Journal of Archaeological Science 52 (2014) 609–620

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

New methods for reconstructing geographical effects on dispersal rates and routes from large-scale radiocarbon databases

Fabio Silva [a, b, *], James Steele [a, c]

[a] Institute of Archaeology, University College London, London WC1H 0PY, UK
[b] School of AHA, University of Wales Trinity Saint David, Lampeter SA48 7ED, UK
[c] SGAES, University of the Witwatersrand, South Africa

ARTICLE INFO

Article history:
Available online 22 May 2014

Keywords:
Radiocarbon
Front propagation
Human dispersal
Fast Marching methods
Phylogeography
Cultural Phylogenetics
Neolithic transition

ABSTRACT

We introduce a methodology for reconstructing geographical effects on dispersal and diffusion patterns, using georeferenced archaeological radiocarbon databases. Fast Marching methods for modelling front propagation enable geographical scenarios to be explored regarding barriers, corridors, and favoured and unfavoured habitat types. The use of genetic algorithms as optimal search tools also enables the derivation of new geographical scenarios, and is especially useful in high-dimensional parameter spaces that cannot be characterized exhaustively due to computer runtime constraints. Model selection is guided by goodness-of-fit statistics for observed and predicted radiocarbon dates.

We also introduce an important additional model output, namely, modelled phylogenies of the dispersing population or diffusing cultural entity, based on branching networks of shortest or 'least cost' paths. These 'dispersal trees' can be used as an additional tool to evaluate dispersal scenarios, based on their degree of congruence with phylogenies of the dispersing population reconstructed independently from other kinds of information.

We illustrate our approach with a case study, the spread of the Neolithic transition in Europe, using a database from the literature (Pinhasi, Fort and Amerman 2005). Our methods find support for a geographical model in which dispersal is limited by an altitudinal cut-off and in which there is a climate-related latitudinal gradient in rate of spread. This model leads to a deceleration in front propagation rate with geodesic distance, which is also consistent with models of the propagation of the Neolithic transition under space competition with pre-existing populations of hunter-gatherers. Our genetic algorithms meanwhile searched the parameter space and found support for an alternative model involving fast spread along the northern Mediterranean coast and the Danube/Rhine riverine corridor. Both these models outperformed the geography-free Great Circle distance model, and both also outperformed another, almost geography-free, model that constrains dispersal to land to and near-offshore coastal waters. The adjusted coefficient of determination for modelled and observed radiocarbon dates for first arrival supports the GA-derived model; the shortest path network analysis, however, gives greater support to the model with altitudinal cut-off and latitudinal gradient in dispersal rate, since it produces branching 'dispersal trees' that are more congruent with these archaeological sites' clade memberships (as defined by archaeological material culture).

1. Introduction

Prehistoric human dispersals, and the spread of major cultural innovations, can often be tracked using radiocarbon dates. Such dates, if accurate, give first observed arrival times at specific locations; by aggregating dates across sufficiently large distances in space and time, patterns can be discerned which indicate the rates and routes of population dispersal or innovation diffusion. Such patterns can then be interpreted in terms of the underlying cultural and demographic dynamics.

Typically, we can also expect that geographical features (barriers, corridors, favourable and unfavourable habitat types) would have affected the pattern of spread of prehistoric populations and/or of cultural innovations. However, models of the underlying dynamics have tended in the past to ignore such variation, and to have had their parameters estimated in relation to some averaged

* Corresponding author.
E-mail addresses: fabio.silva@ucl.ac.uk (F. Silva), j.steele@ucl.ac.uk (J. Steele).

http://dx.doi.org/10.1016/j.jas.2014.04.021

overall observed rate of spread throughout the whole of the studied geographical domain (e.g. Ammerman and Cavalli-Sforza, 1971). This is acceptable as a first approximation, if our aim is to understand at an abstract level the underlying processes that could have given rise to these kinds of archaeological patterns. Nevertheless, it is an inescapable fact that the earth's surface is highly inhomogeneous, and that these inhomogeneities are often of the same spatial scale as the geographical domains on which our studies are focused. Geographical features cannot, therefore, be treated simply as low-level noise and averaged away if we wish to reconstruct the dynamics of any specific dispersal or diffusion episode, even at a continental scale. Recent work has increasingly often recognized this fact, and analyses of large-scale radiocarbon datasets are increasingly focused on variation in rates of spread that may indicate the effects of corridors, barriers, and favourable versus unfavourable habitat types (e.g. Bocquet-Appel et al., 2009, 2012; Baggaley et al., 2012; Russell et al., 2014). Similarly, demographic models of population dispersal are increasingly often geographically explicit, and take such features into account when simulating the underlying dynamics of the spread process (Wirtz and Lemmen, 2003; Ackland et al., 2007; Patterson et al., 2010; Lemmen et al., 2011; Banks et al., 2013; cf. Isern and Fort, 2012; Isern Sardo et al., 2012; Fort et al., 2012). Analogous approaches are increasingly common in models of genetic diversification during and after a dispersal phase (e.g. Ray, 2005; Ray et al., 2005; Kidd and Ritchie, 2006).

Full integration of empirical analysis with forward modelling requires us to be able to estimate the effects of geographical features on spread rates using radiocarbon datasets of first observed arrival times, and to calculate actual spread rates (or front propagation rates, to use a more technical term) across different types of habitat and along different kinds of corridor. Our purpose in this paper is to outline some methodological innovations which should enable such tasks to be more easily undertaken, and which should enable analysts to achieve more reliable results. We summarise below a fast and numerically stable method of modelling front propagation across a geographically-realistic surface, closely related to the cost-surface routines familiar to GIS users; we explain ways of estimating probable effects of geography on dispersal or innovation diffusion, using a coefficient of correlation between dates and distances from an origin point; we discuss how to use automated search algorithms to find the optimal solution to the problem of estimating multiple geographical effects when that search is only weakly constrained, or not at all, by *a priori* hypotheses; and finally, we introduce a method of using independent cultural data to choose between possible geographical scenarios, when the radiocarbon data alone are insufficient to indicate which one is more likely to be correct. We then illustrate the application of these methods, using a well-known published dataset for a well-studied prehistoric case (the Neolithic transition in Europe).

2. Methodology

2.1. Regression estimation of front propagation rate

Regression techniques enable us to discern, and characterize, coherent spatial gradients (where they exist) in observed first arrival times for some archaeological entity, whether this is the appearance of people in an empty continent, of farming in a world of hunter-gatherers, or of some other cultural group or innovation. Such approaches have been used for over 40 years, with radiocarbon date as the time variable and distance from some origin point as the space variable in a bivariate regression analysis (e.g. Ammerman and Cavalli-Sforza, 1971; Russell, 2004; Hazelwood and Steele, 2004; Pinhasi et al., 2005). The correlation coefficient estimates the strength of the relationship (the coherence of the spatiotemporal pattern); the intercept gives the mean expected time of first observed appearance of the entity at the origin point; and the slope coefficient gives the mean rate of spread (the front propagation velocity).

Choice of correlation coefficient should be determined by the nature of the relationship between the two variables: where the relationship is linear and there are no or few strong outliers, Pearson's r may be most appropriate, but where the relationship is monotonic and nonlinear, and/or where there are strong outliers, Spearman's ρ will be the more robust estimator.

Choice of regression line-fitting technique should be determined by the distribution of error among the measured variables. Conventional (ordinary least squares, OLS) regression assumes that measurement error is concentrated in the dependent variable, while reduced major axis (RMA) regression assumes a symmetrical distribution of measurement error between both variables. As we have previously noted (Steele, 2010), simulations (Babu and Feigelson, 1992) have shown that reduced major axis regression, whose slope is the geometric mean of the two ordinary least-squares slopes, performs well in recovering the true functional relationship between two error-prone variables. In practice it is also common for archaeological modellers to estimate front speeds as within the range indicated by the two possible OLS slopes (i.e. with variable 1 and variable 2 each treated as the dependent). Either of these approaches is preferable to using only one of the two possible bivariate OLS model fits without adjusting the slope for error in the independent variable: Cantrell (2008) has used simulations to assess the ability of a single OLS regression to estimate a functional relationship between two variables where each contain error, and where the underlying relationship is unity (a slope of value 1), finding that OLS underestimated the true slope, with a systematic fractional error of underestimation of the order $1 - r$, where r is Pearson's correlation coefficient. An empirical illustration of the points at issue can be seen in Steele (2010, Fig. 6 b,d,f), where reducing the size of an already small archaeological dataset reduces the correlation coefficient and causes the two OLS lines to diverge, while the RMA fit recovers the same underlying slope.

