, it indicates an origin somewhere in the Fertile Crescent.

All these techniques require us to treat radiocarbon dates as point values. However, after calibration, the irregular nature of such dates (seen as probability distributions) makes using the mode or the median rather problematic, as we can see from this example (Figure 10 - the date of a Neolithic site at Pfyn, Switzerland). Our second new visualization technique tries to take account of this problem. We have used a technique similar to that used in the 'ShowTime' animation (which depicts pollen records of the changing distributions of major plant taxa in late glacial and postglacial North America - http://www.ngdc.noaa.gov/paleo/softlib.html).

We have divided the calibrated probability distributions of all our Mesolithic (207 radiocarbon dates) and Neolithic site dates (508 radiocarbon dates) into 100-year 'bins', and plotted these on a series of maps (at 100-year intervals - see Figure 11 for examples). The size of each of the dots is scaled to the probability that the site was occupied in each 100-year interval. We have then stacked the maps up as an animation, with Mesolithic sites plotted in red and Neolithic sites plotted in blue. The animation is available on the Web, as a PowerPoint file (http://www.soton.ac.uk/~mtr/Europe.ppt). In some ways this animation tells the story much more immediately than does any of the preceding methods. However, it does this by addressing the part of our brains that can discern pattern in dynamic visual input, and if we are to work with explicit models then we still need to find ways of translating such perceptions of pattern into verbal form or into simple summary statistics.

Continent-scale models of the Neolithic transition have been less popular with many archaeologists in recent years, because they subsume regional variation and because they often confound observations of pattern with interpretations of process. Better interpretation of process will require not just subdivision of the radiocarbon dataset into more specific sets of cultural associations, but also predictive modelling of the patterns expected as a result of different mechanisms (population expansion, adoption-diffusion, cultural convergence). In this paper our emphasis has been solely on the observation of pattern. The two new techniques represent efforts to meet some of the criticisms of continent-scale modelling, by incorporating statistical uncertainty and regional variation into the visual characterization of a continent-wide diffusion process.

Acknowledgements

We gratefully acknowledge the assistance of Stephen Shennan, the project's lead P.I., and of Marina Gkiasta, the research assistant at UCL who compiled the database. This work has been funded by the AHRB and by the University of Southampton.

References

Ammerman, A.J. and Cavalli-Sforza, L.L. 1971. Measuring the rate of spread of early farming in Europe. *Man n.s.* 6: 674-688.

Ammerman, A.J. and Cavalli-Sforza, L.L. 1984. *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton: Princeton University Press.

Clark, J.G.D. 1965. Radiocarbon dating and the expansion of farming culture from the Near East over Europe. *Proceedings of the Prehistoric Society* 31: 57-73

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2000. *Quantitative Geography*. London: Sage

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2001. *Geographically Weighted Regression.* http://www.ncl.ac.uk/geography/GWR/

Gkiasta, M., Russell, T., Shennan, S. and Steele, J. In press. Origins of European agriculture: the radiocarbon record revisited. *Antiquity*

Glass, C., Steele, and Wheatley, D. 1999. Modelling spatial range expansion across a heterogeneous cost surface. In *Procs. CAA 97, Birmingham. BAR Int Series* 750: 67-72. Oxford: Archaeopress.

Figures

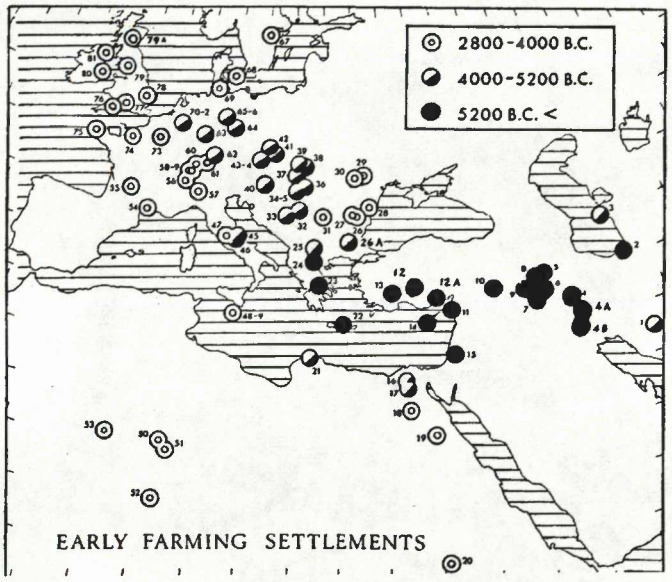


Figure 1. The distribution of early farming sites (Clarke 1965:46). Clarke uses 53 early Neolithic uncalibrated radiocarbon dates in years BC. Sites are grouped together into 1200 year interval bins.

