Modelling Patterns of Archaeological Material Culture using Demographic Simulation

Mirna Kovacevic

May 2017

Thesis submitted for the degree of
Doctor of Philosophy
to
Research Department of Genetics, Evolution and Environment
Centre for Mathematics & Physics in Life Sciences & Experimental Biology (CoMPLEX)
University College London (UCL)
Declaration of Ownership

I, Mirna Kovacevic, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

...........................................
Mirna Kovacevic
**Abstract**

This thesis aims to apply spatially explicit simulation modelling and Approximate Bayesian Computation (Tavare et al. 1997; Beaumont et al. 2002) to explore how human demography and cultural transmission processes have shaped geographic patterns of material culture, as evidenced in the archaeological record. It is an example of how theory developed in different fields can be used to address archaeological and anthropological problems, and formalise interpretations and conclusions drawn from results of standard statistical tests.

A high degree of structuring is seen in the spatial distribution of symbolic artefact types associated with the Aurignacian culture in Upper Palaeolithic Europe, particularly the degree of sharing of ornament types across archaeological sites, and multivariate analyses of these distributions have been interpreted as indicating ethno-linguistic differentiation (Vanhaeren and d’Errico 2006). The primary focus of this thesis is examining whether ethnic structuring is required to explain the spatial distribution of variation in this archaeological material culture data.

This multidisciplinary thesis draws on simulation modelling and statistical techniques in order to test whether the distribution of artefact types reported by Vanhaeren and d’Errico (2006) can be explained by a simple model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving cultural interactions.

To do this, spatially explicit cultural transmission simulation models that generate expectations of a range of spatial statistics describing the distribution of shared ornament types have been developed. These simulated spatial statistics are compared to those observed from archaeological data for Aurignacian Europe, using Approximate Bayesian Computation, in order to test and compare different hypotheses concerning group interaction dynamics for the period. The hypotheses examined include ones where material culture does or does not drive group interaction dynamics.

In addition to the main study, my contributions – which draw on the modelling approaches in the main study – to two collaborative projects are discussed. The first of these examines a hypothesis that Cytochrome P450 3A5 (CYP3A5) is important in salt retention adaptation. The second study is concerned with examining the rate of gene flow over time and space through analyses of craniometric data, with the aim of informing on the extent of past population migration activity.
Table of Contents

Declaration of Ownership 2
Abstract 3
Table of Contents 4
List of Figures 9
List of Tables 18
List of Abbreviations 20
Acknowledgments 21

Chapter 1: Introduction 22
Summary 22
1.1 Overview of Thesis 23
1.2 Archaeological Context 23
1.2.1 Vanhaeren and d’Errico (2006) Study 23
1.2.2 Upper Palaeolithic and Aurignacian 24
1.2.3 Definitions of Ethnicity 25
1.2.4 Ethnicity from Distributions of Material Culture 26
1.2.5 Explaining Distributions of Material Culture 28
1.3 Modelling in Archaeology 30
1.3.1 Studying Cultural Evolution 30
1.3.1.1 Culture as an Evolutionary System 30
1.3.1.2 Transmission Biases 31
1.3.1.3 Transmission Isolating Mechanisms 32
1.3.1.4 Cultural Drift 32
1.3.1.5 ‘Core’ and ‘Package’ Models 33
1.3.1.6 Cultural Isolation-by-Distance 33
1.3.1.7 The Axelrod (1997) Model of Cultural Dissemination: Cultural Interaction Processes 33
1.3.2 Computational Modelling in Archaeology 35
1.3.3 Bayesian Inference 37
1.3.4 Approximate Bayesian Computation (ABC) 44
1.3.4.1 Approximate Bayesian Computation (ABC) Algorithm 44
1.3.4.2 Summary Statistics Choice 45
1.3.4.3 Quantifying Similarity 46
1.3.4.4 Parameter Estimation 46
1.3.4.5 Model Comparison using Bayes Factors 48

1.4 In this thesis… 48
1.4.1 Aims of Study 48
1.4.2 Challenges of Adopted Approaches 48
1.4.3 Thesis Structure 49

Chapter 2: Methods 49
Summary 49

2.1 Simulation Modelling 50
2.1.1 Simulation World 50
2.1.1.1 Geographic Region 50
2.1.1.2 Length of Simulation 50
2.1.1.3 Carrying Capacities and Environmental Scenarios 53
2.1.2 Demographic Processes 53
2.1.2.1 Migratory Processes 53
2.1.2.2 Fission / Extinction Processes and Kernel Density Estimation (KDE) 54
2.1.3 Cultural Processes: Modelling Ethnic Diversity 55
2.1.3.1 The Axelrod (1997) Model of Cultural Dissemination: Defining Culture 55
2.1.3.2 Observed and Simulated Datasets 56
2.1.3.3 Mutation and Drift 57
2.1.3.4 Cultural Interactions 58
2.1.3.5 Collecting Simulated Data (Single Occupancy Sites) 60
2.1.4 The Models: Null Model (NM), Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD) 61
2.1.4.1 Measures of Cultural Similarity 62
2.1.5 Model Versions, Model Parameters and Prior Ranges 64

2.2 Statistical Methods 65
2.2.1 Summary Statistics 65
2.2.1.1 Shared Information (SI) 66
2.2.1.2 Mutual Information (MI) 67
2.2.1.3 Median Absolute Deviation (MAD) 68
2.2.1.4 Spatial Distribution of Sites (DR) 69
2.2.2 Model Comparison 71
2.2.2.1 Relative Marginal Likelihoods 72

Chapter 2: Methods
Summary

2.1 Simulation Modelling
2.1.1 Simulation World
2.1.1.1 Geographic Region
2.1.1.2 Length of Simulation
2.1.1.3 Carrying Capacities and Environmental Scenarios
2.1.2 Demographic Processes
2.1.2.1 Migratory Processes
2.1.2.2 Fission / Extinction Processes and Kernel Density Estimation (KDE)
2.1.3 Cultural Processes: Modelling Ethnic Diversity
2.1.3.1 The Axelrod (1997) Model of Cultural Dissemination: Defining Culture
2.1.3.2 Observed and Simulated Datasets
2.1.3.3 Mutation and Drift
2.1.3.4 Cultural Interactions
2.1.3.5 Collecting Simulated Data (Single Occupancy Sites)
2.1.4 The Models: Null Model (NM), Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD)
2.1.4.1 Measures of Cultural Similarity
2.1.5 Model Versions, Model Parameters and Prior Ranges

2.2 Statistical Methods
2.2.1 Summary Statistics
2.2.1.1 Shared Information (SI)
2.2.1.2 Mutual Information (MI)
2.2.1.3 Median Absolute Deviation (MAD)
2.2.1.4 Spatial Distribution of Sites (DR)
2.2.2 Model Comparison
2.2.2.1 Relative Marginal Likelihoods
Chapter 3: Modelling Patterns of Material Culture in Upper Palaeolithic Europe (I)

Summary

3.1 Results

3.1.1 Exploring Behaviour of Summary Statistics

3.1.1.1 Are the summary statistics correlated?

3.1.1.2 Relationship between Summary Statistics and Model Parameters

3.1.2 Model Comparisons

3.1.2.1 Comparing proposed null models: Null Model (NM) and Null Model with Cultural Interactions (NMCI)

3.1.2.2 Null Model with Cultural Interactions or Culture-Dependent Interactions Model: Which model best explains the observed data?

3.1.3 Parameter Estimation

3.1.4 Exploring Behaviour of Parameters

3.2 Discussion

Chapter 4: Modelling Patterns of Material Culture in Upper Palaeolithic Europe (II)

Summary

4.1 Changes to Simulation Modelling Methods described in Chapter 2

4.1.1 Carrying Capacities and Environmental Scenarios

4.1.1.1 Climate Data

4.1.1.2 The Miami Model: Relating Climate Variables to Net Primary Productivity

4.1.1.3 Calculating Net Primary Productivity (NPP) for the Region of Interest

4.1.2 Fission / Extinction Processes and Distance to Nearest Neighbour as Proxy for Population Density

4.1.2.1 Distance to Nearest Neighbour as Proxy for Population Density

4.1.2.2 Calculating Expected Distance to Nearest Neighbour: 2D Modelled Domain Characterised by Single NPP Value

4.1.2.3 Calculating Expected Distance to Nearest Neighbour: 2D Modelled Domain Characterised by Multiple NPP Values

4.1.2.4 Calculating Expected Distance to Nearest Neighbour: Modelled Domain Representing the Geographic Region of Interest
4.1.3 Migratory Processes and Topographic Roughness 124
  4.1.3.1 Relating Elevation Data and Topographic Roughness 124
  4.1.3.2 Scaling Migratory Processes by Topographic Ruggedness Index (TRI) 125
4.1.4 Collecting Simulated Data (Cumulative Sites) 127
4.1.5 The Models: Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD) 128
4.1.6 Model Versions, Model Parameters and Prior Ranges 128

4.2 Results 137
  4.2.1 Exploring the Behaviour of Summary Statistics 137
    4.2.1.1 Are the summary statistics correlated? 137
    4.2.1.2 Relationship between Summary Statistics and Model Parameters 140
  4.2.2 Model Comparisons 143
    4.2.2.1 Null Model with Cultural Interactions or Culture-Dependent Interactions Model: Which model best explains the observed data? 143
  4.2.3 Parameter Estimation 149
  4.2.4 Exploring Behaviour of Parameters 156

4.3 Discussion 158

Chapter 5: Other Projects 160

Summary 160

5.1 Correlating Cytochrome P450 3A5 Expression with Climate Variables 161
  5.1.1 Summary of Study 161
  5.1.2 My Contribution: Integrating Climate Data 162
    5.1.2.1 Climate Data: Smith and Gregory (2012) 162
    5.1.2.2 Quantifying Aridity 162
    5.1.2.3 Aridity Indices for Geographic Locations of Interest 163
  5.1.3 Summary of Relevant Results 164

5.2 Variation in Human Cranial Morphology 165
  5.2.1 Summary of Study 165
    5.2.1.1 Context 165
    5.2.1.2 Hypothesis 166
    5.2.1.3 Applications to Simulated and Observed Data 169
  5.2.2 My Contribution: The Simulation Model 169
    5.2.2.1 Simulation World 169
    5.2.2.2 Demographic and Evolutionary Processes 170
    5.2.2.3 Simulated Datasets 171
  5.2.3 Summary of Relevant Results 171
List of Figures

Chapter 2

Figure 2.1: European landmass represented as a demic grid at a resolution of 0.2° latitude and 0.2° longitude, totalling 38,656 demes (24,104 land demes and 14,552 sea demes). 50

Figure 2.2: Map of the simulation space, showing the demic distributions of carrying capacities for the environmental scenarios considered:
(A) uniform carrying capacity values (FLAT);
(B) normalised distribution of relative carrying capacity values based on the population density estimates for the Aurignacian period from Bocquet-Appel et al. (2005) (B-A);
(C) (Image and caption edited from Bocquet-Appel et al. (2005) and reproduced with permission from Elsevier) Estimate of the regional distribution of the metapopulation of hunter-gatherers during the Aurignacian period of the Upper Palaeolithic in Europe, superimposed on the IOS3 project maps. The boundaries (in black) of the accretion zones, with the corresponding numbers, account for roughly 90% of the distribution of the local population. 52

Figure 2.3: Map of the simulation space, showing the geographic locations (red dots) of archaeological sites in the observed data (Vanhaeren and d’Errico 2006), which are defined to be the archaeological sites in the model (i.e. locations where simulated data is deposited). 63

Figure 2.4: Flow diagram illustrating the order in which the demographic and cultural processes are performed during each simulation. 66

Figure 2.5: Histograms of the number of occurrences of ornament types (panel (A)) and number of distinct ornament types recovered from individual sites (panel (B)) for the observed data (Vanhaeren and d’Errico 2006). 76

Figure 2.6: Density plots of the DR statistic for the original data (black line) and a random permutation of the same (grey line). 76
Chapter 3

Figure 3.1: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic combinations using 404,808 successful simulations for the NMCI SW FLAT (top panel) and NMCI DIS FLAT (bottom panel) models.

Figure 3.2: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and parameter combinations using 404,808 successful simulations for the NMCI SW FLAT model.

Figure 3.3: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and parameter combinations using 404,808 successful simulations for the NMCI DIS FLAT model.

Figure 3.4: Relative marginal likelihood estimate (y-axis) of each version of Null Model (NM; dashed lines) and Null Model with Cultural Interactions (NMCI; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 480,000 simulations (i.e. 120,000 simulations for each of the 4 models compared).

Figure 3.5: Relative marginal likelihood estimate (y-axis) of each version of Null Model with Cultural Interactions (NMCI; dashed lines) and Culture-Dependent Interactions Model (CD; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 3,200,000 simulations (i.e. 400,000 simulations for each of the 8 models compared).

Figure 3.6: Ranked Euclidean distance values of the best 20,240 simulations (i.e. the closest 5% of 404,808 simulations) for each of the two models plotted in black, with the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.

Figure 3.7: Estimated posterior density distribution of the demographic and evolutionary parameters of interest for the NMCI SW FLAT model, calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations).

Figure 3.8: Estimated posterior density distribution of the demographic and evolutionary parameters of interest for the NMCI DIS FLAT model, calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations).

Figure 3.9: Distributions of the 12 summary statistic values in the 4,000 retained simulations.
simulations (i.e. closest ~1% of 404,808 simulations) for the NMCI SW FLAT model (black lines) and NMCI DIS FLAT model (grey lines).

Figure 3.10: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for the NMCI SW FLAT model.

Figure 3.11: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for the NMCI DIS FLAT model.

Chapter 4

Figure 4.1: Distance between each group and its nearest neighbour, $D(\text{NN})$, is given by the sum of the distances between the two groups and their respective boundaries (i.e. $r$).

Figure 4.2: Six nearest neighbours (shaded blue) of central group, with hexagonal circle packing.

Figure 4.3: Area shaded in blue is lost due to tessellation properties of circles.

Figure 4.4: Domain modelled with two regions characterised by different NPP values. The NPP value of region $A_2$ is twice as high as that of region $A_1$; this translates to the number of groups per unit area in $A_2$ being twice as high as that in $A_1$.

Figure 4.5: Expected distance to nearest neighbour at each location is calculated as the sum of the $r$ at that location and the average $r$ calculated over those locations that are between $r$ and $2r$ of that location (shaded grey area).

Figure 4.6: If a deme is within $r$ of a coastline or a geographic boundary, a proportion of the area specified by a circle of radius $r$ falls on sea demes or outside the geographic domain modelled, respectively (the shaded grey part of circle radius $r$ that is in the uninhabitable region). The circle with radius $r'$ defines a larger circle whose area within the geographic boundary gives the required area (i.e. the shaded grey part of the circle with radius $r'$ accounts for the area of the circle with radius $r$ that is outside of the boundary).

Figure 4.7: In cases where a deme is within $2r$ of a geographic boundary or coastline, there is not enough available area to accommodate six immediately neighbouring groups.
Figure 4.8: Flow diagram illustrating the order in which the demographic and cultural processes are performed during each simulation. The processes grouped by the dashed grey line are iterated at every generation. The parameter that governs each process is indicated in the box specifying that process. The population density estimation process, shaded in grey, is not parameterised.

Figure 4.9: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic combinations using 1,132,411 successful simulations for the NMCI SW FLAT (top panel) and NMCI SW NPPSingarayer (bottom panel) models.

Figure 4.10: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and model parameter combinations using 1,132,411 successful simulations for the NMCI SW FLAT model.

Figure 4.11: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and model parameter combinations using 1,132,411 successful simulations for the NMCI SW NPPSingarayer model.

Figure 4.12: Relative marginal likelihood estimate (y-axis) of each version of Null Model with Cultural Interactions (NMCI; dashed lines) and Culture-Dependent Interactions Model (CD; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 12,000,000 simulations (i.e. 1,000,000 simulations for each of the 12 models compared).

Figure 4.13: Ranked Euclidean distance values of the best 56,621 simulations (i.e. the closest 5% of 1,132,411 simulations) for each of the two models plotted in black, with the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.

Figure 4.14: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the two NMCI models, calculated on the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW FLAT model.

Figure 4.15: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the two NMCI models, calculated on the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW NPPSingarayer model.

Figure 4.16: Distributions of the 12 summary statistic values in the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW FLAT model.
(black lines) and NMCI SW NPPSingerayer model (grey lines).

Figure 4.17: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW FLAT model.

Figure 4.18: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW NPPSingerayer model.

Chapter 5

Figure 5.1: Example plot of time-space Euclidean distance matrix and a morphological distance matrix coefficient ($y$-axis) against the corresponding value of the scaling factor $s$ ($x$-axis).

Figure 5.2: Correlation plots between simulated average migration distance ($x$-axis) and the value inferred by the estimator ($y$-axis) for different values of the rate of mutation parameter $p_{mut}$ (rows) and different rates of fission / extinction parameter $p_{f/e}$ (columns) for simulations of 4,000 generations that produced extra correlation (EC>0).

Figure 5.3: Correlation plots between simulated average migration distance ($x$-axis) and the value inferred by the estimator ($y$-axis) for different values of the rate of mutation parameter $p_{mut}$ (rows) and different rates of fission / extinction parameter $p_{f/e}$ (columns) for simulations of 8,000 generations that produced extra correlation (EC>0).

Appendix A

Figure A.1: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 120,000 successful simulations for the NM SW FLAT model.

Figure A.2: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 120,000 successful simulations for the NM SW B-A model.

Figure A.3: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 120,000 successful simulations for the NM SW B-A model.
black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the NMCI SW FLAT model.

Figure A.4: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the CD SW FLAT model.

Figure A.5: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the NMCI SW B-A model.

Figure A.6: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the CD SW B-A model.

Figure A.7: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the NMCI DIS FLAT model.

Figure A.8: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the CD DIS FLAT model.

Figure A.9: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the NMCI DIS B-A model.

Figure A.10: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient \( R^2 \), blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 400,000 successful simulations for the CD DIS B-A model.

Figure A.11: Ranked Euclidean distance values of the best 20,240 simulations (i.e. the closest 5% of 404,808 simulations) for each of the two models plotted in black, with the 2,000 retained simulations (i.e. closest ~0.5% of 404,808 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.
Figure A.12: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW FLAT model, calculated on the 2,000 retained simulations (i.e. closest ~0.5% of 404,808 simulations).

Figure A.13: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI DIS FLAT model, calculated on the 2,000 retained simulations (i.e. closest ~0.5% of 404,808 simulations).

Figure A.14: Distributions of the 12 summary statistic values in the 2,000 retained simulations (i.e. closest ~0.5% of 404,808 simulations) for the NMCI SW FLAT model (black lines) and NMCI DIS FLAT model (grey lines).

Figure A.15: Ranked Euclidean distance values of the best 20,240 simulations (i.e. the closest 5% of 404,808 simulations) for each of the two models plotted in black, with the 400 retained simulations (i.e. closest ~0.1% of 404,808 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.

Figure A.16: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW FLAT model, calculated on the 400 retained simulations (i.e. closest ~0.1% of 404,808 simulations).

Figure A.17: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI DIS FLAT model, calculated on the 400 retained simulations (i.e. closest ~0.1% of 404,808 simulations).

Figure A.18: Distributions of the 12 summary statistic values in the 400 retained simulations (i.e. closest ~0.1% of 404,808 simulations) for the NMCI SW FLAT model (black lines) and NMCI DIS FLAT model (grey lines).

Figure A.19: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^s$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI SW FLAT model.

Figure A.20: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^s$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD SW FLAT model.

Figure A.21: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^s$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI SW NPP Banks model.
Figure A.22: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD SW NPPBanks model.

Figure A.23: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI SW NPPSingarayer model.

Figure A.24: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD SW NPPSingarayer model.

Figure A.25: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI DIS FLAT model.

Figure A.26: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD DIS FLAT model.

Figure A.27: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI DIS NPPBanks model.

Figure A.28: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD DIS NPPBanks model.

Figure A.29: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the NMCI DIS NPPSingarayer model.

Figure A.30: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise combinations of all summary statistics and model parameters using 1,000,000 successful simulations for the CD DIS NPPSingarayer model.
Figure A.31: Ranked Euclidean distance values of the best 56,621 simulations (i.e. the closest 5% of 1,132,411 simulations) for each of the two models plotted in black, with the 12,000 retained simulations (i.e. closest ~1% of 1,132,411 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.

Figure A.32: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW FLAT model, calculated on the 12,000 retained simulations (i.e. closest ~1% of 1,132,411 simulations).

Figure A.33: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW NPPSingarayer model, calculated on the 12,000 retained simulations (i.e. closest ~1% of 1,132,411 simulations).

Figure A.34: Distributions of the 12 summary statistic values in the 12,000 retained simulations (i.e. closest ~1% of 1,132,411 simulations) for the NMCI SW FLAT model (black lines) and NMCI SW NPPSingarayer model (grey lines).

Figure A.35: Ranked Euclidean distance values of the best 56,621 simulations (i.e. the closest 5% of 1,132,411 simulations) for each of the two models plotted in black, with the 1,200 retained simulations (i.e. closest ~0.1% of 1,132,411 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red.

Figure A.36: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW FLAT model, calculated on the 1,200 retained simulations (i.e. closest ~0.1% of 1,132,411 simulations).

Figure A.37: Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the NMCI SW NPPSingarayer model, calculated on the 1,200 retained simulations (i.e. closest ~0.1% of 1,132,411 simulations).

Figure A.38: Distributions of the 12 summary statistic values in the 1,200 retained simulations (i.e. closest ~0.1% of 1,132,411 simulations) for the NMCI SW FLAT model (black lines) and NMCI SW NPPSingarayer model (grey lines).
List of Tables

Chapter 2

Table 2.1: Model parameters and their prior ranges for the Null Model, Null Model with Interactions and Culture-Dependent Interactions Model. 70

Chapter 3

Table 3.1: Relative marginal likelihood estimates (RML) of each version of Null Model (NM) and Null Model with Cultural Interactions (NMCI). Proportions are calculated on the smallest 2,400 Euclidean distances (i.e. closest 0.5%, which corresponds to x=0.5 in Figure 3.4) of a total of 480,000 simulations (i.e. 120,000 simulations for each of the 4 models compared). 91

Table 3.2: Relative marginal likelihood estimates (RML) of each version of Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD). Proportions are calculated on the smallest 3,200 Euclidean distances (i.e. closest 0.1%, which corresponds to x=0.1 in Figure 3.5) of a total of 3,200,000 simulations (i.e. 400,000 simulations for each of the 8 models compared). 95

Table 3.3: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places. 101

Chapter 4

Table 4.1: Outline of the major processes of the simulation model described in Chapter 2 (Version I) and that model with the modifications and refinements described in section 130
4.1 implemented (Version II).

Table 4.2: Relative marginal likelihood estimates (RML) of each version of Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD). Proportions are calculated on the smallest 60,000 Euclidean distances (i.e. closest 0.5%, which corresponds to $x=0.5$ in Figure 4.12) of a total of 12,000,000 simulations (i.e. 1,000,000 simulations for each of the 12 models compared).

Table 4.3: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places.

**Appendix A**

Table A.1: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 2,000 retained simulations (i.e. closest ~0.5% of 404,808 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places.

Table A.2: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 400 retained simulations (i.e. closest ~0.1% of 404,808 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places.

Table A.3: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 12,000 retained simulations (i.e. closest ~1% of 1,132,411 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places.

Table A.4: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 1,200 retained simulations (i.e. closest ~0.1% of 1,132,411 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Approximate Bayesian Computation</td>
</tr>
<tr>
<td>AMH</td>
<td>anatomically modern humans</td>
</tr>
<tr>
<td>B-A</td>
<td>population densities from Bocquet-Appel et al. (2005)</td>
</tr>
<tr>
<td>CD</td>
<td>Culture-Dependent Interactions Model</td>
</tr>
<tr>
<td>CI</td>
<td>credible interval</td>
</tr>
<tr>
<td>DIS</td>
<td>bounded non-stepwise (discrete) model of cultural variation</td>
</tr>
<tr>
<td>EC</td>
<td>extra correlation</td>
</tr>
<tr>
<td>FLAT</td>
<td>uniform distribution of carrying capacities</td>
</tr>
<tr>
<td>GLM</td>
<td>generalised linear model</td>
</tr>
<tr>
<td>Ka</td>
<td>thousand years ago</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
</tr>
<tr>
<td>NM</td>
<td>Null Model</td>
</tr>
<tr>
<td>NMCI</td>
<td>Null Model with Cultural Interactions</td>
</tr>
<tr>
<td>NPP</td>
<td>Net Primary Productivity</td>
</tr>
<tr>
<td>NPPBanks</td>
<td>NPP calculated from Banks et al. (2008) climate data</td>
</tr>
<tr>
<td>NPPSingarayer</td>
<td>NPP calculated from Singarayer and Burrough (2015) climate data</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least squares</td>
</tr>
<tr>
<td>RML</td>
<td>relative marginal likelihood estimate</td>
</tr>
<tr>
<td>SW</td>
<td>bounded stepwise model of cultural variation</td>
</tr>
<tr>
<td>TRI</td>
<td>Topographic Ruggedness Index</td>
</tr>
</tbody>
</table>
Acknowledgments

Many thanks to my dedicated supervisors Mark Thomas, Kevin Bryson and Stephen Shennan, whose invaluable guidance, ideas and inspiration made the work presented in this thesis possible.

Thanks to all my UCL friends, for sharing this journey with me and keeping me entertained along the way. Particular thanks to Kate Brown, Anke Liebert, Anna Rudzinski, Catherine Walker and the rest of the MACE lab, and everyone at CoMPLEX.

Very special thanks go to Pascale Gerbault for the positive energy and encouragement, for sharing her experiences and helping me climb the stairs even when it seemed impossible, but most importantly for being a great friend.

I owe a debt of gratitude to my family that goes beyond just this thesis. To my parents, Ermina and Ahmed Kovacevic, for all their love, encouragement and never-ending support – I couldn’t have made it this far without you. A big thank you to my brother, Amar Kovacevic, and Lucy Walpole for being there whenever I needed them and always making me giggle.

Thanks also to my non-work friends for reminding me that there is a whole world out there. Special mentions to Clare Craig, Larissa Kunstel-Tabet, Jackie O’Sullivan, Sarah Hickin and Sabrina Terzaga, for countless coffees, cocktails and board game nights, and to Stephanie and Prash Pathak for all the fun times we’ve shared over the years. A big thank you to all of my McTighe family, particularly Mike and Terry, for all they have done for me over the years, and without whose Christmases my world would be a less entertaining place.

I will finish with the most important thanks, to my wonderful and loving husband Henry McTighe, my best friend and the best self-assigned project manager around. Thank you for reminding me to breathe and for believing in me so much that I started to believe in myself. We did it!
Chapter 1: Introduction

Summary

A high degree of structuring is seen in the spatial distribution of symbolic artefact types associated with the Aurignacian culture in Upper Palaeolithic Europe, particularly the degree of sharing of ornament types across archaeological sites. Multivariate analyses of these distributions have been interpreted as indicating ethno-linguistic differentiation (Vanhaeren and d’Errico 2006). This multidisciplinary thesis draws on simulation modelling and statistical techniques in order to test whether the distribution of artefact types reported by Vanhaeren and d’Errico (2006) can be explained by a simple model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that would be expected if material culture was symbolically marking ethnic group identity and homophilically driving cultural interactions.

This chapter aims to introduce the relevant parts of the different disciplines that overlap in this thesis and place the research described here within a broader context of these disciplines. The chapter starts with the archaeological context of the study, providing a brief history of the time period of interest and its characteristics, and goes on to introduce the simulation modelling and statistical approaches adopted in the thesis. The chapter ends with a summary of objectives and structure of the thesis.
1.1 Overview of Thesis

The main theme of the research presented in this multidisciplinary thesis is the development of quantitative spatial modelling methods applied to human evolution and archaeological material culture distributions.

The primary aim is to test whether the observed distribution of Upper Palaeolithic artefact types identified as personal ornaments (Vanhaeren and d’Errico 2006) can be explained by a simple model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that may be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving cultural interactions. The distinction here is that between the existence of inter-group differences arising through cultural mutation and drift (the null model), of which actors are not consciously aware, versus that null model plus the intentional adherence to behavioural norms that imply identity and actively shape interaction processes, and through that, the spread, loss and mixing of culturally inherited traits.

In order to test this, spatially explicit cultural transmission simulation models that generate simulated material culture data under various cultural evolutionary and demographic scenarios have been developed and explored through simulation. The archaeological dataset of Upper Palaeolithic bead types reported by Vanhaeren and d’Errico (2006) is used to assess the validity of each model. Approximate Bayesian Computation (ABC) techniques are used to compare the differences in goodness-of-fit between the observed data and data simulated by different proposed models.

1.2 Archaeological Context

1.2.1 Vanhaeren and d’Errico (2006) Study

In a recent study, Vanhaeren and d’Errico (2006) applied adapted statistical tools to analyse patterns in material culture, and argued that, in such a framework, personal ornaments and beadwork can be used as a proxy for ethno-linguistic identity, and that they offer archaeological advantages over other components of the material record for inferring ethno-linguistic structuring, including their exclusively symbolic function, and the frequency and wide assortment in which they occur at archaeological sites associated with the Upper Palaeolithic (Kuhn et al. 2001; Vanhaeren and d’Errico 2006).

Vanhaeren and d’Errico (2006) considered bead types, identified as personal ornaments, from European Aurignacian sites. Their dataset of personal ornaments was compiled by sourcing data from the literature and direct analysis of published and unpublished archaeological collections.
The dataset consists of 157 distinct ornament types recorded at 98 Aurignacian sites in Europe and the Near East, with records specifying presence / absence of distinct ornament types in sites only. These distinct ornament types are divided according to different raw materials, with “62 representing ornaments made of shells, 31 of teeth, 30 of ivory, 11 of stone, 11 of bone, 7 of deer antler, and one each of belemnite, nummulite, ammonite, sea urchin and amber” (Vanhaeren and d'Errico 2006).

Seriation and correspondence analyses of the data identified geographically non-randomly distributed clusters of sites sharing ornament types. Correspondence analysis is related to principal components analysis and is a method used to identify dimensions of variation in categorical data and rank them by the amount of variance explained. Seriation analysis is a relative dating method used to chronologically order artefacts recovered from different sites and belonging to the same culture, and often uses the first axis of the correspondence analysis (Doran and Hodson 1975; Lipo et al. 1997; Shennan 1997). It is based on the relative chronological order of artefacts and is often applied when absolute dates are not available.

Vanhaeren and d'Errico (2006) argued that the observed variation in spatial distributions was not due to changes over time in personal ornament preference (as evidenced by all identified clusters containing sites “attributed to multiple periods or cultural phases” of the Aurignacian and including “sites with long sequences spanning most of the Aurignacian interlude”) or local availability of raw materials (as evidenced by numerous inconsistencies between ornament types occurring at a number of sites and raw materials available in the regions of those sites), but rather represented cultural differences among the human groups using Aurignacian technologies. They further argued that the identified trends may have reflected ethno-linguistic diversity among Aurignacian populations, and that identified clusters may have represented different language families.

### 1.2.2 Upper Palaeolithic and Aurignacian

The transition from the Middle Palaeolithic to the Upper Palaeolithic period in Europe occurred as early as approximately 44 thousand years ago (Ka) (Kuhn et al. 2001; Bar-Yosef 2002; Mellars 2005; Higham et al. 2012; Banks et al. 2013). This transition is widely seen as marking the appearance of modern human behaviour in Europe, as evidenced in the Upper Palaeolithic material culture by increased and consistent symbolic activity, and other technological and cultural advances (Mellars 1996).

The earliest evidence of anatomically modern humans (AMH) in Europe remains a subject of debate, but is estimated to date to between approximately 45 Ka (Benazzi et al. 2011; Higham et al. 2011) and 40 Ka (Zilhão et al. 2007; Trinkaus and Zilhão 2012). Due to the lack of reliably dated Neanderthal fossils younger than approximately 40 Ka (Pinhasi et al. 2011),
archaeological findings dating to 40 Ka or later are assumed to be mostly the result of activities of AMH populations. Little is known about the migration routes of the first AMH populations into Europe at the onset of the Upper Palaeolithic, the extent of biological, cultural and linguistic diversity among them, or the nature and extent of their interactions with the local Neanderthals (but see, for example, Prüfer et al. (2014)).

Archaeologically, the initial appearance of AMH in the European territory is widely believed to correspond to the Aurignacian (Kuhn et al. 2001; Zilhão 2007). The Aurignacian refers to an archaeological ‘culture’ of the Upper Palaeolithic, located in Europe and south-west Asia, that lasted from approximately 42 Ka to approximately 29 Ka (Higham et al. 2012). It is characterised in the archaeological record by abstract and figurative art, the use of personal ornaments, systematically produced microlithic stone tools, bone, ivory and antler artefacts, and increasingly complex hunting technologies (Mellars 1996). Of these, personal ornaments in particular have been interpreted as being among the first material objects used to communicate social and ethnic identity within and across cultural boundaries (Kuhn et al. 2001). But what is meant by ethnic identity?

1.2.3 Definitions of Ethnicity

The concept of ethnicity has been given different definitions both by different fields in the social sciences (e.g. archaeology, anthropology, sociology) and by different traditions within those fields.

Within archaeology, the definition of ethnicity, and its relationship with archaeological material culture, has evolved dependent on different schools of thought and study frameworks, and is a topic of continuing debate in the field. The history and evolution of the definition of ethnicity and its relationship with material culture is discussed in detail by Shennan (1989) and Jones (1997), and this section provides a brief overview of the salient points for purposes of contextualising the research in this thesis.

In the past, patterns of spatial variation in the archaeological material record have been classified into entities referred to as ‘archaeological cultures’ that were regarded as indicators of different ethnic groups, resulting in normative concepts of culture and ethnicity characterised by homogeneity, order and boundedness. This way of classification, however, has been shown to only be useful as a summary description of patterns of spatial variation and otherwise a misleading concept in analytical approaches, since spatial variation in archaeological material is the result of multiple processes that are not necessarily homogeneous over space and time.