We illustrate this point again in Fig. 1, which shows the effect of adding random noise sampled from a Rayleigh distribution to both variables when the underlying relationship is a perfect linear function. The Rayleigh distribution was chosen because it enables noise to be added in one direction only, so that ages and distances are always underestimated – reflecting an archaeological situation where limited field sampling yields earliest observed sites post-dating the true front passage time, and where distance estimates fail to take account of the finer-grained geographical structure of patches and barriers (cf. Perreault, 2011). As the correlation coefficient decreases due to increasing magnitude of the noise (that is, an increasing Rayleigh scale parameter), so the slopes estimated by each of the two possible OLS models (as it were, space on time and time on space) diverge from the true functional slope and from one another, while the RMA model continues to detect the true relationship.

If the measurement error in both variables were known, and we wished to analyse the remaining (systemic) equation error that is due to effects of some other variable, then (as a number of authors recommend, e.g. McArdle, 2003) we could simply correct the OLS slope to control for that measurement error. However, in the present case, there is unknown measurement error in both variables – the dates will sometimes be inaccurate and are certainly intrinsically imprecise, while the distances are only approximations of the true distances covered by the propagating front at any given date (since the datable archaeological record is both incomplete, and incompletely sampled in space). The fact that these measurement

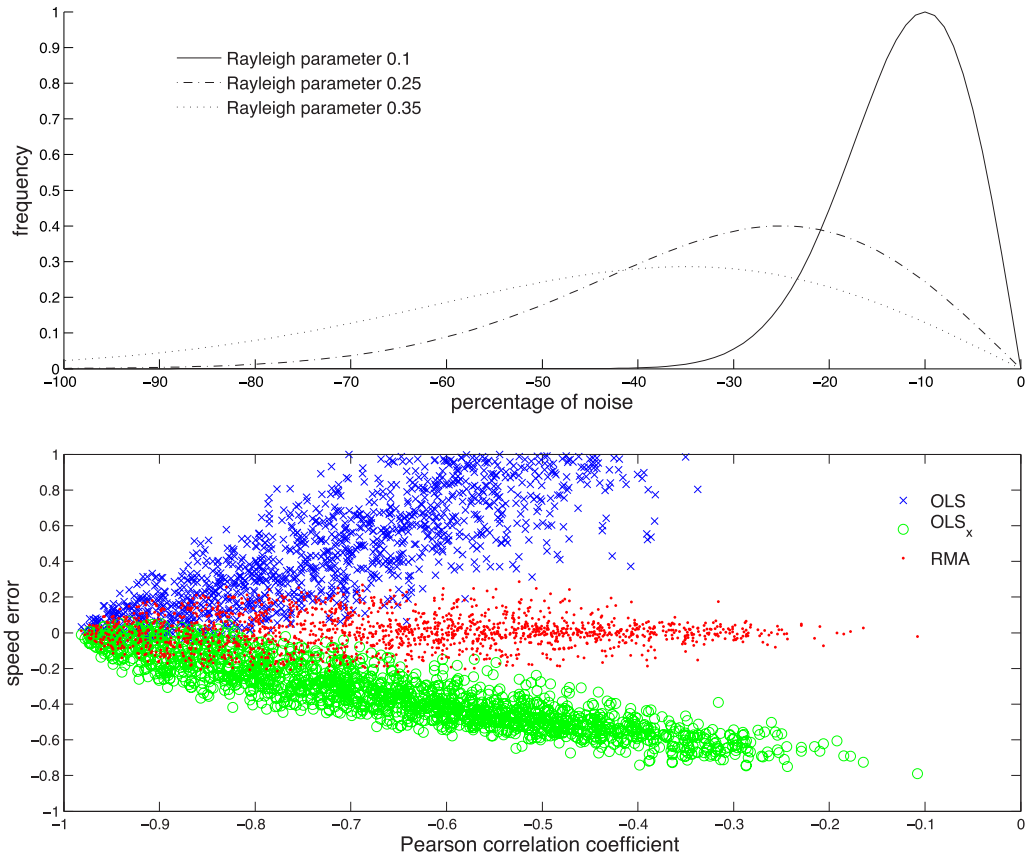

Fig. 1. (Top) Three illustrative Rayleigh distributions for three different scale parameters; (Bottom) Error in fitted speeds (inverse of regression slopes) for different regression methods (OLS and RMA) applied to a simulated dataset with an underlying perfect linear functional relationship, but with added random noise sampled from a Rayleigh distribution (with identical parameters for both variables). Speed error is given by the difference between the fitted value and the expected value (from the underlying rule).

errors are unknown makes it impossible to correct the OLS slope to derive the correct model, and RMA therefore gives a more conservative first approximation solution.

2.2. Distance estimation across heterogeneous surfaces using the Fast Marching algorithm

Calculating such bivariate regressions requires us to obtain values for the calendar date of first appearance of the archaeological entity at each site, and for the distance to each site from some origin location. A set of calendar dates can be obtained by taking the means or medians of the calibrated radiocarbon ages for each site, with the robustness of the obtained regression fit checkable by bootstrapping (e.g. Steele, 2010). In compiling such datasets there are usually also archaeological issues to be resolved to do with sample integrity and sample provenience, where the dated material acts only as a proxy for the true archaeological entity of interest; we do not consider these important issues further here, since our aim is to present a modelling methodology.

Sets of distances have often been calculated for such regression analyses, as a first approximation, as Great Circle distances (geodesic, in the original sense of that word). However, if we wish to explore the possibility that geography influenced directions and rates of spread from some origin point, then we must obtain sets of distances that reflect the influence of geographical features. Cost-surface techniques are most commonly used to obtain such distance estimates (e.g. Glass et al., 1999; Field et al., 2007). In a cost-surface analysis, a front is propagated across a grid of cells from some origin location, with the local rate of propagation in each direction determined by the friction factor assigned to the different cells in the neighbourhood. These friction factors will typically be assigned in a GIS raster layer according to some set of reclassification rules operating on an input layer containing geographical features. For example, to test a specific hypothesis about corridors and barriers, cells containing major rivers might be assigned a low friction value, while cells containing very arid habitat might be assigned a very high friction value, reflecting their different affordances (and thus invasibility) to a dispersing population. When the complete cost-surface has been calculated for the entire domain, values can be read off for the locations of each site, representing the length (in cost units) of the shortest path to that site from the origin location, for a given hypothesis about the frictions afforded by different geographical features. The correlation coefficient for distances (obtained in this way) with dates then becomes an estimator of the explanatory power of that particular geographical hypothesis, with the correlation coefficient obtained using only Great Circle distances providing the baseline (the geographical 'null hypothesis'). The intercept and slope coefficient of the regression model meanwhile enable the cost surface to be reclassified as an arrival time surface, while the slope coefficient and the grid cell resolution enable the rule set (the friction factors) to be translated into a set of modelled front propagation rates across each type of geographical feature in the domain.