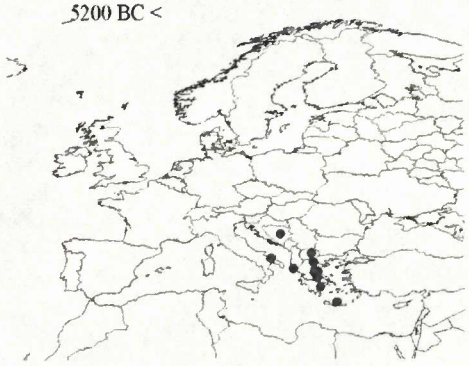
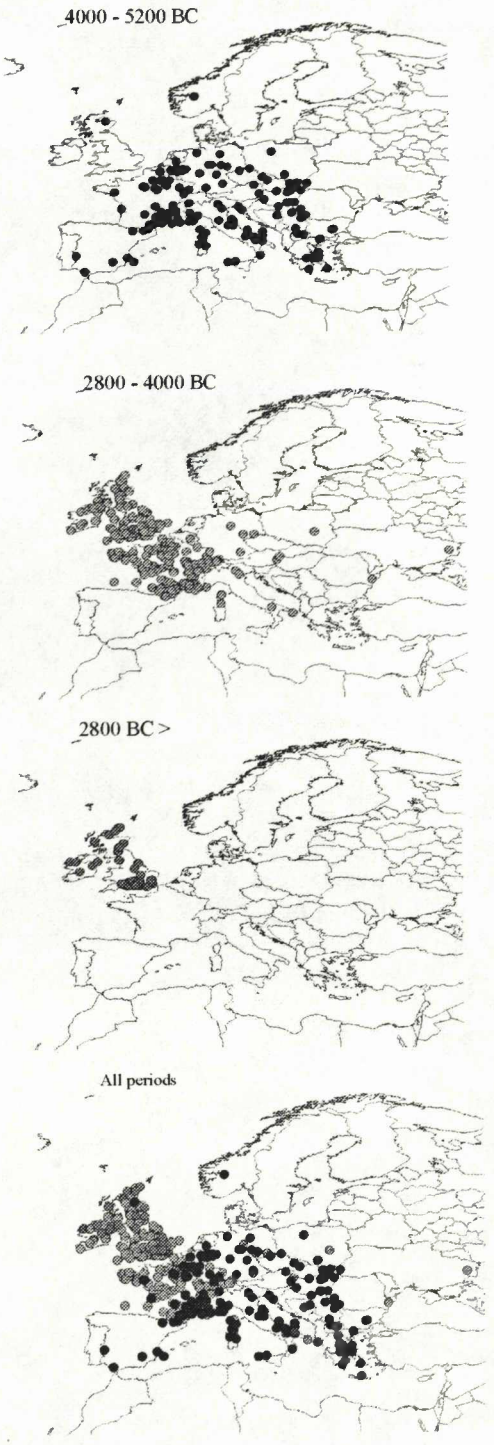


Figure 2: Clarke's method reproduced using 508 early Neolithic dates in uncalibrated radiocarbon years BC. ↑→