As a result, the notion of a one-to-one relationship between material culture groups and ethnic groups was overturned by definitions that instead conceptualise culture as an adaptive system
and the archaeological record as the product of a variety of past processes. The response to this adaptive emphasis was to redefine the idea of ethnic identity in terms of artefact style, and it was argued that stylistic variation in artefacts in part functions as a method of communicating visible messages of ethnic identity that facilitate cultural interactions. Within this framework, a distinction is drawn between spatial variation and ethnic structuring, the latter referring to the decision of individuals, or groups of individuals, to consciously identify with a specific social group (Shennan 1989: 14).

More recently, the idea of ethnic groups as distinct, territorially-based entities has been challenged (Shennan 1989; Jones 1997). It has been argued that the construction of ethnicity is a dynamic process that is dependent on the context of social interaction, in which the material world is used for “communication of similarity and difference that ethnicity inevitably entails” (Jones 2008: 327). These changeable expressions of ethnicity in different social contexts are then reflected in the archaeological record by “a complex web of overlapping styles of material cultures” (Jones 2008: 327).

### 1.2.4 Ethnicity from Distributions of Material Culture

Even with such evolving study frameworks, however, “ethnicity, as well as the relationship between cultures and ethnic groups, remains a problematic area of archaeological analysis” Jones (2008: 328). Invoking ethnic structuring and reconstructing patterns of ethnicity through analyses of material culture data is a challenging task, issues of which have been discussed by several authors (Geary 1983; Shennan 1989; Jones 1997; Brather 2002; Shennan 2002).

In relation to the reasoning that resulted in the evolution of definitions of ethnicity briefly summarised above, Jones (1996) discusses two issues in particular. The first relates to archaeological classification, which often misleadingly assumes that change in material culture is a gradual process occurring in a uniform manner over a spatially homogenous area. Jones (1996: 73) argues that a normative view of culture is inherent in this approach and produces an “illusion of bounded uniform cultural entities” that obscures “information which is of interest in the analysis of ethnicity”.

The second issue is one of contextualising material culture. While non-random distributions of material culture may relate to expressions of ethnicity, a wide range of different processes may be involved in producing the observed variation (e.g. coping with different environments, organisation of material production, distribution of social power), and it is therefore necessary to contextualise material culture using independent evidence (Shennan 1989: 16-17), since “the significance of material culture in terms of ethnicity is culturally and historically specific” (Jones 1996: 73). Furthermore, in the context of more recent archaeological time periods (e.g. Late Antiquity, early middle ages), it has been argued that ethnic identity is a subjective and
largely arbitrary phenomenon that changes depending on context, and that it is therefore misleading to assume that ‘peoples’ described in sources necessarily correspond to self-conscious identification with particular groups, a link that is essential to the definition of ethnicity (Geary 1983). Thus, even where multiple lines of evidence are available, it has been argued that it is not possible to infer ethnic identities from archaeological material culture data (Brather 2002).

Another aspect of spatial variation in archaeological material culture is style. One definition of stylistic variation in artefacts is that it functions to communicate information that facilitates social interaction, however the nature and significance of stylistic variation has been widely debated (Shennan 1989). The most prominent debate has been between Wiessner (1985, 1983, 1984) and Sackett (1982, 1985).

Wiessner (1984, 1983, 1985) has argued for an active view of style with emblemic and assertive aspects, defining emblemic style as “variation in material culture that has a distinct referent and transmits a clear message to a defined target population about conscious affiliation or identity” (Wiessner 1983: 257), which therefore carries information about group identity rather than the extent of interaction between those groups. Assertive style is defined as “variation in material culture which is personally based and which carries information supporting individual identity” and provides a “measure of interpersonal contact” (Wiessner 1983: 258). Sackett (1985, 1982), on the other hand, has argued that stylistic variation in artefacts is passive and carries ethnic symbolism because it is created in an ethnically bounded context, and refers to it instead as isochrestic variation.

Following these initial definitions, Wiessner (1989) has integrated emblemic and assertive style with isochrestic variation in a framework based on “identification via comparison” such that which of these definitions is applicable to a particular artefact is dependent on the context and conditions within which that artefact is considered. As Shennan (1989: 21) surmises, it is therefore “impossible to regard what goes on within social groups as independent of what happens in the relations between them”, which relates back to the point discussed earlier that it is necessary to contextualise material culture using independent evidence.

In addition to these frameworks for considering stylistic variation in artefacts, experimental (Efferson et al. 2008) and modelling (Ihara 2011) research on coordination games has suggested that initially meaningless symbolic markers can evolve to play a key role in cultural group formation. In other words, such symbolic markers can evolve to signal information about socially critical variables that are not otherwise observable, and therefore facilitate the formation of cultural groups consisting of individuals sharing behaviour expectations and symbolic markers that signal group affiliation.

Given the discussed considerations regarding contextualisation of material culture, it is clear that drawing conclusions about ethnic identity and structure for prehistoric populations is
difficult since there are no data in the material record for which there is independent evidence that they relate to individuals’ conscious identification. Furthermore, invoking ethnic structure for the Upper Palaeolithic in Europe is a particularly challenging task given the paucity of both material culture and other data for the period.

However, ethnic and linguistic identity and structuring are universals in the modern world and ethno-linguistic groups are generally believed to have existed in the past. Identifying the earliest appearance of ethnic structuring is an issue of general importance for the history of human evolution that has implications for the emergence of language, and may inform on the evolutionary dynamics of human populations, as well as the role of identity construction, in people today.

1.2.5 Explaining Distributions of Material Culture

Two principal frameworks are adopted in studies of material culture distributions, focussing either on explanations based on shared descent or those based on shared function. The key distinction between the two is whether similar cultural traits reflect shared history between those traits (i.e. shared descent), or whether they represent similar solutions to similar problems (i.e. shared function).

Shared descent is the focus of much work in the cultural evolution literature and focuses on descriptive approaches to studies of material culture distributions. The main idea of this approach is that similarities between material culture traits represent a shared history between them; thus “spatial or chronological uniformities” observed in material culture distributions represent “the traditions of human groups”, with “major changes in the archaeological record” occurring solely through the replacement of one group by another (Shennan 2002: 69). The details of the evolution of culture historical approaches in archaeology are discussed in detail by Shennan (2002: Chapter 4).

Shared function, on the other hand, has more recent origins and is the focus of the functional approach promoted initially by Binford (Binford and Binford 1966). This approach was to a considerable extent a response to culture historical archaeology that had previously dominated archaeological thinking. This is particularly evident in Lewis Binford’s classic debate with Francois Bordes on culture history versus function, the details of which are discussed by Wargo (2009). At the heart of the shared function approach is the idea that similarities between material culture traits represent shared function that arises as a result of adaptation. In contrast to shared descent which takes a descriptive approach to explaining material culture distributions, shared function is explanatory and as such represents the start of the transition in archaeological thinking towards being an explanatory science.
An approach to studying material culture distributions with the idea of shared function at its core is experimental archaeology, in which the manufacturing techniques and function of material culture objects are explored (e.g. Mesoudi and O’Brien 2008a, b; Vanhaeren et al. 2013), in order to understand the processes that have led to the observed archaeological record. In addition to this, the spread, or diffusion, of material culture traits is also often considered (e.g. Bar-Yosef 2002; Vanhaeren and d’Errico 2006), in an attempt to understand the resulting material culture distributions.

It is important to note that neither shared history nor shared function should be privileged a priori as an explanation for observed material culture distributions – as Shennan (2002: 73) points out, “every practice that is socially learned, whether” it is “obviously functional or not, will have a history of descent”. Both explanations should therefore be the subject of empirical investigation on a case-by-case basis, an example of which is the discussion of Clovis projectile points by O’Brien et al. (2014: 114).

Modern evolutionary archaeology, parts of which are discussed in section 1.3.1, revisits aspects of culture history, particularly the potential importance of history, but with new ideas and tools, including those from functional archaeology, at its disposal (Shennan 2008).
1.3 Modelling in Archaeology

1.3.1 Studying Cultural Evolution

This section focuses on introducing elements of cultural evolution theory relevant to the current problem, which is outlined in section 1.1 and further described in section 1.4, in order to contextualise the processes that make up the developed modelling framework, detailed in Chapter 2 and Chapter 4. Of particular interest are cultural transmission processes in relation to the Axelrod model of cultural dissemination (Axelrod 1997), as well as cultural drift and isolation-by-distance processes.

1.3.1.1 Culture as an Evolutionary System

Cultural evolution theory draws parallels with genetic evolution in order to study spatial and temporal changes in culture (for a recent review of the field of cultural evolution, see Mesoudi (2015)). In this context, ‘culture’ is defined to be information that is transmitted socially between individuals. In cultural evolution theory, this socially transmitted information forms an evolutionary system that follows Darwinian principles of evolution.

Cultural evolution theory views culture as a dynamic population-level result of social transmission processes, akin to the study of genetic evolution in the field of population genetics. The mode of cultural transmission is not as simple as genetic transmission, however, and a wide array of transmission mechanisms have been proposed (see Henrich and McElreath (2003) and Laland (2004) for reviews). In contrast to genetic transmission, which describes a strict parent-to-child inheritance of genes, cultural information in the form of cultural variants can be transmitted in a number of ways, collectively labelled transmission mechanisms. These mechanisms include inheritance of cultural variants from a genetic parent (vertical transmission), from older non-parental individuals (oblique transmission), from peers (horizontal transmission), or some combination of these modes.

In addition to the necessary requirement of transmission of information, in order to form an evolutionary system, it is also required that variation is generated, and that such variation exhibits differential probabilities of transmission, either by affecting transmission rates directly, or by affecting the survival of agents carrying culturally transmitted traits. The variation requirement of an evolutionary system is satisfied since cultural variation can arise through transformation of cultural information during transmission, as a result of random error (analogous to genetic drift) or guided variation (intentional modification, recombination or innovation). Lastly, differential probabilities of transmission can arise by several mechanisms, including guided variation, transmission biases and cultural drift (analogous to genetic drift).
In this context it is worth noting the difference between stylistic and functional variation in artefacts from a cultural evolution perspective. As noted by Neiman (1995), style refers to variation in which the fitness values associated with the variants are effectively the same, such that the variants are selectively neutral. Conversely, in functional variation, different variants have different fitness values. In terms of cultural evolution, this implies that spatial and temporal variation in stylistic variant frequencies may be subject to evolutionary forces that introduce selectively neutral cultural variation, such as random mutation and cultural drift, while functional variant frequencies may in addition be subject to adaptive forces, such as transmission biases.

1.3.1.2 Transmission Biases

The group of mechanisms termed transmission biases describe the processes by which cultural variants are preferentially adopted during transmission; such biases increase or decrease the frequency of preferred cultural variants. Henrich and McElreath (2003) have categorised transmission biases into content-specific and context-specific biases. Content biases, also termed direct biases (Boyd and Richerson 1985), refer to the preferential adoption of a cultural trait dependent on its intrinsic properties. Conversely, context biases concern the source of cultural information, and are not necessarily dependent on its content. Context biases can further be divided into two groups with different mechanisms, namely model-based biases and frequency-dependent biases.

Model-based biases are dependent on choice of individual from whom to learn, examples including success bias and prestige bias. Success bias involves copiers choosing a model by making an evaluation of the (apparent) success of all potential cultural models within the cultural domain of interest and choosing the most successful one, while prestige bias causes copiers to preferentially choose a prestigious individual for reasons that may not necessarily be related to the cultural domain of interest (Henrich and Gil-White 2001).

Frequency-based biases, on the other hand, occur in what Henrich and McElreath (2003) refer to as information-poor situations – instances when no correlation is observed between success or prestige and cultural behaviour. In such cases, one solution is to treat frequencies of cultural variants in the population of all potential cultural models as cues of their respective adaptive qualities and preferentially copy cultural variants at high frequency (conformist bias) or at low frequency (anti-conformist, or rarity, bias). Such biases appear to evolve under a wide range of conditions (Henrich and Boyd 1998) and have been shown to be advantageous in heterogeneous environments (Boyd and Richerson 1985; Henrich and Boyd 1998).

It has also been argued that conformist biases underpin the evolution of large scale altruistic (Boyd and Richerson 1985, 2009; Henrich 2004a) or potentially parochial (Choi and Bowles
norms of behaviour during the Late Pleistocene, arguing that conformism reduces within-group cultural variation, and so serves to establish greater inter-group variation due to the possibility of multiple stable cultural equilibria (Boyd and Richerson 2009). Cultural group-selection pressures, possibly involving violence (e.g. Choi and Bowles 2007; Bowles 2009) or not (Boyd and Richerson 2002, 2009), then act on this inter-group level variation and favour groups with more, or stronger, pro-social norms, possibly including parochial behaviour (Bowles 2009).

1.3.1.3 Transmission Isolating Mechanisms

In addition to these transmission processes that may give rise to cultural variation, cultural transmission between groups is often constrained by transmission isolating mechanisms that are analogous to reproductive isolating mechanisms that prevent gene flow. Examples of such mechanisms include environmental, linguistic and cultural barriers, and it has been suggested that the processes involved in generating between-group cultural differences are fundamentally similar to the processes that give rise to speciation (Durham 1990, 1992).

1.3.1.4 Cultural Drift

In contrast to the above mentioned models of non-random or biased transmission of cultural information, some authors have considered the evolution of cultural traits under neutral models of random copying (Neiman 1995; Shennan and Wilkinson 2001; Bentley et al. 2004), which are directly analogous to the neutral model in population genetics (Kimura and Crow 1964; Kimura 1983). Under the neutral model in population genetics, genetic drift is a process by which allele frequencies change as a result of random sampling of genes in reproduction. In finite populations genetic drift corresponds to the intergeneration sampling error (see, for example, Tishkoff and Verrelli (2003)). Cultural drift is a direct analogue to genetic drift that acts on cultural variant frequencies. In populations of finite size, cultural drift results in stochastic fluctuations over time in cultural variant frequencies due to sampling error, which can lead to drastic changes in frequencies, including loss or fixation, of cultural traits.

It may be worth noting here the difference between models of neutral random copying and frequency-dependent biased copying as described above. In the neutral model, individuals copy the cultural variant of interest from a randomly selected individual, which results in frequency-independent change in variant frequencies (i.e. cultural drift). Frequency-dependent biases, on the other hand, are defined by an intentional, non-random choice of individual to copy from, which is dependent on the frequency of the cultural variant of interest that that individual carries.
1.3.1.5 ‘Core’ and ‘Package’ Models

Some authors have also drawn a distinction between what they term ‘core’ and ‘peripheral’ traits (Boyd et al. 1997). It is posed that core traits, which are thought to stick together over time to form an isolated and consistent basis of societal norms, are likely to be subject to transmission bias mechanisms, while peripheral traits, those that are not closely tied to the core and are therefore less tightly constrained, are suggested to be subject to cross-cultural borrowing and may best be described by stochastic processes such as cultural drift. Furthermore, it is suggested that cultural groups may be defined as hierarchically integrated systems with both core and peripheral elements, or possibly as assemblages of multiple ‘packages’ – series of separate groups of cultural elements – that are subject to different evolutionary histories.

1.3.1.6 Cultural Isolation-by-Distance

Another idea in cultural evolution theory that has been adopted from population genetics is the isolation-by-distance model, which states that, under geographically limited dispersal, the extent of genetic differentiation (between individuals or populations) increases with geographic distance (Wright 1943; Slatkin 1993). Similar principles can be applied to cultural differentiation, where a simple model of cultural isolation-by-distance states that the frequency of cultural transmission is inversely proportional to geographic proximity, such that the occurrence of transmission events, and therefore similarity of cultural traits, decreases with increased geographic distance. This effectively corresponds to the neutral model with a spatial dimension, since the only transmission isolating mechanism in this case is distance.

1.3.1.7 The Axelrod (1997) Model of Cultural Dissemination: Cultural Interaction Processes

In addition to the idea of cultural isolation-by-distance, another process that may give rise to distinct spatial patterns in cultural variation is cultural interaction that is dependent on the extent of similarity between interacting entities. Such a process can potentially amplify small differences in cultural variation that are the result of cultural drift, and may lead to the formation of cultural boundaries and distinct clusters of cultural traits (Schelling 1969; Axelrod 1997; Centola et al. 2007).

As a result, if differences in cultural preferences structure the extent of cultural interaction, one expectation might be that entities will preferentially interact with other culturally similar entities, in other words homophily. Homophily is the tendency of people with similar traits (including, but not limited to, physical and cultural traits) to preferentially interact with each other rather than with people with dissimilar traits, in part because “communication is most
effective between similar people” (Axelrod 1997: 205). Homophily is considered to be a prominent explanation for the persistence of cultural diversity (Centola et al. 2007).

Using a simple simulation model, Axelrod (1997) investigated the persistence of cultural heterogeneity as a result of interactions between individual agents that are dependent on the extent of cultural similarity between those agents, and showed that differences between agents persist despite the fact that repeated interactions between agents cause them to become more alike. Rooted in the field of political science, Axelrod’s study aims to build on preceding approaches to social influence and research on maintenance of cultural differences, in relation to concepts such as state formation and development of transnational institutions. The model offers a “way of looking at the dynamic process of social influence” that is based on the “fundamental principle of human communication”, namely that “communication is most effective between similar people” (Axelrod 1997: 205-206). In Axelrod’s model, “culture is taken to be what social influence influences” and “is assumed to satisfy two simple premises: people are more likely to interact with others who share many of their cultural attributes, and interactions between two people tend to increase the number of attributes they share” (Axelrod 1997: 206-207). In using Axelrod’s framework, and this concept in particular, as a starting point, an inherent assumption of the models in this thesis is that “similar individuals are more likely to influence each other than dissimilar individuals” (Axelrod 1997: 207).

Axelrod’s idea that “similarity leads to interaction, and interaction leads to still more similarity” (Axelrod 1997: 205) can potentially be extended: for example, while an interaction between two culturally similar populations may result in exchange or sharing of cultural traits between the two, an interaction between two culturally different populations may result in subsequent mutual avoidance, physical conflict or cultural coercion, dependent on the extent of the cultural difference between them. This model offers a different way of investigating the dynamic processes that may have resulted in observed spatial patterns in material culture, and is taken as the basis of the models developed in this thesis.

Relating this idea back to the brief discussion of ethnicity in sections 1.2.3 and 1.2.4, it is possible to treat ethnicity as a form of homophily (or, conversely, consider the effects of homophily built on symbolically marked ethnicity), such that, if ethnicity exists and is reflected by similarities and differences in material culture, then it may be expected that ethnicity has a homophilic drive on cultural interaction processes. In this case, it may be informative to treat homophily as a bridge between ethnicity and cultural interaction processes: if populations are ethnically structured and the degree of ethnic similarity or difference (reflected in the degree of similarity and difference in material culture) affects homophily (which is actioned through cultural interaction processes), then investigating the effect of cultural interaction processes that are homophilically driven by similarities and differences in material culture provides an approach to investigating ethnic structuring.
1.3.2 Computational Modelling in Archaeology

The field of cultural evolution as briefly described above is characterised by mathematical and computational models that attempt to address a variety of questions concerning the evolution of culture. More generally, over the last 30 years, there has been a movement from qualitative towards quantitative approaches to the study of archaeological material culture, and a progression towards modelling approaches to understanding past processes (Gerbault et al. 2014). Archaeologists are now more widely postulating explicit hypotheses to explain the material culture records recovered from archaeological sites, and developing various methods to test these hypotheses. As a result, the field of archaeology has become increasingly systematic and multidisciplinary; this is marked by an increase in archaeological studies employing computational modelling approaches and statistical techniques to address longstanding archaeological and anthropological questions that are difficult to formally test through interpretation of archaeological data alone (for a recent comprehensive review of modelling in archaeology, see Lake (2014)).

A number of different simulation modelling approaches have been used in archaeological studies (Gerbault et al. 2014; Lake 2014). Lake (2014) categorises these into the broad groups of reaction-diffusion models, dynamical systems models, and agent-based models. Reaction-diffusion models have primarily been used to model hominin dispersal and colonisation events (for a review see Steele (2009)), while dynamical systems modelling has focussed on interactions between societal and environmental change. Agent-based simulation modelling has more widely been used to model the evolution and spread of cultures, with particular focus on processes of cultural innovation and the transmission and accumulation of cultural traits (Neiman 1995; Shennan 2001; Henrich 2004b; Powell et al. 2009), the emergence of large-scale cultural complexity effects (Powell et al. 2009, 2010; Premo and Kuhn 2010) and for studies of hominin dispersal (Hughes et al. 2007; Mithen and Reed 2002; Nikitas and Nikita 2005).

It is important to note here the distinction in the level of complexity of simulation models used for theory building (i.e. hypothesis generating) and those used for hypothesis testing in archaeology. Simulation models used for hypothesis testing have tended to be relatively complex, modelling many aspects of the problem being addressed and invoking numerous parameters in an attempt to include a high degree of realism into the simulations. This is particularly true of models concerned with long-term human-environment interactions, such as those developed by Kohler et al. (2012), aiming to understand variability in settlement patterns, and those published by Barton et al. (2010a) and Barton et al. (2010b), addressing socioecological dynamics of Mediterranean landscapes. In contrast, models used for theory building (i.e. hypothesis generating) have tended to be relatively simple, focussing on abstract simulations with few parameters. Such models are often useful in the theory building process as rigorously exploring the full parameter space serves to highlight both the variables of highest importance and assumptions or formulations of the problem that are inadequate (Premo et al. 2005). This relatively simple simulation modelling approach has been widely applied in
research on cultural transmission, for example to develop and test theory on different modes of cultural transmission (Neiman 1995; Bentley et al. 2004; Mesoudi and Lycett 2009) and to explore the emergence of cultural complexity (Powell et al. 2009, 2010; Premo and Kuhn 2010). An important contribution of this thesis to the field of simulation modelling in archaeology is that it combines aspects of the two different modelling approaches and promotes a method of hypothesis testing using relatively simple models with few parameters.

Simulation modelling is a powerful approach that is well suited to archaeological study. The approach requires explicit assumptions and formulations of systems under study, and allows the incorporation of stochasticity (variation in demographic and cultural processes arising from random events) into the models. Simulation modelling, and computational modelling in general, also allows researchers to account for sample sizes and the spatial distribution of sample sites, effectively incorporating sampling error and some archaeological bias in inferences on the past (Shennan et al. 2013; Gerbault et al. 2014).

Nevertheless, inferring evolutionary processes that may be responsible for the observed patterns in distributions of material culture remains a challenge for two main reasons. First is the issue of equifinality as a result of the large number of different mechanisms that may have resulted in the observed patterns (Premo 2010; Gerbault et al. 2014), which can be addressed using simulation modelling approaches. Secondly, the relatively poor quality of the archaeological record, particularly for prehistoric time periods, limits the comparisons between observed data and different evolutionary models. Coupling simulation modelling with Bayesian approaches, particularly the Approximate Bayesian Computation framework, provides a way to address these issues. With such methods, archaeologists are able to develop robust frameworks that allow a qualitative comparison of alternative modelled scenarios with each other and with observed material culture records, in effect creating virtual experiments to test the effect of varying parameter values on the similarity between simulated and observed material culture data.

One prominent example of the value of simulation modelling coupled with Approximate Bayesian Computation is the application to the study of evolution and spread of lactase persistence. While details of the archaeological context and the findings of the studies on lactase persistence are not directly relevant to the research presented in this thesis and will not be discussed here in detail (for a summary of modelling approaches to and recent studies on lactase persistence, see Gerbault et al. (2011)), there are direct parallels in the value and applicability of simulation modelling approaches employed in both.

In the context of lactase persistence, simulation studies (see, for example, Itan et al. (2009) and Gerbault et al. (2011)) have two main advantages over purely genetic or purely archaeological studies. First, they have allowed the integration of information obtained from different fields of study, including genetics, archaeology and ecology. Second, the simulation modelling approach has provided a formal comparison of different possible scenarios, meaning that many scenarios can be tested and evaluated statistically, and extreme scenarios potentially excluded.
Furthermore, ABC methods have proved to be very useful when evaluating the fit of data simulated under complex scenarios to observed data, adding power to simulation studies. Studies using demographic simulations at the continental scale have also served to further understanding of the general features and processes of the European Neolithic transition, and to explore different possibilities that may have led to observed patterns of genetic diversity.

1.3.3 Bayesian Inference

In brief, Bayesian inference is a branch of statistics that uses particular datasets to infer the probability that a proposed hypothesis, or a parameter value of that hypothesis, is true. In contrast to frequentist statistics, where the hypothesis is fixed and variation in outcomes (data) is explored, in Bayesian inference the data becomes fixed and some space of possible explanations (hypotheses) is explored (Gerbault et al. 2014). This means that Bayesian approaches are naturally well suited to archaeological inference since observed data from the past is fixed but only one of a number of possible outcomes of a set of stochastic processes of interest.

In Bayesian inference, models with set numbers of parameters are proposed, and the posterior probability distributions of these parameters are inferred using information from prior probability distributions of the parameters and information provided by the observed data.

Under the Bayesian paradigm, the parameters of a model, denoted $\theta$, are assumed to have probability distributions that reflect the uncertainty in their true values. The distribution of model parameters before data $D$ has been observed is referred to as the prior distribution and denoted $P(\theta)$, and allows the incorporation of prior information about the parameters that is not related to data $D$. After data $D$ has been observed, information contained within it can be used to update the prior knowledge about the parameters; this conditional distribution of parameters given the data is referred to as the posterior distribution and denoted $P(\theta|D)$. The posterior distribution of parameters is defined by Bayes’ theorem:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)},$$

(1.1)

where $P(D|\theta)$ specifies the likelihood of observing data $D$ given parameters $\theta$, referred to as the likelihood function, and $P(D)$ is the unconditional marginal likelihood of the data, referred to as the prior probability of the data, which does not depend on the parameters $\theta$ of the model and is effectively a normalising constant.

Inferences on parameters $\theta$ can be derived from their posterior distributions. The mode of the posterior distribution of each parameter is often taken as a point estimate of the value that that parameter is likely to take, with the credible intervals (CI) constructed using the tail-quantiles or
the highest posterior density interval of the posterior distribution. While the Bayesian CI is analogous to the frequentist confidence interval, its interpretation differs. An X% Bayesian CI can be interpreted as containing the true parameter with probability X%, while the frequentist X% confidence interval specifies the percentage of confidence intervals (each created identically by repeated sampling) that contain the true parameter value.

As in many other disciplines, Bayesian methods in archaeology have surged in popularity in recent years, primarily in the context of dating to, for example, integrate stratigraphic information with radiocarbon date estimates (Buck 2001), and in analyses of ancient and modern DNA data (Fagundes et al. 2007; Gerbault et al. 2014). Although Bayesian methods have not been extensively implemented in archaeological research, more and more studies are being published that apply Bayesian methods developed in physical and biological sciences to archaeological problems (Kandler and Laland 2013; Crema et al. 2014).

While Bayesian approaches offer the advantage of incorporating prior information about, and the uncertainty in, model parameters into the analysis, a central issue is the necessary estimation of the likelihood function that is often difficult to formulate or computationally expensive to calculate, especially with increasing model complexity. Approximate Bayesian Computation (ABC), a relatively recent and still-developing application of the Bayesian paradigm, neatly circumvents the need to explicitly formulate the likelihood function.

1.3.4 Approximate Bayesian Computation (ABC)

The principles of Approximate Bayesian Computation (ABC) were first outlined by Tavare et al. (1997) and later formalised by Pritchard et al. (1999) into what is now commonly referred to as the ABC rejection algorithm.

The simple rejection algorithm, outlined in section 1.3.4.1, involves simulating a large number of datasets under a model assuming different, randomly chosen, parameter values from within prior ranges. Those parameter sets that result in simulated data that emulates the observed data are retained in the posterior probability distributions of the parameters, which can then be used to make inferences about the values, and confidence in these values, of the model parameters. ABC methods are therefore used to infer posterior distributions of the parameters without explicitly computing the likelihood, and instead estimate it by comparison between the observed and simulated data. Additionally, this approach allows the researcher to postulate a number of alternative hypotheses and, provided that they are sufficiently well defined to allow data to be simulated, test which of these hypotheses are more likely given the observed data.

Given the algorithm as outlined below, the ABC framework requires the user to make three choices: summary statistics that sufficiently describe the observed dataset, a distance measure
used to quantify the similarity between observed and simulated data, and the threshold with which simulated datasets are accepted into the retained set of simulations (i.e. those most similar to observed data).

1.3.4.1 Approximate Bayesian Computation (ABC) Algorithm

Let $M$ denote the proposed model and the parameters of $M$ be $\theta = (\theta_1, \ldots, \theta_m)$. Let $S = (s_1, \ldots, s_n)$ and $S' = (s_1', \ldots, s_n')$ denote the values of the summary statistics for the observed and simulated datasets, respectively. Values $S = (s_1, \ldots, s_n)$ are referred to as the target values for each of the summary statistics. The ABC algorithm is applied as follows:

1. define a set of summary statistics that capture relevant information contained in the observed dataset;
2. compute summary statistics values $S = (s_1, \ldots, s_n)$ for the observed dataset – these are the target values;
3. sample parameters $\theta^* = (\theta_1^*, \ldots, \theta_m^*)$ from an appropriate prior distribution;
4. simulate data by using parameter $\theta^*$ set with model $M$;
5. compute summary statistics values $S' = (s_1', \ldots, s_n')$ for the simulated data;
6. compute $\delta(S, S')$, where $\delta$ is an appropriately chosen distance measure;
7. for a chosen acceptance threshold $\varepsilon$, retain parameter set $\theta^*$ in the posterior distribution of $\theta$ if $\delta(S, S') < \varepsilon$;
8. repeat steps 1-7 until the desired number of parameter sets have been accepted to the posterior distribution.

1.3.4.2 Summary Statistics Choice

When dealing with large datasets and complex models, several authors have suggested reducing the full observed dataset to a set of statistics that summarise all relevant information contained in the observed data and represent the maximum amount of this information in the simplest possible form (Tavare et al. 1997; Weiss and von Haeseler 1998; Pritchard et al. 1999). These are robust statistics that sufficiently describe the full properties of the data, and are referred to as summary statistics.
The choice of which, and how many, summary statistics to use is pivotal in the accuracy of the inferences made. Using a single or very few statistics to represent complex datasets is likely a crude representation of the data and often results in capturing a limited amount of information, which in turn produces biases in the analyses (Marjoram et al. 2003). Therefore, at least in principle, the more summary statistics are used to describe the observed data, the more of the information contained in the data is captured. However, including many statistics that provide little information about the parameters of interest introduces stochastic noise and reduces the number of simulations retained (Beaumont et al. 2002). This can be because a particular statistic may provide very little or no information on the parameters of the model, which in turn means that it will have little or no inferential power on the posterior distributions of the parameters, or, alternatively, two or more summary statistics may be correlated (i.e. contain similar or the same information about the observed data), which could lead to a posterior biased by the over-representation of the aspect of the observed data captured by correlated statistics.

Ideally, in order to ensure that the inferences are robust, the summary statistics used should capture the relevant features of the data in a sufficient manner. This means that the posterior probability of a parameter calculated using the summary statistics should be the same as its posterior distribution calculated using the complete observed dataset (Marjoram and Tavare 2006). There is currently no universal rule to define which, or how many, summary statistics should be used. Empirical studies often use between 5 and 20 summary statistics (Bertorelle et al. 2010), and most applications of ABC have used summary statistics chosen on the basis of established practice rather than for their role in providing a good approximation to the posterior distribution (Burr and Skurikhin 2013).

To address the issue of summary statistic choice, Joyce and Marjoram (2008) proposed performing a leave-one-out validation procedure to identify the most informative of a large set of summary statistics. This approach involves scoring each proposed summary statistic based on how much its inclusion in the analysis improves the quality of the inference, and retaining only those summary statistics that significantly modify the posterior distributions of parameters.

As an alternative, Wegmann et al. (2009) proposed a dimension reduction on a large number of proposed summary statistics using a partial least squares (PLS) regression approach. Similarly to principal component analysis, their method involves extracting a number of orthogonal components from the set of proposed summary statistics that explain the variation of parameters. After a PLS analysis, the large number of proposed, potentially correlated, summary statistics is therefore reduced to a smaller number of independent and informative (i.e. explain the variation in the parameters) components, which can then be used in place of the original set of summary statistics.
1.3.4.3 Quantifying Similarity

When comparing simulated and observed datasets, it is possible to only retain those simulations whose summary statistics values match exactly the observed summary statistics values.

However, since simulating data that is identical to the observed data is unlikely for anything but the simplest models and would require extremely large numbers of simulations, an alternative is to accept simulations with summary statistics values that are considered sufficiently close to the target values. In order to do this, it is necessary to define a distance measure (used to quantify similarity between observed and simulated data) and an acceptance threshold (used to identify simulated datasets to be retained).

As suggested by Beaumont et al. (2002), the similarity $\delta$ between observed data, $S$, and simulated data, $S'$, is often quantified as the normalised Euclidean distance of individual summary statistics:

$$\delta(S, S') = \sqrt{\sum_{i=1}^{n} \frac{(s_i - s'_i)^2}{\sigma(s'_i)^2}},$$

where $s$ and $s'$ are values of each of the summary statistics for the observed and simulated datasets, respectively, subscript $i$ denotes the $i$th of $n$ statistics, subscript $j$ denotes the $j$th of $N$ simulations and $\sigma(s'_i)$ is the standard deviation of the $i$th statistic over all $N$ simulations.

Since different summary statistics values may be on different scales (i.e. have different magnitudes), in order to prevent bias in the amount of influence each summary statistic has on the value of the Euclidean distance, it is necessary to normalise each by its variance so that they are on the same interval and equally weighted (Beaumont et al. 2002).

There is no systematic rule to define the value of the acceptance threshold $\varepsilon$ (Bertorelle et al. 2010), and so it is often defined to be a small quantile of the distribution of distances between the observed and simulated data, $\delta(S, S')$. It may be possible to identify a sensible acceptance threshold by plotting Euclidean distance values in ascending order; in some cases (Kamberov et al. 2013), this procedure identifies a noticeable increase in Euclidean distance values that normally begins after a plateau in the same, and so the simulations whose Euclidean distance values are smaller than that at the plateau can be considered as the accepted simulations. Simulations that are accepted are the best simulations – those generating data most similar to the observed data.