In GIS applications, cumulative cost surfaces are typically calculated using Dijkstra's algorithm (Dijkstra, 1959), which solves the single-source shortest-path problem when all edges have non-

negative weights. This algorithm starts from the source and works outwards, starting at each iteration with the cell with the lowest cumulative cost value among the cells that the front has already reached and adding this to the values of the cells in its neighbourhood. The neighbourhood is typically defined on a regular grid or raster as including the cells that are reachable in a single step in a Rook's pattern (4-cell neighbourhood), a Queen's pattern (8-cell neighbourhood) or a Knight's pattern (16-cell neighbourhood). In our own work, we use a closely-related approach known as the 'Fast Marching' method (Sethian, 1996, 1999), implemented in MATLAB. Dijkstra's algorithm with a graph-based (i.e. grid cell neighbourhood) update is prone to introduce artefactual 'staircasing' into least cost paths. The Fast Marching method instead overcomes these constraints by replacing the graph update with a local resolution of the Eikonal equation (in our case, by a second-order finite-difference approximation; Silva and Steele, 2012, their Eq. 4). This produces a more accurate treatment of the underlying continuous spatial surface. Instead of the Dijkstra update algorithm, where D is the distance from cell j to the source of the dispersal evaluated according to the above iteration rule, and dx and dy are the distance from the nearest neighbour on the x and y axis respectively to the same source:

$$D(j) = \min(dx + W(j),\ dy + W(j));$$

We use an Eikonal update:

$$\Delta = 2{*}W(j) - (dx - dy)^2;$$

if $\Delta \geq 0$

$$D(j) = \left(dx + dy + \sqrt{\Delta}\right)\Big/2;$$

```
else
```

$$D(j) = \min(dx + W(j),\ dy + W(j))$$

where W is the neighbourhood metrical weight (Baerentzen, 2000; Sethian, 1999). The most time consuming aspect of Dijkstra-like approaches is the management of the list of cells for which cumulative costs have been computed. The Fast Marching method streamlines this process by singly focussing on the narrow band that encloses the propagating front, and using only known upwind values to estimate the cumulative costs (Sethian, 1999). We have introduced elsewhere a generalization of this approach to model multiple competing fronts with different origin locations, onset times and propagation rates, where each cell in a grid is populated by the descendants of one or other source population accord to a first arrival rule (Silva and Steele, 2012).

It is also common in Fast Marching implementations to express the metrical weight as a function of the local speed of the propagating front. This allows for a new generalization, introduced here to model dispersals over heterogeneous domains. For each individual cell, W_j can be multiplied by a friction factor, boosting or inhibiting the speed with which the front will propagate locally: a friction value of 0.5 on a given cell will mean the front propagates locally at half its base speed. Implementation of this generalization requires the construction of friction raster layers covering the computational domain. For current purposes these were obtained by reclassifying freely available present-day biogeographical distributions according to some rule set (see below), using GRASS GIS, and then exporting them to Matlab ready to be used by our Fast Marching algorithm.

The Fast Marching algorithm outputs a raster where each cell value is given by the shortest-path distance to the source of dispersal according to the constraints of the model (landcover, friction layers). This cumulative cost-surface is then queried for the cost-distance values of each site in the archaeological database, yielding the distances that are used for the regression analysis. As a derivative of the regression models which are then estimated from dates and cost distances, we obtain a predicted set of spatially heterogeneous, feature-specific local front propagation rates which can then be interpreted in terms of Fisher-KPP reaction-diffusion theory.

2.3. Parameter space search using genetic algorithms

We have now outlined a procedure for estimating sets of distances from some origin point to each site in an archaeological database, as a function of the intervening geographical features and their affordances to movement (of people in a dispersal case, or of ideas in a case of innovation diffusion). There may however be many such geographical features whose influence needs to be evaluated, and many different possible relative friction weights assignable to each such feature. The problem is one of optimization, i.e. of finding the set of parameter values that maximizes a fitness function: in this case the correlation coefficient. If we wish to obtain the parameter set that provides the best fit to the radiocarbon dataset independently of prior hypotheses in the literature, a comprehensive characterization of that parameter space may be computationally intractable (since for a parameter space with n independent geographical features, each with k possible friction weights, the number of possible combinations will be of the order k^n). Where fully exploring the parameter space is not an option, we need to deploy some kind of search heuristic.

We have therefore implemented a Genetic Algorithm (GA). GAs are optimization and search techniques, based on an analogy with the evolution of gene frequencies under natural selection. They mimic the natural processes of reproduction, including selection, mating with crossover, and mutation, in order to 'evolve' a best-fit parameter set out of a random population of parameters. GAs were developed originally by Holland (1975) and, particularly since the 1980s, have increased in popularity due to their usefulness for function optimization and other applications. They have since become standard techniques in several disciplines, including bioinformatics, computational science, mathematics and engineering (Haupt and Haupt, 2004). A typical GA run starts with a random population of models (i.e. a set of models with random values for the parameters – in our case, friction weights for the different features in the map layers) whose fitness is evaluated by some function (in our case, the coefficient of correlation between date and cost-distance). The best-fit models are then copied to the next generation unscathed (cloned), whereas less fit models are discarded. To keep the population size constant, the best-fit models are also allowed to reproduce. This involves the genetic principle of crossover, in which both parent models give only a part of their parameter set to the child model. Mutation can then occur on any model of the new generation, except for the very best one. This process is iterated until some convergence criterion is satisfied. Crossover and mutation are controlled by fixed rates and are essential to ensure that the GA does not become confined to a local maximum of the fitness function, but instead samples enough of the parameter space to locate a global maximum. After some generations the population begins to converge on the parameter set that maximizes the fitness function.

This application of GAs was developed specifically for the kinds of archaeological problems outlined here, and implemented in MATLAB. The GA parameters were as follows: population size was kept constant at 10; each generation kept 50% of the models from the previous one, and the other half was populated by heuristic crossover (Haupt and Haupt, 2004: 58); and a mutation rate of 20%

was used. The GA was allowed to run for 100 generations; convergence was then confirmed by checking the variation in the best-fit models over the later generations, and checking that the parameter space had been sufficiently sampled. We use Spearman's rank correlation coefficient (which is robust to nonlinearity and to strong outliers) to explore the parameter space. This correlation is independent of any subsequent regression model fitting. The best-fitting solution (or set of rules for reclassifying the geographical surface by friction), optimised using a correlation coefficient for dates and cost distances, can then be analysed using regression techniques and an interpretation developed in terms of the underlying ecological and behavioural dynamics.

However, we may subsequently wish to evaluate the fit of alternative regression models. Where we have multiple possible models (for example, those based on scenarios in the literature and those derived in an unconstrained or less constrained parameter space search using GAs), we can compare them formally using model selection tools such as adjusted R^2 or Akaike's Information Criterion (Akaike, 1974; Burnham and Anderson, 2002). These indices allow models with a different number of parameters to be compared on equal grounds. Here, where different geographical scenarios impose differing samples sizes, we use the adjusted R^2:

$$R^2_{\text{adj}} = 1 - \frac{n-1}{n-p} * \left(1 - R^2\right),$$

where n = sample size and p = number of parameters, and

$$R^2 = 1 - \frac{\Sigma(y-\hat{y})^2}{\Sigma(y-\bar{y})^2}$$

with y = observed value of 'dependent' variable, $\hat{y}$ = its fitted value and $\bar{y}$ = the mean observed value. Note that in our analyses we calculate the residuals perpendicular to the RMA line (cf. McArdle, 1988), and thus R^2 is not equivalent to the square of Pearson's r.

2.4. Shortest Path Trees

Having obtained a model of the effects of geographical features on the spread of some archaeologically-documented phenomenon, we may then wish to evaluate our results against other independent evidence. Where our modelling aims to reconstruct a pattern of population dispersal, it will often be appropriate to compare our results with other evidence of phylogenetic branching processes associated with the same episode. That evidence may come from historical linguistics, from genetics, or from archaeological studies of traditions of material culture (e.g. Rogers et al., 2009; Gray et al., 2010; Currie et al., 2013). If our best-fit dispersal model, conditioned as it is by archaeological radiocarbon dates, predicts a geographical branching pattern that is congruent with the clustering of variation into clades found independently in associated cultural or genetic histories, then this provides independent support for our model. Where our search has found multiple closely-comparable local optima in the parameter space, then the degree of congruence with this independent evidence may also indicate which solution is more likely to be the more empirically correct one.

To make such comparisons, we must obtain a phylogenetic summary of the model output. This can be done by representing the modelled dispersal pattern as a network of least-cost paths leading from the source location (or dispersal origin) to each point in space for which we have relevant independent evidence, such as cultural affiliation by archaeological pottery style. Points where such paths branch, can be considered as nodes on a tree. With a single source location, the network will necessarily have a branching tree topology; we therefore need to calculate accurately the shortest paths across the modelled cost surface, and then extract the tree in a format that will enable comparison with trees and/or lists of clade memberships obtained elsewhere from other forms of evidence.