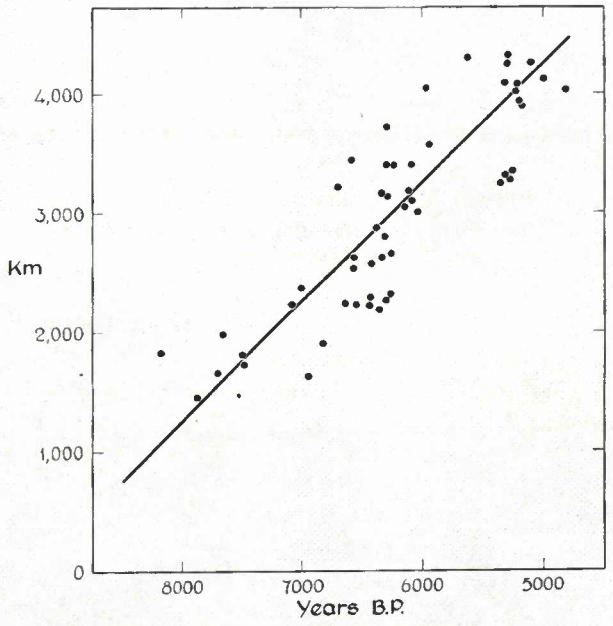


Figure 3. Regression analysis of the age of early farming sites with their distance from an origin in Jericho (Ammerman & Cavalli-Sforza 1984:53). This uses 53 uncalibrated radiocarbon dates.

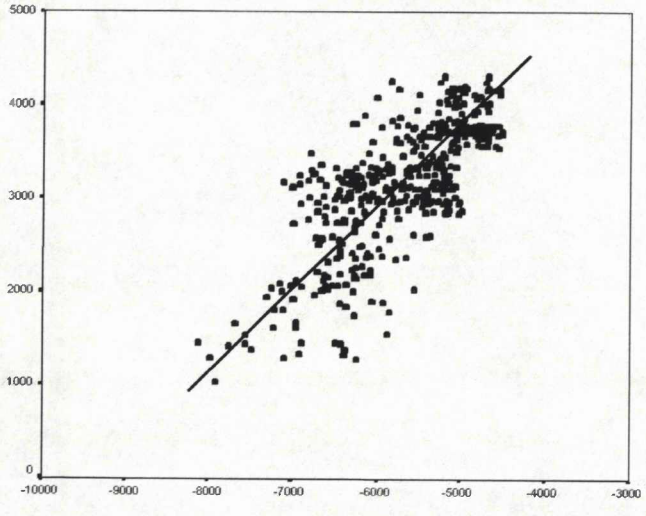


Figure 4. Regression analysis of geodesic distance from an assumed origin at Jericho (kms, y-axis) with the site age (years bp, x-axis). 508 uncalibrated radiocarbon dates (for early Neolithic sites) are used in the analysis.

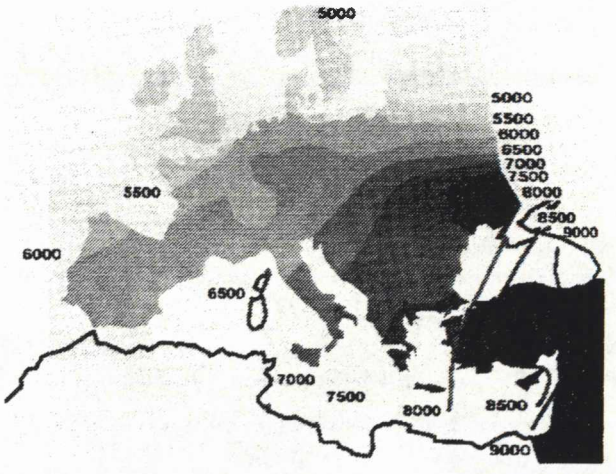


Figure 5. Isochron map of the spread of the first farmers to Europe (Ammerman & Cavalli-Sfroza 1984). 106 sites are used and dates are uncalibrated years BP. Isochrons are drawn at 500 year intervals.

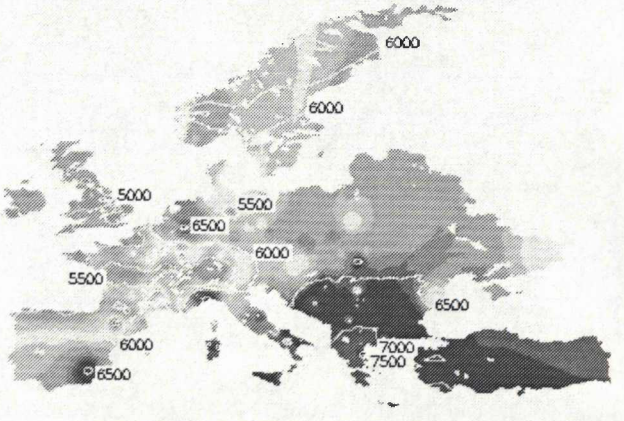


Figure 6. Isochron map for the distribution of early Neolithic sites using 508 radiocarbon dates. Dates are uncalibrated years BP. Isochrons are created at 500 year intervals.