1.3.4.4 Parameter Estimation

As an improvement on the simple rejection algorithm for parameter estimation, other approaches for estimating parameters have been suggested.
Beaumont et al. (2002) proposed performing a locally weighted linear regression in order to explicitly account for the distance (i.e. goodness-of-fit) between the observed data and each simulated dataset, and showed that this regression step improved the estimations of parameters. Their approach assumes linear dependence between the parameters and the summary statistics, and uses this assumption to adjust the posterior parameter values estimated from the retained simulations to account for the distance between the observed and simulated datasets.

As an alternative, Leuenberger and Wegmann (2010) proposed a generalised linear model (GLM) approach. In contrast to the approach of Beaumont et al. (2002) which assumes that the sample of accepted values is generated by a linear model, the method of Leuenberger and Wegmann (2010) assumes that it is generated by a normal linear model.

Since all simulations are independent in the ABC approach, the exploration of parameter space is not efficient – sequential randomly chosen sets of parameters may produce very different simulated data, and while the first may be similar to the observed and therefore useful for inference, the next may not. In an attempt to improve the efficiency of the ABC approach, Marjoram et al. (2003) proposed an algorithm that explores the parameter space iteratively by linking simulations along a Markov chain Monte Carlo (MCMC). In this approach, parameter values that result in simulated datasets close to the observed data are chosen preferentially, resulting in a higher acceptance ratio and therefore a more efficient scheme, but the independence between simulations is lost. This algorithm, referred to as ABC-MCMC, raises two sets of issues: the first, common to all MCMC approaches, includes determining the length of the chain and assessing the convergence of the chain to a stationary state (Marjoram and Tavare 2006), while the second, specific to ABC-MCMC, is the choice of the acceptance threshold, which may prevent the chain from escaping regions of low likelihood, due to the correlated nature of the acceptance rate and the likelihood of the parameters (Wegmann et al. 2009).

An approach that overcomes some of the issues in ABC-MCMC, first proposed by Sisson et al. (2007) and further developed by (Toni et al. 2009), is rooted in sequential Monte Carlo (SMC) methods and involves importance sampling arguments, a standard statistical method for improving the efficiency of Monte Carlo methods by concentrating simulation and computational effort on regions of parameter space that are more consistent with the data. In this scheme, preliminary runs are used to explore a large set of randomly chosen parameter vectors, referred to as particles. Those particles that decrease the acceptance threshold are investigated further, while those that do not are rejected. In theory, the iterative process results in the particles converging to a sample that forms the posterior distribution of parameters.
1.3.4.5 Model Comparison using Bayes Factors

Another useful feature of the ABC approach is the ability to formally compare the relative performance of different models using Bayes Factors (Kass and Raftery 1995). A Bayes Factor is a summary of the evidence provided by the data in favour of one model over another; it is a direct estimator of the marginal posterior probability of a model, which allows for straightforward and intuitive comparison of the relative performance of proposed models (Kass and Raftery 1995).

Formally, given models $M_0$ and $M_1$, not necessarily with the same number of parameters, Bayes Factor $B$ is given by:

$$B = \frac{P(D|M_1)}{P(D|M_0)} = \frac{P(M_1|D)P(M_0)}{P(M_0|D)P(M_1)},$$  \hspace{1cm} (1.3)

where $P(M_i)$ is the prior probability of model $M_i$, $P(D|M_i)$ is the probability of data $D$ given model $M_i$ and $P(M_i|D)$ is the posterior probability of the model, defined using Bayes’ theorem as stated in equation (1.1).

An important advantage of this form of model comparison is that it is independent of the parameters for each model, and instead estimates the likelihood of the model considering all possible parameter values. As is further explained in section 2.2.2.1, it follows that, in cases where models with different numbers of parameters are compared, this method automatically and correctly penalises model complexity and therefore prevents overfitting (i.e. invoking parameters to explain aspects of the data that are in fact due to randomness).

It is worth noting that this probability may not be very accurate if the number of retained simulations is very small, which is dependent on the acceptance threshold $\epsilon$, or if these include simulated datasets not closely matching the observed data (Robert et al. 2011), which is dependent on the choice of summary statistics. Alternative methods for estimating the relative marginal posterior probabilities of proposed models include adjusting the Bayes factor using a weighted multinomial logistic regression (Fagundes et al. 2007; Beaumont 2008; Cornuet et al. 2008), and using sequential Monte Carlo methods for model selection (Toni et al. 2009).
1.4 In this thesis...

1.4.1 Aims of Study

As explained in section 1.1, in this study, spatially explicit cultural transmission simulation models that generate simulated material culture data under various cultural evolutionary and demographic scenarios have been developed and explored through simulation. The archaeological dataset of Upper Palaeolithic bead types identified as personal ornaments published by Vanhaeren and d'Errico (2006) is used to assess the validity of each model. Approximate Bayesian Computation (ABC) techniques are used to compare the differences in goodness-of-fit between the observed data and data simulated by different proposed models.

The underlying principle of this approach is that conditions under which the simulated data is very similar to the observed archaeological data – as reflected in a range of spatial statistics describing the distribution of artefact types – are more likely to be true than conditions under which the simulated data is unlike the observed data. Once the best-performing model is identified, parameter values under which that model generates datasets closest to the observed data form a sample of the posterior probability distributions of the parameters, from which inferences are made on the likely values of the parameters of interest.

The primary aim is to test whether the distribution of artefact types reported by Vanhaeren and d’Errico (2006) can be explained by a simple model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that may be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving cultural interactions. The distinction here is that between the existence of inter-group differences arising through cultural mutation and drift (the null model), of which actors are not consciously aware, versus that null model plus the intentional adherence to behavioural norms that imply identity and actively shape interaction processes, and through that, the spread, loss and mixing of culturally inherited traits.

1.4.2 Challenges of Adopted Approaches

Taking the simulation modelling approach described in section 1.4.1 necessitates reduced models and these, by definition, will never fully describe the complexity of the true processes that shaped the material culture data. Despite these necessary simplifications, this is a formal scientific approach that proposes a model with an explicit prediction of the distribution of variation in material culture data, and tests this formally by comparing summary statistics describing the simulated data to those describing the observed data for validation. This is an essential step in interpretation of the distribution of this material culture data since, as Kuhn (2013: 208) points out, apparently complex large-scale phenomena can arise “as a function of
simple transmission rules operating on bounded social networks”, thus other, and simpler, processes accounting for the observed patterning need to be considered and rejected. Simulation modelling within the Bayesian ABC framework provides a means of doing this.

One of the challenges is defining the processes that may have given rise to the observed distribution of material culture. Beyond inferring ethnic structuring, Vanhaeren and d’Errico (2006) did not make any explicit speculation about the specifics of the processes leading to the observed pattern. However, in order to be able to test different hypotheses as explanations of the observed variation, such processes need to be sufficiently well defined so that data can be simulated. The approach adopted here is to draw on the idea that ethnicity may be expected to homophilically drive cultural interaction processes, and build on the definitions proposed by Axelrod (1997) (briefly introduced in section 1.3.1.7) to explicitly define the nature of these interactions.

Another challenge of this approach is the quality and sourcing of raw material. The large majority of sites in the personal ornament database used in this study “were excavated long ago” (Vanhaeren and d’Errico 2006), and a fundamental issue is the poor quality of older excavations, including the absence of stratigraphic and spatial context (White 2007). Additionally, given the complexity of the problems that these data are used to address, the amount of material culture and other data for the period is relatively small.

A further related drawback of the approach adopted here is that it does not explicitly incorporate temporal information for the time period studied. The modelling framework does not ignore the fact that there is temporal variation – it simply does not use temporal information to inform on the outcomes of the models due to the lack of temporal resolution in the observed data and for the time period in general. The effects of temporal variability on model outcomes are briefly discussed in Chapter 6 in relation to variability over time in model processes, and in section 4.1.4 in relation to accumulation of material culture over time.

Despite these drawbacks, the marriage of simulation modelling and ABC approaches is one that enables testing and comparison of a range of hypotheses concerning group interaction dynamics for the period of interest in a formal manner, and subsequently facilitates the estimation of key demographic and evolutionary parameters.

1.4.3 Thesis Structure

There is a historical component to the structure of this thesis that is the result of the model development process; development of and refinements to the modelling framework were made as a consequence of results of initial versions of the same. The thesis is therefore separated into three main chapters (Chapter 2, Chapter 3 and Chapter 4).
Chapter 2 describes the underlying model developed to test and compare a range of hypotheses concerning group interaction dynamics during the Aurignacian, as well as the implementation of the ABC framework. The results of the first applications of this modelling framework are presented in Chapter 3.

Following on from the results in Chapter 3, substantive methodological developments and model refinements, and the results of these, are presented in Chapter 4. These focus on incorporating climatic and topographic information into the modelling framework.

Chapter 5 discusses my contributions to two collaborative projects; my involvement in these studies was in part a by-product of the main research that this thesis is concerned with. The first study examines a hypothesis that cytochrome P450 3A5 (CYP3A5), an enzyme involved in the metabolism of many therapeutic drugs, is important in salt retention adaptation. My contribution to this work focussed on integrating climatic information into the existing analysis framework. The second study is concerned with examining the rate of gene flow over time and space through analyses of craniometric data, with the aim of informing on the extent of past population migration activity. My contribution to this project involved the development of a simple spatiotemporally explicit model, based on the simulation modelling approaches described in Chapter 2 and Chapter 4, to simulate variation in craniometric measures.

Finally, Chapter 6 presents a general discussion and some conclusions of the work, highlighting how the approaches contribute to understanding and interpretations of distributions of material culture.
Chapter 2: Methods

This chapter is an expanded version of the methods section of the article attached in Appendix B, which has been published as:


The modelling framework presented in the article was developed and implemented by Mirna Kovacevic, with guidance from Mark G. Thomas and Stephen Shennan in a supervisory capacity. Marian Vanhaeren and Francesco d’Errico provided published archaeological data (Vanhaeren and d'Errico 2006).

Summary

This chapter focuses on describing the underlying spatially explicit cultural transmission simulation framework developed in this thesis, and the implementation of Approximate Bayesian Computation (ABC) for integrating the modelling framework with observed material culture data. The results of the simulation models described in this chapter are presented in Chapter 3.

Please note that there are additional methods in Chapter 4 that describe further refinements following on from the results in Chapter 3. These are substantive methodological advances that were developed and implemented after the initial application of the model described in this chapter, and have therefore been placed in a discrete methods section in Chapter 4.

Unless otherwise specified, the simulation modelling framework was implemented in the Python programming language (http://www.python.org/). ABC analyses were implemented in the R programming language (http://www.r-project.org/).
2.1 Simulation Modelling

The models described in this chapter are spatially explicit models of cultural identity-by-descent with modification and isolation-by-distance, and are made up of parameterised demographic (i.e. migration and fission / extinction) and cultural (i.e. cultural mutation and, in some models, cultural interaction) group level processes iterated forward in time within a geographically realistic domain.

2.1.1 Simulation World

2.1.1.1 Geographic Region

The geographic region considered in this study is the range of latitudes and longitudes corresponding to the European territory. The longitude, $\varphi$, ranges from $-11^\circ$ to $30^\circ$, which relative to modern day country boundaries is approximately the area from the western Irish boundary to the western Russian boundary at the Urals. The latitude, $\lambda$, ranges from $35^\circ$ to $60^\circ$, which is approximately the area from the northern boundary of Africa to the northern boundary of Scotland.

In the simulations described in this thesis, the geographic region is modelled as a demic grid at a resolution of 0.2° latitude and 0.2° longitude. The grid is therefore made up of 38,656 demes, 24,104 of which are land demes and 14,552 of which are sea demes, arranged to approximate the European landmass, as shown in Figure 2.1.

Although it may be possible to incorporate changes in sea levels through time by using available bathymetry data, dramatic geostatic rebound for northern latitudes makes it difficult to accurately estimate changes in coastlines for northern Europe. For this reason, modern coastlines are used in simulations.

2.1.1.2 Length of Simulation

Each simulation is initialised at the onset of the Aurignacian period, approximately 42 Ka, and simulated forward in time to the end of the Aurignacian period, approximately 29 Ka (Higham et al. 2012). Each simulation therefore spans a total of 13,000 years, or 520 generations assuming a 25 year generation time (Tremblay and Vezina 2000; Thomas et al. 2006).

Each simulation includes a 1,000 year, or 40 generation, burn-in period at the start of the simulation during which no simulated data is collected. Results from early stages of model development indicate that the global carrying capacity (i.e. the total maximum number of groups that the modelled domain can sustain, specified by the $N_{groups}$ parameter) is reached within a relatively small number of generations (in the majority of simulations within 40
generations from the onset of the simulation). Since this study is concerned with material culture distributions and the underlying carrying capacity distributions (rather than the migration routes of the first anatomically modern human populations into Europe at the onset of the Upper Palaeolithic, about which little is known), including a burn-in period that lasts until the global carrying capacity stabilises allows for uncertainties in the initial locations of simulated groups.
Figure 2.1: European landmass represented as a demic grid at a resolution of 0.2° latitude and 0.2° longitude, totalling 38,656 demes (24,104 land demes and 14,552 sea demes).
2.1.1.3 Carrying Capacities and Environmental Scenarios

Within the defined region, each geographic location is assigned a local carrying capacity. The carrying capacity of a location determines the habitability, and therefore potential population density, of that location; a zero carrying capacity corresponds to an uninhabitable region, for example sea or ice covered land. The local carrying capacity values are normalised by the maximum observed local carrying capacity value for the modelled domain to generate a distribution of relative carrying capacity values.

The potential, or target, population density for each location is calculated as the product of the relative carrying capacity at that location and the $N_{\text{groups}}$ parameter, which specifies the total maximum number of groups that the modelled domain can sustain. This parameter is treated as unknown and a range of values are explored.

In order to estimate the local carrying capacities for geographic locations in the modelled domain, two distinct environmental scenarios have been considered; each simulation is conditioned on only one of these two environmental scenarios.

2.1.1.3.1 Uniform Carrying Capacities (FLAT)

The first is a simple scenario in which Europe is assumed to be a flat space. This corresponds to a uniform distribution of relative carrying capacity values across all locations within the modelled domain (Figure 2.2, panel (A)).

2.1.1.3.2 Population Density Estimates from Bocquet-Appel et al. (2005) (B-A)

In the second of the environmental scenarios, instead of treating Europe as a flat space, information on estimated population densities during the Aurignacian has been taken from Bocquet-Appel et al. (2005), to inform on carrying capacities for the modelled domain. Bocquet-Appel et al. (2005) used databases of archaeological sites corresponding to the Upper Palaeolithic period, together with simulated climatic variables and ethnography of hunter-gatherers, to estimate the distribution of hunter-gatherer populations in Upper Palaeolithic Europe.

The original distribution of estimated population densities for the Aurignacian period (Figure 2.2, panel (C)) was not made available for this study, so the distribution used is approximated based on the original figure. Since this study is concerned with relative distributions rather than exact numbers estimated in the original study (Bocquet-Appel et al. 2005), this estimate is normalised by the maximum observed value to give relative distributions. The normalised distribution (Figure 2.2, panel (B)) is used in simulations as the relative carrying capacity value for each location.
Figure 2.2: Map of the simulation space, showing the demic distributions of carrying capacities for the environmental scenarios considered: (A) uniform carrying capacity values (FLAT); (B) normalised distribution of relative carrying capacity values based on the population density estimates for the Aurignacian period from Bocquet-Appel et al. (2005) (B-A); (C) (Image and caption edited from Bocquet-Appel et al. (2005) and reproduced with permission from Elsevier) Estimate of the regional distribution of the metapopulation of hunter-gatherers during the Aurignacian period of the Upper Palaeolithic in Europe, superimposed on the IOS3 project maps. The boundaries (in black) of the accretion zones, with the corresponding numbers, account for roughly 90% of the distribution of the local population.
2.1.2 Demographic Processes

Each simulation is initialised with a fixed number of groups \((N_{in, groups}=20)\) placed in randomly chosen habitable locations in the modelled domain; all attributes and processes are defined at the level of the group, rather than individuals in that group, and groups are assumed to be the same size. Groups migrate locally and undergo fission/extinction processes. These demographic processes are analogous to an isolation-by-distance model in population genetics (Wright 1943; Slatkin 1993).

2.1.2.1 Migratory Processes

At each generation, groups are subjected to migratory processes modelled as parameterised Gaussian random walks; these occur in continuous, rather than demic, space. Modelling migratory processes in continuous, rather than demic, space is a trade-off between computational resources on the one hand and archaeological material culture data availability and realism on the other. In the case of this study, movement on a demic level would be an oversimplification and, while it is marginally more computationally expensive to model, movement in continuous space is more natural. Additionally, it would be difficult to resolve the material culture distributions at individual archaeological sites in the context of demic movement since there are several clusters of sites in which all sites belong to a single deme at the resolution used in this study (0.2° latitude and 0.2° longitude as detailed in section 2.1.1.1); similar problems arise when modelling cultural interactions detailed in section 2.1.3.4 in the context of demic movement.

The distance each group traverses in a migration process is calculated based on two independent picks from a normal distribution with mean \(\mu_{mig}\) and standard deviation \(\sigma_{mig}=d_{mig}\). Positive and negative values picked from the distribution correspond to movement in opposite directions, namely East and West, and North and South, respectively. The mean of the distribution is therefore set to \(\mu_{mig}=0\) to ensure that movement in opposite directions is equally likely. Parameter \(d_{mig}\) corresponds to the standard deviation, or width, of the normal distribution and specifies the range of values that the migration distance is most likely to take in each of the East-West and North-South directions. Parameter \(d_{mig}\) is treated as unknown and a range of values are explored.

The distances travelled by each group at each generation in the East-West and North-South directions, denoted \(\Delta x\) and \(\Delta y\) respectively, are represented as independent picks from the above-described normal distribution. The resultant distance, \(d\), and direction, \(\theta\), that define each group's movement are then calculated as:

\[
d = \sqrt{(\Delta x)^2 + (\Delta y)^2}, \text{ and } (2.1)
\]
where $\text{arctan2}$ corresponds to a variant of the $\text{arctan}$ function that takes into account the sign of both vectors in question (i.e. $\Delta x, \Delta y$) and distinguishes diametrically opposite directions (i.e. East and West, North and South), therefore specifying unique angle values in the range $(0, 2\pi)$.

The new proposed location of each group is then calculated based on the group’s current location, the distance, $d$, and the direction, $\theta$, of movement. If the longitude and latitude of the group’s current location are $\lambda_{\text{current}}$ and $\phi_{\text{current}}$, respectively, and the same for the group’s new location are $\lambda_{\text{new}}$ and $\phi_{\text{new}}$, respectively, then:

$$
\lambda_{\text{new}} = \sin^{-1} \left( \sin(\lambda_{\text{current}}) \cdot \cos \left( \frac{d}{R} \right) + \cos(\lambda_{\text{current}}) \cdot \sin \left( \frac{d}{R} \right) \cdot \theta \right),
$$

(2.3)

$$
\phi_{\text{new}} = \phi_{\text{current}} + \text{arctan2} \left( \cos \left( \frac{d}{R} \right) + \cos(\lambda_{\text{current}}) \cdot \sin \left( \frac{d}{R} \right) \cdot \theta, \cos \left( \frac{d}{R} \right) - \sin(\lambda_{\text{current}}) \cdot \sin(\lambda_{\text{new}}) \right).
$$

(2.4)

These formulae allow for the curvature of the Earth when calculating the new group locations. Although the curvature of the Earth has little effect in the current framework, as the migration distance $d$ is small relative to the radius of the Earth (denoted $R$ and assumed to be constant at 6,371 km), using these formulae ensures that the model can be applied accurately with arbitrarily large migration distances.

Once calculated, the viability of the proposed location is checked; if the proposed location is within a land deme, it is accepted as the group’s new location, however, if it is within a sea deme, the group remains at its current location.

### 2.1.2.2 Fission / Extinction Processes and Kernel Density Estimation (KDE)

In addition to the migratory process undergone at each generation, each group also undergoes a fission / extinction process with parameterised probability. The probability that a group undergoes a fission / extinction process is given by the probability of fission / extinction parameter, $p_{f/e}$; this parameter is treated as unknown and a range of values are explored. The type of process (i.e. fission or extinction) that a selected group undergoes is determined by the difference between target and current local population density at the group’s location. The population density at the current generation is estimated from the group locations using kernel density estimation (Wand and Jones 1995), specifically the $\text{bkde2D()}$ function from R’s $\text{KernSmooth}$ package. For each group, this function uses a bivariate Gaussian kernel centred on that group’s location to estimate the local density as a result of that group, with the global distribution of densities calculated from these local density estimates.
The difference between target and current local population density at the group’s location is an indicator of potential for local population growth: a positive value indicates that the location is below carrying capacity (i.e. the target local population density is greater than the current local population density – the location is under-populated and so there is potential for growth) and therefore results in a fission event, while a negative value indicates that the location is above carrying capacity (i.e. the target local population density is smaller than the current local population density – the location is over-populated and so there is no potential for growth) and therefore results in an extinction event.

An extinction event results in the group being deleted from the simulation, while a fission event results in a replication such that two groups, the parent and offspring, are present in the next generation. The offspring group retains the cultural traits of the parent group (i.e. the offspring group is an exact replica of the parent group except for any mutation events as described in section 2.1.3.3), analogous to identity-by-descent in population genetics. In subsequent generations, the parent and offspring groups migrate and undergo fission / extinction processes independently, and their respective cultures also evolve independently.

2.1.3 Cultural Processes: Modelling Ethnic Diversity

2.1.3.1 The Axelrod (1997) Model of Cultural Dissemination: Defining Culture

The models developed in this thesis simulate innovation in culture and so require the concept of culture to be mathematically defined. For this purpose, an adapted version of Axelrod’s definition, proposed in the study briefly discussed in section 1.3.1.7 (Axelrod 1997), has been used. In Axelrod’s definition, the culture of an agent (an individual or a group of individuals) is defined to be a set of attributes that are subject to social influence; the culture of an agent consists of some number of these attributes, referred to as cultural features, and each can assume one of a predefined number of values, referred to as traits, thus, each agent is monomorphic for each cultural feature. In this definition, the culture of an agent is then described as a list of digits, with the position of a digit corresponding to the feature and the value of a digit specifying the current trait for that feature. In Axelrod’s definition, the trait – or value that a feature takes – is assigned at the start of the simulation and is only influenced by social interactions (i.e. it does not undergo any mutation processes).

In Axelrod’s formulation, social interactions are constrained to occur only between agents that are immediate neighbours. The simulations occur on a square lattice with agents arrayed at discrete points over the lattice. Most agents therefore have four immediate neighbours, with those on the edge of the lattice having three and those in the corners having two immediate neighbours. Also in Axelrod’s model, the probability of an interaction between two agents is proportional to the cultural similarity between them. This similarity is quantified as the
proportion of their features that have the same trait. The interaction then consists of an agent, and an immediate neighbour to that agent, being chosen at random. A single feature on which the chosen agent’s and neighbour’s cultures differ is selected at random, and the value of this feature (trait) in the chosen agent’s culture is set to the value of the same feature in the neighbour’s culture.

This formulation is a good basis; however, it is very limited in diversity of cultural features and traits and is inadequate to capture the high dimensionality of the observed data used in the current study (Vanhaeren and d’Errico 2006). In addition, over long chronological periods such as those simulated here, it is necessary to consider the effect of cultural mutation and drift processes. This definition must therefore be modified so that it can be applied to the current problem.

2.1.3.2 Observed and Simulated Datasets

The observed dataset (Vanhaeren and d'Errico 2006) consists of 157 distinct ornament types recorded at 98 Aurignacian sites in Europe and the Near East, with records specifying presence / absence of distinct ornament types in sites only. These distinct ornament types are divided between 11 features according to different raw materials, with “62 representing ornaments made of shells, 31 of teeth, 30 of ivory, 11 of stone, 11 of bone, 7 of deer antler, and one each of belemnite, nummulite, ammonite, sea urchin and amber” (Vanhaeren and d’Errico 2006).

For the purposes of the models developed in this study, Axelrod’s definition of culture, described in section 2.1.3.1, has been adapted so that each agent, in this case a group, is polymorphic for each cultural feature. To allow for this, each group carries a parameterised number of items, or ornaments, in its cultural repertoire, specified by the $N_{\text{items}}$ parameter, treated as unknown and chosen at the onset of each simulation from a pre-defined range of values. The number of items each group carries is selected once per simulation for two main reasons. Firstly, changing the value of this parameter during this simulation implies that groups’ item capacity changes over time; there is no justification for this change since there is little information in the archaeological record that can act as a direct proxy for the value of this parameter, or for when and how many times the parameter value may have changed during the time period modelled. Secondly, the design choice is partly the result of computational constraints, as incorporating an update of the parameter value becomes computationally intractable relatively quickly. The prior range for this parameter is set to be relatively wide with 1,000 distinct values explored (see Table 2.1 for details of prior ranges of parameters), and changing the value during the simulation would have a multiplicative effect on the size of the parameter space, increasing the number of values explored to 1,000,000 if one update of the parameter value is made, to 1,000,000,000 if two updates are made, and so on.
The chosen number of items are then divided between the 11 features probabilistically (using a multinomial function), such that the probability of an item being assigned to a particular feature is proportional to the ratio of unique items observed in that feature and the total number of unique items observed (39.5% shells, 19.7% teeth, 19.1% ivory, 7% stone, 7% bone, 4.5% deer antler, and 0.6% each for belemnite, nummulite, ammonite, sea urchin and amber). Within each feature, each item can then take one of a number of unique possible values (traits), corresponding to the number of distinct ornament types for that feature in the observed data (Vanhaeren and d'Errico 2006). The number of items allocated to each feature is divided between the feature traits probabilistically (using a multinomial function) such that each trait is equally likely at the onset of the simulation.

2.1.3.3 Mutation and Drift

The culture of each group undergoes mutation and drift processes at each generation, such that the culture of each group at the next generation will be the combined result of mutation and drift processes acting on the culture of that group at the current generation. What is referred to here as mutation corresponds to innovation in the terminology of cultural evolution theory. In addition to testing various environments as described in section 2.1.1.3, two different models of cultural variation have been considered; in each simulation data is simulated under only one of these two cultural variation models.

It is important to note here that, other than the observed data used in the current study (Vanhaeren and d’Errico 2006), there is little archaeological information available that can be used to inform on Aurignacian ornament types and the features that they categorise into. As a result, an underlying assumption of the models described in this thesis is that the personal ornament dataset published by Vanhaeren and d’Errico (2006) is representative of the diversity and relative proportions of ornament types that were present across the geographic domain during the time period considered, and that the features that these types categorise into represent static and discrete classes of categorical material culture data. This assumption is incorporated into the models by dividing the number of items each group carries (i.e. $N_{items}$) between the 11 features in the observed data probabilistically at the onset of each simulation, as described in section 2.1.3.2, and subsequently restricting mutation to within each feature, as described in this section, in order to maintain these proportions.

2.1.3.3.1 Bounded Stepwise Mutation Model (SW)

In the first model of cultural variation, mutation is modelled according to the bounded stepwise model often used to model mutations at microsatellite loci in population genetics (Kimura and Ohta 1978; Valdes et al. 1993), and occurs at each generation for each item in each group’s culture with probability proportional to the $p_{mu}$ parameter. This parameter, which specifies the probability of mutation, is treated as unknown and a range of possible values are explored.
Under this stepwise mutation model, a cultural trait in a particular feature at the current generation is constrained to mutate to one of the cultural traits on either side of it, within that feature, at the next generation – mutation therefore changes the frequency with which each trait occurs in the next generation.

The cultural trait data considered in this study is necessarily categorical. Since categorical distinctions are often arbitrarily made and cultural change may in fact be scalable, the stepwise mutation model is intended to represent a midway between discrete and continuous variation, in an attempt to relax the assumption of categorical difference. In this mutation model, cultural traits are assumed to be ordered in such a way that adjacent traits are more similar than traits that are further apart in the sequence. Since cultural traits considered are discrete and fixed (i.e. one trait cannot morph into another trait), mutation under this model corresponds to a group being more likely to add an item to its cultural repertoire that is morphologically similar to one that is already present in its cultural repertoire than one that is very different. Similarly to population genetics, cultural mutation has the effect of increasing diversity.

2.1.3.3.2 **Bounded Non-Stepwise (Discrete) Mutation Model (DIS)**

In the second model of cultural variation, mutation is discrete within the bounds of each feature. Similarly to the stepwise mutation model, in this bounded non-stepwise (discrete) model mutation occurs at each generation for each item in each group’s culture with probability proportional to the \( p_{\text{mut}} \) parameter. This parameter again specifies the probability of mutation; it is treated as unknown and a range of possible values are explored. Under this mutation model, however, a cultural trait in a particular feature at the current generation is permitted to mutate to any of the other cultural traits within that feature with equal probability at the next generation. The mutation process again changes the frequency with which each trait occurs in the next generation and has the effect of increasing diversity.

2.1.3.3.3 **Drift**

Drift has the opposite effect to mutation and decreases the amount of diversity in each group’s culture. Here, cultural drift is modelled as a direct analogue to genetic drift, where allele frequencies change as a result of random differences in reproduction. In finite populations genetic drift corresponds to the intergeneration sampling error (see, for example, Tishkoff and Verrelli (2003)); analogously, in populations of finite size, cultural drift results in stochastic fluctuations over time in cultural variant frequencies due to sampling error, which can lead to drastic changes in frequencies, including loss, of cultural traits.

The drift process is modelled by using a multinomial function to sample the traits of each cultural feature independently. This implementation takes into account frequencies of cultural traits in the current generation, such that, for a particular group, cultural traits that are at higher frequencies in the group’s culture at the current generation are more likely to be present in the group’s culture at the next generation.
2.1.3.4 Cultural Interactions

The nature of cultural interaction – representing the hypothesis of culture symbolically marking ethnic identity and homophilically driving cultural interactions – is at the crux of this study.

In this thesis, ethnicity and cultural interaction processes are linked through homophily, the tendency of people with similar traits to preferentially interact with each other rather than with people with dissimilar traits. As explained in section 1.3.1.7, this implies that, if populations are ethnically structured and the extent of ethnic similarity (reflected by the extent of similarity in material cultures) affects homophily (which is in turn actioned through cultural interaction processes), then investigating the effect of cultural interaction processes that are homophilically driven by similarities and differences in material culture allows the investigation of ethnic structuring.

Thus, if ethnicity exists and is reflected by similarities and differences in material culture, then similarities and differences in groups’ cultures signify intentional adherence to behavioural norms that imply identity and actively shape interaction processes, and through that, the spread, loss and mixing of culturally inherited traits.

For instance, interactions between culturally similar groups, examples of which include exchange of material culture and cultural convergence, would be expected to cause those groups to become overall more culturally similar. In order to account for these homogenising processes and their effects, a process referred to as ‘sharing’ is modelled as described in section 2.1.3.4.2. Conversely, to account for interaction processes between culturally different groups, examples of which include physical conflict, cultural coercion and cultural dominance, a process referred to as ‘conflict’ is modelled as described in section 2.1.3.4.3.

The interaction processes modelled here represent a multitude of different possible interaction processes and their effects, and the names ‘sharing’ and ‘conflict’ are effectively only labels for these groups of processes.

2.1.3.4.1 Interaction Distance

A pair of groups will interact if they are within a parameterised geographical distance, $d_{int}$, of each other. While it is possible that a functional relationship exists between interaction distance and the underlying carrying capacity values, in that changes in the carrying capacity values would effect changes in the value of interaction distance, there is little information in the archaeological record about the extent and dynamics of this relationship. The interaction distance is therefore parameterised as a random variable – it is treated as unknown and a range of values are explored.
The distance between each pair of groups is quantified using the geodesic distance measure, which is the aerial path between two points, also called the as-the-crow-flies, great-circle or orthodromic distance. To account for curvature of the Earth, geographic distances are calculated using the Haversine Formula (Sinnott 1984). This calculates the great-circle distance between two points on a sphere given their respective longitudes and latitudes. If the longitude and latitude of the points are $\phi_1$ and $\lambda_1$ for point one and $\phi_2$ and $\lambda_2$ for point two, respectively, and:

\[ \Delta \phi = \frac{\phi_1 - \phi_2}{2}, \]  
\[ \Delta \lambda = \frac{\lambda_1 - \lambda_2}{2}, \]

then the distance, $D$, between the two points is calculated as:

\[ D = 2R \sin^{-1}\left( \sin^2(\Delta \phi) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2(\Delta \lambda) \right)^{1/2}, \]

where $R$, the radius of the Earth, is assumed to be constant at 6,371 km.

Two cultural interaction processes are modelled in this framework – ‘sharing’ and ‘conflict’ – and a pair of interacting groups will undergo one of these two processes.

2.1.3.4.2 ‘Sharing’

One of the interaction processes considered is ‘sharing’ of cultures between interacting groups. ‘Sharing’ is modelled as a permutation, whereby the cultural traits of the two interacting groups are pooled, permuted and then divided between the two. This is analogous to culturally similar groups swapping cultural traits.

This process is derived from the interaction process modelled by Axelrod (1997) as discussed in sections 1.3.1.7 and 2.1.3.1. It is based on the ideas that “people are more likely to interact with others who share many of their cultural attributes, and interactions between two people tend to increase the number of attributes they share” (Axelrod 1997: 206-207).

2.1.3.4.3 ‘Conflict’

The other interaction process considered is ‘conflict’. While the ‘sharing’ process discussed in section 2.1.3.4.2 is derived directly from the cultural interaction process modelled by Axelrod (1997), the ‘conflict’ process is introduced in order to account for interaction processes that may have occurred between culturally different groups.

While Axelrod’s model is abstract (i.e. it does not specifically relate to any particular time period or geographic area) and its principles can be applied to cultural interactions within a wide
variety of contexts, the basis of the model is social influence, the definition of which is likely different depending on the spatial, temporal and social context considered. Nuances of context from which Axelrod’s model stems relate to concepts that are more relevant to recent time periods than that considered in this thesis, such as state formation and development of transnational institutions. The cultural interaction model described by Axelrod therefore needs to be modified so that it is applicable to the current problem; this is achieved by introducing the concept of the ‘conflict’ interaction process between groups that are culturally different, which aims to represent boundary-forming cultural interactions that are more relevant to the time period modelled.