In our work, the cumulative cost-surface is used to derive these shortest paths: the shortest-path from any point on the surface to the dispersal centre can be traced by analogy to water flowing downslope, from a point of high elevation (cost-distance) to low elevation (the source of dispersal). The closest GIS analogy is with algorithms that route downslope flows across a raster elevation map (such as GRASS's *r.drain*; GRASS Development Team, 2012). We calculate the paths using MATLAB algorithms, and extract and convert the set of such paths and their branch points into the required format again using a purpose-built MATLAB algorithm. Paths estimated in a raster grid, that begin at the terminus point (for example, an archaeological site) and incrementally derive a shortest path across a cumulative cost surface to the source location using only direct neighbours (Rook's or Queen's pattern, see above), are prone to strong discretization effects. Consider, for example, a plane uniformly inclined and sloping at an angle of 20 degrees to the y-axis of the grid. A path calculated from any point on the plane using only information from the Queen's pattern neighbours will run parallel to the y-axis until it reaches the edge of the domain, when it will turn and run along the domain boundary until it reaches the lowest corner. This is clearly not the optimal solution. In our own algorithm this is done using a Knight's pattern neighbourhood. The slope of all cells in this neighbourhood with respect to the central one is calculated, with appropriate metric correction. The next cell of the path will then be the one with the greatest downslope. To ensure the shortest path links together a continuous stream of raster cells, if a knight's move is selected as the next move, the nearest inner diagonal is also tagged as an intermediate step.

We turn now to an empirical illustration of the application of these methods.

3. Case study: the Pinhasi et al. (2005) dataset for the European Neolithic transition

To illlustrate these methods we now apply them to a published dataset previously used to estimate the rate of propagation of the European Neolithic transition (Pinhasi et al., 2005). Radiocarbon dates for the earliest Neolithic occupation from the earliest-dated levels of 765 sites in the Near East, Europe, and Arabia were collated from four pre-existing online databases (Pinhasi et al., 2005; Table S1). The collated database recorded single radiocarbon assays in each case, recording for each such site/phase the oldest date which had a standard error no greater than ±200 ^{14}C years, and which had not previously been flagged by archaeological consensus as anomalous. The vast majority of these dates are conventional radiocarbon measurements: Pinhasi et al. (2005) acknowledge that this might introduce extra uncertainty, but argue that the size of the sample and its large geographical coverage should still enable accurate discernment of the major trend. We have excluded 30 sites in Arabia, as this is outside the geographic domain of interest for our re-study, leaving 735 sites for analysis (Fig. 2; we should also note that Pinhasi, Fort and Ammerman's data table includes no value for the standard error of the radiocarbon measurement for 44 of the retained dates).

The dates were calibrated using OxCal ver.4.2b (Bronk Ramsey, 2009) and the INTCAL09 calibration curve (Reimer et al., 2009), and a value was obtained in each case for the mean calibrated age BP from an MCMC sample (cf. Steele, 2010). Pinhasi et al. (2005) had used CalPal (Weninger and Joris, 2004) and the CalPal 2004_Jan calibration curve; their and our results are essentially identical

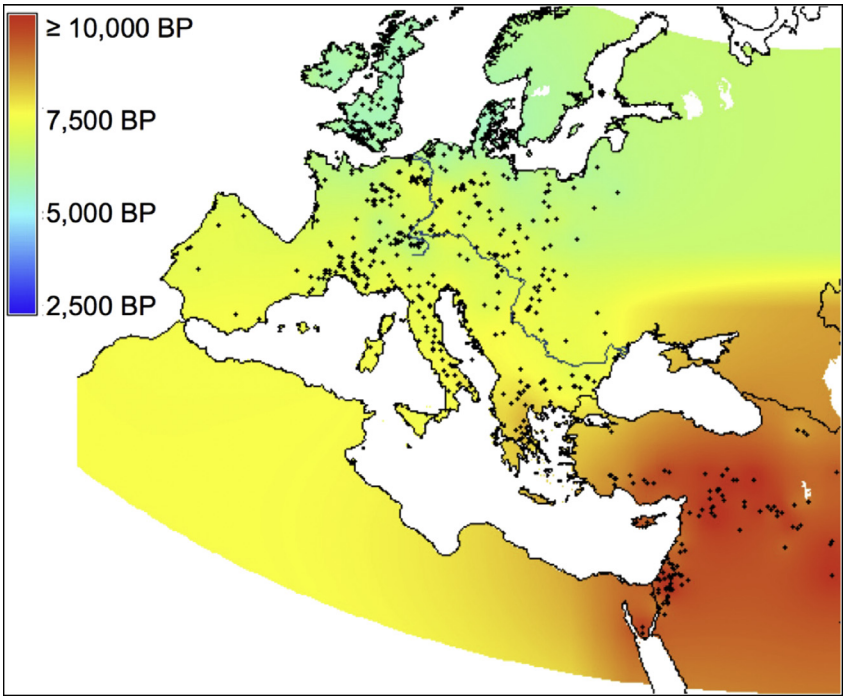

Fig. 2. Map showing the 735 sites in Pinhasi, Fort and Ammerman's (2005) dataset that were used in the present analysis, with their calibrated dates spatially interpolated in GRASS GIS.

(Pearson's $r = 0.9999$ for the correlation between their and our point values for the 735 calibrated ages; note that the INTCAL04 and INTCAL09 curves are identical for the relevant period).

Using a set of archaeological sites as possible dispersal origin points (note that this is a heuristic device to anchor the spatial modelling, and should not be taken literally), Pinhasi et al. (2005) compared the values for Pearson's correlation coefficient for calibrated age and distance in the above sample of 735 dated sites using Circle distances, and shortest path distances along land- and near-offshore based dispersal routes. Çayönü in southern Turkey was the archaeological site whose location as an origin gave the best fit for the terrain-dependent shortest path lengths ($r = -0.823$, an apparent improvement on their Great Circle distance fit for the same origin point of $r = -0.793$); using the relevant data from their Table S1, those values are confirmed and the respective values for Spearman's ρ are also obtained as $\rho = -0.775$ and $\rho = -0.751$.

The original authors had therefore already explored the possibility of explanatory gains from including basic geography. In Pinhasi, Fort and Ammerman's methodology (2005: Supplementary information), land- and near-offshore shortest path distances from Çayönü were estimated by rule of thumb. For 128 sites in the Near East, Anatolia and Asia, unadjusted Great Circle distances were used. For 594 sites in non-Iberian Europe, the paths are described as having been calculated as the sums of the Great Circle distances from Çayönü to a point on the Dardanelles Strait, and from that intermediate point to the site itself (but in a minor calculation error, a constant of 967.05 kms, the Great Circle distance from Çayönü to that point on the Dardanelles was instead added to the Great Circle distances from Çayönü to all 594 sites, see their Table S1; we have not checked or recalculated land-based distances from other possible centres of origin in that Table, since our intention here is to illustrate our new methods and not to exhaustively replicate this earlier study). For the 13 Iberian sites, an additional intermediate point to that at the Dardanelles was described as having been introduced on the Spanish/French border; however, the obtained differences from the Great Circle distances without any intermediate points are in the range 1.65–168 kms (their Table S1), which is a smaller adjustment than might be expected. We have therefore checked and confirmed the Great Circle distances from Çayönü using the set of site coordinates given by Pinhasi et al. (2005), and will use the associated correlation coefficient values henceforward for the baseline (no geography) model ($r = -0.793$; $\rho = -0.751$). Meanwhile we have also obtained the correlation coefficients for the geographically-constrained shortest path distances from Çayönü (but this time with those lengths calculated correctly using the above-described rule-of-thumb), obtaining values of $r = -0.800$ and $\rho = -0.770$; the value for r now shows no significant improvement in fit as a result of including that element of geography. These adjustments are minor in the context of Pinhasi, Fort and Ammerman's (2005) important main result, which was to confirm an overall average front propagation rate that was robust to factors such as whether or not the radiocarbon dates were uncalibrated or calibrated; but for our purposes here, where we wish to consider the extra explanatory power gained by allowing for geography, it is useful to set out clearly the results already obtained by other authors.

For our base maps (Fig. 3), we have calculated areas representing the Mediterranean and other coastal corridors; the Mediterranean, temperate forest and other present-potential biomes; the Danube-Rhine river corridor; and polygons defining inaccessible areas with a 1100 m altitudinal cut-off. These were obtained using public-domain GIS map layers (rivers from ESRI World Rivers shapefile, elevations from the ETOP05 Digital Elevation Model, and biomes from the Terrestrial Ecoregions of the World shapefile compiled by Olson et al. (2001) and projected into a Lambert Azimuthal Equal Area projection centred at 45 N, 45 E using GRASS GIS (GRASS Development Team, 2012).