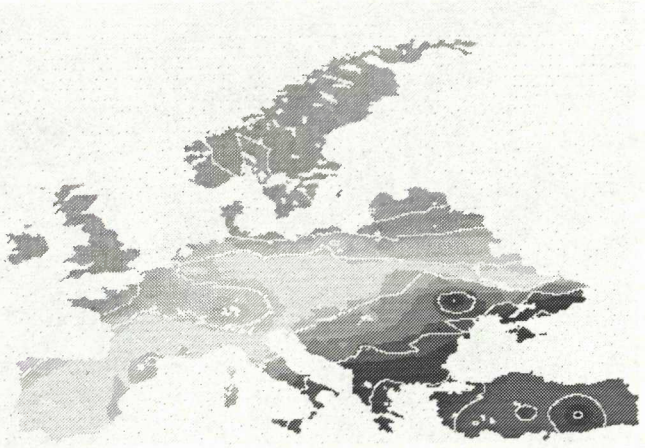


Figure 7. A control map. This isochron map is generated using our method of interpolation and contour fitting coupled with Ammerman and Cavalli-Sforza'a data (106 uncalibrated radiocarbon dates). The difference between this Figure and Figure 5 lies in the methods of interpolation and contour fitting, as the input data is the same in both cases.

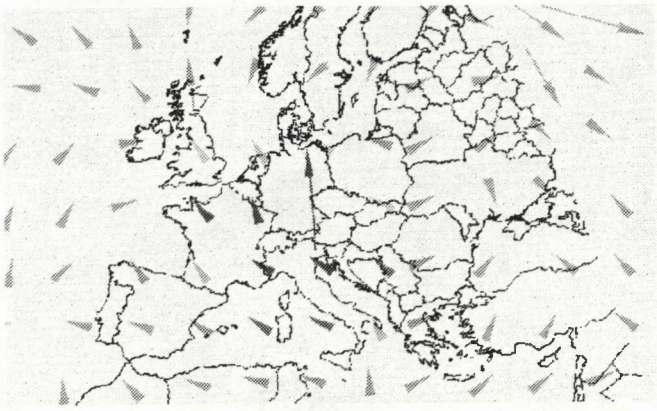


Figure 8. Results of a locally-weighted linear trend analysis for the ages of early Neolithic sites, for each of a regular grid of points. We have used projected co-ordinates, and we have only used the earliest sites in each 60km–by-60km quadrat of a sampling grid. The orientations of the arrows show the direction in which sites get younger. The lengths of the arrows scale to the rate of spread. The colours of the arrows indicate the strength of the locally-weighted linear trend (where red is a well-fitting model, and blue is a poorly-fitting model). Bandwidth = 400.