The outcome of a ‘conflict’ process is the replacement of the culture of one group by that of the other. To model this, one in each pair of interacting groups is assigned as a ‘winning’, and one as a ‘losing’ group, and the culture of the ‘losing’ group is replaced entirely by that of the ‘winning’ group. The decision on which is assigned to be the ‘winning’ group, and which the ‘losing’ group, is made at random due to the assumption that the aspects of material culture considered (personal ornamentation) do not have an effect on, and are not a proxy for, group fitness (it should be noted here that group fitness can be defined in a multitude of different ways, including, for example, as the average individual fitness in a group). Additionally, since groups are modelled such that they are the same size, group size cannot be used as a proxy for group fitness. The ‘conflict’ interaction process is analogous to a group imposing its culture upon another group, or, alternatively, assimilating the group into their own, followed by a fission process.

The adopted approach is only one way to model the outcome of interactions between culturally different groups. An alternative that could be considered is anti-conformism, whereby culturally different groups become more different upon interaction, for example through moving their trait values in opposite directions. This outcome of an interaction process would represent an expression of individual identity, which is potentially a form of assertive style as discussed in section 1.2.4. This is an alternative mechanism of cultural boundary formation, and is suggested in Chapter 6 as a possible extension of the work presented in this thesis.

### 2.1.3.5 Collecting Simulated Data (Single Occupancy Sites)

Locations of archaeological sites in the model are defined to correspond to the locations of the archaeological sites in the observed data (Vanhaeren and d’Errico 2006), shown in Figure 2.3. A group will deposit its culture at a site when within a specified geographic distance of that site. This catchment distance is initially set to be equal for all sites, and updated iteratively for each site to be the distance of the group that deposited material closest to that site, with the further constraint that if two groups are within the catchment distance then the group closest to the site will be the one to deposit its culture there. The distance between group locations and
archaeological sites is quantified using the geodesic distance measure, and the Haversine Formula (Sinnott 1984), as described in section 2.1.3.4.1, is used to account for curvature of the Earth.

There is little chronological resolution within the time period considered for the Vanhaeren and d’Errico (2006) dataset; the authors explain that this is partly because the majority of sites in the database are older excavations with no stratigraphic context, and partly due to the low resolution of radiocarbon dating for the time period. As a result, little account is taken of the temporal sequence of deposition in the model: each site is considered to be single occupancy – only the material culture of the group that deposited material closest to that particular site is considered. A group arriving at a site at which a deposit has previously been made will deposit its culture at that site and overwrite the existing deposit if it comes within closer proximity to that site than the last group that deposited its culture there. (The inconsistency with the cumulative aspect of the archaeological record of the assumption that each site is single occupancy is addressed in section 4.1.4, by considering the accumulation of material culture deposits at archaeological sites).

Each site is assigned an item capacity, which corresponds to the number of items recovered from that site as reported in the original study (Vanhaeren and d’Errico 2006). When a group comes within the specified distance for a particular site, a number of unique items, equal to the item capacity for that site, are selected probabilistically (using a multinomial function so that trait frequencies are taken into account) from the group’s entire culture to be deposited at the site – this corresponds to one copy of each cultural trait that is selected being deposited at the site. The original dataset contains presence / absence records of distinct ornament types only, so specifying the item capacity and ‘uniqueness’ of items deposited at each site in simulations should theoretically minimise archaeological and sampling bias.

Simulated material culture data deposited at the locations of the archaeological sites in the observed data (Vanhaeren and d’Errico 2006) are collected at the end of each simulation, which corresponds to the end of the Aurignacian period.
Figure 2.3: Map of the simulation space, showing the geographic locations (red dots) of archaeological sites in the observed data (Vanhaeren and d’Errico 2006), which are defined to be the archaeological sites in the model (i.e. locations where simulated data is deposited).
2.1.4 The Models: Null Model (NM), Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD)

The model initially considered as the null model for this framework is the Null Model (NM), a model of cultural identity-by-descent with modification and isolation-by-distance, which is made up of the demographic and cultural processes described in sections 2.1.2 and 2.1.3, respectively, with the exception of the interaction processes described in section 2.1.3.4 (i.e. no interaction processes are modelled). However, for reasons explained below, it was decided that the Null Model with Cultural Interactions (NMCI) is a fairer choice of null model for testing the proposed hypothesis under the described framework.

The Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD) are both models of cultural identity-by-descent with modification and isolation-by-distance, and are made up of the demographic and cultural processes described in sections 2.1.2 and 2.1.3, respectively. The difference between the two models lies in the role played by material culture in determining the nature of interactions between groups. At the suggestion of Bill Croft (personal communication), both models specify that group interactions occurred, but in the NMCI model the outcomes of these interactions are independent of similarities or differences in groups’ ornamental material cultures, while in the CD model their outcomes are determined based on the extent of these similarities or differences.

Figure 2.4 is a flow diagram illustrating the order in which the demographic and cultural processes described in sections 2.1.2 and 2.1.3, respectively, are performed during each simulation. The processes grouped by the dashed grey line are iterated at every generation.

The rationale behind choosing the NMCI model as the more suitable null model is that this study is testing whether the observed data supports the hypothesis that inter-group interactions were driven by cultural similarities and differences between interacting groups. The appropriate null model for this hypothesis is therefore one in which interactions are not driven by cultural similarities and differences between groups, but are instead random. Additionally, while it could be argued that an appropriate null model could either be a scenario in which group interactions do not occur at all (as is the case in the Null Model), or one in which interactions are not driven by cultural similarities and differences between groups (as is the case in the Null Model with Cultural Interactions), it is important to make the null model as similar to the model including the processes under investigation as possible, but without including those processes. This is considered to be the Null Model with Cultural Interactions.

The difference between the NMCI and CD models is formalised in the process of deciding which type of interaction will occur between two interacting groups. In the NMCI model, the type of interaction is decided at random; groups are equally as likely to ‘share’ material culture as they are to undergo ‘conflict’. This therefore models a scenario in which group interactions are independent of similarities or differences in groups’ ornamental material cultures. Conversely,
in the CD model, the type of interaction is decided probabilistically and depends on the extent of cultural similarity between the two interacting groups. The extent of culturally similarity therefore homophilyically drives interaction processes such that groups that are relatively culturally similar are more likely to ‘share’ cultures while those that are relatively culturally different are more likely to undergo ‘conflict’.

The overall aim is to test which of these models best explains the spatial distribution of ornament types observed in the archaeological record, and the CD model is intended to represent the effects of ethnic structuring on the spatial distribution of material culture. Intuitively, it is expected that the CD model will generate simulated datasets with greater spatial structuring of material culture – if ‘conflict’, and therefore loss of cultural variation, is more likely to occur at boundaries of cultural regions, then this would be expected to lead to further spatial structuring. Additionally, loss of variation due to ‘conflict’ will increase the effect of cultural drift, and so generate further regional differentiation (Henrich and Boyd 1998; Henrich 2004b; Bell 2010). From an archaeological perspective, it has been suggested that representations of cultural difference involved in the articulation of ethnicity may be expected to “disrupt regular spatio-temporal patterning” (Jones 1996: 73) and give rise to boundaries that may be spatially and temporally discontinuous. Greater spatial structuring is therefore expected in simulated datasets generated by the model in which these representations of cultural difference directly influence the type of cultural interaction that occurs (i.e. the CD model).
Figure 2.4: Flow diagram illustrating the order in which the demographic and cultural processes are performed during each simulation. The processes grouped by the dashed grey line are iterated at every generation. The parameter that governs each process is indicated in the box specifying that process. The population density estimation process, shaded in grey, is not parameterised.
2.1.4.1 Measures of Cultural Similarity

As explained in the previous section, in the CD model, the type of interaction (i.e. ‘sharing’ or ‘conflict’) that occurs between two interacting groups depends on the extent of cultural similarity between them; it is therefore necessary to formally quantify the extent of this similarity.

The extent of cultural similarity between a pair of interacting groups is quantified differently depending on which of the models of cultural variation described in section 2.1.3.3 is considered. For both models, the cultural similarity measure \((\delta \mu)^2\) for the bounded stepwise mutation model, \(F_{ST}\) for the bounded non-stepwise (discrete) mutation model) can take values between 0 and 1 and is treated as a probability to decide which interaction process (i.e. ‘sharing’ or ‘conflict’) occurs between the two interacting groups; a value of 0 indicates that the two groups have identical cultural repertoires and are therefore more likely to ‘share’ cultures, while a value of 1 indicates complete cultural difference and means that the two groups are more likely to undergo ‘conflict’.

For both models of cultural variation, the cultural similarity measure captures trait frequency differences rather than feature differences. This is because, as explained in section 2.1.3.3, an underlying assumption of the simulation models in this thesis is that features in the observed data (Vanhaeren and d’Errico 2006) represent static and discrete classes of categorical material culture data, since there is little other archaeological information available that can be used to inform on Aurignacian ornament types and the features that they categorise into. Given that the features are assumed to be static, measuring feature differences would be uninformative as these differences would be constant; dynamics of trait frequency differences are therefore considered. In order to model feature differences, it would be necessary to model changes in preference for different features, and as there is no information available in the archaeological record relating to preference for different features, considering dynamics of trait frequency differences is a more parsimonious approach.

2.1.4.1.1 Bounded Stepwise Mutation Model (SW)

In simulations that follow the stepwise mutation model, the extent of cultural similarity is quantified using a measure akin to the \((\delta \mu)^2\) measure used to quantify the genetic similarity between populations using microsatellite data (Goldstein et al. 1995a; Goldstein et al. 1995b). This cultural \((\delta \mu)^2\) measure is defined as:

\[
(\delta \mu)^2 = \Sigma_i \Sigma_j (i - j)^2 x_i y_j - \frac{1}{2} \left[ \Sigma_i \Sigma_j (i - j)^2 x_i x_j + \Sigma_i \Sigma_j (i - j)^2 y_i y_j \right],
\] (2.8)

where \(x_i\) and \(y_j\) are frequencies of traits \(i\) and \(j\) in (interacting) groups \(x\) and \(y\) respectively. This measure therefore quantifies the cultural similarity between the two interacting groups by taking into account the frequencies with which all traits occur in their respective cultural
repertoires – this measure does not discriminate between differences in cultural features. The calculated value of \((\delta \mu)^2\) is normalised by the maximum \((\delta \mu)^2\) recorded up to that generation of that simulation, giving a measure of cultural similarity that is relative to the maximum measured cultural similarity.

2.1.4.1.2 Bounded Non-Stepwise (Discrete) Mutation Model (DIS)

In simulations that follow the non-stepwise (discrete) mutation model, the extent of cultural similarity between a pair of interacting groups is quantified using a measure akin to the \(F_{ST}\) measure used to quantify the genetic similarity between populations using allele frequency data (Wright 1978; Cavalli-Sforza et al. 1996). The cultural \(F_{ST}\) measure is defined as:

\[
F_{ST} = \frac{H_T - H_S}{H_T},
\]  
(2.9)

In this definition, \(H_T\) is the amount of variation in all traits in the whole population (considered to be the two interacting groups) and is defined as:

\[
H_T = 1 - \sum_i \bar{p}_i^2,
\]  
(2.10)

where \(\bar{p}_i\) is the average frequency of trait \(i\) calculated over the two interacting groups. \(H_S\) is the amount of variation between traits within each group (calculated separately for each of the two interacting groups); \(\bar{H}_S\) is the average of \(H_S\) calculated over the two interacting groups. \(H_S\) is defined as:

\[
H_S = 1 - \sum_i p_i^2,
\]  
(2.11)

where \(p_i\) is the frequency of trait \(i\). Similarly to the \((\delta \mu)^2\) measure, the \(F_{ST}\) measure takes into account the frequencies with which all traits occur in the cultural repertoires of the two interacting groups – as with the \((\delta \mu)^2\) measure, this measure does not discriminate between differences in cultural features. It is therefore an estimate of the proportion of the total variation in a set of traits that is the result of between-group differences (Bell et al. 2009). Similarly again to \((\delta \mu)^2\), the calculated value of \(F_{ST}\) is normalised by the maximum \(F_{ST}\) recorded up to that generation of that simulation, giving a measure of cultural similarity that is relative to the maximum measured cultural similarity.

2.1.5 Model Versions, Model Parameters and Prior Ranges

A total of 10 model versions are simulated under the described modelling framework. For the Null Model (NM), data are simulated under pairwise combinations of the bounded stepwise
SW) model of cultural variation and one of two environmental scenarios (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)); this is a total of two model versions. For each of the Null Model with Cultural Interactions (NMCI) and the Culture-Dependent Interactions Model (CD), data is simulated under all pairwise combinations of one of two environmental scenarios (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and one of two models of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)); this is a total of eight model versions.

In total there are 5 parameters that govern the processes considered in the NM model, and 6 parameters that govern the processes considered in the NMCI and CD models. All three models have 3 key processes: migration, fission / extinction and cultural mutation, governed by 3 parameters: $d_{mig}$, $p_{f/e}$ and $p_{mut}$, respectively. In addition to these, the NMCI and CD models have a fourth process, namely cultural interaction, governed by the $d_{int}$ parameter. There are 2 further parameters in all three models, namely the maximum number of groups, $N_{groups}$, and the number of items in each group’s cultural repertoire, $N_{items}$.

There is little information in the archaeological record that can act as a direct proxy for the values that these parameters may take. Each parameter is therefore constrained to a uniform prior range, with the value of each parameter in each simulation randomly assigned from this uniform prior and fixed for the duration of the simulation. It is possible that indirect information on these parameters is contained in the Vanhaeren and d’Errico (2006) data, in which case their estimation may be possible using the Approximate Bayesian Computation machinery as described in section 2.2.3. However, if that information is not present, then constraining each parameter to a uniform prior range serves to integrate uncertainty in the values of the parameters.

Prior ranges for each parameter within each model are listed in Table 2.1. The ranges of parameters $d_{mig}$, $d_{int}$, $N_{items}$ and $N_{groups}$, respectively specifying the migration distance, interaction radius, total number of items in each group’s cultural repertoire and maximum number of groups, were defined based on a combination of limited information available in the literature and personal communication with several leading experts in Aurignacian archaeology (Andrew Garrard, Francesco d’Errico, Marian Vanhaeren, Stephen Shennan). The ranges of probability parameters $p_{mut}$ and $p_{f/e}$, respectively specifying the probability of cultural mutation and the probability of fission / extinction, were initially both set to the maximum range (i.e. 0-1). Results from early stages of model development indicated that the range of values that the probability of cultural mutation parameter takes is very narrow with an upper bound of less than 0.1. As a result, in order to minimise unnecessary use of computational resources (i.e. by limiting the exploration of parameter space that would not give a good fit to the observed data), the range for this parameter was reduced to its current range as detailed in Table 2.1, which is still conservative given the initial results.
Table 2.1: Model parameters and their prior ranges for the Null Model, Null Model with Interactions and Culture-Dependent Interactions Model. The shaded cells represent parameters and their prior ranges that are the same for all three models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Null Model</th>
<th>Null Model with Interactions</th>
<th>Culture-Dependent Interactions Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum</td>
<td>maximum</td>
<td>minimum</td>
</tr>
<tr>
<td>$p_{\text{mut}}$: probability of cultural mutation</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>$p_{\text{f/e}}$: probability of fission / extinction</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$d_{\text{mig}}$: migration distance (km)</td>
<td>1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>$N_{\text{items}}$: total number of items in each group’s cultural repertoire</td>
<td>500</td>
<td>1500</td>
<td>500</td>
</tr>
<tr>
<td>$N_{\text{groups}}$: maximum number of groups</td>
<td>50</td>
<td>1000</td>
<td>50</td>
</tr>
<tr>
<td>$d_{\text{int}}$: interaction radius (km)</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>
2.2 Statistical Methods

Once a large number of simulations have been performed under each proposed model, the objective of the data analysis is to quantify the extent of similarity between the simulated and observed material culture data, in order to (1) compare the proposed models and evaluate which best describes the observed data, and, for the best-performing model, (2) estimate the values of the demographic and cultural evolutionary parameters of interest.

To do this, the Approximate Bayesian Computation (ABC) framework has been implemented in the form of a simple rejection algorithm, as described in section 1.3.4.1. As noted in section 1.3.4, the ABC framework requires the user to make three choices: the summary statistics that best describe the various aspects of the data, the threshold with which simulated datasets are accepted into the retained set of simulations (i.e. those most similar to observed data), and the distance measure used to quantify the similarity between observed and simulated data (Burr and Skurikhin 2013). In this study, the normalised Euclidean distance, defined in equation (1.2), is used as the distance measure between observed and simulated datasets. As explained previously, there is no universal rule to define which, or how many, summary statistics should be used, or the value of the acceptance threshold (Bertorelle et al. 2010; Csillery et al. 2010; Burr and Skurikhin 2013). The summary statistics chosen for the dataset used in this study are described in section 2.2.1, followed by an explanation of the methods used for model comparison (section 2.2.2) and parameter estimation (section 2.2.3). The threshold values used vary between analyses and each is specified in the corresponding section of the results chapters (Chapter 3 and Chapter 4).

2.2.1 Summary Statistics

As explained in section 1.3.4.2, to be able to compare simulated and observed datasets using ABC methods, summary statistics that capture the information contained in the observed data must be developed. These should be robust statistics that sufficiently describe the near-full properties of the observed dataset of interest, and have to be defined dependent on the observed dataset and its relevant properties.

In this study, the observed dataset of interest is that published by Vanhaeren and d’Errico (2006), which consists of personal ornaments from European Aurignacian sites. As already noted, this dataset contains 157 distinct ornament types recorded at 98 Aurignacian sites in Europe and the Near East, with records specifying presence / absence of distinct ornament types in sites only. These distinct ornament types are divided between 11 features according to different raw materials, with “62 representing ornaments made of shells, 31 of teeth, 30 of ivory, 11 of stone, 11 of bone, 7 of deer antler, and one each of belemnite, nummulite, ammonite, sea urchin and amber” (Vanhaeren and d’Errico 2006).
As explained in section 1.2.1, seriation and correspondence analyses of this data identified geographically non-randomly distributed clusters of sites sharing ornament types, which were subsequently interpreted as evidence of ethno-linguistic structuring (Vanhaeren and d’Errico 2006). It is therefore necessary to define summary statistics that capture these properties in the observed data, specifically quantifying the extent of cultural diversity between both sites and ornament types, as well as the spatial distribution of sites sharing ornament types.

Cultural diversity is quantified using the mean and variance of both shared information and mutual dependence between, as well as the median absolute deviation in, both ornament types and sites. Since the geographic locations of the archaeological sites in both the observed data and the model are fixed, the spatial distribution of sites is quantified in terms of the shared ornament types between them, with both mean and variance considered. A total of 12 summary statistics, the details of which are given below, are therefore used in the analyses of data.

### 2.2.1.1 Shared Information (SI)

Shared information, denoted $SI$, is a statistic that measures the extent of similarity between two variables. For measuring the shared information between ornament types, $SI$ is defined to be:

$$SI(t_i, t_j) = \frac{f_if_j}{f^2} \log \frac{r(t_i) + r(t_j)}{r(t_i, t_j)}, \quad (2.12)$$

where $r(t_i)$ and $r(t_j)$ respectively denote the ratio of the number of sites in which ornament types $i$ and $j$ occur to the total number of sites, $r(t_i, t_j)$ is the ratio of the number of sites in which at least one of ornament types $i$ and $j$ occur to the total number of sites, $f_i$ and $f_j$ represent the number of sites in which ornament types $i$ and $j$ occur, respectively, and $f$ is the average number of times any ornament type occurs over all sites. In this case, $SI$ measures the similarity between pairwise ornament types in terms of which sites they are present in. When two ornament types never occur in the same site,

$$r(t_i) + r(t_j) = r(t_i, t_j), \quad (2.13)$$
$$SI(t_i, t_j) = 0. \quad (2.14)$$

A similar equation can be used to measure the shared information between sites:

$$SI(s_i, s_j) = \frac{g_ig_j}{g^2} \log \frac{r(s_i) + r(s_j)}{r(s_i, s_j)}, \quad (2.15)$$
where \( r(s_i) \) and \( r(s_j) \) respectively denote the ratio of the number of ornament types present in sites \( i \) and \( j \) to the total number of ornament types, \( r(s_i, s_j) \) is the ratio of the number of distinct ornament types occurring in the two sites \((i \text{ and } j)\) to the total number of ornament types, \( g_i \) and \( g_j \) represent the number of ornament types present in sites \( i \) and \( j \), respectively, and \( \bar{g} \) is the average number of ornament types occurring per site. In this case, \( SI \) measures the extent of similarity between pairwise sites in terms of ornament types present in those sites. When two sites have no ornament types in common,

\[
r(s_i) + r(s_j) = r(s_i, s_j), \text{ and } \]

\[
SI(s_i, s_j) = 0. \quad (2.16)
\]

A total of 4 shared information statistics are considered: the mean and variance of \( SI:ornaments \) and \( SI:sites \).

### 2.2.1.2 Mutual Information (MI)

The mutual information, \( MI \), between two random variables \( X \) and \( Y \) is a measure of the mutual dependence between them. It is defined as:

\[
MI(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)}, \quad (2.18)
\]

where \( p(x, y) \) denotes the joint probability of \( x \) and \( y \) (the probability of \( x \) and \( y \) occurring together), and \( p(x) \) and \( p(y) \) denote the marginal probabilities of \( x \) and \( y \) respectively (the probabilities of the specified values of \( x \) and \( y \) occurring). If \( X \) and \( Y \) are independent, then:

\[
p(x, y) = p(x) \cdot p(y), \text{ and } \]

\[
MI(X; Y) = 0. \quad (2.19)
\]

For the observed dataset in this study, setting \( X = t_i \) and \( Y = t_j \), where \( t_i \) and \( t_j \) respectively correspond to the number of sites in which ornament types \( i \) and \( j \) occur, allows the mutual information between all pairs of ornament types to be computed. Analogously, setting \( X = s_i \) and \( Y = s_j \), where \( s_i \) and \( s_j \) correspond to the total number of ornament types present in sites \( i \) and \( j \) respectively, allows the mutual information between all pairs of sites to be computed.

In contrast to the \( SI \) statistic, which only examines the common presences between sites or ornament types, the \( MI \) statistic examines both the common presences and common absences. It therefore represents the dependence between the pairwise vectors in question.
A total of 4 mutual information statistics are considered: the mean and variance of $MI:ornaments$ and $MI:sites$.

### 2.2.1.3 Median Absolute Deviation (MAD)

The observed dataset shows large fluctuations both in the number of ornament types recovered at individual sites, and the number of times each particular ornament type occurs, as shown in Figure 2.5. Assuming that this is not the result of archaeological bias, these differences could be attributed to cultural wealth at sites, and the preference for particular ornament types, respectively. To quantify this, the median absolute deviation statistic, $MAD$, is used. It is a measure of the variability of a random sample, and is defined to be:

$$MAD = \text{median}(|X_i - \text{median}(X_i)|).$$

(2.21)

The $MAD$ statistic is similar to standard deviation, however, it is used here instead of the standard deviation as it is more resilient to outliers. In the calculation of standard deviation, the distances from the mean are squared and so large deviations from the mean are weighted more heavily, whereas using the median average in the calculation of $MAD$ limits the impact of outliers.

Letting $X_i = T = \frac{f_i}{\bar{f}}$, where $f_i$ represents the number of sites in which ornament type $i$ occurs and $\bar{f}$ is the average number of times any ornament type occurs over all sites, the $MAD$ statistic is a measure of the variability in the number of occurrences of different ornament types. This can be thought of as a measure of variability in the popularity of, or preference for, ornament types.

Letting $X_i = S = \frac{g_i}{\bar{g}}$, where $g_i$ represents the total number of ornament types present in site $i$ and $\bar{g}$ is the average number of ornament types occurring per site, the $MAD$ statistic measures the variability in the number of distinct ornament types recovered from each site. This can be thought of as a measure of cultural diversity, or variability in the cultural wealth recovered from sites.

Two median absolute deviation statistics are considered: $MAD:ornaments$ and $MAD:sites$.

### 2.2.1.4 Spatial Distribution of Sites (DR)

The extent to which sites share ornament types may be a function of the distance between those sites. It is logical to expect that sites that are located near to each other share ornament types
more frequently than those that are far apart. The spatial distribution of sites can be explored by considering the average distance between sites sharing ornament type \( i \), \( d_i \), in relation to the average distance between all sites, \( d \), as follows:

\[
DR_i = \frac{d_i}{d}.
\]  

\( DR \) therefore quantifies the spatial distribution of sites in terms of the shared ornament types between them. The mean and variance of \( DR \) are considered as summary statistics.

Figure 2.6 shows the density of \( DR \) for the original observed dataset (black line) and a random permutation of the same (grey line). The obvious shift to the right in the density of \( DR \) for the permuted dataset implies that the distance between sites sharing a particular ornament type is on average larger if ornament types are randomly assigned to sites. For the original observed dataset this implies that sites that are located closer to one another on average share ornament types more frequently with each other than with sites that are further away, as expected.

Two spatial distribution statistics are considered: the mean and variance of \( DR \).
Figure 2.5: Histograms of the number of occurrences of ornament types (panel (A)) and number of distinct ornament types recovered from individual sites (panel (B)) for the observed data (Vanhaeren and d'Errico 2006).

Figure 2.6: Density plots of the DR statistic for the original data (black line) and a random permutation of the same (grey line).
2.2.2 Model Comparison

As discussed in section 1.3.4.5, when comparing the relative performance of two models, a Bayes Factor is defined to be the ratio of probabilities of the two models, and is used to evaluate the relative support of one model in relation to another given the observed data (Csillery et al. 2010). It is a direct estimator of the posterior probability of a model, allowing for straightforward and intuitive comparison of the relative performance of proposed models (Kass and Raftery 1995).

Intuitively, when considering the retained set of simulations (i.e. those most similar to observed data), the most probable model should be over-represented, while the less probable models should be under-represented or absent. The probability of a model is therefore proportional to the relative frequency of the simulated data sets it generated that are in the retained set of simulations – the higher the frequency of simulated datasets in the retained set of simulations generated under a particular model, the higher the probability of that model, and the better the relative performance of the model.

It is important to note that, using this method of model comparison, it is not possible to rate how good a model is absolutely and any model considered will only be relatively better or worse than any other model considered. Methods have been suggested that go some way towards testing whether an observed dataset is a typical outcome of a proposed model (Voight et al. 2005), however such methods have not been not implemented in this thesis.

2.2.2.1 Relative Marginal Likelihoods

In order to implement the approach described in section 1.3.4.5 to evaluate which of the proposed models best explains the observed data, the relative marginal likelihood of each is estimated by taking into account the extent of similarity between observed data (Vanhaeren and d’Errico 2006) and data simulated under each proposed model. More explicitly, to compare models $M_1$ and $M_2$, their respective relative marginal likelihoods $l_1$ and $l_2$ are calculated as:

$$l_1 = \frac{N_1}{N},$$

and

$$l_2 = \frac{N_2}{N},$$

where $N_1$ and $N_2$ are the number of accepted simulations that come from models $M_1$ and $M_2$, respectively, and $N (= N_1 + N_2)$ is the total number of simulations considered. The relative marginal likelihood of each model is defined to be the proportion of the total number of accepted simulations that come from that model. This is therefore a measure of which model explains the observed data better, given that $N$ simulations are considered.
As mentioned previously in relation to Bayes Factors (section 1.3.4.5), this form of model comparison is independent of the number of parameters for each model, and instead estimates the likelihood of the model considering all possible parameter values. In cases where models with different numbers of parameters are compared, this method automatically and correctly penalises model complexity and consequently prevents overfitting (i.e. invoking parameters to explain aspects of the data that are in fact due to randomness). For models with a large number of parameters there is a larger parameter space to explore and so it is more difficult to find the parameter sets that generate data similar to the observed data. Models with more parameters are therefore penalised for the increased complexity compared to simpler models, resulting in a comparison weighted by model complexity.

2.2.2.2 Implementation

The process for evaluating the relative performance of each proposed model involves the following:

(1) pool together the same number of simulated datasets generated by each proposed model;

(2) calculate the normalised Euclidean distance between each of these simulated datasets and the observed data using the individual distances between observed and simulated summary statistics defined in section 2.2.1;

(3) order the calculated normalised Euclidean distances in ascending order (i.e. smallest first);

(4) retain the closest simulated datasets (i.e. those with the smallest normalised Euclidean distances) into the retained set of simulations – these are the best simulations, those generating data most similar to the observed data;

(5) calculate the relative marginal likelihood of each proposed model by considering the frequency, in the retained set of simulations, of simulated datasets generated under each model.

It should be noted that the simulated datasets used here correspond to ‘successful’ simulations (i.e. those simulations in which at least one group was present at the end of the simulation) and do not include ‘extinct’ simulations (i.e. those simulations in which all groups became extinct prior to the end of the simulation).
2.2.2.3 The Dirichlet Distribution

Since there is no systematic rule for choosing the value of the acceptance threshold, as explained in section 1.3.4.3, this is done on a comparison-by-comparison basis. The value of the threshold used for each comparison is given in the corresponding section of the results chapters (Chapter 3 and Chapter 4).

In order to estimate uncertainty in the model comparison, the discrete nature of the data from which the relative marginal likelihoods of the models are calculated (i.e. the number of simulations coming from each model) is coupled with the properties of the Dirichlet distribution.

A Dirichlet distribution, denoted \( \text{Dir}(\alpha_1+1, \alpha_2+1, ..., \alpha_n+1) \), is the conjugate prior of the multinomial distribution and generates the probabilities of \( n \) rival events occurring given the number of observations of each event, denoted \( \alpha_i \) for event \( i \). The addition of 1 to each number of observations denotes a uniform prior. In our case, at a particular acceptance threshold, the \( n \) rival events correspond to the \( n \) different proposed models and \( \alpha_i \) corresponds to the number of simulated datasets in the retained set of simulations that are generated by model \( i \).

The Dirichlet distribution can therefore be used to provide a measure of stochasticity of, and so confidence in, the calculated relative marginal likelihood of each model at different acceptance thresholds. Picks from the Dirichlet distribution defined at a particular acceptance threshold effectively represent the noise in the data (i.e. signal-to-noise ratio) – this is particularly important when there are very few observations (i.e. when considering only the few very best simulations) and the data is very stochastic.

2.2.3 Parameter Estimation

For those models identified as best explaining the observed data, posterior distributions of parameters are generated, which, for each model independently, involves the following:

1. calculate the normalised Euclidean distance between each simulated dataset and the observed data using the individual distances between observed and simulated summary statistics defined in section 2.2.1;

2. order the calculated normalised Euclidean distances in ascending order (i.e. smallest first);

3. determine the acceptance threshold by considering the general distribution of normalised Euclidean distance values;
(4) retain the closest simulated datasets (i.e. those with normalised Euclidean distance values smaller than the threshold determined in (3)) into the retained set of simulations – these are the best simulations, those generating data most similar to the observed data;

(5) generate posterior distributions of parameters from the retained set of simulations – these are estimates of parameter values that generate data most similar to the observed data.

It should again be noted that the simulated datasets used here correspond to ‘successful’ simulations (i.e. those simulations in which at least one group was present at the end of the simulation) and do not include 'extinct' simulations (i.e. those simulations in which all groups became extinct prior to the end of the simulation). It should also be noted that posterior distributions of parameters are estimated using the rejection algorithm as described in section 1.3.4.1, and no regression weighting, discussed in section 1.3.4.4, is performed in this estimation.

### 2.2.4 Correlation Analyses

For purposes of critical evaluation of the models and the adopted approach, three sets of correlation analyses are performed for each model version.

First, the correlation between each pair of proposed summary statistics using all simulations is considered. If two summary statistics are highly correlated and so carry the same, or very similar, information about the dataset of interest, they can be considered redundant and a dimensional reduction of statistics can be made by excluding one of the pair of correlated statistics from further analyses. This is beneficial both to ensure that a particular aspect of the data is not over-represented in the analysis and for computational economy reasons.

The second correlation considered is that between pairwise summary statistic and model parameter combinations using all simulations. This analysis is done in order to assess which statistics inform on which parameters, or, conversely, which parameters drive which statistics.

Finally, the correlation between each pair of model parameters using retained simulations is considered. This analysis is done to assess whether parameters that govern the various aspects of the demographic and cultural processes simulated are correlated in the accepted simulations. One interpretation of such correlations is that changes in one parameter require changes in another to achieve a good fit to the observed data.

In this thesis, all correlation analyses are performed using Spearman’s rank correlation coefficient $R^2$ to assess the statistical dependence between the considered variables. The test is
performed using the function `cor.test(method="spearman")` from R’s `stats` package. It is important to note here that Spearman’s rank correlation coefficient is a non-parametric measure of statistical dependence, or association, between two variables, rather than correlation in the conventional sense. The reported correlation value (i.e. $R^2$) is a measure of the common variance between the two variables (i.e. the proportion of variance in one of the variables accounted for in the other).

When reporting the results of the correlation analyses, a pairwise scatterplot, correlation coefficient $R^2$ and associated p-value are given. Due to the large number of data points, and the significant overlap of these data points, the scatterplots are generated using the function `hexplom()` from R’s `hexbin` package, which generates a scatterplot with bivariate binning into hexagonal cells; these hexagonal cells are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells, such that, for each individual scatterplot, the cell with the highest number of points is the darkest. Additionally, since it can be misleading to plot a regression line on a graph analysed with rank correlation, the scatterplots do not have trend lines overlaid.

Although p-values are reported, they are not a particularly meaningful measure in this case and may be somewhat misleading. This is because the p-value is proportional to the sample size and increasing the sample size would increase the likelihood of the correlation being identified as significant (i.e. the larger the sample size, the more likely a significant result). It follows that the likelihood of the correlations performed here being identified as significant is high, since a large number of simulations are considered in the analysis. Since the number of simulations considered in the analysis is not set by any property of the model, but rather by the amount of computational time available, the p-values reported are not particularly meaningful. Thus, the reported correlation coefficient $R^2$ values should be considered as the most informative indicator of correlation.
Chapter 3: Modelling Patterns of Material Culture in Upper Palaeolithic Europe (I)

This chapter is an expanded version of the results section of the article attached in Appendix B, which has been published as:


The analyses presented in the article were performed by Mirna Kovacevic, with guidance from Mark G. Thomas and Stephen Shennan in a supervisory capacity. Marian Vanhaeren and Francesco d'Errico provided published archaeological data (Vanhaeren and d'Errico 2006).