To determine the optimal set of reclassification rules (friction factors), we analysed the radiocarbon dataset in relation to several

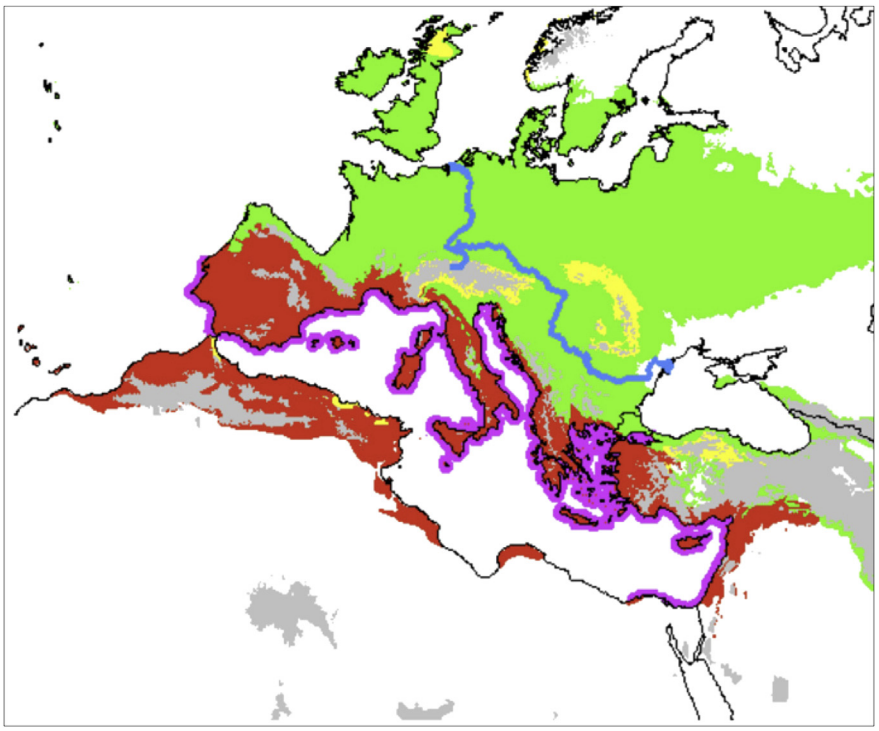

Fig. 3. Geographical features included in one or more of the models. The different Olson biomes are identified by colours: red for Mediterranean Forests, Woodland & Scrub; green for Temperate Broadleaf & Mixed Forests; and yellow for Temperate Conifer Forests. The Danube-Rhine corridor is marked in blue, whereas the northern Mediterranean coastal corridor is marked in purple. Greyed-out areas exceed the 1100 m above sea level altitudinal cut-off. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

existing hypotheses in the literature, as well as using an unconstrained GA search. In all models only land is colonisable, subject to a near-offshore coastal buffer of 45 kms which represents the bridging potential of maritime transport. Except for the baseline model, we used the Fast Marching method to derive shortest-path distances on a cost-surface given by the model's contraints. Although different centres might yield higher correlation coefficients for the different models, we are interested in exploring the effects of adding biogeography and therefore, for direct comparison, we have used Çayönü as the source of dispersal on all models.

Model 1 (the baseline model) uses Great Circle distances only (geography-free).

Model 2 uses shortest paths across land surface only, but with near-offshore bridges created by buffering out from the coastline.

Model 3 gives the Mediterranean coastal corridor a 10-fold acceleration factor and the Danube-Rhine corridor a 5-fold acceleration factor compared to the rest of the land surface, based on early empirical observations by Ammerman and Cavalli-Sforza (1971; cf. Davison et al., 2006: 642).

Model 4 has a uniform base speed which however decreases as a linear function of increasing latitude, from a base value at Mediterranean latitudes (40°N) to half of that by the latitude of Denmark (55°N), and an altitudinal cut-off of 1100 m above sea level, based on a scenario from Davison et al. (2006).

Model 5 is a combination of the rules for Model 3 and for Model 4, based again on a scenario from Davison et al. (2006).

Model 6 is the best-fit solution found in an unconstrained search by GAs of a parameter space in which the following corridors and biome types are all free to vary: Northern Mediterranean coasts; all other coasts in the modelled domain; the Danube-Rhine corridor; Mediterranean Forests, Woodland & Scrub; Temperate Broadleaf & Mixed Forests; Temperate Conifer Forests.

As an independent check on the fit of our models (as visualised using Shortest Path Trees), we have identified a subset of the dated sites in the Pinhasi et al. (2005) dataset that are associated in that database with pottery of the LBK and Cardial Ware traditions. Colouring the branches of the Shortest Path Trees by these cultural affiliations will enable us to visually determine which of our models best segregate these sites into their cultural 'clades'.

4. Results for the case study

Table 1 gives the results for the six models. Consistent with earlier results (Pinhasi et al., 2005), Models 1 and 2 (Great Circle and land-based distances, respectively) are able to account for up to 60% of the variation in dates in this dataset, based on the adjusted R^2 values. Pinhasi et al. (2005) had obtained confidence intervals of 0.6–1.0 and 0.7–1.1 km/yr, respectively, for front speeds in a Great Circle and a land-based model. Our fitted front speeds for these geography-free, or near-geography-free models are near to or slightly lower than the middle of their ranges (we obtain 0.72 and 0.74 km/yr for Models 1 and 2 respectively), reflecting the different choice of line-fitting technique, the different choice of dispersal origin for the Great Circle analysis and, for the shortest paths by land, an overestimation of distances from Çayönü to the majority of the sites in the original study (see preceding Section).

Among our more geography-rich models, Model 4 (which stipulates an altitudinal cut-off, and a linear decrease in front propagation rate with latitude) yields a linear association between date and cost distance, with a front speed in southern parts of the domain of 1.05 km/yr, decreasing to 0.525 km/yr in more northerly latitudes. This model accounts for 63.4% of the variation in dates, which is an improvement on Models 1 and 2.

In contrast to those models, we find that Models 3, 5 and 6 – all involving rapid dispersal along coastal and riverine corridors –

Table 1
Reclassification rule sets (friction weights), fitted regression models, geography-dependent front speeds, and goodness-of-fit statistics for the six models. Dates are treated as the *y* variable, and distances as the *x* variable (model 1 uses Great Circle distance, all others use cost distances).

Model	Parameters and relative friction weights	*N* sites	RMA equation	Front speed (s), km/yr	Pearson's r[a]	Spearman's ρ	R^2_{adj}
1	Great Circle distances	735	$y = 10{,}522 - 1.387x$	0.72	−0.793	−0.751	0.586
2	Land only	734	$y = 10{,}504 - 1.359x$	0.74	−0.801	−0.768	0.601
3	NMed. coasts 10: Danube/Rhine 5: rest of land 1	734	$\log10(y) = 4.874 - 0.351\ \log10(x)$	N/A	−0.745 (−0.708)	−0.786	0.488
4	Latitude gradient; altitude cutoff at 1100 m	688	$y = 10{,}062 - 0.950x$	1.05–0.53	−0.818	−0.815	0.634
5	Models 3 and 4 combined	688	$\log10(y) = 4.672 - 0.278\ \log10(x)$	N/A	−0.776 (−0.683)	−0.794	0.548
6	Best-fit GA[b]	734	$\log10(y) = 4.806 - 0.266\ \log10(x)$	N/A	−0.834 (−0.783)	−0.856	0.665

[a] (In brackets: Pearson's *r* for the untransformed variables, for Models 3, 5 & 6).
[b] Best-fit GA – relative friction weights: northern Mediterranean coasts 2.0: Danube/Rhine 0.5: Temperate Broadleaf & Mixed Forest, and Mediterranean Forest, Woodland & Scrub biomes 0.2: other coasts, and Temperate Coniferous Forest biome 0.1: rest of land 1.0.

yield nonlinear relationships between date and cost distance, with the front accelerating as a function of time. This nonlinearity is apparent from the scatterplots (Fig. 4), and also from the offset between Pearson's *r* and Spearman's ρ for the untransformed variables (ρ having the higher value). We have therefore computed the regressions on the log-transformed values of both variables for these three Models. In these cases, it is not possible to fit a fixed front propagation speed across each type of geographical feature, because the front propagation rate also has a time dependence. Nevertheless it is striking that Models 3 and 5, that impose coastal and riverine corridoring effects, do not have as much explanatory power as the geography-free, or the altitude and latitude-governed geography-dependent models. One of the possible reasons behind this is that the parameters of these models were taken from the existing literature and not allowed to vary in order to find the ones that provide the best fit to the data. This highlights the importance of using optimisation algorithms, such as the proposed GA methodology. Model 6, the best-fit model obtained by the GAs, has the highest value for adjusted R^2 of any model considered. It explains 66.5% of the variation in the dates, which is proportionally a 10% increase in explanatory power over the geography-free or near-geography-free Models 1 and 2. The relative friction weights recovered by the GAs for Model 6 suggest an important accelerating role for a northern Mediterranean coastal dispersal corridor, a significant but less marked accelerating role for the Danube/Rhine corridor, and a decelerating effect of forested biomes away from these corridor, most markedly in the higher-altitude temperate coniferous forest biome.