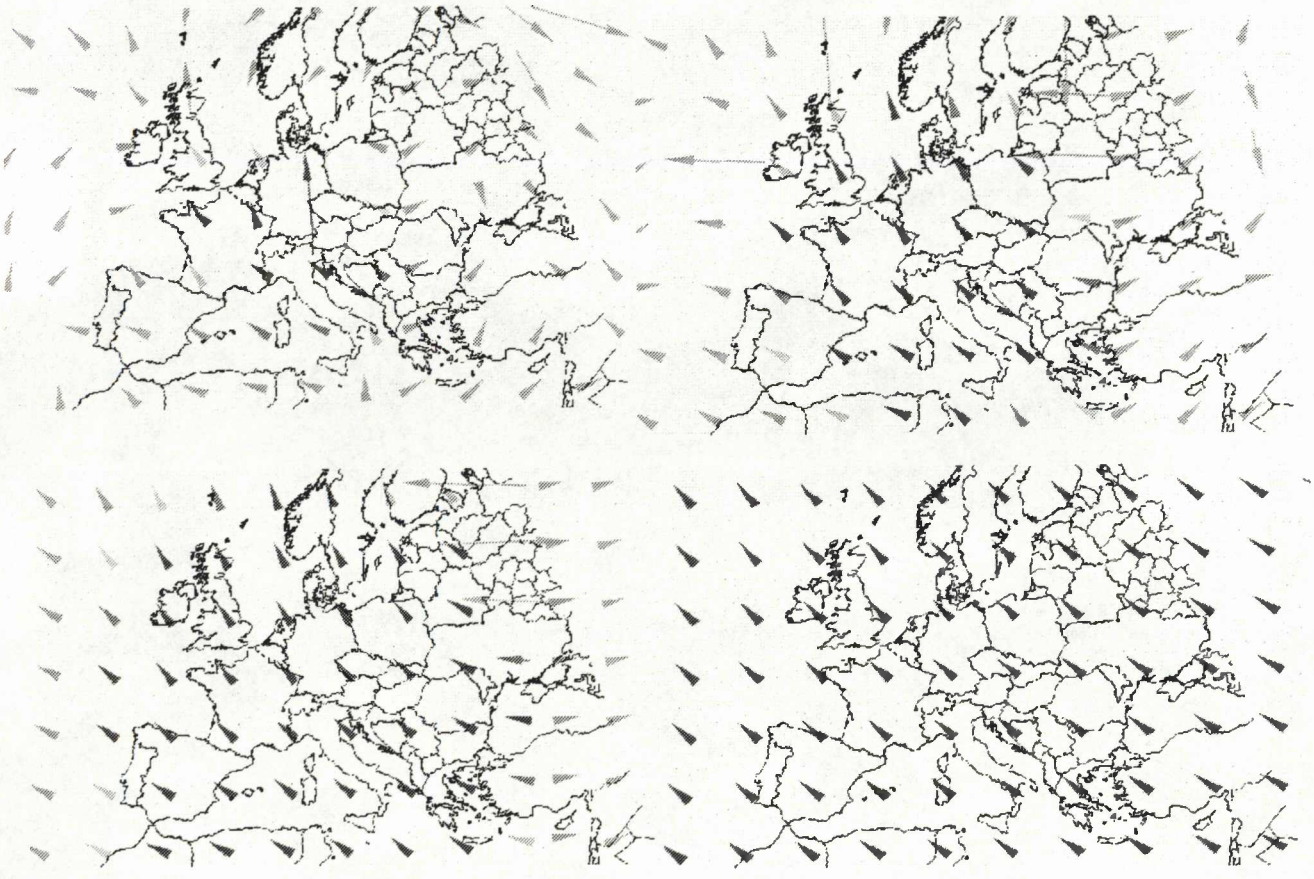
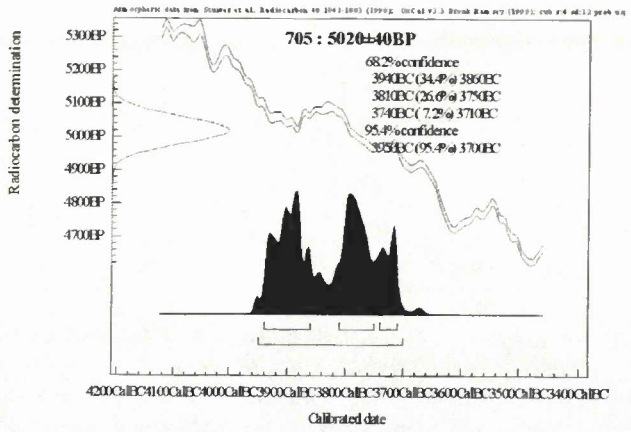


Figure 9. Results of a locally-weighted linear trend analysis for the ages of early Neolithic sites, for each of a regular grid of points. Key as for Figure 8. Bandwidth = 400, 650, 950, 2000.