Summary

This chapter presents the results of the spatially explicit cultural transmission simulation models described in Chapter 2, followed by some conclusions and a critical assessment of the modelling framework. The results of model comparison analyses (as described in section 2.2.2), parameter estimation analyses (as described in section 2.2.3) and correlation analyses (as described in section 2.2.4) are presented.

Approximate Bayesian Computation (ABC) analyses presented here were performed, and results visualised, using the R programming language (http://www.r-project.org/).
3.1 Results

3.1.1 Exploring Behaviour of Summary Statistics

3.1.1.1 Are the summary statistics correlated?

In this analysis, the correlation (i.e. dependence) between each pair of proposed summary statistics is examined using simulated data, in order to check whether any pair of summary statistics carry the same information about the dataset considered. As explained in section 2.2.4, if two statistics are correlated, they can be considered redundant and one of the pair of correlated statistics can be excluded from further analyses.

A correlation analysis of simulated summary statistics values was performed for each of the 10 models as listed in section 2.1.5, considering 400,000 simulations generated by each model (with the exception of the two Null Models (NM), where 120,000 simulations for each model were considered) and using Spearman’s rank correlation coefficient $R^2$ to assess the statistical dependence between each pair of summary statistics. The results for each model are shown in Appendix A.1. All models showed similar patterns in pairwise summary statistic correlations. The results presented and discussed in this section are used as examples of the full set of results shown in Appendix A.1.

For the results presented here, the analysis was performed using 404,808 simulations generated by the Null Model with Cultural Interactions (NMCI) under each of the two best scenarios as identified by the model comparison analysis in 3.1.2.2 (bounded stepwise (SW) and bounded non-stepwise (DIS) models of cultural variation, and environmental scenario with uniform distribution of carrying capacities (FLAT)). The results are shown in Figure 3.1, which shows the pairwise binned scatterplots (below diagonal), correlation coefficients $R^2$ (above diagonal; black) and associated p-values (above diagonal; blue) for the NMCI SW FLAT (upper panel) and NMCI DIS FLAT (lower panel). As explained in section 2.2.4, the p-values are to an extent determined by the number of simulations considered, which is only constrained by the amount of computational time available and is therefore arbitrary, and so the reported correlation coefficient $R^2$ values are considered to be the most informative indicator of correlation.

The scatterplots and correlation values shown in Figure 3.1 indicate that correlations between proposed summary statistics do exist, however, in most cases these correlations are weak ($R^2 \leq 0.25$). Cases where the strongest correlations are observed (e.g. S7: mean, and S8: variance, of mutual information between sites) are between different measures (i.e. mean and variance) of the same summary statistic. Even in the cases of the strongest correlations (i.e. between summary statistics S7 and S8), the summary statistics are not completely correlated (NMCI SW FLAT: $R^2 = 0.837$; NMCI DIS FLAT: $R^2 = 0.757$). It can therefore be inferred that these two summary statistics, although related, are carrying different information to some extent. Since similar conclusions can be made for all pairs of statistics, all 12 proposed summary statistics are included in further analyses of the data.
Figure 3.1: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic combinations using 404,808 successful simulations for the NMCI SW FLAT (top panel) and NMCI DIS FLAT (bottom panel) models. The cells in each pairwise scatterplot (below diagonal) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (above diagonal) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value ≥ 0.05). S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>S12</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.413</td>
<td>0.035</td>
<td>0.031</td>
<td>0.092</td>
<td>0.063</td>
<td>0.002</td>
<td>0.015</td>
<td>0.045</td>
<td>0.001</td>
<td>0.067</td>
<td>0.006</td>
<td>0.421</td>
</tr>
<tr>
<td>S2</td>
<td>0.035</td>
<td>0.114</td>
<td>0.053</td>
<td>0.031</td>
<td>0.057</td>
<td>0.027</td>
<td>0.251</td>
<td>0.03</td>
<td>0.079</td>
<td>0.052</td>
<td>0.265</td>
<td>S11</td>
</tr>
<tr>
<td>S3</td>
<td>0.031</td>
<td>0.141</td>
<td>0.039</td>
<td>0.031</td>
<td>0.001</td>
<td>0.036</td>
<td>0.074</td>
<td>0.002</td>
<td>0.003</td>
<td>0.092</td>
<td>0.002</td>
<td>0.027</td>
</tr>
<tr>
<td>S4</td>
<td>0.045</td>
<td>0.116</td>
<td>0.053</td>
<td>0.031</td>
<td>0.057</td>
<td>0.027</td>
<td>0.251</td>
<td>0.03</td>
<td>0.079</td>
<td>0.052</td>
<td>0.265</td>
<td>S10</td>
</tr>
<tr>
<td>S5</td>
<td>0.053</td>
<td>0.191</td>
<td>0.048</td>
<td>0.031</td>
<td>0.057</td>
<td>0.027</td>
<td>0.251</td>
<td>0.03</td>
<td>0.079</td>
<td>0.052</td>
<td>0.265</td>
<td>S9</td>
</tr>
<tr>
<td>S6</td>
<td>0.053</td>
<td>0.204</td>
<td>0.026</td>
<td>0.176</td>
<td>0.035</td>
<td>0.035</td>
<td>0.237</td>
<td>0.042</td>
<td>0.04</td>
<td>0.036</td>
<td>0.035</td>
<td>S8</td>
</tr>
<tr>
<td>S7</td>
<td>0.053</td>
<td>0.426</td>
<td>0.227</td>
<td>0.121</td>
<td>0.007</td>
<td>0.002</td>
<td>0.029</td>
<td>0.032</td>
<td>0.025</td>
<td>0.006</td>
<td>0.002</td>
<td>S7</td>
</tr>
<tr>
<td>S8</td>
<td>0.053</td>
<td>0.251</td>
<td>0.037</td>
<td>0.121</td>
<td>0.007</td>
<td>0.002</td>
<td>0.029</td>
<td>0.032</td>
<td>0.025</td>
<td>0.006</td>
<td>0.002</td>
<td>S6</td>
</tr>
<tr>
<td>S9</td>
<td>0.042</td>
<td>0.133</td>
<td>0.013</td>
<td>0.027</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.014</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>S5</td>
</tr>
<tr>
<td>S10</td>
<td>0.042</td>
<td>0.096</td>
<td>0.013</td>
<td>0.041</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.014</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>S4</td>
</tr>
<tr>
<td>S11</td>
<td>0.042</td>
<td>0.133</td>
<td>0.013</td>
<td>0.027</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.014</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>S3</td>
</tr>
<tr>
<td>S12</td>
<td>0.042</td>
<td>0.133</td>
<td>0.013</td>
<td>0.027</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.014</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>S2</td>
</tr>
</tbody>
</table>

S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.
3.1.1.2 Relationship between Summary Statistics and Model Parameters

In this analysis, the dependence between pairwise summary statistic and model parameter combinations is examined using simulated data, in order to assess which statistics inform on which parameters, or, conversely, which parameters drive which statistics.

A correlation analysis of each pairwise combination of summary statistic and model parameter was performed, for each of the 10 models as described in section 2.1.5, considering 400,000 simulations generated by each model (with the exception of the two Null Models (NM), where 120,000 simulations for each model were considered). Spearman’s rank correlation coefficient $R^2$ was again used to assess the statistical dependence between each summary statistic and each model parameter. The results for each model are shown in Appendix A.1. All models showed similar patterns in correlations between pairwise combinations of summary statistic and model parameter. The results presented and discussed in this section are examples of the full set of results shown in Appendix A.1.

The results presented here are of the correlation analysis of each pairwise combination of summary statistic and model parameter performed using 404,808 simulations generated by the NMCI SW FLAT (Figure 3.2) and NMCI DIS FLAT (Figure 3.3) models. The two figures give the pairwise binned scatterplots (top panel), the correlation coefficients $R^2$ (bottom panel; black) and associated p-values (bottom panel; blue) for each model respectively. Again it should be noted that the reported correlation coefficient $R^2$ values are considered to be the most informative indicator of correlation since the p-values are to an extent determined by the number of simulations considered, which is effectively arbitrary.

For each of the two models presented here, and indeed all models, weak ($R^2 \leq 0.25$) correlations exist for most pairwise summary statistic and model parameter combinations. This suggests that a combination of model parameters drive each summary statistic, or, conversely, this can be interpreted as a combination of summary statistics informing on each of the model parameters.
Figure 3.2: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and parameter combinations using 404,808 successful simulations for the NMCI SW FLAT model.

The cells in each pairwise scatterplot (top panel) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (bottom panel) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value $\geq 0.05$). P1: probability of cultural mutation; S2: probability of fission / extinction; P3: migration distance (km); P4: total number of items in each group’s cultural repertoire; P5: maximum number of groups; P6: interaction radius (km); S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.
Figure 3.3: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and parameter combinations using 404,808 successful simulations for the NMCI DIS FLAT model.

The cells in each pairwise scatterplot (top panel) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (bottom panel) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value ≥ 0.05). P1: probability of cultural mutation; S2: probability of fission / extinction; P3: migration distance (km); P4: total number of items in each group’s cultural repertoire; P5: maximum number of groups; P6: interaction radius (km); S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites. model.
3.1.2 Model Comparisons

In this analysis, the relative marginal likelihood of each of the models described in section 2.1.5 is estimated, conditioned on the extent of similarity between a range of spatial statistics describing the observed data (Vanhaeren and d'Errico 2006) and data simulated under each proposed model. As explained in section 2.2.2, it is important to note that any scenario considered will only be relatively better or worse than any other scenario considered, and, using the methods of model comparison employed here, it is not possible to rate how good a scenario is absolutely; as noted previously, some methods have been suggested for testing whether an observed dataset is a typical outcome of a proposed model (Voight et al. 2005), however such methods have not been implemented in this thesis.

3.1.2.1 Comparing proposed null models: Null Model (NM) and Null Model with Cultural Interactions (NMCI)

The first model comparison compares the relative performance of the two proposed null models – one with and one without cultural interactions – as described in section 2.1.4.

A total of 4 models are compared; data considered is that generated by the Null Model (NM) and the Null Model with Cultural Interactions (NMCI), where each data set is simulated under the bounded stepwise (SW) model of cultural variation and one of two environmental scenarios (either uniform carrying capacities (FLAT) or population densities from Bocque-Appel et al. (2005) (B-A)). For each of the 4 models compared, 120,000 simulations are considered. For the relative marginal likelihood estimation, results for the 4 models are considered together; a total of 480,000 simulations are therefore taken into account.

Figure 3.4 is a plot of the estimated relative marginal likelihood of each version of the NM model (dashed lines) and the NMCI model (solid lines) at different thresholds. It shows what proportion of the best simulations (y-axis) – those generating data most similar to the observed data – are coming from each model for each percentage of closest simulations (x-axis). The colours refer to the combination of environmental scenario and model of cultural variation considered. At each percentage of closest simulations (x-axis), each of the ten grey points represents a pick made from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison. As explained in section 2.2.2.3, at each threshold, the distribution of these points is a measure of stochasticity in the data at that threshold, and provides a proxy for confidence in the value of the estimated relative marginal likelihoods – the tighter the grey points fall around the lines, the more confidence there is in the model comparison; the relative marginal likelihoods are only considered as different when the distributions of grey points from different models do not overlap.
Since the plot depicts proportions, for any particular percentage of closest simulations (i.e. for any particular value on the x-axis), the sum of the proportions of the closest Euclidean distances coming from each model (i.e. the sum of the values on the y-axis) will always be 1. The relative marginal likelihood estimates of each model calculated on the smallest 2,400 Euclidean distances (i.e. closest 0.5% of the considered simulations, which corresponds to x=0.5 in Figure 3.4) are detailed in Table 3.1.

Figure 3.4 and Table 3.1 indicate that, within each scenario modelled – scenario here referring to an environmental scenario, thus either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A) – the best fits of simulated to observed data are generated by the Null Model (NM; dashed lines).

Comparing the scenarios, however, indicates that, for both the Null Model (NM; dashed lines) and the Null Model with Cultural Interactions (NMCI; solid lines), data simulated under the environmental scenario in which Europe is assumed to have a uniform distribution of carrying capacities (FLAT; blue lines) provide a better fit to observed data than those simulated under the environmental scenario that takes information on population densities from Bocquet-Appel et al. (2005) (B-A; orange lines). This suggests that the choice of environmental scenario is important in how well the simulated data perform when compared to observed data.

Overall, these results indicate that the more simple Null Model (NM), a model of cultural identity-by-descent with modification and isolation-by-distance, generates a better fit to the observed data compared to the more complex Null Model with Cultural Interactions (NMCI) that, in addition, takes into account cultural interactions that are independent of similarities or differences in groups’ ornamental material cultures. Considering the intricacies of the Approximate Bayesian Computation approach adopted here, this is likely the result of the relatively small number of simulations (i.e. 120,000) considered in the comparison, which may not be enough to adequately explore the parameter space considered. This is particularly true for the NMCI model, which is more complex than the NM model since invoking cultural interactions requires an extra parameter; this has a multiplicative effect on the number of simulations required to explore the full parameter space, as explained in section 2.2.2.1. It is therefore possible that the results observed with this relatively small number of simulations are somewhat misleading.

Although the Null Model (NM) provides a better fit to observed data than the Null Model with Cultural Interactions (NMCI) within each scenario, for reasons explained in section 2.1.4, the latter is a fairer choice of null model for testing the proposed hypothesis under the described framework, and is therefore used for all subsequent analyses.
Figure 3.4: Relative marginal likelihood estimate (y-axis) of each version of Null Model (NM; dashed lines) and Null Model with Cultural Interactions (NMCI; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 480,000 simulations (i.e. 120,000 simulations for each of the 4 models compared). At each percentage of closest simulations (x-axis), each of the ten grey points represents a pick made from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison; this is a measure of stochasticity in the data and provides a proxy for confidence in the value of the estimated relative marginal likelihoods.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and a model of cultural variation (bounded stepwise (SW)).
Table 3.1: Relative marginal likelihood estimates (RML) of each version of Null Model (NM) and Null Model with Cultural Interactions (NMCI). Proportions are calculated on the smallest 2,400 Euclidean distances (i.e. closest 0.5%, which corresponds to $x=0.5$ in Figure 3.4) of a total of 480,000 simulations (i.e. 120,000 simulations for each of the 4 models compared). The shaded cells represent the best two models.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and a model of cultural variation (bounded stepwise (SW)).

<table>
<thead>
<tr>
<th>Model</th>
<th>Cultural Mutation Model</th>
<th>Environmental Scenario</th>
<th>RML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model (NM)</td>
<td>SW</td>
<td>FLAT</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.184</td>
</tr>
<tr>
<td>Null Model with Cultural Interactions (NMCI)</td>
<td>SW</td>
<td>FLAT</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.018</td>
</tr>
</tbody>
</table>
3.1.2.2 Null Model with Cultural Interactions or Culture-Dependent Interactions Model: Which model best explains the observed data?

By determining which of the proposed models best fits the observed data, this model comparison aims to test the main hypothesis – whether the observed distribution of artefact types reported by Vanhaeren and d’Errico (2006) can be explained by a model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving interaction processes.

A total of 8 models are compared; data considered is that generated by the Null Model with Cultural Interactions (NMCI) and the Culture-Dependent Interactions Model (CD) under a pairwise combination of one of two environmental scenarios (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and one of two models of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)). For each model version (i.e. each of the 8 models compared), 400,000 simulations are considered. For the relative marginal likelihood estimation, results for the 8 models are considered together; a total of 3,200,000 simulations are therefore taken into account.

Figure 3.5 is a plot of the estimated relative marginal likelihood of each version of the NMCI model (dashed lines) and the CD model (solid lines) at different thresholds. Similarly to Figure 3.4, it shows what proportion of the best simulations (y-axis) – those generating data most similar to the observed data – are coming from each model for each percentage of closest simulations (x-axis). The colours again refer to the combination of environmental scenario and model of cultural variation considered. At each percentage of closest simulations (x-axis), ten picks from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison are used as a proxy for confidence in the value of the estimated relative marginal likelihoods. The relative marginal likelihood estimates of each model calculated on the smallest 3,200 Euclidean distances (i.e. closest 0.1% of the considered simulations, which corresponds to x=0.1 in Figure 3.5) are detailed in Table 3.2.

Figure 3.5 and Table 3.2 indicate that, for all scenarios modelled – scenario here referring to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)) – there is little difference in how well the NMCI models (dashed lines) and CD models (solid lines) perform. This implies that, when compared to the NMCI models (i.e. simple scenario of cultural identity-by-descent with modification and isolation-by-distance), the added complexity of the CD models (i.e. invoking cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving interaction processes) does not improve the fit of simulated to observed data. The comparison also indicates that scenario choice affects which of the two performs best;
the NMCI model seems to perform better under some scenarios (i.e. SW FLAT), while the CD model performs better under others (i.e. DIS FLAT).

For both the NMCI and CD models, the best fits of simulated to observed data are generated under the scenario that combines the bounded stepwise mutation model (SW) and the environmental scenario in which Europe is assumed to have a uniform distribution of carrying capacities (FLAT) (blue lines).

That the best fits of simulated to observed data are generated by this combination of environmental scenario and cultural mutation model, and that these are a far better fit than any of the other scenarios considered, implies that both the assumed mutation model and the assumed environmental scenario strongly influence the fit of simulated to observed data. It is interesting to note that the assumed environmental scenario appears to makes a bigger difference than the assumed mutation model, as evidenced by the fact that the choice of cultural mutation model does not affect the fit of simulated to observed data for simulations conditioned on the unrealistic distribution from the Bocquet-Appel et al. (2005) study (i.e. under this environmental scenario, the two cultural mutation models considered perform equally poorly).

When considering the two environmental scenarios used to condition the demography of the simulation space, the comparison indicates that, for each of the two mutation models considered, simulations in which the demography is conditioned on a uniform distribution of carrying capacities generate a better fit to the observed data; there is no improvement in the fit of simulated to observed data when conditioning simulations on the distribution from the Bocquet-Appel et al. (2005) study, and moreover simulations conditioned on this distribution perform more poorly. Since a uniform distribution of carrying capacities is clearly not realistic, this result could be interpreted as indicating that both environmental scenarios used to condition the demography of the simulation space in this study are unrealistic; this is further discussed as a caveat of the current modelling framework in section 3.2, along with suggestions for possible improvements.

For simulations conditioned on a uniform distribution of carrying capacities, data simulated under the stepwise mutation model (blue lines) generate a better fit to observed data than those simulated under the non-stepwise (discrete) model (red lines). It could be hypothesised that this observation suggests the existence of scaling of item similarity in material culture repertoires, implying that the extent of similarity of a particular item to items already in the repertoire would have affected how that item was perceived. In other words, a group is more likely to add an item to its cultural repertoire that is morphologically similar to one that is already present in its repertoire than one that is very different. As described in section 2.1.3.3.1, the stepwise mutation model relaxes the assumption that the cultural trait data considered in this study is necessarily categorical since categorical distinctions are often arbitrarily made.
Figure 3.5: Relative marginal likelihood estimate (y-axis) of each version of Null Model with Cultural Interactions (NMCI; dashed lines) and Culture-Dependent Interactions Model (CD; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 3,200,000 simulations (i.e. 400,000 simulations for each of the 8 models compared). At each percentage of closest simulations (x-axis), each of the ten grey points represents a pick made from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison; this is a measure of stochasticity in the data and provides a proxy for confidence in the value of the estimated relative marginal likelihoods.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)).
Table 3.2: Relative marginal likelihood estimates (RML) of each version of Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD). Proportions are calculated on the smallest 3,200 Euclidean distances (i.e. closest 0.1%, which corresponds to $x=0.1$ in Figure 3.5) of a total of 3,200,000 simulations (i.e. 400,000 simulations for each of the 8 models compared). The shaded cells represent the best four models.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT) or population densities from Bocquet-Appel et al. (2005) (B-A)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)).

<table>
<thead>
<tr>
<th>Model</th>
<th>Cultural Mutation Model</th>
<th>Environmental Scenario</th>
<th>RML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model with Cultural Interactions (NMCI)</td>
<td>SW</td>
<td>FLAT</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>DIS</td>
<td>FLAT</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.014</td>
</tr>
<tr>
<td>Culture-Dependent Interactions Model (CD)</td>
<td>SW</td>
<td>FLAT</td>
<td>0.330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>DIS</td>
<td>FLAT</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B-A</td>
<td>0.010</td>
</tr>
</tbody>
</table>
3.1.3 Parameter Estimation

In order to obtain posterior estimates of parameters of interest, 404,808 simulations generated by the Null Model with Cultural Interactions (NMCI) under each of the two best scenarios (bounded stepwise (SW) and bounded non-stepwise (DIS) models of cultural variation and environmental scenario with uniform distribution of carrying capacities (FLAT)) were used. The rationale behind this choice of models is that, since NMCI and CD models perform similarly well under each scenario, it is more interesting to examine how scenario choice affects the posterior parameter distributions. NMCI is chosen over CD because it is a simpler model – since there is no improvement in fit of simulated to observed data with the added complexity of the CD model, there is no justification for adding this complexity.

In order to choose the number of closest simulations (i.e. the threshold level) to consider in the parameter estimation, the Euclidean distances corresponding to the closest 5% of 404,808 simulations (i.e. 20,240 simulations) for each of the two models were plotted in ascending order. As discussed in section 1.3.4.3, in some cases (e.g Kamberov et al. 2013), this procedure has been used to identify a sensible threshold by way of a noticeable increase in Euclidean distance values that normally begins after a plateau in the same. This is not the case here, as illustrated in Figure 3.6 by the smooth increase in Euclidean distance values for both models. A threshold value of ~1% (best 4,000 of 404,808 simulations considered) is therefore used for purposes of parameter estimation; the parameter estimation analysis has also been performed with ~0.5% (best 2,000 of 404,808 simulations considered) and ~0.1% (best 400 of 404,808 simulations considered), and these results are shown in Appendix A.2. The lack of a clear choice for the value of the threshold in this case illustrates why no systematic rule exists for choosing this value, and why it should be evaluated and defined separately for each simulation model (Joyce and Marjoram 2008; Wegmann et al. 2009; Bertorelle et al. 2010).

Figure 3.7 and Figure 3.8 show the estimated posterior density distributions of the model parameters for the NMCI SW FLAT and NMCI DIS FLAT models, respectively, calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for each model. The boundaries of the equal-tailed 95% (i.e. the upper and lower 2.5%) credible intervals (CIs) of each distribution are indicated by shading. The posterior parameter distributions for each model are summarised in Table 3.3, by giving the mode, 2.5% and 97.5% quantiles of each distribution. Figure 3.9 shows the performance of the two models in explaining the observed data. It is a plot of the distribution of each statistic generated by the 4,000 retained simulations for each model, and shows that simulations coming from NMCI SW FLAT (black lines) are closer to their target values (i.e. summary statistic values calculated on the observed data, indicated by red vertical lines) than those coming from NMCI DIS FLAT (grey lines); this is to be expected since the NMCI DIS FLAT model was shown in the model comparison analysis in section 3.1.2.2 to generate a poorer fit to observed data.
As Figure 3.7, Figure 3.8 and Table 3.3 show, for some parameters, particularly (B) probability of fission / extinction, and (C) migration distance, the 95% CIs of the posterior estimates are similar to their prior ranges, which implies that little information could be obtained about these parameters using the models and observed data. With the exception of the probability of cultural mutation parameter (panel (A) of the two figures), the 95% CIs of the posterior parameter estimates are on the same order of magnitude for the two models, and wider for the NMCI DIS FLAT model than the NMCI SW FLAT model. For the probability of cultural mutation parameter (panel (A) of the two figures), the 95% CIs of the posterior estimate is wider for the (better-performing) NMCI SW FLAT model.

The posterior parameter analyses did appear informative for some key parameters. Firstly, the 95% CI of the posterior estimate of the probability of cultural mutation (panel (A) in Figure 3.7 and Figure 3.8), is considerably narrower (NMCI SW FLAT: 0.0018-0.0413, mode=0.0059; NMCI DIS FLAT: 0.0006-0.0083; mode=0.0024) than its prior range (0-0.2).

Secondly, the 95% CI of the posterior estimate of the maximum number of groups (panel (E) of the two figures), is narrower (NMCI SW FLAT: 454-960, mode=571; NMCI DIS FLAT: 458-964; mode=530) than its prior range (50-1000), though little further inference on the likely value of this parameter within the estimated posterior range can be made.

Considering now only the (better-performing) NMCI SW FLAT model, Figure 3.7 indicates that the 95% CI of the posterior estimate of the interaction radius parameter (panel (F)) is similar to but that it falls in the lower end (1-46, mode=4.4031) of its prior range (0-50). Since this parameter specifies the geographic distance within which a pair of groups must be in order to interact, a low value is to be expected. Similarly, the posterior estimate of the total number of items in each group’s cultural repertoire parameter (panel (D)) is similar to but falls in the lower end (531-1416, mode=571) of its prior range (500-1500). This estimate may be of interest as it suggests that groups’ cultural repertoires did not contain a large number of items (ornaments).

Although the 95% CIs of the posterior estimates of these two parameters ((D) total number of items in each group’s cultural repertoire, and (F) interaction radius) are similar to their prior ranges, the fact that the estimated posterior density distributions peak suggests that the observed data may contain some information about these parameters.

It is interesting to consider the extinct density distributions (i.e. those simulations in which all groups became extinct prior to the end of the simulation), shown by the dashed grey lines in the plots, of the demographic parameters (B) probability of fission / extinction, (C) migration distance, and (E) maximum number of groups. These distributions indicate that a very high rate of fission / extinction, as well as a low migration distance and low maximum number of groups in the domain, readily result in a truncation of the simulation (i.e. all groups become extinct prior to the end of the simulation), as is to be expected. Additionally, these demographic parameters are the only parameters that have an effect on viability of simulations (i.e. whether all groups became extinct prior to the end of the simulation); this is reasonable as the cultural
parameters (panels (A) probability of cultural mutation, (D) total number of items in each group's cultural repertoire, and (F) interaction radius) would not be expected to have an effect on the demography of the simulation space, as confirmed by the extinct density distributions of these cultural parameters mirroring their respective prior distributions.
Figure 3.6: Ranked Euclidean distance values of the best 20,240 simulations (i.e. the closest 5% of 404,808 simulations) for each of the two models plotted in black, with the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red. The two panels correspond to the two models of interest: NMCI SW FLAT (top panel) and NMCI DIS FLAT (bottom panel).
Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the two NMCI models, calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for each model. The boundaries of the equal-tailed 95% credible intervals (i.e. the upper and lower 2.5%) of each distribution are indicated by shading; these are also summarised in Table 3.3. The solid and dashed grey lines represent the prior and extinct (i.e. those simulations in which all groups became extinct prior to the end of the simulation) density distributions of parameters, respectively.

Figure 3.7: NMCI SW FLAT model.

Figure 3.8: NMCI DIS FLAT model.
Table 3.3: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places. The letter in the far left column corresponds to the panels in Figure 3.7 (NMCI SW FLAT) and Figure 3.8 (NMCI DIS FLAT).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Range</th>
<th>SW FLAT</th>
<th>DIS FLAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>minimum</td>
<td>maximum</td>
<td>mode</td>
</tr>
<tr>
<td>(A) ( p_{mut} ): probability of cultural mutation</td>
<td>0</td>
<td>0.2</td>
<td>0.0059</td>
</tr>
<tr>
<td>(B) ( p_{f/e} ): probability of fission / extinction</td>
<td>0</td>
<td>1</td>
<td>0.8591</td>
</tr>
<tr>
<td>(C) ( d_{mig} ): migration distance (km)</td>
<td>1</td>
<td>100</td>
<td>75.9295</td>
</tr>
<tr>
<td>(D) ( N_{items} ): total number of items in each group’s cultural repertoire</td>
<td>500</td>
<td>1500</td>
<td>588</td>
</tr>
<tr>
<td>(E) ( N_{groups} ): maximum number of groups</td>
<td>50</td>
<td>1000</td>
<td>571</td>
</tr>
<tr>
<td>(F) ( d_{int} ): interaction radius (km)</td>
<td>0</td>
<td>50</td>
<td>4.4031</td>
</tr>
</tbody>
</table>
Figure 3.9: Distributions of the 12 summary statistic values in the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for the NMCI SW FLAT model (black lines) and NMCI DIS FLAT model (grey lines). The title of each panel corresponds to the summary statistic as discussed in section 2.2.1. The red vertical line indicates the target value of each summary statistic (i.e. the value of that statistic calculated from the observed data).
3.1.4 Exploring Behaviour of Parameters

Similarly to considering the dependence between proposed summary statistics in section 3.1.1, the dependence between the model parameters can be examined to assess whether parameters that govern the various aspects of the demographic and cultural processes simulated are correlated in the accepted simulations.

To do this, a correlation analysis of parameter values for both the NMCI SW FLAT (Figure 3.10) and NMCI DIS FLAT (Figure 3.11) models was performed, using the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for each. Spearman’s rank correlation coefficient $R^2$ was again used to assess the statistical dependence between the parameters. For each of the two models and each pairwise combination of model parameters, the binned scatterplot (below diagonal), correlation coefficient $R^2$ (above diagonal; black) and associated p-value (above diagonal; blue) are given. The reported correlation coefficient $R^2$ values are again considered to be the most informative indicator of correlation since the p-values are to an extent determined by the number of simulations considered, which is effectively arbitrary. Note that marginal views of the scatterplots shown in Figure 3.10 and Figure 3.11 correspond to the density plots in Figure 3.7 and Figure 3.8, respectively, as expected.

Weak ($R^2 \leq 0.25$) correlations are observed between all pairwise combinations of the model parameters. This suggests that the parameters considered affect the various aspects of the demographic and cultural processes simulated relatively independently.
Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 4,000 retained simulations (i.e. closest ~1% of 404,808 simulations) for the NMCI SW FLAT (top panel, i.e. Figure 3.10) and NMCI DIS FLAT (bottom panel, i.e. Figure 3.11) models. The cells in each pairwise scatterplot (below diagonal) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (above diagonal) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value $\geq$ 0.05).

Figure 3.10: NMCI SW FLAT model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R^2$</th>
<th>p-value</th>
<th>Parameter</th>
<th>$R^2$</th>
<th>p-value</th>
<th>Parameter</th>
<th>$R^2$</th>
<th>p-value</th>
<th>Parameter</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>migration distance (km)</td>
<td>0.116</td>
<td>0.004</td>
<td>total number of items in each group's cultural repertoire</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probability of cultural mutation</td>
<td>0.002</td>
<td>0.001</td>
<td>interaction radius (km)</td>
<td>0.001</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probability of fission / extinction</td>
<td>0.013</td>
<td>0.004</td>
<td>maximum number of groups</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probability of cultural mutation</td>
<td>0.002</td>
<td>0.008</td>
<td>total number of items in each group's cultural repertoire</td>
<td>0.001</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>migration distance (km)</td>
<td>0.001</td>
<td>0.116</td>
<td>interaction radius (km)</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>probability of cultural mutation</td>
<td>0.001</td>
<td>0.001</td>
<td>maximum number of groups</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.11: NMCI DIS FLAT model.
3.2 Discussion

The results of the model comparison analyses reported here (section 3.1.2.2) do not support the hypothesis that Aurignacian populations in Early Upper Palaeolithic Europe were ethnically structured in a manner related to ornamental material culture. The spatially explicit simulation models and ABC analyses, conditioned on the data presented by Vanhaeren and d’Errico (2006), indicate that – in terms of fit between simulated and observed data – there is little difference between the simple scenario of cultural identity-by-descent with modification and isolation-by-distance, and the more complex one that, in addition, invokes cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving interaction processes.

It should be noted that, for the ABC approach adopted here, the number of simulations per model is relatively small and may not be enough to adequately explore the parameter space considered; for this reason, the number of simulations performed under each combination of environmental scenario and cultural variation model should be systematically increased. However, this is something that must be balanced against the computational cost of performing the simulations.

As explained in section 2.1.1.2, each simulation is initialised at the onset of, and simulated forward in time to the end of, the Aurignacian period, spanning a total of 13,000 years, or 520 generations. Since this may be an overestimate of the length of the Aurignacian period (Zilhão and Pettitt 2006), data was also collected when each simulation reached 10,000 years, or 400 generations. The results of the model comparison analysis on these data (not included here) showed similar results to those reported in section 3.1.2.2, suggesting that the difference in length of the simulations considered does not affect the fit of simulated to observed data. As a result, further analyses were not performed on data collected at 400 generations.

As detailed in the description of the modelling framework in section 2.1.3.5, each group in the simulation deposits its material culture at the locations of the archaeological sites in the observed data (Vanhaeren and d’Errico 2006) and overwrites any existing deposits in the site if it comes within closer proximity to that site than the last group that deposited its culture there. The simulated material culture data is therefore a collection of items selected from different groups’ material cultures (each of which is the result of mutation, drift and cultural interaction processes) and deposited at different points throughout the time period of interest; the process of a group depositing its culture is only dependent on the geographic distance between the group and the location of the site and deposits are made with equal probability throughout the simulation. Each site is considered to be single occupancy – only the material culture of the group that deposited material closest to that particular site is considered. Assuming that each site is single occupancy may be misleading since the observed data (Vanhaeren and d’Errico 2006) cannot be chronologically resolved and some sites may feature multiple layers that were deposited thousands of years apart within the period of interest.
Furthermore, the two environmental scenarios used to condition the demography of the simulation space in this study are not realistic. In the first scenario, Europe is assumed to have a flat distribution of carrying capacities; this is clearly a simplistic and unrealistic view since topographic and climatic variation within the geographic region considered during the time period of interest would have had an impact on differences in habitability, and therefore the carrying capacity values, of geographic locations at different points in time. In the second scenario, information on estimated population densities is taken from the Bocquet-Appel et al. (2005) study to inform on carrying capacities. The reported geographic distribution and relative estimates of Upper Palaeolithic population size are an indicative starting point; however, the study itself could be considered somewhat controversial since the millennial scale climatic variation observed during the time periods that are considered is not taken into account. The geographic region during the time period of interest in the current study is characterised by a number of rapid climatic changes (Banks et al. 2008) and it is therefore unrealistic to consider the environment, and the resulting potential population densities, as static for the entire duration of a simulation. Since these environmental scenarios are unrealistic, it is important to consider how the results would be affected when simulations are conditioned on modelled environments that take into account the climatic variability across the geographic region considered during the time period of interest (Banks et al. 2008).