We turn now to independent checks on the accuracy of our models, using culture-group colourings of the Shortest Path Trees.

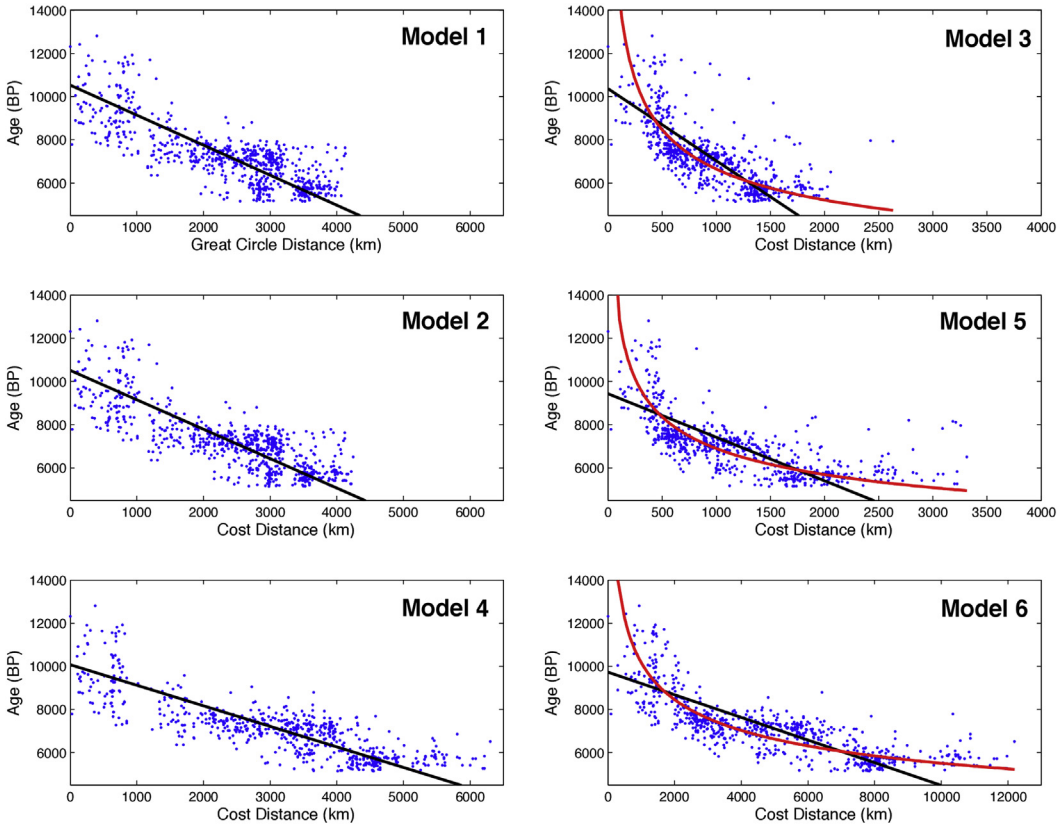

Fig. 4. Scatterplots of dates versus cost distances for the six models (model number in upper right corner of each graph), with their linear regression lines in black. The nonlinear relationship between the two variables in Models 3, 5 and 6 is apparent. Nonlinear regression lines, obtained by log transforming both variables, also plotted in red for these three cases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

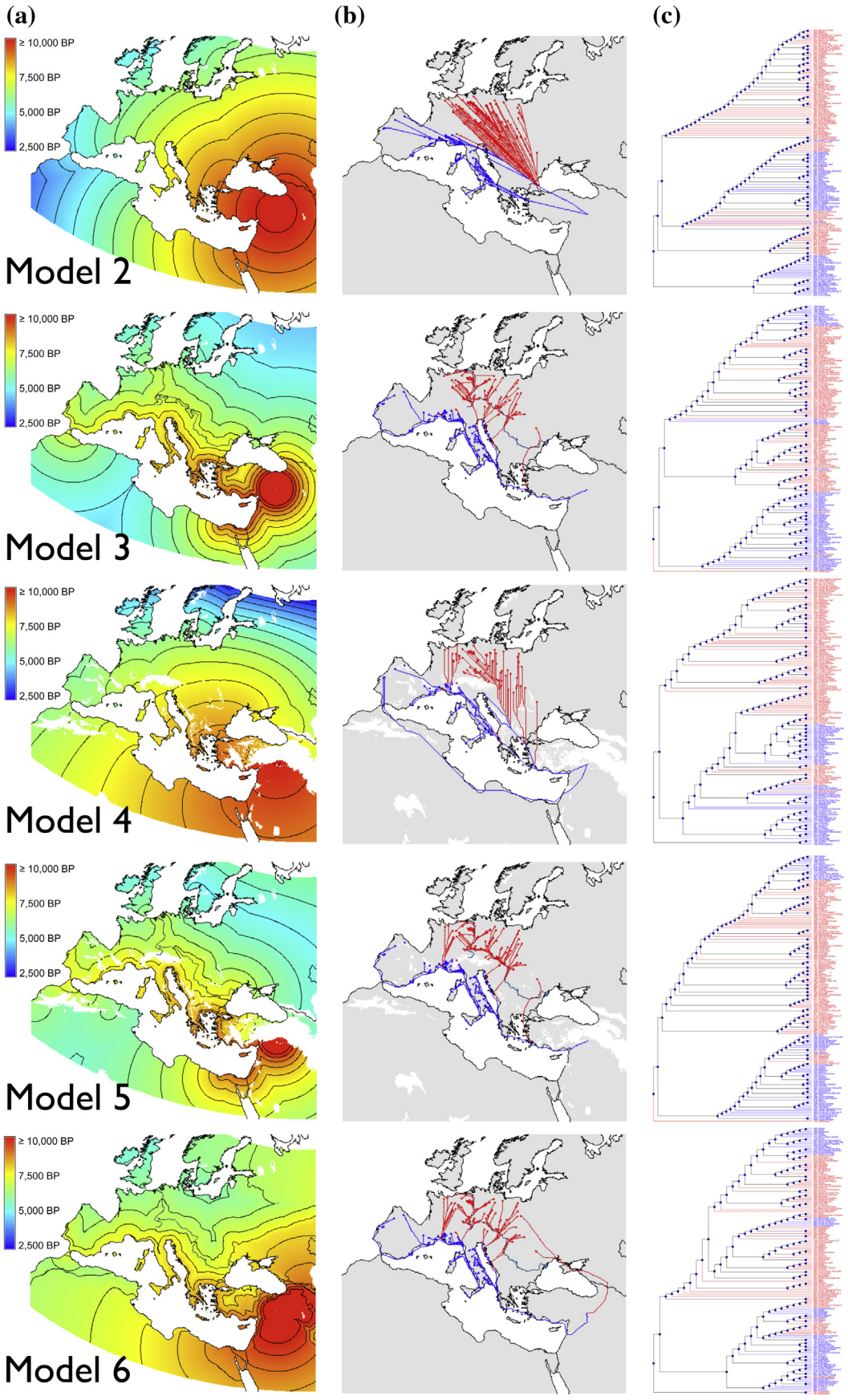

Fig. 5. Results for the five geography-based models (parameters as in Table 1; 'geography-free' Great Circle model not shown). Top to bottom: results successively for Model 2, 3, 4, 5 and 6. Left to right: (a) cost surface, reclassified as an arrival time surface using the results of the RMA regression (Table 1); (b) routing of the shortest path tree, for a subset of sites with cultural affiliations (red = LBK sites, blue = Cardial); (c) shortest path tree represented ultrametrically for the same sub-sample, with branch colours as for (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 shows for each of the geographically explicit models the modelled arrival times obtained by applying the regressed equations (Table 1 and Fig. 4) to the cost-distance surface of each model. Also shown for each model is the network of shortest path routes from Çayönü to those sites in the database that were labelled as belonging to the LBK or Cardial Ware ceramic traditions, and a chart with that shortest path network represented as a (coloured) ultrametric tree.