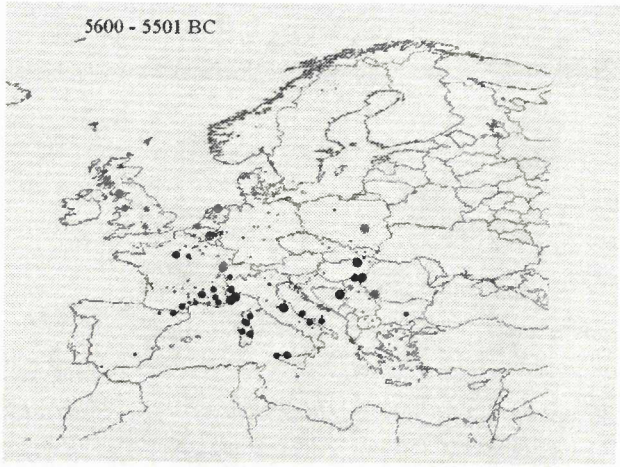
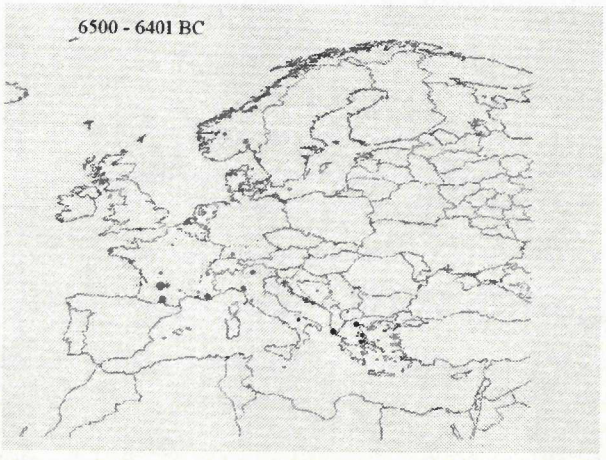
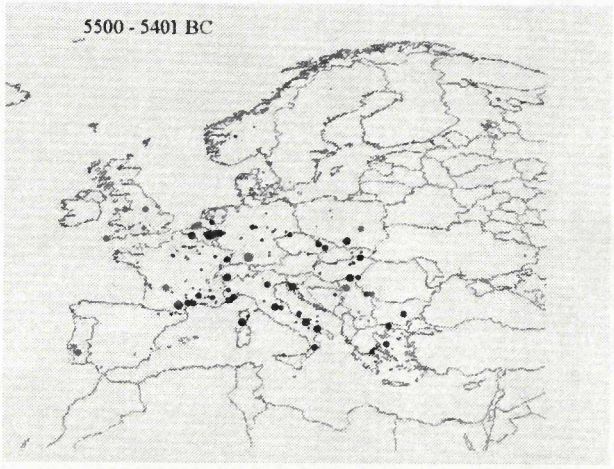
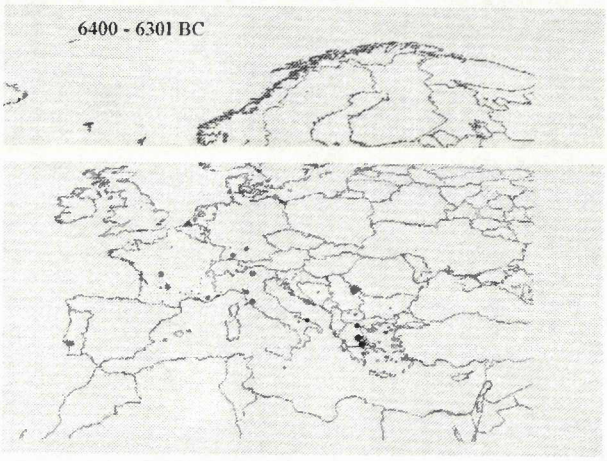


Figure 10: The calibrated probability curve for the Neolithic site of Pfyn in Switzerland to show the bimodal shape of the calibrated radiocarbon date.







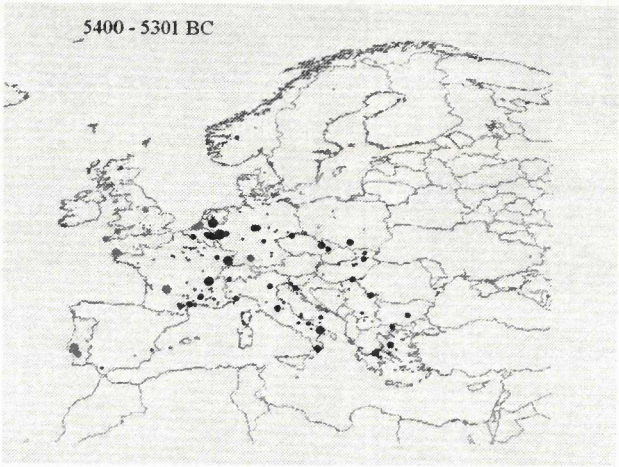


Figure 11. The distribution of early Neolithic sites. The fluctuation in the probability that the site is occupied at each 100 year interval (as calculated from the area under the calibrated probability curve of the radiocarbon date) is reflected in the size of the circular symbol marking a site's location. The probability that a site is occupied at a given time varies from less than 10 % (smallest symbol area) to greater than 80% (largest symbol area).