Additionally, this model does not take into account variation in elevation and the heterogeneous topography across the geographic area of interest. Just as the climatic fluctuations during the time period considered would have affected carrying capacities, geographic variation would have had an effect on demographic processes such as population migrations. Since this is the case, it is important to consider how the results would be affected when this spatial variation is taken into account in the model.

In addition to these caveats, the modelling framework in this chapter is relatively computationally inefficient, an important issue in the context of the Approximate Bayesian Computation framework (details of which are discussed in section 1.3.4); since a very large number of simulations are required, it is essential that the models are scalable in terms of computational efficiency. Furthermore, there are some issues with the kernel density methods used to model the fission / extinction process (detailed in section 2.1.2.2). These issues are discussed in section 4.1.2, but it is important to note here that the calculation is not transparent or intuitive, and is computationally inefficient.

Since the number of parameter sets considered in this chapter is relatively small in the context of the ABC approach, it could be argued that exploring simple models (such as those described in this chapter) more fully before moving to more realistic models (such as those described in Chapter 4) would allow for more theory building (i.e. understanding the model dynamics more fully) and more extensive exploration of the large parameter space. However, given the extent of the necessary modifications and additions to the current modelling framework as described in
the previous paragraphs, the focus of this thesis is on refining the modelling framework by addressing each of the caveats discussed above as detailed in Chapter 4.
Chapter 4: Modelling Patterns of Material Culture in Upper Palaeolithic Europe (II)

Summary

This chapter details the methodological refinements that have been made to the simulation model described in section 2.1 following the results and observations in Chapter 3. These refinements are made in an attempt to address each of the caveats discussed in section 3.2, and in part focus on integrating multiple sources of data to more realistically represent the conditions that characterise the geographic area and time period of interest in this study. Palaeoclimate data have been integrated in order to condition the demography of the simulation space, while elevation data have been integrated to condition the extent of group migrations. The methods of integrating each of these sources are described in this chapter. Additionally, methodological developments of the process of estimating population density, namely using the distance between groups as a proxy for population density, as well as the process of depositing cultures at archaeological sites, in order to account for the cumulative aspect of the archaeological record, are described.

The chapter then goes on to present the results of the updated simulation model. The results of model comparison analyses (as described in section 2.2.2), parameter estimation analyses (as described in section 2.2.3) and correlation analyses (as described in section 2.2.4) are presented.

Unless otherwise specified, the simulation modelling framework was implemented in the Python programming language (http://www.python.org/). Approximate Bayesian Computation (ABC) analyses were performed, and results visualised, using the R programming language (http://www.r-project.org/).
4.1 Changes to Simulation Modelling Methods described in Chapter 2

This section details the methodological refinements that have been made to the simulation model described in section 2.1. Unless specifically discussed below, the model processes are as described in section 2.1.

4.1.1 Carrying Capacities and Environmental Scenarios

As mentioned in section 3.2, the geographic region during the time period of interest in the current study is characterised by a number of rapid climatic changes (Banks et al. 2008), which means that it is unrealistic to consider the environment, and the resulting carrying capacities, as static for the entire duration of a simulation, as is done in the previously described framework (section 2.1.1.3). It is therefore important to consider how the results of the simulation model are affected when simulations are conditioned on modelled environments that take into account the climatic variability across the geographic region considered during the time period of interest.

Additionally, since choice of environment has a strong effect on fitting simulated to observed data, as evidenced by the results in section 3.1.2.2, in order to test the proposed hypothesis (i.e. examine the effect of cultural interactions informed by differences in material culture on this fit) and have confidence in the results, it stands to reason that environmental scenarios used to condition the demography of the simulation space should reflect as closely as possible the climatic changes that occurred.

In this study, this is realised by using simulated Palaeoclimate data to inform on the relative carrying capacity values, and therefore potential population densities, of locations in the region of interest. Since Palaeoclimate data are available at different time points throughout the time period of interest, this approach allows the observed climatic variability to be integrated by updating the carrying capacities in the modelled domain throughout each simulation.

Following the precedent set by Eriksson et al. (2012), Palaeoclimate is linked to human demography by using climatic variables to estimate the annual Net Primary Productivity (NPP) for each location in the region of interest. NPP provides a general proxy for food availability and has been shown to be a predictor of demographic patterns in ecological studies (Binford 2001; Luck 2007); it is therefore an informative proxy for carrying capacity values, and thus potential population densities.

It should be noted here that the NPP resulting from marine habitats is not explicitly taken into account, which is a potential caveat of the current model. However, while marine habitats can generate higher NPP compared to terrestrial habitats, there is limited evidence for dependence of Aurignacian populations on marine resources, so this choice is in part justified.
4.1.1.1 Climate Data

Each simulation is conditioned on one of three environmental scenarios: either a uniform distribution of relative carrying capacity values across all locations within the modelled domain and over the length of the simulation (Figure 2.2, panel (A)), or one of two climate data sets as described below.

4.1.1.1.1 Singarayer and Burrough (2015)

The first of the two climate data sets is an updated version of the climate data used in the Eriksson et al. (2012) study for the purpose of linking climate to human demography. In their study, Eriksson et al. (2012) developed a spatially explicit model of the expansion of anatomically modern humans (AMH) that uses Palaeoclimate data to quantify the effects of climate on human demography. They conditioned simulations on genetic variation in modern human populations using Approximate Bayesian Computation (ABC) in order to estimate past demographic parameters.

The climate data sets, kindly provided by Joy Singarayer, are outputs of snap-shot simulations that form detailed global climate reconstructions for the last 120,000 years, performed with the Hadley Centre Climate Model (HadCM3) by Singarayer and Burrough (2015). The HadCM3 is a general circulation model consisting of linked interactive atmospheric, interactive ocean and static sea-ice models (Gordon et al. 2000; Pope et al. 2000) that has been shown to have the capability of capturing historical climate change (Stott et al. 2000). The data sets used here are outputs of a version of the HadCM3 model originally described by Singarayer and Valdes (2010) that also includes an interactive vegetation model; the data used in the Eriksson et al. (2012) study were generated by a version of the same model that did not include interactive vegetation. The climate reconstructions cover the last 120,000 years in 62 snap-shots distributed at intervals of every 1,000 years between the (pre-industrial) present day and 21 Ka (Last Glacial Maximum), every 2,000 years between 21 Ka and 80 Ka, and every 4,000 years between 80 Ka and 120 Ka. The spatial resolution of each data set is 2.5° latitude and 3.75° longitude.

4.1.1.1.2 Banks et al. (2008)

As an alternative to this climate data set, climate data generated for a study that considered the current geographic region and time period of interest were used; these data were kindly provided by William Banks. These climatic simulations Banks et al. (2008) were created using the LMDZ3.3 Atmospheric General Circulation Model (Jost et al. 2005) and focus on three climatic phases during which the majority of AMH colonisation of Europe is believed to have occurred, namely the Greenland Interstadials 9-11 (from 43.3 Ka to 40.2 Ka), Heinrich Event 4 (from 40.2 Ka to 38.6 Ka) and Greenland Interstadial 8 (from 38.6 Ka to 36.5 Ka), as detailed by Banks et al. (2008). The spatial distribution of the climate data points for these three phases is not even across the geographic region of interest, but data are available at an approximate resolution of 50 km.
4.1.1.2 The Miami Model: Relating Climate Variables to Net Primary Productivity

The aspects of the climate reconstructions of interest in this study were mean annual temperature and total annual precipitation. Net Primary Productivity (NPP) was estimated from these variables using the Miami Model that was first proposed by Lieth and Whittaker (1975: Ch 12) and subsequently described by Adams et al. (2004). This model does not explicitly deal with seasonality or the effects of carbon dioxide, light capture or humidity, but instead provides a simple estimate of NPP values based on observed values of key climatic variables. It therefore provides a convenient method for quantifying the effects of major climatic variables on productivity.

In the Miami Model, the annual NPP, measured in g/m²/year and denoted \( NPP(T, P) \), is estimated to be the lower of the productivity values independently predicted from temperature and precipitation variables, denoted \( NPP(T) \) and \( NPP(P) \) respectively, such that:

\[
NPP(T, P) = \min\{NPP(T), NPP(P)\}.
\]

(4.1)

Each independently-estimated productivity value is calculated using an empirical function appropriate to that climatic variable. The temperature-dependent productivity level, denoted \( NPP(T) \), is defined to be:

\[
NPP(T) = \frac{k_1}{1 + e^{k_2 - k_3 T}},
\]

(4.2)

where \( T \) is the mean annual temperature in degrees Celsius and \( k_i \) are temperature response parameters defined to take values \( k_1=3000, k_2=1.315 \) and \( k_3=0.119 \), respectively (values of temperature response parameters estimated by Lieth and Whittaker (1975) from graphic analysis of the correlation between observed NPP and temperature values). This function describes a sigmoidal increase in NPP with temperature, such that the strongest increase in productivity is between 10°C and 25°C, and maximum productivity at 25°C.

The precipitation-dependent productivity level, denoted \( NPP(P) \), is defined to be:

\[
NPP(P) = k_4 (1 - e^{-k_5 P}),
\]

(4.3)

where \( P \) is the total annual precipitation in millimetres and \( k_i \) are precipitation response parameters defined to take values \( k_4=3000 \) and \( k_5=0.000664 \), respectively (values of precipitation response parameters estimated by Lieth and Whittaker (1975) from graphic analysis of the correlation between observed NPP and precipitation values). This function describes an asymptotic increase in NPP with precipitation, with positive NPP at all non-zero precipitation values, assuming maximum productivity of 3,000 g/m²/year.
4.1.1.3 Calculating Net Primary Productivity (NPP) for the Region of Interest

To obtain NPP data for the time period and geographic area of interest in this study, it was first necessary to use the Miami Model equations described above to calculate global NPP values at the available spatial resolution for each snap-shot of interest (i.e. the snap-shots that fall between 29 Ka and 42 Ka as described in section 2.1.1.2, namely 7 snap-shots in the first climate data set, corresponding to 28 Ka, 30 Ka, 32 Ka, 34 Ka, 36 Ka, 38 Ka and 40 Ka, and all 3 snap-shots in the second climate data set, corresponding to the three climatic phases detailed in section 4.1.1.1.2), and then interpolate the global NPP data calculated for each of these snap-shots to the required spatial resolution (i.e. 0.2° latitude and 0.2° longitude as described in section 2.1.1.1). Following interpolation, the NPP values for those longitude and latitude values corresponding to the geographic area of interest (i.e. longitudes ranging from -11° to 30° and latitudes ranging from 35° to 60° as described in section 2.1.1.1) were extracted. NPP was calculated prior to data interpolation to ensure more accurate values – if individual components of the climate data (i.e. temperature and precipitation) are interpolated individually and NPP calculated from interpolated values of these variables, any uncertainties that arise during the interpolation would be amplified when combining these variables to calculate NPP.

The interpolations were performed using the function smooth.2d() from R’s fields package, with the default Gaussian kernel and a bandwidth of 0.55. The interpolation was performed under a range of bandwidths and the outputs compared using various metrics (maximum, minimum, mean, standard deviation, variance) – a value of 0.55 was chosen for the bandwidth as the data interpolated under this value had minimum NPP value closest to 0 (which is a physical/biological constraint) and maximum NPP value closest to maximum of input NPP data. The function interp() from R’s akima package was also considered, however when comparing the interpolated NPP values output by the two functions, smooth.2d() showed a better correlation with input data at locations where input data was available, and was therefore favoured.

Differences in NPP values between consecutive snap-shots were assessed in order to determine whether it is necessary to consider all snap-shots within the time period of interest. If differences were relatively small, it would be possible to justify using only those snap-shots in which NPP values are significantly different as, over millions of simulations, this would reduce computation time. In this case, differences between the snap-shots were relatively large and so all the snap-shots were used to update the productivity values that condition the demography of the simulation space at appropriate time points over the length of the simulation.

Observed hunter-gatherer data (Binford 2001) were used to determine physical / biological constraints, primarily the minimum value of NPP that corresponds to non-zero population density. This value, determined from the observed hunter-gatherer data (Binford 2001) to be 47.9927, was used to adjust the NPP estimates such that all locations with NPP values estimated to be below this observed minimum were deemed to be uninhabitable.
4.1.2 Fission / Extinction Processes and Distance to Nearest Neighbour as Proxy for Population Density

In this chapter, in order to condition the demography of the simulation space on Net Primary Productivity (NPP) data as discussed in section 4.1.1, it is necessary to develop methods for calculating the potential, or target, population density of each location as well as the population density at each generation.

In the version of the modelling framework described in section 2.1, the potential, or target, population density of each location is calculated as the product of the relative carrying capacity value at that location and the \( N_{\text{groups}} \) parameter. As detailed in section 2.1.2.2, at each generation, the difference between this potential population density and the current population density, estimated from the group locations using kernel density estimation (Wand and Jones 1995), is used to determine whether a particular group, selected with parameterised probability \( p_{f/e} \), undergoes a fission or an extinction process.

However, there are some issues with the kernel density methods used. In particular, the density estimate is heavily dependent on the bandwidth of the kernel, an arbitrary value for which there is little information available. Additionally, the calculation is not transparent or intuitive, and is computationally inefficient.

4.1.2.1 Distance to Nearest Neighbour as Proxy for Population Density

Because of the points discussed above, it is necessary to introduce an alternative method for calculating local population densities and target population densities for each location. This method should be intuitive and free of arbitrary parameters or non-transparent functions. In order to achieve this, a method has been developed that uses the distance between groups as a proxy for population density. As in the models described in section 2.1, the population density determines which process (i.e. fission or extinction) each group selected to undergo a fission / extinction process (with parameterised probability \( p_{f/e} \)) actually undergoes.

Given the aim of conditioning human demography on NPP, and the fact that the maximum number of groups that the modelled domain can sustain is parameterised (since there is little information in the archaeological record relating to the precise values that this parameter may take), using the distance to nearest neighbour is an intuitive method for estimating local population density. Intuitively, if a selected group is too close to its nearest neighbour according to the expected distance at that group’s location, then the local population density is too high (i.e. the location is over-populated) and this results in an extinction event. Conversely, if a selected group is further away from its nearest neighbour than expected, then the local population density is low (i.e. the location is under-populated) and this results in a fission event.
The challenge in using distance to nearest neighbour as a proxy for population density is determining the distance at each location that would be expected given a particular underlying distribution of carrying capacity values (i.e. NPP values) and a particular value of the \( N_{\text{groups}} \) parameter. The aim is therefore to generate a distribution of expected distances that, through migratory and fission / extinction processes, gives rise to the correct number of groups (i.e. \( N_{\text{groups}} \)) and the correct distribution of those groups in space (i.e. mirroring the input NPP distribution). The approach used to calculate the expected distance values, along with the necessary assumptions, is described below.

Once the expected distance to nearest neighbour at each location in the modelled domain is known, the difference between the expected and current distance to nearest neighbour can be used as an indicator of potential for local population growth; a negative value indicates that the location is below carrying capacity (i.e. the expected distance is smaller than the current distance – the location is under-populated and so there is potential for growth) and therefore results in a fission event, while a positive value indicates that the location is above carrying capacity (i.e. the expected distance is greater than the current distance – the location is over-populated and so there is no potential for growth) and therefore results in an extinction event.

### 4.1.2.2 Calculating Expected Distance to Nearest Neighbour: 2D Modelled Domain Characterised by Single NPP Value

Although conceptually transparent, the calculation of expected distance values has several steps. To explain the approach, a simple scenario of a two-dimensional land plane over which the distribution of NPP values, and therefore carrying capacity values, is uniform is considered first (translated to the modelled domain, this corresponds to the environmental scenario in which Europe is assumed to be a flat space). Since each location is equally habitable, the expected distance between a group and its nearest neighbour should be the same at all locations, which in turn means that the groups are evenly distributed in space.

Because the groups are evenly distributed in space, the total land area is evenly divided among the groups. The amount of land area available to each group, denoted \( A_{\text{group}} \), can therefore be calculated from the total available land area, denoted \( A_{\text{total}} \), and the total number of groups the domain can sustain, given by the \( N_{\text{groups}} \) parameter: the more groups there are, the smaller the proportion of the total area available to each group and so the smaller the expected distance between groups. Since the total available land area is divided evenly amongst the groups, this can be expressed as:

\[
A_{\text{group}} = \frac{A_{\text{total}}}{N_{\text{groups}}} \tag{4.4}
\]
This area can then be used to calculate the expected distance to nearest neighbour at each location, given some assumptions. The distance to nearest neighbour defines a perimeter around each group that is a border between areas of fission and extinction. For a particular group selected to undergo a fission / extinction process, this perimeter represents the area within which the presence of another group will result in the extinction of the original group; conversely, if there are no other groups within this perimeter then the group would undergo fission. Assuming that the distance from a particular group to its perimeter is the same in each direction inherently leads to each group occupying a circular area.

From knowing the amount of land area available to each group, denoted \( A_{\text{group}} \), and assuming that this area is represented by a circle, the expected distance between that group and the boundary of its respective area, denoted \( r \), is given by:

\[
A_{\text{group}} = \frac{A_{\text{total}}}{N_{\text{groups}}} = \pi r^2,
\]  

(4.5)

which, rearranged, gives:

\[
r = \sqrt{\frac{A_{\text{group}}}{\pi}} = \sqrt{\frac{A_{\text{total}}}{\pi N_{\text{groups}}}}.
\]  

(4.6)

This is the distance between each group and the boundary of its respective area. The distance between each group and its nearest neighbour is given by the sum of the distances between the two groups and their respective boundaries, as shown in Figure 4.1. Since in this case the groups are distributed evenly, the \( r \) for each group is the same, and so the distance to nearest neighbour for each group is given by:

\[
D(\text{NN}) = 2 \cdot r = 2 \cdot \sqrt{\frac{A_{\text{total}}}{\pi N_{\text{groups}}}}.
\]  

(4.7)

This gives an expected distance to nearest neighbour, which the distance calculated at each generation in the simulation can be compared against in order to determine which of the two processes (i.e. fission or extinction) each group in the simulation undergoes.

### 4.1.2.2.1 Number of Nearest Neighbours

At each generation in the simulation, the current distance to nearest neighbour for each group is calculated as the average of the distances to that group’s six nearest neighbours, as shown in Figure 4.2. This is done instead of using the true distance to the nearest neighbour, which is a very noisy measure that uses limited information and is therefore not entirely representative of the true density and spatial distribution of groups. Additionally, assuming that each group occupies a circular area results in an hexagonal close packing pattern (analogous to the mathematical problem of packing equally sized circles on a planar surface). From this
observation, it is intuitive to consider the six nearest neighbours in the calculation of the distance to nearest neighbour.

### 4.1.2.2.2 Area Loss due to Circle Tessellation

Assuming that each group occupies a circular area, however, results in a proportion of the total available area remaining unoccupied; this is due to the tessellation properties of circles, as shown in Figure 4.2. This in turn means that the number of groups in the simulation will always be less than that specified by the \( N_{\text{groups}} \) parameter.

In order to account for the area lost due to circle tessellation, the amount of area per group is increased by a factor of \( \frac{2\sqrt{3}}{\pi} \). This factor arises from the ratio of the occupied area and the total available area, and can be calculated by considering the triangle shown in Figure 4.3. The area of the whole triangle is given by:

\[
A_{\text{triangle}} = r \cdot 2r \cos \frac{\pi}{6} = 2r^2 \frac{\sqrt{3}}{2} = r^2 \sqrt{3},
\]

while the area of the triangle occupied by the circles is given by:

\[
A_{\text{circles}} = 3 \cdot \frac{\pi r^2}{6} = \frac{\pi r^2}{2}.
\]

The ratio of these two areas gives the proportion of the area occupied:

\[
\frac{A_{\text{circles}}}{A_{\text{triangle}}} = \frac{\frac{\pi r^2}{2}}{r^2 \sqrt{3}} = \frac{\pi}{2\sqrt{3}}.
\]

Adjusting (increasing) the available area by a factor of \( \frac{2\sqrt{3}}{\pi} \) in the above calculation of expected distance to nearest neighbour accounts for the effects of circle tessellation and ensures that the entire available area is available for occupation.
Figure 4.1: Distance between each group and its nearest neighbour, $D(\text{NN})$, is given by the sum of the distances between the two groups and their respective boundaries (i.e. $r$).

Figure 4.2: Six nearest neighbours (shaded blue) of central group, with hexagonal circle packing.

Figure 4.3: Area shaded in blue is lost due to tessellation properties of circles.
4.1.2.3 Calculating Expected Distance to Nearest Neighbour: 2D Modelled Domain Characterised by Multiple NPP Values

Similar principles can be applied in a scenario where the carrying capacity values are not uniformly distributed, in order to calculate distances that will give rise to a distribution of groups that reflects the underlying NPP values.

To explain how these principles are applied to a more complex scenario, consider a two-dimensional land plane divided into two regions, as shown in Figure 4.4, one of which, denoted $A_2$, has an NPP value, and therefore carrying capacity, that is twice as high as that of the other, denoted $A_1$. This translates to the number of groups per unit area in $A_2$ being twice as high as that in $A_1$.

In order to account for this in the calculation of area per group, it is necessary to take into account the NPP value as well as the amount of area with this NPP value. The area of each region is therefore weighted by its respective NPP value and the total weighted area is then divided among the groups specified by the $N_{\text{groups}}$ parameter. The total weighted area, denoted $wA_{\text{total}}$, is calculated as:

$$wA_{\text{total}} = \sum_i A_i \cdot NPP_i,$$  \hspace{1cm} (4.11)

where $A_i$ and $NPP_i$ are the area and corresponding NPP value of region $i$, with $i$ iterating over the range of different areas in the modelled domain. Note that $\sum A_i$ gives the total land area of the modelled domain.

The number of groups in a particular region is the same fraction of the total number of groups in the domain, given by the $N_{\text{groups}}$ parameter, as the fraction of total weighted area in that region is of the total weighted area in the domain. In region $j$ whose area and NPP values are $A_j$ and $NPP_j$, respectively, the number of groups, denoted $N_j$, can therefore be expressed as:

$$N_j = \frac{A_j \cdot NPP_j}{wA_{\text{total}}} \cdot N_{\text{groups}} = \frac{A_j \cdot NPP_j}{\sum_i A_i \cdot NPP_i} \cdot N_{\text{groups}}.$$  \hspace{1cm} (4.12)

Thus the geographic area per group in region $j$, denoted $A_{\text{group}}(j)$, is given by:

$$A_{\text{group}}(j) = \frac{A_j}{N_j} = \frac{A_j}{\sum_i A_i \cdot NPP_i} \cdot \frac{NPP_j}{N_{\text{groups}}} = \sum_i \frac{A_i \cdot NPP_i}{N_{\text{groups}}}.$$  \hspace{1cm} (4.13)

Assuming that each group occupies a circular area as before, this can be used to calculate the expected distance to nearest neighbour in each part of the domain. For area $j$, this can be expressed as:
\[ A_{\text{group}}(j) = \frac{\sum_i A_i \cdot NPP_i}{NPP(j) \cdot N_{\text{groups}}} = \pi r(j)^2, \quad (4.14) \]

which can be rearranged to give:

\[ r(j) = \sqrt{\frac{\sum_i A_i \cdot NPP_i}{NPP(j) \cdot N_{\text{groups}} \cdot \pi}}. \quad (4.15) \]

As in the previous example, this is the distance between each group and the boundary of its respective area. The distance between each group and its nearest neighbour is given by the sum of the distances between the two groups and their respective boundaries.

In the example in section 4.1.2.2, each location is equally habitable, which results in the distance between each group and the boundary of its respective area (i.e. \( r \)) being the same, and so the distance to nearest neighbour for each group can be calculated simply by doubling the \( r \). This is not the case in the scenario where the modelled domain is characterised by multiple NPP values, and the resulting variation in the distance between each group and its boundary at different locations needs to be taken into account when calculating the expected distance to nearest neighbour. To do so, the expected distance to nearest neighbour for each group is calculated as the sum of the group’s \( r \) (i.e. the distance between the group and the boundary of its respective area), and the average \( r \) calculated over those locations that are between \( r \) and \( 2r \) of the group, as shown in Figure 4.5.

The distance calculated at each generation in the simulation can then be compared against this expected distance in order to determine which of the two processes (i.e. fission or extinction) each group in the simulation undergoes.
Figure 4.4: Domain modelled with two regions characterised by different NPP values. The NPP value of region $A_2$ is twice as high as that of region $A_1$; this translates to the number of groups per unit area in $A_2$ being twice as high as that in $A_1$.

Figure 4.5: Expected distance to nearest neighbour at each location is calculated as the sum of the $r$ at that location and the average $r$ calculated over those locations that are between $r$ and $2r$ of that location (shaded grey area).
4.1.2.4 Calculating Expected Distance to Nearest Neighbour: Modelled Domain
Representing the Geographic Region of Interest

As explained in section 2.1.1.1, the geographic region of interest in this study is modelled as a
demic grid made up of 38,656 demes, 24,104 of which are land demes and 14,552 of which are
sea demes, arranged to approximate the European landmass. When applying the method
described in section 4.1.2.3 to this geographic area, NPP data is available at the resolution of the
demes and so the expected distance to nearest neighbour is calculated on a demic level.

Since the Earth is approximately spherical, while distance per degree of latitude remains
approximately constant with distance from the equator at ~111.5 km, distance per degree of
longitude decreases from ~111.5 km at the equator to 0 km at the North and South poles. It
follows that the area per degree of latitude and longitude decreases with increasing distance
from the equator. Since the geographic region of interest is represented as a demic grid with a
constant resolution of 0.2° latitude and 0.2° longitude, it follows that deme area is not constant
across the range of latitude and longitude values representing the modelled domain. When
calculating the expected distance to nearest neighbour, deme area is taken into account in
equation (4.11) (i.e. the calculation of weighted area), with \( i \) taking values between 1 and 38,656
since there are 38,656 demes in the demic grid.

4.1.2.4.1 Geographic Boundaries and Coastlines

There are two special cases that need to be considered. Firstly, if the distance between a deme
and any of the four domain boundaries is less than the expected distance to nearest neighbour
(i.e. \( r \)) for that deme, as shown in Figure 4.6, the area specified by a circle of radius \( r \) that is
within the geographic boundary will be less than required since some of the area will be outside
the geographic boundary (grey area falling in the uninhabitable region in Figure 4.6). It is
therefore necessary to adjust (i.e. increase) \( r \) so that it defines a larger circle whose area within
the geographic boundary gives the required area (circle with radius \( r' \) in Figure 4.6, with the
shaded grey part of it accounting for the area of the smaller circle that is outside of the
boundary).

Similarly, if a proportion of the area specified by a circle of radius \( r \) for a particular deme falls on
sea demes (i.e. it is near or on a coastline), again depicted by Figure 4.6, it is necessary to adjust
(i.e. increase) \( r \) such that the area on land gives the required area. To do this, the proportion of
the circle radius \( r \) that is on land is calculated, and \( r \) is increased so that it defines a larger circle
that, with the same proportion of area on land demes, gives the required area.

4.1.2.4.2 Number of Nearest Neighbours

At each generation in the simulation, the current distance to nearest neighbour for each group is
calculated as the average of the distances to that group’s six nearest neighbours. In both the
cases described in section 4.1.2.4.1, there is not enough available area to accommodate six
immediately neighbouring groups, as shown in Figure 4.7. The number of nearest neighbours considered in the calculation of current distance to nearest neighbour is therefore proportionally scaled to reflect the unavailable area (e.g. if a particular group is exactly \( r \) away from the boundary, there is only enough room to accommodate 4 nearest neighbours, and so 4 nearest neighbours will be considered in the calculation).

**4.1.2.4.3 Temporal Variation in Climate**

As explained previously, the two climate data sets are made up of multiple snap-shots that represent variations in climatic variables, and therefore NPP, over the time period of interest. To account for this variation when calculating the expected distance to nearest neighbour, it is assumed that the snap-shot with the maximum total carrying capacity (i.e. maximum total weighted area as calculated in equation (4.11)) can accommodate the maximum possible number of groups, as given by the \( N_{groups} \) parameter, and that the maximum number of groups in other snap-shots is in the same proportion to \( N_{groups} \) as that snap-shot’s total carrying capacity is to the maximum total carrying capacity.
Figure 4.6: If a deme is within $r$ of a coastline or a geographic boundary, a proportion of the area specified by a circle of radius $r$ falls on sea demes or outside the geographic domain modelled, respectively (the shaded grey part of circle radius $r$ that is in the uninhabitable region). The circle with radius $r'$ defines a larger circle whose area within the geographic boundary gives the required area (i.e. the shaded grey part of the circle with radius $r'$ accounts for the area of the circle with radius $r$ that is outside of the boundary).

Figure 4.7: In cases where a deme is within $2r$ of a geographic boundary or coastline, there is not enough available area to accommodate six immediately neighbouring groups.
4.1.3 Migratory Processes and Topographic Roughness

As explained in section 3.2, topographic variation of the geographic region of interest is expected to have had an effect on demographic processes such as population migrations. This effect is not accounted for in the model described in section 2.1. The changes to the simulation model described here attempt to address this issue by modelling the effect on group migrations of the observed variation in the elevation, and the heterogeneous topography, across the geographic area of interest.

In order to account for the effects of elevation on population migrations, the initial idea was to use elevation values directly such that migrations would be relatively more difficult at high elevations than at low elevations (i.e. a migration distance value would be scaled such that it is shorter at high elevations than at low elevations).

There is ample evidence, however, of human populations living at and traversing high elevation domains; the traditional mountain passes from Tibet to Nepal are suggestions for extremes of height and the Nangpa La mountain pass from Tibet into the Khumbu, at an elevation of 5,806m, has been in use for hundreds of years (Cynthia Beall, personal communication).

The more challenging and pertinent variable for migrations is topographic roughness (Mark Maslin, personal communication). Topographic roughness is a concept from geology and geomorphology that attempts to characterise small-scale (less than 3m in height) terrain features for terrain typing and traversability determination, and can be estimated from high-resolution elevation data (Hobson 1972; Hoffman and Krotkov 1989). In this context, a terrain that is relatively rough will be more difficult to traverse than one that is relatively smooth, regardless of the elevation (i.e. a terrain at a low elevation that is relatively rough will be more difficult to traverse than one at a high elevation that is relatively smooth). In the context of migrations modelled here, estimates of topographic roughness, rather than elevation, are used to scale the extent of group migrations, as explained in section 4.1.3.2.

An elevation threshold of 5806m (corresponding to the elevation of the Nangpa La mountain pass from Tibet into the Khumbu) is enforced such that all locations at elevations greater than this threshold would be considered uninhabitable and migrations would be impossible; the elevation values of the geographic area of interest are in the range [0, 3761 m], thus no locations are excluded as a result of this elevation threshold.

4.1.3.1 Relating Elevation Data and Topographic Roughness

While there are several methods of quantifying topographic roughness, the approach here uses the Topographic Ruggedness Index (TRI) described by Riley et al. (1999), which quantifies
topographic roughness as the sum change in elevation between a grid cell and its eight neighbour grid cells.

Using high-resolution (90m or 3 arc-sec) elevation data (Jarvis et al. 2008) for the geographic area of interest, TRI values were calculated and averaged to the required resolution in QGIS, an open-source geographic information system application. After averaging to the required resolution, the TRI values of the geographic area of interest were in the range [0.00125, 96.875]. The elevation data used here are derived from NASA’s Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) data, which are freely available through the CGIAR Consortium for Spatial Information (CGIAR-CIS; http://srtm.csi.cgiar.org/).

4.1.3.2 Scaling Migratory Processes by Topographic Ruggedness Index (TRI)

The TRI value calculated for each location is used to scale the distance that a group can travel at that location; at locations with low values of the index (low topographic roughness) migrations are relatively easier, while at locations with high values of the index (high topographic roughness) migrations are relatively more difficult.

As in the model described in Chapter 2, in the current model, groups are subjected to migratory processes, modelled as parameterised Gaussian random walks, at each generation. The standard deviation of the normal distribution, whose value is given by the unknown parameter \(d_{\text{mig}}\), specifies the range of values that the migration distance is most likely to take in each of the East-West and North-South directions.

As before, the distances travelled by each group at each generation in the East-West and North-South directions are represented as independent picks from the above-described normal distribution, and the resultant distance and direction of movement are calculated from these distances using equations (2.1) and (2.2).

The proposed resultant location of each group is then calculated based on the group’s current location, the resultant distance and the direction of movement. In addition to the proposed resultant location of each group, proposed locations corresponding to independent movement in the East-West and North-South directions, respectively, are calculated using the corresponding distances picked from the above-described normal distribution. Should the resultant location be deemed uninhabitable (i.e. location corresponds to sea deme), independent movement in the East-West and North-South directions can be considered. Equations (2.3) and (2.4) are again used to calculate the longitude and latitude of each group’s three proposed locations (corresponding to movement in each of the resultant, East-West and North-South directions) from the current location and respective distances and directions of travel.
Differently to the method described in Chapter 2 (section 2.1.2.1), however, once the three proposed locations are calculated, TRI values are sampled at every 1% of the distance along the path between the current location and each proposed location, and used to calculate the mean TRI value along each path. This mean TRI value, denoted $\overline{TRI}$, is then used to calculate the scaling factor for each proposed movement as follows:

$$scaling\ factor = \frac{90}{\sqrt{90^2 + TRI^2}}.$$

(4.16)

The values used in equation (4.16) stem from the resolution of the high-resolution elevation data, which is available at a resolution of 90m (3 arc-sec). The scaling factor is therefore an estimate of the percentage decrease in migration distance as a result of topographic roughness along the path of travel, and is used to scale the distance of movement in each direction. Scaling the movement distance ensures that the roughness of the path is accounted for; if $\overline{TRI}$ is small (i.e. the path is relatively smooth), the scaling factor is approximately 1 and the scaled distance is approximately equal to the true distance, while, if $\overline{TRI}$ is large (i.e. the path is relatively rough), the scaling factor is very small and the scaled distance is a small fraction of the true distance.