One can readily see, from the ultrametric trees, that Model 4 does very well in recovering a deep split between the paths leading to the majority of the LBK sites (coloured in red), and those leading to the Cardial Ware sites (coloured in blue). The main discrepancy is a group of LBK sites nested within the blue clade, which are located in northern France and Belgium and which Model 4 predicts would have been reached northwards from the Mediterranean coast of France—as opposed to westwards from the Early LBK zone of Central Europe. Model 6, obtained by GAs, has a higher value for the adjusted R^2; however, its shortest path network structure appears less congruent with the pottery-based clades. In addition to replicating the 'misplacement' of the northwestern branch of the LBK, Model 6 conflates 'blue' sites in the Eastern Adriatic within the main LBK clade because it routes the LBK-directed dispersal paths along that seaboard and then north to the middle Danube, whereas Model 4 routes the LBK-directed dispersal paths inland through eastern Europe towards the lower Danube corridor. Thus, independent evidence of branching patterns – here, in early Neolithic material cultural traditions – can provide an extra tool for model selection, suggesting in this case that while the GA-derived model may have greater power to predict overall arrival times, it is less successful than Model 4 at recovering the axes of major Neolithic dispersal pathways and their branching points inferred from material cultural traditions. It would be interesting in the future to test for model congruence with palaeoeconomy datasets (showing patterns of plant and animal exploitation), such as those already used to explore the effects of historical divergence and ecological convergence on interassemblage variability in early Neolithic Europe (Coward et al. 2008, Manning et al. 2013).

5. Discussion

We have introduced a set of methods for reconstructing geographical effects on dispersal and diffusion patterns, using georeferenced archaeological radiocarbon databases. These methods enable geographical scenarios to be explored regarding barriers, corridors, and favoured and unfavoured habitat types. The use of genetic algorithms as optimal search tools also enables the derivation of new geographical scenarios, and is especially useful in high-dimensional parameter spaces that cannot be characterized

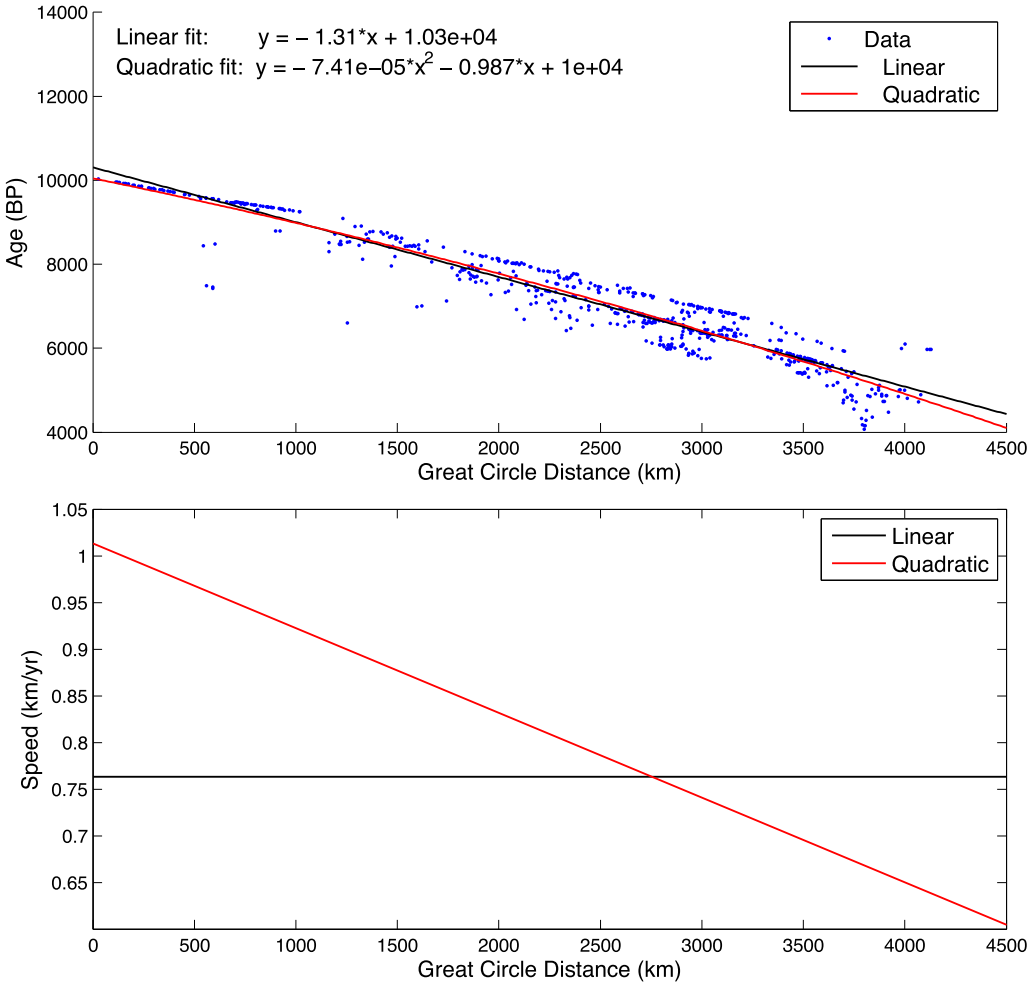

Fig. 6. Predicted arrival time plotted against geodesic distance from Cayonu for Model 4 (altitudinal cut-off, and latitude gradient in front speed). The curves added are for linear (black) and quadratic (red) fits, with the latter showing a linear decline in front speed with geodesic distance of the same order as the south-north gradient in front speeds modelled in the underlying Model 4 cost-distance analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exhaustively due to computer runtime constraints. Model selection is guided by goodness-of-fit statistics for observed and predicted radiocarbon dates, but we have also introduced an important additional model output, namely, modelled phylogenies of the dispersing population or diffusing cultural entity, based on branching networks of shortest or 'least cost' paths. These 'dispersal trees' can be used as an additional tool to evaluate dispersal scenarios, based on their degree of congruence with phylogenies of the dispersing population reconstructed independently from other kinds of information.

We have then illustrated our approach with a case study, the spread of the Neolithic transition in Europe. We have chosen this case study because it is well-known and well-studied, and because a suitable georeferenced radiocarbon dataset already exists in the peer-reviewed published literature. Our methods find support for a geographical model outlined by Davison et al. (2006), in which dispersal is limited by an altitudinal cut-off and in which there is a latitudinal gradient in rate of spread (due – they had suggested – to the limiting effects of the harsher northern climate). Interestingly, this model also predicts a deceleration in front propagation rate with geodesic distance (Fig. 6), which is consistent with models of front delay and deceleration proposed by Isern and Fort (2012; cf. Isern Sardó et al., 2012) where there is competition with pre-existing populations of hunter-gatherers.

Our GAs searched the parameter space and found support for an alternative model with fast spread along specific coastal and riverine corridors. This model required a nonlinear curve-fit for the effect of cost-distance on archaeological first arrival date; the structure of the modelled arrival time surface (Fig. 5, Model 6) suggests that this is due to an initially slow front propagation rate, which increases once the northern Mediterranean coastal corridor is reached and enables locally rapid long-distance dispersal.

Both these models outperformed the geography-free Great Circle distance model, and both also outperformed another, almost geography-free, model which constrains dispersal to land and near-offshore coastal waters. The coefficient of correlation for modelled distances and archaeological radiocarbon dates supports the GA-derived model; the Shortest Path Trees, however, give greater support to the model with altitudinal cut-off and latitudinal gradient in dispersal rate, since the latter produces branching 'dispersal trees' that are more congruent with the terminal nodes' (i.e. the archaeological sites') clade memberships as defined by material culture.