Once the scaled distance of movement in each direction has been calculated, equations (2.3) and (2.4) are used once more in order to calculate the longitude and latitude of each group’s proposed location from the current location and respective scaled distances and directions of travel.

Once calculated, the viability of each proposed location is checked. If the proposed resultant location is within a land deme, it is accepted as the group’s new location. However, if the proposed resultant location is within a sea deme, independent movements in the East-West and North-South directions are considered. If both result in locations that are within land demes, then one is picked at random and the proposed location corresponding to that movement is accepted as the group’s new location. If only one results in a location that is within a land deme, then the proposed location corresponding to that movement is accepted as the group’s new location. Should all three movements result in a location that is within a sea deme, the group remains at its current location.

One drawback of this method is that the scaling factor is calculated along the whole path between the current location and the initial proposed location, rather than incrementally accounting for the change in TRI along the path. This means that the TRI of a section of the path near to the initial proposed location is taken into account even when this section would not be reached due to distance scaling. This simplification is made for practical and computational economy reasons.
4.1.4 Collecting Simulated Data (Cumulative Sites)

The simulation model described in Chapter 2 (section 2.1) assumes that each archaeological site is single occupancy (i.e. only the material culture of the group that deposited material closest to that particular site is considered). As explained previously, assuming that each site is single occupancy may be misleading since the observed data (Vanhaeren and d’Errico 2006) cannot be chronologically resolved and some sites may feature multiple layers that were deposited thousands of years apart within the period of interest. The changes to the simulation model described here therefore attempt to address the inconsistency of this assumption with the cumulative aspect of the archaeological record.

To do this, the depositing process described in section 2.1.3.5 has been modified such that, instead of overwriting previous deposits at a particular site, a group depositing its culture at that site simply adds its entire culture, including information on the frequency of each trait, to the existing deposits. At the end of the simulation, which corresponds to the end of the Aurignacian period, a number of unique items, equal to that recovered from the site as reported in the original study (Vanhaeren and d’Errico 2006), are selected probabilistically for each site (using a multinomial function so that trait frequencies are taken into account), such that the probability of an item being selected is proportional to the frequency with which it occurs in that site.

The remainder of the depositing process has been kept the same as described in section 2.1.3.5. The locations of archaeological sites in the model are defined to correspond to the locations of the archaeological sites in the observed data (Vanhaeren and d’Errico 2006), shown in Figure 2.3. A group will deposit its culture at a site when within a specified geographic distance of that site. This catchment distance is initially set to be equal for all sites, and updated iteratively for each site to be the distance of the group that deposited material closest to that to that site, with the further constraint that if two groups are within the catchment distance then the group closest to the site will be the one to deposit its culture there. As described in section 2.1.3.5, the distance between group locations and archaeological sites is quantified using the geodesic distance measure, and the Haversine Formula (Sinnott 1984) is used as described in section 2.1.3.4.1 to account for curvature of the Earth in the calculation.
4.1.5 The Models: Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD)

The Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD) are both models of cultural identity-by-descent with modification and isolation-by-distance, and are made up of the demographic and cultural processes described in section 2.1, with the changes to the framework implemented as detailed above. As in the models described in Chapter 2, the difference between the NMCI and CD models lies in the role played by material culture in determining the nature of interactions between groups.

As a reminder, the overall aim is to test which of these models best explains the spatial distribution of ornament types observed in the archaeological record, and the CD model is intended to represent the effects of ethnic structuring on the spatial distribution of material culture.

Figure 4.8 is a flow diagram illustrating the order in which the demographic and cultural processes described above are performed during each simulation. The processes grouped by the dashed grey line are iterated at every generation. Table 4.1 summarises the processes of the simulation model described in Chapter 2 and that model with the modifications and refinements described in section 4.1 above implemented, highlighting the major differences between the two.

4.1.6 Model Versions, Model Parameters and Prior Ranges

A total of 12 models are simulated under the modelling framework described in section 4.1: for each of the NMCI and CD models, data are simulated under all pairwise combinations of one of three environmental scenarios (either uniform carrying capacities (FLAT), or NPP calculated from Banks et al. (2008) climate data (NPPBanks), or NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)) and one of two models of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)).

As for the modelling framework described in section 2.1, the NMCI and CD models in this chapter have 4 key processes (although some of these processes are modelled differently, as described above): migration, fission / extinction, cultural mutation and cultural interaction, governed by 4 parameters: $d_{mig}$, $P_{f/e}$, $P_{mut}$ and $d_{int}$, respectively. As before, there are 2 further parameters in the models, namely the maximum number of groups, $N_{groups}$, and the number of items in each group’s cultural repertoire, $N_{items}$. Prior ranges for each parameter within each of the NMCI and CD models are kept the same as in the previous version of the modelling framework, and are listed in Table 2.1.
Figure 4.8: Flow diagram illustrating the order in which the demographic and cultural processes are performed during each simulation. The processes grouped by the dashed grey line are iterated at every generation. The parameter that governs each process is indicated in the box specifying that process. The population density estimation process, shaded in grey, is not parameterised.
Table 4.1: Outline of the major processes of the simulation model described in Chapter 2 (Version I) and that model with the modifications and refinements described in section 4.1 implemented (Version II). The shaded area represents features that are common to both models. When an entire process is common to both models, details of the process are given once and the corresponding columns for the two models are merged. Italicised text in some cells attempts to further highlight the differences between the two models.

<table>
<thead>
<tr>
<th>Major Processes</th>
<th>Version I</th>
<th>Version II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initialise simulation world:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>geographic area modelled</td>
<td>European territory (with modern coastlines due to dramatic geostatic rebound for northern latitudes):</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- latitude, $\lambda$, ranges from 35° to 60° (relative to modern day country boundaries, this is approximately the area from the northern boundary of Africa to the northern boundary of Scotland)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- longitude, $\phi$, ranges from -11° to 30° (relative to modern day country boundaries, this is approximately the area from the western boundary of Ireland to the western boundary of Russia at the Urals)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- resolution – 1 deme = 0.2° latitude × 0.2° longitude (38656 demes: 24104 land demes, 14552 sea demes)</td>
<td></td>
</tr>
<tr>
<td>length of simulation</td>
<td>13,000 years (including 1000 year burn-in period at start of simulation), i.e. 520 generations (including 40 generation burn-in period at start of simulation) assuming a 25 year generation time (Tremblay and Vezina 2000; Thomas et al. 2006)</td>
<td></td>
</tr>
<tr>
<td>carrying capacity / target population density of each deme</td>
<td>Each simulation is conditioned on ONE of two environmental scenarios:</td>
<td>Each simulation is conditioned on ONE of three environmental scenarios:</td>
</tr>
<tr>
<td></td>
<td>- FLAT (constant over length of simulation): all</td>
<td>- FLAT (constant over length of simulation): all</td>
</tr>
<tr>
<td>Major Processes</td>
<td>Version I</td>
<td>Version II</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>locations have equal carrying capacity values</td>
<td>locations have equal carrying capacity values</td>
</tr>
<tr>
<td></td>
<td>- <strong>Bocquet-Appel</strong> (constant over length of simulation): normalised estimated population densities during the Aurignacian from Bocquet-Appel et al. (2005)</td>
<td>- <strong>NPPBanks</strong> (constant over length of simulation): NPP values for each location calculated following the Miami Model (Lieth and Whittaker 1975: Chapter 12) using values of climatic variables generated by climate model from Banks et al. (2008)</td>
</tr>
<tr>
<td></td>
<td>Target <em>population density</em> calculated as product of normalised environmental data and <em>N</em>_{groups} parameter.</td>
<td>- <strong>NPPSingarayer</strong> (constant over length of simulation): NPP values for each location calculated following the Miami Model (Lieth and Whittaker 1975: Chapter 12) using values of climatic variables generated by climate model from Singarayer and Burrough (2015)</td>
</tr>
<tr>
<td></td>
<td>Population density at current generation estimated from group locations using kernel density estimation (Wand and Jones 1995).</td>
<td>Expected <em>distance to nearest neighbour</em> calculated from environmental data and <em>N</em>_{groups} parameter, and used as proxy for population density.</td>
</tr>
<tr>
<td></td>
<td>Expected <em>distance to nearest neighbour</em> calculated from environmental data and <em>N</em>_{groups} parameter, and used as proxy for population density.</td>
<td>Distance to nearest neighbour at current generation calculated as average distance between each group and its six nearest neighbours.</td>
</tr>
</tbody>
</table>

2. **Select value for each parameter of interest from pre-defined prior range (fixed for current simulation):**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>d</em>_{mig}</td>
<td>migration distance (km): value selected from discrete flat uniform prior, range = [1, 100]</td>
</tr>
<tr>
<td>Major Processes</td>
<td>Version I</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>$d_{int}$</td>
<td>interaction distance (km): value selected from discrete flat uniform prior, range = [0, 50]</td>
</tr>
<tr>
<td>$p_{f/e}$</td>
<td>probability of fission / extinction: continuous flat uniform prior, range = [0, 1]</td>
</tr>
<tr>
<td>$p_{mut}$</td>
<td>probability of cultural mutation: continuous flat uniform prior, range = [0, 0.2]</td>
</tr>
<tr>
<td>$N_{groups}$</td>
<td>maximum number of groups modelled area can sustain: discrete flat uniform prior, range = [50, 1000]</td>
</tr>
<tr>
<td>$N_{items}$</td>
<td>number of items in each group’s cultural repertoire: discrete flat uniform prior, range = [500, 1500]</td>
</tr>
</tbody>
</table>

3. Iterate population processes over defined number of generations:

<p>| group migration | Group migrations are parameterised Gaussian random walks, with the distance travelled by each group in the East-West and North-South directions represented as independent picks from a parameterised Gaussian distribution ($\sigma_{mig}=d_{mig}$; $\mu_{mig}=0$, i.e. positive and negative values picked from the distribution correspond to movement in opposite directions). Proposed location (corresponding to resultant movement) calculated from current location and distance moved in each of East-West and North-South directions. | Group migrations are parameterised Gaussian random walks, with the distance travelled by each group in the East-West and North-South directions represented as independent picks from a parameterised Gaussian distribution ($\sigma_{mig}=d_{mig}$; $\mu_{mig}=0$, i.e. positive and negative values picked from the distribution correspond to movement in opposite directions). Initial proposed locations (corresponding to resultant movement, independent movement in East-West direction and independent movement in North-South direction) calculated from current location and distance moved in |</p>
<table>
<thead>
<tr>
<th>Major Processes</th>
<th>Version I</th>
<th>Version II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viability of proposed location checked:</td>
<td></td>
<td>each of East-West and North-South directions.</td>
</tr>
<tr>
<td>- proposed location on land: proposed location accepted as new location</td>
<td></td>
<td>Distance of movement in each of the three directions (resultant, East-West, North-South) then scaled by <strong>topographic roughness</strong> along path between current location and each initial proposed location (i.e. proposed location calculated with no distance scaling).</td>
</tr>
<tr>
<td>- proposed location on sea: group remains at current location</td>
<td></td>
<td>Proposed locations (corresponding to resultant movement, independent movement in East-West direction and independent movement in North-South direction) calculated from current location and scaled distance moved in each direction, respectively.</td>
</tr>
</tbody>
</table>

Viability of each proposed location checked:

| | | |
| - proposed resultant location on land: proposed resultant location accepted as new location | | |
| - proposed resultant location on sea BUT independent movement in East-West AND/OR North-South directions results in a location on land: viable location accepted as new location | | |
| - proposed resultant location on sea AND independent movement in East-West AND North-South directions results in a location on sea: group remains at current location | | |

**cultural interaction**

unknown parameter: $d_{int}$

prior range: [0, 50]

If two groups are within $d_{int}$ of each other, they will undergo ONE of two interaction processes:
<table>
<thead>
<tr>
<th>Major Processes</th>
<th>Version I</th>
<th>Version II</th>
</tr>
</thead>
<tbody>
<tr>
<td>- ‘conflict’ – replacement of the culture of one group by that of the other; analogous to a group imposing its culture on a group they have defeated / assimilating the defeated group into their own, followed by a fission process</td>
<td>- ‘conflict’ – replacement of the culture of one group by that of the other; analogous to a group imposing its culture on a group they have defeated / assimilating the defeated group into their own, followed by a fission process</td>
<td></td>
</tr>
<tr>
<td>- ‘sharing’ – sharing of cultures between interacting groups, modelled as a permutation; analogous to groups swapping cultural traits</td>
<td>- ‘sharing’ – sharing of cultures between interacting groups, modelled as a permutation; analogous to groups swapping cultural traits</td>
<td></td>
</tr>
<tr>
<td>fission / extinction unknown parameter: $p_{f/e}$ prior range: [0, 1]</td>
<td>Each group undergoes a fission / extinction process with parameterised probability ($p_{f/e}$):</td>
<td>Each group undergoes a fission / extinction process with parameterised probability ($p_{f/e}$):</td>
</tr>
<tr>
<td></td>
<td>- <strong>extinction</strong> – group is deleted from the simulation</td>
<td>- <strong>extinction</strong> – group is deleted from the simulation</td>
</tr>
<tr>
<td></td>
<td>- <strong>fission</strong> – replication resulting in two copies of the group (parent and offspring) in the following generation; offspring group retains cultural traits of parent group (i.e. offspring group is an exact copy of parent group, except for any mutation events)</td>
<td>- <strong>fission</strong> – replication resulting in two copies of the group (parent and offspring) in the following generation; offspring group retains cultural traits of parent group (i.e. offspring group is an exact copy of parent group, except for any mutation events)</td>
</tr>
<tr>
<td>Type of process is determined by difference between target and current local population density at the group’s location (an indicator of potential for local population growth):</td>
<td>Type of process is determined by difference between expected and current average distance to nearest neighbour at the group’s location (an indicator of potential for local population growth):</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- $(\text{target} - \text{current}) &gt; 0$ – location below carrying capacity (under-populated) and so there is potential for growth $\Rightarrow$ <strong>fission</strong></td>
<td>- $(\text{expected} - \text{current}) &lt; 0$ – location below carrying capacity (under-populated) and so there is potential for growth $\Rightarrow$ <strong>fission</strong></td>
</tr>
<tr>
<td></td>
<td>- $(\text{target} - \text{current}) &lt; 0$ – location above carrying capacity (over-populated) and so there is no potential for growth $\Rightarrow$ <strong>extinction</strong></td>
<td>- $(\text{expected} - \text{current}) &gt; 0$ – location above carrying capacity (over-populated) and so there is no potential for growth $\Rightarrow$ <strong>extinction</strong></td>
</tr>
<tr>
<td>Major Processes</td>
<td>Version I</td>
<td>Version II</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>cultural mutation</td>
<td>Each item in each group’s culture undergoes a mutation process with parameterised probability ($p_{\text{mut}}$). Each simulation follows ONE of two mutation models:</td>
<td></td>
</tr>
<tr>
<td>unknown parameter: $p_{\text{mut}}$</td>
<td>- <strong>SW – discrete stepwise</strong>: cultural trait in a particular feature at the current generation is constrained to mutate to one of the cultural traits on either side of it, with equal probability, at the next generation.</td>
<td></td>
</tr>
<tr>
<td>prior range: [0, 0.2]</td>
<td>- <strong>DIS – discrete non-stepwise</strong>: cultural trait in a particular feature at the current generation can mutate to any of the other cultural traits within that feature, with equal probability, at the next generation.</td>
<td></td>
</tr>
<tr>
<td>cultural drift</td>
<td>Traits of each cultural feature are sampled independently taking into account their frequencies (analogous to genetic drift, where allele frequencies change as a result of random differences in reproduction).</td>
<td></td>
</tr>
<tr>
<td>depositing cultures in sites</td>
<td>A group deposits its culture at a site (locations of the archaeological sites defined in Vanhaeren and d’Errico (2006)) if it comes within closer proximity to that site than the last group that deposited its culture there. A deposit at a site will overwrite any existing deposit, i.e. sites are considered to be <strong>single occupancy</strong>. At the <strong>time of each deposit</strong>, a number of unique items, equal to the item capacity for that site (number of unique items recovered from the site as reported by Vanhaeren and d’Errico (2006)) are selected probabilistically from the group’s entire culture.</td>
<td>A group deposits its culture at a site (locations of the archaeological sites defined in Vanhaeren and d’Errico (2006)) if it comes within closer proximity to that site than the last group that deposited its culture there. A deposit at a site will add the group’s entire culture to the existing deposits, i.e. deposits are <strong>cumulative</strong>. At the <strong>end of the simulation</strong>, a number of unique items, equal to the item capacity for that site (number of unique items recovered from the site as reported in Vanhaeren and d’Errico (2006)) are selected probabilistically from the entire collection of deposits made at that site over the length of the simulation.</td>
</tr>
</tbody>
</table>
4. Record data of interest and save to output files:

<table>
<thead>
<tr>
<th>Major Processes</th>
<th>Version I</th>
<th>Version II</th>
</tr>
</thead>
<tbody>
<tr>
<td>record parameter values</td>
<td>value assigned to each parameter of interest</td>
<td></td>
</tr>
<tr>
<td>record simulated summary statistics</td>
<td>mean and variance of pairwise site similarity ((SI:sites) and (MI:sites))</td>
<td>mean and variance of pairwise ornament type similarity ((SI:ornaments) and (MI:ornaments))</td>
</tr>
<tr>
<td></td>
<td>mean and variance of pairwise ornament type similarity ((SI:ornaments) and (MI:ornaments))</td>
<td>variability in the number of occurrences of each ornament type ((MAD:ornaments))</td>
</tr>
<tr>
<td></td>
<td>variability in ornament types recovered from sites ((MAD:sites))</td>
<td>variability in ornament types recovered from sites ((MAD:sites))</td>
</tr>
<tr>
<td></td>
<td>mean and variance of spatial distribution of sites sharing each ornament types ((DR))</td>
<td>mean and variance of spatial distribution of sites sharing each ornament types ((DR))</td>
</tr>
<tr>
<td>record simulation output</td>
<td>bead types present and absent at each site</td>
<td></td>
</tr>
</tbody>
</table>
4.2 Results

This section presents the results of the spatially explicit cultural transmission simulation models described in section 2.1, with the refinements and changes described in section 4.1 implemented. The analyses performed and described in section 3.1 are repeated for the results from the refined modelling framework as described in this chapter.

As explained in section 2.1.1.2, each simulation is initialised at the onset of, and simulated forward in time to the end of, the Aurignacian period, spanning a total of 13,000 years, or 520 generations. Since this may be an overestimate of the length of the Aurignacian period (Zilhão and Pettitt 2006), data was also collected when each simulation reached 10,000 years, or 400 generations. The results of the model comparison analysis on these data (not included here) showed similar results to those reported in section 3.1.2.2, which suggests that the difference in length of the simulations considered does not affect the fit of simulated to observed data. As a result, further analyses were not performed on data collected at 400 generations.

4.2.1 Exploring the Behaviour of Summary Statistics

4.2.1.1 Are the summary statistics correlated?

As in section 3.1.1.1, in order to examine whether any pair of proposed summary statistics carry the same information about the dataset considered, a correlation analysis of simulated summary statistics values was performed. This was done for each of the 12 models as listed in section 4.1.6, considering 1,000,000 simulations generated by each model and again using Spearman’s rank correlation coefficient $R^2$ to assess the statistical dependence between each pair of summary statistics. The results for each model are shown in Appendix A.3. As for the results presented in section 3.1.1.1, all models showed similar patterns in pairwise summary statistic correlations. The results presented and discussed in this section are used as examples of the full set of results shown in Appendix A.3.

For the results presented here, the analysis was performed using 1,132,411 simulations generated by the Null Model with Cultural Interactions (NMCI) under each of the two best scenarios as identified by the model comparison analysis in section 4.2.2 (bounded stepwise (SW) model of cultural variation, and environmental scenarios with uniform distribution of carrying capacities (FLAT) and NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)). The results are shown in Figure 4.9, reporting the pairwise binned scatterplots (below diagonal), correlation coefficients $R^2$ (above diagonal; black) and associated p-values (above diagonal; blue) for the NMCI SW FLAT (upper panel) and NMCI SW NPPSingarayer (lower panel). As explained in section 2.2.4, the p-values are determined to an extent by the number of simulations considered, which is not is not set by any property of the
model and is therefore effectively arbitrary, and so the reported correlation coefficient $R^2$ values are considered to be the most informative indicator of correlation.

The scatterplots and correlation values shown in Figure 4.9 again indicate that correlations between proposed summary statistics do exist. In most cases these correlations are weak ($R^2 \leq 0.25$). Cases where strongest correlations are observed (i.e. S7: mean, and S8: variance, of mutual information between sites, and S3: mean, and S4: variance, of shared information between sites) are between different measures (i.e. mean and variance) of the same summary statistic. As noted before, even in the cases of these strongest correlations, the summary statistics are not completely correlated (NMCI SW FLAT: $R^2$(S3,S4) = 0.688, $R^2$(S7,S8) = 0.636; NMCI SW NPPSingarayer: $R^2$(S3,S4) = 0.683, $R^2$(S7,S8) = 0.592). It can therefore again be inferred that these pairs of summary statistics, although related, are carrying different information to some extent. As for results reported in section 3.1.1.1, similar conclusions can be made for all pairs of statistics, and so all 12 proposed summary statistics are included in further analyses.
Figure 4.9: Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic combinations using 1,132,411 successful simulations for the NMCI SW FLAT (top panel) and NMCI SW NPPSingarayer (bottom panel) models. The cells in each pairwise scatterplot (below diagonal) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (above diagonal) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value $\geq 0.05$). S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>S12</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.47</td>
<td>0.07</td>
<td>0.05</td>
<td>0.15</td>
<td>0.03</td>
<td>0.04</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>S2</td>
<td>0.07</td>
<td>0.36</td>
<td>0.17</td>
<td>0.06</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>S3</td>
<td>0.05</td>
<td>0.17</td>
<td>0.06</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>S4</td>
<td>0.10</td>
<td>0.27</td>
<td>0.14</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>S5</td>
<td>0.03</td>
<td>0.07</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>S6</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>S7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>S8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>S9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>S10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>S11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.
4.2.1.2 Relationship between Summary Statistics and Model Parameters

Similarly to the analysis performed in section 3.1.1.2, this analysis checks the dependence between pairwise summary statistic and model parameter combinations using simulated data, in order to inform on the relationship between summary statistics and model parameters.

A correlation analysis of each pairwise combination of summary statistic and model parameter was performed, for each of the 12 models as listed in section 4.1.6, considering 1,000,000 simulations generated by each model. Spearman’s rank correlation coefficient $R^2$ was used to assess the statistical dependence between each summary statistic and each model parameter, and the results for each model are shown in Appendix A.3. All models again showed similar patterns in correlations between pairwise combinations of summary statistic and model parameter. The results presented and discussed in this section are used as examples of the full set of results shown in Appendix A.3.

Presented here are the results of the correlation analysis of each pairwise combination of summary statistic and model parameter performed using 1,132,411 simulations generated by the NMCI SW FLAT (Figure 4.10) and NMCI SW NPPSingarayer (Figure 4.11) models. The two figures give the pairwise binned scatterplots (top panel), the correlation coefficients $R^2$ (bottom panel; black) and associated p-values (bottom panel; blue) for each model respectively. Again it should be noted that the reported correlation coefficient $R^2$ values are considered to be the most informative indicator of correlation since the p-values are to an extent determined by the number of simulations considered, which is only constrained by the amount of computational time available and is therefore effectively arbitrary.

As observed for the results in section 3.1.1.2, each of the two models presented here, and the results of all models considered, show weak ($R^2 \leq 0.25$) correlations for most pairwise summary statistic and model parameter combinations. This again suggests that a combination of model parameters drive each summary statistic, or, conversely, that a combination of summary statistics inform on each of the model parameters.
Figure 4.10: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and model parameter combinations using 1,132,411 successful simulations for the NMCI SW FLAT model.

The cells in each pairwise scatterplot (top panel) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (bottom panel) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value ≥ 0.05). P1: probability of cultural mutation; S2: probability of fission / extinction; P3: migration distance (km); P4: total number of cultural items in each group’s cultural repertoire; P5: maximum number of groups; P6: interaction radius (km); S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.
Figure 4.11: Scatterplots (top panel) and correlation values (bottom panel; black: Spearman's rank correlation coefficient $R^2$, blue: associated p-value) of pairwise summary statistic and model parameter combinations using 1,132,411 successful simulations for the NMCI SW NPP Singarayer model.

The cells in each pairwise scatterplot (top panel) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (bottom panel) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value $\geq 0.05$). P1: probability of cultural mutation; S2: probability of fission / extinction; P3: migration distance (km); P4: total number of items in each group’s cultural repertoire; P5: maximum number of groups; P6: interaction radius (km); S1: mean(SI:ornaments); S2: var(SI:ornaments); S3: mean(SI:sites); S4: var(SI:sites); S5: mean(MI:ornaments); S6: var(MI:ornaments); S7: mean(MI:sites); S8: var(MI:sites); S9: mean(DR); S10: var(DR); S11: MAD:ornaments; S12: MAD:sites.
4.2.2 Model Comparisons

4.2.2.1 Null Model with Cultural Interactions or Culture-Dependent Interactions Model: Which model best explains the observed data?

As in section 3.1.2.2, by determining which of the proposed models best fits the observed data, this model comparison aims to test the main hypothesis – whether the observed distribution of artefact types reported by Vanhaeren and d’Errico (2006) can be explained by a model of cultural identity-by-descent with modification and isolation-by-distance, or whether it is necessary to invoke cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving interaction processes.

In this analysis, the relative marginal likelihood of each of the 12 models described in section 4.1.6 is estimated, conditioned, as before, on the extent of similarity between a range of spatial statistics describing the observed data (Vanhaeren and d’Errico 2006) and data simulated under each proposed model. As in section 3.1.2, please note that, using this method, it is only possible to assess the relative performance of any scenario considered compared to the other scenarios considered, and it is not possible to rate how good a scenario is absolutely; as noted previously, some methods have been suggested for testing whether an observed dataset is a typical outcome of a proposed model (Voight et al. 2005), however such methods have not been implemented in this thesis. It is also important to note here that it would be inappropriate to compare the models described in section 2.1 to those described in section 4.1. Although the two model versions generate expectations of the same data (i.e. the spatial distribution of symbolic artefact types associated with the Aurignacian culture in Upper Palaeolithic Europe), the processes modelled are rather different in their details and so the comparison would not be between like-for-like models; the defined set of summary statistics may therefore be differently efficient in capturing information about the parameters of the different model versions, which can potentially lead to misleading results.

All 12 models described in section 4.1.6 are compared; the data considered is that generated by the Null Model with Cultural Interactions (NMCI) and the Culture-Dependent Interactions Model (CD) under each pairwise combination of one of three environmental scenarios (either uniform carrying capacities (FLAT), or NPP calculated from Banks et al. (2008) climate data (NPPBanks), or NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)) and one of two models of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)). For each model version (i.e. each of the 12 models compared), 1,000,000 simulations are considered. For the relative marginal likelihood estimation, results for the 12 models are considered together; a total of 12,000,000 simulations are therefore taken into account.

Figure 4.12 is a plot of the estimated relative marginal likelihood of each version of the NMCI model (dashed lines) and the CD model (solid lines) at different thresholds. It shows what
proportion of the best simulations (y-axis) – those generating data most similar to the observed data – are coming from each model for each percentage of closest simulations (x-axis). The colours again refer to the combination of the environmental scenario and the model of cultural variation considered. As in the previous chapter, at each percentage of closest simulations (x-axis), each of the ten grey points represents a pick made from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison. At each threshold, the distribution of these points is a measure of stochasticity in the data at that threshold, and provides a proxy for confidence in the value of the estimated relative marginal likelihoods, with the relative marginal likelihoods only considered as different when the distributions of grey points from different models do not overlap.

As explained before, since the plot depicts proportions, for any particular percentage of closest simulations (i.e. for any particular value on the x-axis), the sum of the proportions of the closest Euclidean distances coming from each model (i.e. the sum of the values on the y-axis) will always be 1. The relative marginal likelihood estimates of each model calculated on the smallest 60,000 Euclidean distances (i.e. closest 0.5% of the considered simulations, which corresponds to x=0.5 in Figure 4.12) are detailed in Table 4.2.

Figure 4.12 and Table 4.2 indicate that, for all scenarios modelled – scenario here referring to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT), or NPP calculated from Banks et al. (2008) climate data (NPPBanks), or NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)) – there is very little difference in how well the NMCI models (dashed lines) and CD models (CD) perform. As noted in section 3.1.2.2, this indicates that, when compared to the NMCI models (i.e. simple scenario of cultural identity-by-descent with modification and isolation-by-distance), the added complexity of the CD models (i.e. invoking cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophically driving interaction processes) does not improve the fit of simulated to observed data. In each scenario, the NMCI model performs equally well, or better, compared to the CD model.

For both the NMCI and CD models, the best fits of simulated to observed data are again generated under the scenario that combines the bounded stepwise mutation model (SW) and the environmental scenario in which Europe is assumed to have a uniform distribution of carrying capacities (FLAT) (blue lines). However, differently to the results reported in section 3.1.2.2, in the case of the models in this chapter, there is not as big a difference in how well each of the scenarios perform.

Since the best fits of simulated to observed data are generated by this combination of environmental scenario and cultural mutation model, this indicates that both the assumed mutation model and the assumed environmental scenario influence the fit of simulated to observed data.
observed data, as observed in section 3.1.2.2. Differently to the results in section 3.1.2.2, however, in this case the assumed mutation model appears to makes a bigger difference than the assumed environmental scenario – for all three environmental scenarios, data simulated under the stepwise mutation model (blue, green and yellow lines) generate a better fit to observed data than those simulated under the non-stepwise (discrete) model (red, orange and purple lines). Again, this observation may indicate the existence of scaling of item similarity in material culture repertoires. As explained in section 2.1.3.3.1, scaling of item similarity implies that a group is more likely to add an item to its cultural repertoire that is morphologically similar to one already present in its repertoire than one that is very different. The stepwise mutation model relaxes the assumption that the cultural trait data considered in this study is necessarily categorical since categorical distinctions are often arbitrarily made and cultural change may in fact be scalable.

Within each of the two cultural mutation models considered, simulations in which the demography is conditioned on a uniform distribution of carrying capacities generate a better fit to the observed data, as already noted; there is no improvement in the fit of simulated to observed data when conditioning simulations on the NPP distributions calculated from climate data (Banks et al. (2008) and Singarayer and Burrough (2015)). The fit of simulated to observed data is similar for the models in which the demography is conditioned on the two climate datasets (i.e. simulations conditioned on each of the two climate datasets perform almost equally well); this is to be expected since these climate datasets are derived from similar climate models, as described in section 4.1.1.1.

The fact that simulations conditioned on these more realistic scenarios perform less well than those conditioned on a uniform distribution is surprising – the geographic region during the time period of interest in the current study is characterised by a number of rapid climatic changes that would be expected to have had an effect on potential population densities, and in turn on observed patterns in material culture. One possible reason for simulations conditioned on a uniform distribution of carrying capacities performing better than those conditioned on NPP distributions calculated from climate data could be that Aurignacian hunter-gatherer populations may have been resource-stressed, irrespective of differential carrying capacity values. In this case, NPP distributions calculated from climate data may not be a representative proxy for the true population density distributions, with the less realistic uniform distribution of carrying capacities providing a closer representation of the same, which would account for the observed difference in performance between simulations conditioned on these carrying capacity distributions.

Another reason for the observed difference in performance could be the difference in number of material culture deposits in sites. In simulations conditioned on NPP distributions, groups gravitate to areas of high NPP. Since these areas also correspond to where the majority of sites in the observed data are located – as is to be expected since hunter-gatherer populations are more likely to settle in resource rich areas – this results in more frequent material culture
deposits in these sites. This dynamic results in fewer deposits, or indeed none, in sites located in areas with low NPP. This is intuitively correct since resource rich areas would be expected to have higher population densities and therefore richer material culture distributions compared to resource poor areas. In simulations conditioned on a uniform distribution of carrying capacities, visits to sites are equally likely irrespective of their location within the geographic domain modelled, which would result in more even distributions of material culture across the sites. Thus, since the distribution of sites in the conditioning data (i.e. observed data published by Vanhaeren and d’Errico (2006)) does not reflect well the carrying capacities inferred from net primary productivity data, simulations conditioned on a uniform distribution of carrying capacities generate a better fit to observed data.

A further possible reason for the difference in performance could be a weak relationship between NPP distributions calculated from climate data and the potential carrying capacity distributions. As mentioned in section 4.1.1, this study follows the precedent set by Eriksson et al. (2012) in using climatic variables to estimate annual NPP distributions across the region of interest, which are in turn a proxy for potential population densities. While Eriksson et al. (2012) showed that there is a broad positive relationship between NPP and population density values that is qualitatively similar to that reported in ethnographic studies of modern hunter gatherers by Binford (2001), this relationship is an incredibly noisy one (Figure 1 in Eriksson et al. 2012) and therefore NPP distributions calculated from climate data may not be entirely representative of the true population density distributions.
Figure 4.12: Relative marginal likelihood estimate (y-axis) of each version of Null Model with Cultural Interactions (NMCI; dashed lines) and Culture-Dependent Interactions Model (CD; solid lines) for each percentage of closest simulations (x-axis), taking into consideration a total of 12,000,000 simulations (i.e. 1,000,000 simulations for each of the 12 models compared). At each percentage of closest simulations (x-axis), each of the ten grey points represents a pick made from a Dirichlet distribution whose parameters are the number of simulations coming from each of the models considered in the comparison; this is a measure of stochasticity in the data and provides a proxy for confidence in the value of the estimated relative marginal likelihoods.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT), or NPP calculated from Banks et al. (2008) climate data (NPPBanks), or NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)).
Table 4.2: Relative marginal likelihood estimates (RML) of each version of Null Model with Cultural Interactions (NMCI) and Culture-Dependent Interactions Model (CD). Proportions are calculated on the smallest 60,000 Euclidean distances (i.e. closest 0.5%, which corresponds to x=0.5 in Figure 4.12) of a total of 12,000,000 simulations (i.e. 1,000,000 simulations for each of the 12 models compared). The shaded cells represent the best six models.

Version here refers to a pairwise combination of an environmental scenario (either uniform carrying capacities (FLAT), or NPP calculated from Banks et al. (2008) climate data (NPPBanks), or NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer)) and a model of cultural variation (either bounded stepwise (SW) or bounded non-stepwise (DIS)).