Further work on the database and its geographical analysis should enable some of these residual uncertainties to be resolved. We have meanwhile also applied the same methods elsewhere to a new dataset of georeferenced radiocarbon dates associated with the so-called Bantu language/farming dispersal episode in sub-Saharan Africa (Russell et al., 2014). Future work should also address demographic interpretations of these geographical effects and of the possible nonlinearities in front propagation rate.

Acknowledgements

We are very grateful to Bob Kelly for the invitation to participate in this highly interesting 2013 SAA symposium, and to two anonymous referees for their helpful comments on an earlier version.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jas.2014.04.021.

References

Ackland, G.J., Signitzer, M., Stratford, K., Cohen, M.H., 2007. Cultural hitchhiking on the wave of advance of beneficial technologies. Proc. Natl. Acad. Sci. 104 (21), 8714–8719.
Akaike, Hirotugu, 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19 (6), 716–723.
Ammerman, A.J., Cavalli-Sforza, L.L., 1971. Measuring the rate of spread of early farming in Europe. Man 6 (4), 674–688.
Babu, J.G., Feigelson, E.D., 1992. Analytical and Monte Carlo comparisons of six different linear least squares fits. Commun. Stat. Simul. Comput. 21 (2), 533–549.
Baerentzen, J.A., 2000. On the Implementation of Fast Marching Methods for 3D Lattices. Technical Report IMM-REP-2001-13, DTU.IMM. Online: http://www2.imm.dtu.dk/pubdb/views/publicationdetails.php?id=841.
Baggaley, A.W., Sarson, G.R., Shukurov, A., Boys, R.J., Golightly, A., 2012. Bayesian inference for a wave-front model of the neolithization of Europe. Phys. Rev. E 86 (1), 016105.
Banks, W.E., Antunes, N., Rigaud, S., d'Errico, F., 2013. Ecological constraints on the first prehistoric farmers in Europe. J. Archaeol. Sci. 40, 2746–2753.
Bocquet-Appel, J.P., Naji, S., Linden, M.V., Kozlowski, J.K., 2009. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of ^{14}C dates. J. Archaeol. Sci. 36 (3), 807–820.
Bocquet-Appel, J.P., Naji, S., Vander Linden, M., Kozlowski, J., 2012. Understanding the rates of expansion of the farming system in Europe. J. Archaeol. Sci. 39 (2), 531–546.
Bronk Ramsey, C., 2009. Bayesian analysis of radiocarbon dates. Radiocarbon 51 (1), 337–360.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: a Practical Information-theoretic Approach, second ed. Springer-Verlag.
Cantrell, C.A., 2008. Technical note: review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems. Atmos. Chem. Phys. 8, 5477–5487.
Coward, F., Shennan, S., Colledge, S., Conolly, J., Collard, M., 2008. The spread of Neolithic plant economies from the Near East to northwest Europe: a phylogenetic analysis. J. Archaeol. Sci. 35 (1), 42–56.
Currie, T.E., Meade, A., Guillon, M., Mace, R., 2013. Cultural phylogeography of the Bantu languages of sub-Saharan Africa. Proc. R. Soc. B Biol. Sci. 280 (1762).
Davison, K., Dolukhanov, P., Sarson, G.R., Shukurov, A., 2006. The role of waterways in the spread of the Neolithic. J. Archaeol. Sci. 33 (5), 641–652.
Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1 (1), 269–271.
Field, J.S., Petraglia, M.D., Lahr, M.M., 2007. The southern dispersal hypothesis and the South Asian archaeological record: examination of dispersal routes through GIS analysis. J. Anthropol. Archaeol. 26, 88–108.
Fort, J., Pujol, T., Vander Linden, M., 2012. Modelling the Neolithic transition in the Near East and Europe. Am. Antiq. 77 (2), 203–219.
Glass, C., Steele, J., Wheatley, D., 1999. Modelling spatial range expansion across a heterogeneous cost surface. In: Procs. CAA 97, Birmingham. BAR Int. Series 750ArcheoPress, Oxford, pp. 67–72.
GRASS Development Team, 2012. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.2. Open Source Geospatial Foundation. http://grass.osgeo.org.
Gray, R.D., Bryant, D., Greenhill, S.J., 2010. On the shape and fabric of human history. Philos. Trans. R. Soc. B Biol. Sci. 365 (1559), 3923–3933.
Haupt, R.L., Haupt, S.E., 2004. Practical Genetic Algorithms. John Wiley & Sons.
Hazelwood, L., Steele, J., 2004. Spatial dynamics of human dispersals: constraints on modelling and archaeological validation. J. Archaeol. Sci. 31 (6), 669–679.
Holland, J.H., 1975. Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. U Michigan Press.
Isern, N., Fort, J., 2012. Modelling the effect of mesolithic populations on the slowdown of the Neolithic transition. J. Archaeol. Sci. 39, 3671–3676.
Isern Sardó, N., Fort, J., Vander Linden, M., 2012. Space competition and time delays in human range expansions. Application to the Neolithic transition. PloS One 7 (12), 51106.
Kidd, D.M., Ritchie, M.G., 2006. Phylogeographic information systems: putting the geography into phylogeography. J. Biogeogr. 33, 1851–1865.
Lemmen, C., Gronenborn, D., Wirtz, K.W., 2011. A simulation of the Neolithic transition in Western Eurasia. J. Archaeol. Sci. 38, 3459–3470.
Manning, K., Downey, S.S., Colledge, S., Conolly, J., Stopp, B., Dobney, K., Shennan, S., 2013. The origins and spread of stock-keeping: the role of cultural and environmental influences on early Neolithic animal exploitation in Europe. Antiquity 87 (338), 1046–1059.
McArdle, B.H., 1988. The structural relationship: regression in biology. Can. J. Zool. 66, 2329–2339.
McArdle, B.H., 2003. Lines, models, and errors: regression in the field. Limnol. Oceanogr. 48 (3), 1363–1366.
Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on earth. BioScience 51, 933–938.
Patterson, M.A., Sarson, G.R., Sarson, H.C., Shukurov, A., 2010. Modelling the Neolithic transition in a heterogeneous environment. J. Archaeol. Sci. 37 (11), 2929–2937.
Perreault, C., 2011. The impact of site sample size on the reconstruction of culture histories. Am. Antiq. 76 (3), 547–572.

Pinhasi, R., Fort, J., Ammerman, A.J., 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biol. 3 (12), e410.
Ray, N., 2005. PATHMATRIX: a geographical information system tool to compute effective distances among samples. Mol. Ecol. Notes 5 (1), 177–180.
Ray, N., Currat, M., Berthier, P., Excoffier, L., 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. Genome Res. 15 (8), 1161–1167.
Reimer, P.J., Baillie, M.G.L., Bard, E., Weyhenmeyer, C.E., 2009. IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. Radiocarbon 51 (4), 1111–1150.
Rogers, D.S., Feldman, M.W., Ehrlich, P.R., 2009. Inferring population histories using cultural data. Proc. R. Soc. B Biol. Sci. 276 (1674), 3835–3843.
Russell, T.M., 2004. The Spatial Analysis of Radiocarbon Databases. In: BAR International Series S1294. ArcheoPress, Oxford.
Russell, T., Silva, F., Steele, J., 2014. Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. PLoS One 9 (1), e87854. http://dx.doi.org/10.1371/journal.pone.0087854.
Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts. Proc. Natl. Acad. Sci. 93 (4), 1591–1595.
Sethian, J.A., 1999. Level Set Methods and Fast Marching Methods, second ed. Cambridge University Press, Cambridge.
Silva, F., Steele, J., 2012. Modeling boundaries between converging fronts in prehistory. Adv. Complex Syst. 15 (01n02).
Steele, J., 2010. Radiocarbon dates as data: quantitative strategies for estimating colonization front speeds and event densities. J. Archaeol. Sci. 37 (8), 2017–2030.
Weninger, B., Joris, O., 2004. Glacial radiocarbon calibration. The CalPal program. In: Higham, T., Bronk Ramsey, C., Owen, C. (Eds.), Radiocarbon and Archaeology, Fourth International Symposium. Oxford, 2002, Oxford University School of Archaeology, Monograph 62, pp. 9–15.
Wirtz, K.W., Lemmen, C., 2003. A global dynamic model for the Neolithic transition. Clim. Change 59 (3), 333–367.