<table>
<thead>
<tr>
<th>Model</th>
<th>Cultural Mutation Model</th>
<th>Environmental Scenario</th>
<th>RML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Model with Cultural Interactions</td>
<td>SW</td>
<td>FLAT</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPBanks</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPSingarayer</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>DIS</td>
<td>FLAT</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPBanks</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPSingarayer</td>
<td>0.051</td>
</tr>
<tr>
<td>Culture-Dependent Interactions Model</td>
<td>SW</td>
<td>FLAT</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPBanks</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPSingarayer</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>DIS</td>
<td>FLAT</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPBanks</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NPPSingarayer</td>
<td>0.051</td>
</tr>
</tbody>
</table>
4.2.3 Parameter Estimation

To obtain the posterior estimates of parameters of interest, 1,132,411 simulations generated by the Null Model with Cultural Interactions (NMCI) under the bounded stepwise (SW) model of cultural variation and each of two environmental scenarios – those where the demography is conditioned on a uniform distribution of carrying capacities (FLAT) and on the distribution of NPP calculated from Singarayer and Burrough (2015) climate data (NPPSingarayer) – were used. These models were chosen because the NMCI and Culture-Dependent Interactions (CD) models perform similarly well under each combination of environmental scenario and model of cultural variation, and therefore it is more interesting to examine how scenario choice affects the posterior parameter distributions. As explained in section 3.1.3, since there is no improvement in fit of simulated to observed data with the added complexity of the CD model, there is no justification for adding this complexity, and the NMCI model is considered over the CD model.

Similarly to the method employed in section 3.1.3, in order to choose the number of closest simulations (i.e. the threshold level) to consider in the parameter estimation, the Euclidean distances corresponding to the closest 5% of 1,132,411 simulations (i.e. 56,621 simulations) for each of the two models were plotted in ascending order. As explained in section 1.3.4.3, in some cases (e.g. Kamberov et al. 2013), a noticeable increase in Euclidean distance values is observed after a plateau in the same, and simulations for which the Euclidean distance values are smaller than that at the plateau (i.e. those simulations for which the Euclidean distance values fall before the increase) can be regarded as the retained simulations and used in the posterior parameter estimation. This again was not the case here, which can be seen from the smooth increase in Euclidean distance values for both models in Figure 4.13, and so a threshold value of ~0.5% (best 6,000 of 1,132,411 simulations considered) is used for purposes of parameter estimation for each of the two models. Since there is not a clear choice for the value of the acceptance threshold, the parameter estimation analysis has also been performed with ~1% (best 12,000 of 1,132,411 simulations considered) and ~0.1% (best 1,200 of 1,132,411 simulations considered), and these results are shown in Appendix A.4.

Figure 4.14 and Figure 4.15 show the estimated posterior density distributions of the model parameters for the NMCI SW FLAT and NMCI SW NPPSingarayer models, respectively, calculated on the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for each model. The boundaries of the equal-tailed 95% (i.e. the upper and lower 2.5%) credible intervals (CIs) of each distribution are indicated by shading. The posterior parameter distributions for each model are summarised in Table 4.3, by giving the mode, 2.5% and 97.5% quantiles of each distribution. Figure 4.16 shows the performance of the two models in explaining the observed data. It is a plot of the distribution of each statistic generated by the 6,000 retained simulations for each model, and shows that simulations coming from NMCI SW FLAT (black lines) are marginally closer to their target values (i.e. summary statistic values calculated on the observed data, indicated by red vertical lines) than those coming from NMCI SW NPPSingarayer (grey lines); this is to be expected since the NMCI SW FLAT model was
shown in the model comparison analysis in section 3.1.2.2 to generate a better fit (but only slightly better) to observed data.

Figure 4.14, Figure 4.15 and Table 4.3 show that all posterior parameter estimates are on the same order of magnitude for the two models, and that they have similar mode and 95% CIs values. This indicates that the differences in scenario choice between these two models has little effect on the posterior parameter distributions. The following observations and interpretations are therefore made by considering the (marginally better-performing) NMCI SW FLAT model, though similar observations and interpretations can be drawn for the NMCI SW NPPSingarayer model.

For some parameters (particularly (B) probability of fission / extinction, and (F) interaction radius), the 95% CIs of the posterior estimates are similar to their prior ranges with no obvious peak at a particular value, which implies that little information could be obtained about these parameters using the models and observed data.

The posterior parameter analyses appear informative for some key parameters. Similarly to the results reported in section 3.1.3, the 95% CI of the posterior estimate of the probability of cultural mutation (panel (A)) is considerably narrower (0.0003-0.0402, mode=0.0020) than its prior range (0-0.2).

For some parameters (particularly (C) migration distance, (D) total number of items in each group's cultural repertoire, and (E) maximum number of groups), although the 95% CIs of the posterior estimates are similar to their prior ranges, the fact that the estimated posterior density distributions peak suggests that the observed data may contain some information about these parameters.

Firstly, the posterior estimate of the total number of items in each group's cultural repertoire parameter (panel (D)) is similar to but falls in the lower end (510-1436, mode=582) of its prior range (500-1500). As observed in section 3.1.3, this estimate may be of interest as it may indicate that groups' cultural repertoires did not contain a large number of items (ornaments). It should be noted that this may not be a truly informative estimate since the diversity of ornaments in the model is constrained to that in the observed dataset as explained in section 2.1.3.2.

Similarly, the posterior estimate of the migration distance parameter (panel (C)) is similar to but falls in the lower end (9-86, mode=16.2427) of its prior range (0-100). This estimate may be an interesting result as it suggests that the distance travelled by groups at each generation was relatively low.

Lastly, the posterior estimate of the maximum number of groups parameter (panel (E)) shows a similar pattern – the posterior estimate is similar but falls in the lower end (61-965, mode=124)
of its prior range (50-1000) – however the peak observed in this case is not as strong as for the previous two parameters. This may be an interesting result as it suggests that the observed patterns in the distribution of personal ornaments may be the result of a relatively small number of groups, however, this needs further investigation since this peak is not observed when parameter estimation is performed with a smaller acceptance threshold, as is done in Appendix A.4.2.

As has previously been observed for the results in section 3.1.3, the extinct density distributions (i.e. those simulations in which all groups became extinct prior to the end of the simulation), shown by the dashed grey lines in the plots, of the demographic parameters (i.e. (B) probability of fission / extinction, (C) migration distance, and (E) maximum number of groups) indicate that a very high rate of fission / extinction, as well as a low migration distance and low maximum number of groups in the domain, readily result in a truncation of the simulation (i.e. all groups become extinct prior to the end of the simulation), as is to be expected. It is again evident that these demographic parameters are the only parameters that have an effect on viability of simulations (i.e. whether all groups become extinct prior to the end of the simulation). As explained before, this is reasonable since the cultural parameters (panels (A) probability of cultural mutation, (D) total number of items in each group’s cultural repertoire, and (F) interaction radius) would not be expected to have an effect on the demography of the simulation space, and this is reflected in the extinct density distributions of these cultural parameters mirroring their respective prior distributions. It is also interesting to note that, for the migration distance (panel (C)) and maximum number of groups (panel (E)) parameters, the peaks in the posterior distributions fall in the same part of the range as the peaks of the extinct density distributions.
Figure 4.13: Ranked Euclidean distance values of the best 56,621 simulations (i.e. the closest 5% of 1,132,411 simulations) for each of the two models plotted in black, with the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) used for estimating the posterior parameter distributions of parameters highlighted in red. The two panels correspond to the two models of interest: NMCI SW FLAT (top panel) and NMCI SW NPPSingarayer (bottom panel).
Estimated posterior density distributions of the demographic and evolutionary parameters of interest for the two NMCI models, calculated on the 6,000 retained simulations (i.e. closest ∼0.5% of 1,132,411 simulations) for each model. The boundaries of the equal-tailed 95% credible intervals (i.e. the upper and lower 2.5%) of each distribution are indicated by shading; these are also summarised in Table 3.3. The solid and dashed grey lines represent the prior and extinct (i.e. those simulations in which all groups became extinct prior to the end of the simulation) density distributions of parameters, respectively.

Figure 4.14: NMCI SW FLAT model.

Figure 4.15: NMCI SW NPPSingarayer model.
Table 4.3: Prior ranges and posterior estimates of parameters for the two versions of interest of the Null Model with Cultural Interactions. For each model, the posterior parameter ranges are calculated on the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) and expressed by giving the mode, 2.5% and 97.5% quantiles, expressed to 4 decimal places. The letter in the far left column corresponds to the panels in Figure 4.14 (NMCI SW FLAT) and Figure 4.15 (NMCI SW NPPSingarayer).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Range</th>
<th>SW FLAT</th>
<th>SW NPPSingarayer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>minimum</td>
<td>maximum</td>
<td>mode</td>
</tr>
<tr>
<td>(A) $p_{mut}$: probability of cultural mutation</td>
<td>0</td>
<td>0.2</td>
<td>0.0020</td>
</tr>
<tr>
<td>(B) $p_{f/e}$: probability of fission / extinction</td>
<td>0</td>
<td>1</td>
<td>0.0724</td>
</tr>
<tr>
<td>(C) $d_{mig}$: migration distance (km)</td>
<td>1</td>
<td>100</td>
<td>16.2427</td>
</tr>
<tr>
<td>(D) $N_{items}$: total number of items in each group's cultural repertoire</td>
<td>500</td>
<td>1500</td>
<td>582</td>
</tr>
<tr>
<td>(E) $N_{groups}$: maximum number of groups</td>
<td>50</td>
<td>1000</td>
<td>124</td>
</tr>
<tr>
<td>(F) $d_{int}$: interaction radius (km)</td>
<td>0</td>
<td>50</td>
<td>6.3601</td>
</tr>
</tbody>
</table>
Figure 4.16: Distributions of the 12 summary statistic values in the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW FLAT model (black lines) and NMCI SW NPPSingarayer model (grey lines). The title of each panel corresponds to the summary statistic as discussed in section 2.2.1. The red vertical line indicates the target value of each summary statistic (i.e. the value of that statistic calculated from the observed data).
4.2.4 Exploring Behaviour of Parameters

As in the analysis of results reported in section 3.1.4, in this analysis, the dependence between the model parameters is examined to assess whether parameters that govern the various aspects of the demographic and cultural processes simulated are correlated in the accepted simulations. As a reminder, a possible interpretation of such correlations is that changes in one parameter require changes in another to achieve a good fit to the observed data.

A correlation analysis of parameter values for both the NMCI SW FLAT (Figure 4.17) and NMCI SW NPPSingarayer (Figure 4.18) models was performed, using the retained 6,000 simulations (i.e. closest ~0.5% of 1,132,411 simulations) for each model. Spearman's rank correlation coefficient $R^2$ was used to assess the statistical dependence between the parameters, and for each of the two models and each pairwise combination of model parameters, the binned scatterplot (below diagonal), correlation coefficient $R^2$ (above diagonal; black) and associated p-value (above diagonal; blue) are given. As in previous correlation analyses, the reported correlation coefficient $R^2$ values are considered to be the most informative indicator of correlation since the p-values are to an extent determined by the number of simulations considered, which is effectively arbitrary. As noted in section 3.1.4, the marginal views of the scatterplots shown in Figure 4.17 and Figure 4.18 correspond to the density plots in Figure 4.14 and Figure 4.15, respectively, as expected.

Weak ($R^2 \leq 0.25$) correlations are again observed between all pairwise combinations of the model parameters, which suggests that the parameters considered affect the various aspects of the demographic and cultural processes simulated relatively independently.
Binned scatterplots (below diagonal) and correlation values (above diagonal; black: Spearman’s rank correlation coefficient $R^2$, blue: associated p-value) of pairwise parameter combinations using the 6,000 retained simulations (i.e. closest ~0.5% of 1,132,411 simulations) for the NMCI SW FLAT (top panel, i.e. Figure 4.17) and NMCI SW NPPSingarayer (bottom panel, i.e. Figure 4.18) models. The cells in each pairwise scatterplot (below diagonal) are shaded according to the number of data points they contain, with darker cells containing a larger number of data points than lighter cells. Pairwise panels giving the correlation values (above diagonal) are shaded according to $R^2$ value and the text greyed out for those that are not significant (p-value $\geq 0.05$).

Figure 4.17: NMCI SW FLAT model.

Figure 4.18: NMCI SW NPPSingarayer model.
4.3 Discussion

The refinements described in this chapter have integrated multiple sources of data into the modelling framework described in section 2.1 in order to more realistically represent the conditions that characterise the geographic area and time period of interest in this study. Environmental conditions have been shown to affect human migrations and demography and are consequently important to consider when modelling human population migrations and interactions. In this chapter, Palaeoclimate data have been integrated using a biologically relevant measure, namely Net Primary Productivity (NPP), in order to condition the demography of the simulation space (section 4.1.1). Elevation data have been integrated in an anthropologically relevant way, by using Topographic Ruggedness Index (TRI) to condition the extent of group migrations (section 4.1.3).

Significant methodological developments of the process of estimating population density, namely using the distance between groups as a proxy for population density, have been implemented (section 4.1.2), in addition to modifications of the process of depositing cultures at archaeological sites in order to account for the cumulative aspect of the archaeological record (section 4.1.4).

By making these refinements and integrating multiple sources of data, the models, and thereby confidence in the conclusions drawn, have been improved. Implementing these modifications has also served to improve the computational efficiency of the simulations through simultaneous optimisation of the code written to model the processes. While there is a multitude of models of cultural evolution in the literature (see Lake (2014) and Mesoudi (2015) for recent reviews), they often do not include spatial and temporal environmental heterogeneity based on existing data (such as climate and elevation data). In this respect, the models presented here provide an improvement to previous existing work.

Similarly to the results presented in Chapter 3, the results of the model comparison analyses reported here (section 4.2.2) do not support the hypothesis that Aurignacian populations in Early Upper Palaeolithic Europe were ethnically structured in a manner related to ornamental material culture homophilically driving cultural interaction processes. The spatially explicit simulation models and ABC analyses, conditioned on the data presented by Vanhaeren and d’Errico (2006), again indicate that there is little difference between the simple scenario of cultural identity-by-descent with modification and isolation-by-distance, and the more complex one that, in addition, invokes cultural group interaction processes that would be expected if differences in material culture were symbolically marking differences between ethnic groups and homophilically driving cultural interactions.

Additionally, the results of the model comparison analyses indicate that data simulated under the bounded stepwise mutation model provide a better fit to observed data than that simulated under the bounded non-stepwise (discrete) mutation model. These analyses also indicate a
better fit between data simulated under a uniform distribution of carrying capacities than under distributions in which carrying capacities are conditioned on Palaeoclimate data. These observations and their implications are further discussed in Chapter 6.

The correlation analyses, the results of which are reported in section 4.2.1, go some way to critically evaluating the choices made in adopting the Approximate Bayesian Computation (ABC) approach, and are further discussed in Chapter 6. It is important to note here that, while the low correlations observed between pairwise summary statistic and model parameter combinations can be interpreted as a combination of model parameters driving each summary statistic, or, conversely, a combination of summary statistics informing on each of the model parameters, these correlations may also indicate that the summary statistics used are not sufficient to capture the relevant features of the observed data. A set of summary statistics is considered sufficient when no other statistic calculated on the observed data provides any additional information on the value of a parameter; the posterior probability of a parameter calculated using a sufficient set of summary statistics should be the same as its posterior distribution calculated using the complete observed dataset (Marjoram and Tavare 2006). Since the use of sufficient summary statistics is pivotal in the ABC approach adopted here, it may be necessary to redefine the set of summary statistics considered in future work.
Chapter 5: Other Projects

Summary

This chapter describes two additional, collaborative projects that I have made a substantial contribution to. My involvement in each of these studies was in part a by-product of the main research that this thesis is concerned with.

The first study examines a hypothesis that cytochrome P450 3A5 (CYP3A5), an enzyme involved in the metabolism of many therapeutic drugs, is important in salt retention adaptation. My contribution to this work focussed on integrating climatic information into the existing analysis framework.

The second study is concerned with examining the rate of gene flow over time and space through analyses of craniometric data, with the aim of informing on the extent of past population migration activity. My contribution to this project involved the development of a simple spatiotemporally explicit model, based on the simulation modelling approaches in Chapter 2 and Chapter 4, to simulate variation in craniometric measures under known migratory activity rates, in order to test an estimator of this parameter.

Unless otherwise specified, the simulation modelling framework and data processing were implemented in the Python programming language (http://www.python.org/).
5.1 Correlating Cytochrome P450 3A5 Expression with Climate Variables

This study, designed and carried out by Ripudaman K. Bains (BioMed Central) as part of her PhD project, used a multidisciplinary approach to characterize cytochrome P450 3A5 (CYP3A5) variation in geographically and ethnically diverse populations from in and around Africa, and to infer the evolutionary processes that have shaped patterns of diversity in this gene. My contribution to this work, detailed in section 5.1.2, focussed on obtaining and integrating the relevant past and present ecological data, namely climatic information, into the existing analysis framework, in order to allow examination of the hypothesis that CYP3A5 is important in salt retention adaptation.

The study has been published (article attached in Appendix C) as:


5.1.1 Summary of Study

Cytochrome P450 3A5 (CYP3A5) is an enzyme involved in the metabolism of many therapeutic drugs. CYP3A5 gene expression levels vary between individuals and populations, and this contributes to adverse clinical outcomes. Variable expression is largely attributed to four alleles in the CYP3A5 gene, namely CYP3A5*1 (expresser allele); CYP3A5*3 (rs776746), CYP3A5*6 (rs10264272) and CYP3A5*7 (rs41303343) (low/non-expresser alleles).

Studies of genes encoding drug metabolizing enzymes, such as those in the Cytochrome P450 (CYP450) super-family to which CYP3A enzymes belong, have identified variation which affects the safety and efficacy of therapeutic drugs. However little is known about intra-African variation at these loci. Africa is heavily burdened with common and infectious diseases (Coleman 1998), which are treated with multiple drugs. Studies of intra-African variation at genes encoding drug metabolizing enzymes are likely to be beneficial to clinicians, geneticists and researchers within the emerging field of evolutionary medicine (Nesse et al. 2010; Stearns 2012). They are also likely to have great potential for minimizing the risk of adverse clinical outcomes in patients with recent African ancestry (Zhou et al. 2009).

There is considerable inter-ethnic variability in CYP3A5 expression levels (Lee et al. 2003). A previous study reported a strong positive correlation between CYP3A5*3 allele frequencies and latitude (Thompson et al. 2004). Latitude is a correlate of multiple ecological variables that are associated with functional markers of genes involved in heat adaptation, such as angiotensinogen (AGT), G protein β3 subunit (GNB3) and epithelial sodium channel γ (ENaCγ) (Young et al. 2005). In this study, tests were performed for correlations between frequencies of low/non-expresser CYP3A5 alleles, and inferred expresser phenotypes, with latitude and ecological variables, namely temperature and precipitation. Additionally, tests were performed
for correlations with aridity indices calculated from temperature and precipitation data using the de Martonne aridity index (de Martonne 1920, 1941; Croitoru et al. 2013), which allowed the combined effect of temperature and precipitation on CYP3A5 phenotypes to be considered. Correlations were performed using Mantel and Partial Mantel tests, and estimated using ecological data for the present day, and inferred for 10,000 years ago (Holocene) and 50,000 years ago (Late Pleistocene).

5.1.2 My Contribution: Integrating Climate Data

5.1.2.1 Climate Data: Smith and Gregory (2012)

The climate data used in this study originate from the ALL-5G dataset Smith and Gregory (2012) generated by the Fast Met Office and UK Universities Simulator (FAMOUS) (Jones et al. 2005; Smith et al. 2008; Smith and Gregory 2012), which is associated with the Quaternary QUEST (Quantifying and Understanding the Earth System) Project (http://researchpages.net/qq/). The data are archived at the British Atmospheric Data Centre (BADC; http://badc.nerc.ac.uk/) for use by the climate research community, and available by request to the BADC.

FAMOUS is a low-resolution atmosphere-ocean general circulation model (AOGCM) that is based on the widely used high resolution Hadley Centre coupled model (HadCM3) (Gordon et al. 2000; Pope et al. 2000). As mentioned in section 4.1.1.1, HadCM3 is a general circulation model consisting of linked interactive atmospheric, interactive ocean and static sea-ice models (Gordon et al. 2000; Pope et al. 2000) that has been shown to have the capability of capturing historical climate change (Stott et al. 2000).

The ALL-5G dataset (Smith and Gregory 2012) has been compiled from transient climate runs of the FAMOUS model and describes global climatic conditions of the atmosphere and ocean through the whole last glacial cycle, covering the last 120,000 years. For each year, monthly global temperature and precipitation data are available at a resolution of 5 degrees latitude and 7.5 degrees longitude.

5.1.2.2 Quantifying Aridity

In addition to temperature and precipitation data, the combined effect of these two climate variables on CYP3A5 phenotypes was of interest in this study. To investigate this effect, a measure referred to as an aridity index, calculated from temperature and precipitation data, was used. This index provides a convenient method of quantifying the combined effects of major climatic variables.
Aridity is generally defined as the amount of ground or air moisture at a particular location. It is quantified using so-called aridity indices, calculated as the ratio of precipitation to evaporation, expressed in the same units, at the geographic location of interest. For the purposes of this study, aridity is quantified according to the de Martonne Aridity Index, \( I_{DM} \), defined as:

\[
I_{DM} = \frac{P}{T_a + 10},
\]

(5.1)

where \( P \) is the total annual precipitation in millimetres and \( T_a \) is the mean annual air temperature in degrees Celsius (de Martonne 1920, 1941; Croitoru et al. 2013). In the case of the de Martonne Aridity Index, evaporation is approximated as 10 degrees higher than the mean annual temperature in degrees Celsius, and is expressed in millimetres. An obvious disadvantage of this aridity index is the pole at -10°C, which gives rise to negative \( I_{DM} \) values at temperatures lower than 10°C with \( I_{DM} \) undefined at 10°C. For this reason, the de Martonne Aridity Index is subjected to the following additional condition (as detailed by Grieser et al. in an un-reviewed publication of the Food and Agricultural Organization of the United Nations: http://www.fao.org/nr/climpag/globgrids/kc_commondata_en.asp):

\[
I_{DM} = \begin{cases} 
100 & \text{if } T_a < 9.9^\circ\text{C} \\
\frac{P}{T_a + 10} & \text{otherwise}
\end{cases}
\]

(5.2)

Low values of \( I_{DM} \) indicate a low level of precipitation compared to evaporation and an arid climate, while high \( I_{DM} \) values indicate a high level of precipitation compared to evaporation and therefore a humid climate.

### 5.1.2.3 Aridity Indices for Geographic Locations of Interest

To obtain the values of each of the climatic variables (mean annual temperature in degrees Celsius and total annual precipitation in millimetres) at the geographic locations of interest, it was first necessary to interpolate the available data for each time slice of interest (i.e. present day, 10,000 years ago (Holocene) and 50,000 years ago (Late Pleistocene)) to the required resolution, and then extract the values at the required geographic locations (i.e. those locations where DNA was sampled).

In addition, prior to performing the interpolations, it was necessary to obtain aridity estimates at these locations by calculating the de Martonne Aridity Index \( I_{DM} \), using the definition described in section 5.1.2.2, at the available spatial resolution for each time of interest. The aridity index was calculated prior to data interpolation to ensure more accurate values – if components of the climate data (i.e. temperature and precipitation) are interpolated individually and \( I_{DM} \) calculated from interpolated values of these variables, any uncertainties
that arise during the interpolation would be amplified when combining these variables to calculate $I_{DM}$.

The global precipitation, temperature and aridity index distributions at 0 (present day), 10,000 years ago (Holocene) and 50,000 years ago (Late Pleistocene) were then interpolated to a resolution of 1 degree latitude and 1 degree longitude, and the value of each extracted at the required geographic co-ordinates for use in the correlation analyses. The interpolations were performed using the `smooth.2d()` function from R’s `fields` package, with the default Gaussian kernel and a bandwidth of 0.9.

### 5.1.3 Summary of Relevant Results

Temperature and precipitation data associated with the Quaternary QUEST Project ([http://researchpages.net/QQ/](http://researchpages.net/QQ/)) suggest that changes in precipitation over the past 50,000 years are greater than those in temperature; however, no significant correlation was found between precipitation values and $CYP3A5$ allele frequencies. Negative correlations were observed between inferred expression phenotypes (assuming $CYP3A5^*$6 is a low/non-expressor allele) and aridity values for the Holocene (10,000 years ago) [Spearman’s Rho = $-0.465$, p-value = 0.004] and Late Pleistocene (50,000 years ago) [Spearman’s Rho = $-0.379$, p-value = 0.02]. Under the de Martonne aridity index, this means that high frequencies of high-activity alleles are positively correlated with arid and semi-arid environments (Croitoru et al. 2013). This finding is consistent with the hypothesis that high-activity $CYP3A5$ alleles may be adaptive in regions where there are frequent water shortages, by aiding the rapid retention of water (Thompson et al. 2004). However, stronger correlations were found with temperature alone. Although further work will be needed to confirm these ecological correlations, the strong correlation with temperature is consistent with what is expected for functional variation of genes involved in heat adaptation (Young et al. 2005).

Full details of the study, including the results, can be found in the publication (attached in Appendix C).
5.2 Variation in Human Cranial Morphology

This study is part of an on-going project in collaboration with Liisa Loog (School of Archaeology, University of Oxford), and a manuscript is currently being prepared for publication. The study focuses on development of a novel approach for estimating past rates of migratory activity from variation data that follows identity-by-descent with modification and isolation-by-distance models (e.g. genetic, morphological, cultural data). My contribution to this work, detailed in section 5.2.2, involved the development of a simple spatiotemporally explicit simulation model to generate simulated data that could be used to test the robustness of the developed estimator prior to its application to observed data.

5.2.1 Summary of Study

5.2.1.1 Context

In recent years, methods developed in the fields of quantitative and population genetics have been widely used by anthropologists to study the evolution of, and observed variation in, human cranial morphology (e.g. Relethford 1994, 2002, 2004; Weaver et al. 2007; Betti et al. 2009; von Cramon-Taubadel 2009a, b; Betti et al. 2010). Using such methods allows researchers to postulate testable hypotheses and develop robust frameworks to test these hypotheses in the study of phenotypic variation.

Many of these studies have shown that patterns of variation in human cranial form are similar to those in human genetic variation and largely follow a model of neutral evolution (Relethford 1994, 2002, 2004; Gonzalez-Jose et al. 2004; Roseman and Weaver 2004; Hubbe et al. 2009). This implies that the slight variations observed in human cranial form within and between populations can be explained by underlying, presumably neutral, genetic variation. If this is the case, patterns of human cranial variation can be used as a proxy for neutral genetic data and consequently to infer and study past population histories (von Cramon-Taubadel and Weaver 2009). This is useful in cases where genetic data are unavailable since cranial morphology data are more readily available and can be accessed over greater time ranges than ancient DNA technologies generally permit.

It has also been shown that, similarly to human genetic variation, isolation-by-distance (the idea from population genetics that states that populations that are in geographically proximate locations share more genetic markers than those that are further apart (Wright 1943; Slatkin 1993)) plays an important role in explaining observed differences in cranial form between modern human populations (Relethford 2004; Betti et al. 2010). This implies that geographically mediated gene flow can, to a certain extent, explain these observed differences, since geographically mediated gene flow can be thought of as isolation-by-distance. Similarly, it can be assumed that temporal distance between individuals can, to a certain extent, explain these observed differences, since it is expected that the further apart individuals are in time the less morphologically similar they are.
5.2.1.2 Hypothesis

Since population migrations can affect patterns of genetic, and therefore craniometric, variation through the action of gene flow (Relethford 2004), it may be possible to obtain some information about past migrations through analyses of variation in craniometric data over space and through time. The aim of this study was to develop an estimator of the rate of gene flow through time that is particularly suited to sparsely sampled data.

The idea is that information about the rate of gene flow can be retrieved from craniometric measurements by considering the extent to which spatial and temporal factors can explain observed variation in human cranial form. If this is the case, it should be possible to retrieve this information by correlating a joint time-space Euclidean distance matrix and a morphological distance matrix between samples. The key point here is that combining a spatial distance matrix and a temporal distance matrix to produce a joint time-space Euclidean distance matrix requires a scaling factor – a parameter relating spatial distances (in km) to temporal distances (in years). The aim is to find a value for this parameter that maximises the correlation between the joint time-space Euclidean distance matrix and the morphological distance matrix. This value is then treated as an estimator of the mean migratory activity in the geographic region and time period considered.

If these assumptions hold for analyses of craniometric data, there is potential in using this method to retrieve information about migratory activity from any type of trait data that follows identity-by-descent with modification and isolation-by-distance models, such as genetic or cultural data. As for craniometric data, information about the extent of migratory activity would be contained at the point of maximum correlation of the time-space Euclidean distance matrix and the trait distance matrix between samples.

For the purposes of this study, the time-space Euclidean distance, $D$, between a pair of samples $u$ and $v$ is calculated as follows:

$$D(u, v) = [(G(u, v) \cdot s)^2 + T(u, v)^2]^{\frac{1}{2}},$$

(5.3)

where $G(u, v)$ is the geographic distance between samples $u$ and $v$ in kilometres calculated as the geodesic distance from geographic coordinates (the Haversine Formula (Sinnott 1984) is used to account for curvature of the Earth), $T(u, v)$ is the temporal distance between samples $u$ and $v$ in years and $s$ is a scaling factor. The morphological distance, $M$, between a pair of samples $u$ and $v$ is calculated using the Euclidean distance measure as follows:

$$M(u, v) = \sqrt{\sum_{i=1}^{n}(v_i - u_i)^2},$$

(5.4)

where $u_i$ and $v_i$ are the values of craniometric measurement $i$ for samples $u$ and $v$, respectively, and $n$ is the number of craniometric measurements considered.
Once the time-space Euclidean distance $D$ and the morphological distance $M$ have been calculated for each pair of samples in the dataset considered, they can be combined in the time-space Euclidean distance matrix and morphological distance matrix, respectively, and the two matrices correlated. An example correlation is shown in Figure 5.1.

The value of interest is the value of the scaling factor $s$ in equation (5.3) that generates the maximum correlation between the time-space Euclidean distance matrix and the morphological distance matrix (vertical red line in Figure 5.1). Since it is assumed that the value of the scaling factor $s$ at the point of maximal correlation between the two matrices is the estimator of the underlying gene flow rate, this value consequently provides information about the relative importance of the temporal and spatial factors in explaining the observed morphological variation. In cases where the temporal factor has more effect relative to the spatial factor, lower rates of gene flow between populations, and therefore a lower level of migratory activity, can be inferred, while in cases where the spatial factor has more effect relative to the temporal factor, the opposite is true and points to higher rates of gene flow and levels of migratory activity.

In addition to the value of the scaling factor $s$, the extra correlation (EC) that results from considering the combined time-space Euclidean distance matrix (arrow in Figure 5.1) is calculated. It is defined to be the difference between the maximum correlation of the time-space Euclidean distance matrix with the morphological distance matrix, and the higher of the correlations of the geographic distance matrix and temporal distance matrix with the morphological distance matrix, respectively. This extra correlation is considered to be a proxy for the strength of the signal of migratory activity.
Figure 5.1: Example plot of time-space Euclidean distance matrix and a morphological distance matrix coefficient (y-axis) against the corresponding value of the scaling factor $s$ (x-axis). The values of interest is the value of the scaling factor $s$ that generates the maximum correlation (red vertical line), and the extra correlation (EC) that results from considering the combined time-space Euclidean distance matrix (arrow).
5.2.1.3 Applications to Simulated and Observed Data

The ultimate aim of this study is to apply these methods to observed human craniometric data in order to infer rates of past migratory activity. The observed dataset used consists of 25 linear caliper measurements taken from crania of archaeological remains of 3,666 individuals (10,809 individuals including incomplete cases). The specimens cover all major regions of the world (Europe; Africa; North, Central and South America; Asia; Australia and Oceania) and range temporally from Early Upper Pleistocene to recent times (Lahr 1996). All observed data were collected by Dr Marta Lahr (Department of Archaeology and Anthropology, University of Cambridge).

Prior to applying these methods to the observed craniometric data, it was important to test whether the method is able to reliably retrieve some information from craniometric datasets about the effect of the spatial and temporal factors on human cranial form. For this purpose, a spatiotemporally explicit simulation model that generates simulated craniometric data was developed. Since values of parameters under which simulated data has been generated are known, it is possible to apply the estimator to the simulated data in order to check whether the migratory activity values simulated can be reliably and accurately retrieved using the approach described in section 5.2.1.2.

5.2.2 My Contribution: The Simulation Model

The simple spatiotemporally explicit simulation model described in this section simulates craniometric data under different, but known, parameter ranges, and was developed to test the robustness of the approach described in section 5.2.1.2.

5.2.2.1 Simulation World

5.2.2.1.1 Simulation Space

Since the current model has been developed to test the robustness of the method described section 5.2.1.2, the spatial region currently used does not map onto a geographical region, though it would be trivial to perform the simulations in a geographically-realistic domain. The spatial domain currently modelled represents a flat surface, 4,000 by 4,000 units in size, ranging from -2000 to 2000 units in each of the x and y directions.

5.2.2.1.2 Length of Simulation

Each simulation spans a previously assigned number of generations assuming a 25 year generation time (Tremblay and Vezina 2000; Thomas et al. 2006). The lengths of simulation considered are 2,000, 4,000, 6,000, 8,000 and 10,000 generations (corresponding to 50,000,
100,000, 150,000, 200,000 and 250,000 years, respectively). Each simulation includes a burn-in period of 1,500 generations (corresponding to 37,500 years) at the start of the simulation during which no data is collected; this is included to allow variation to arise through the demographic and evolutionary processes described in the next section.

5.2.2.2 Demographic and Evolutionary Processes

Each simulation is initialised with one entity placed in a randomly chosen location in the modelled domain. At each generation, entities migrate locally and undergo fission / extinction processes. These demographic processes are analogous to an isolation-by-distance model in population genetics (Wright 1943; Slatkin 1993).

5.2.2.2.1 Migratory Processes

At each generation entities are subjected to migratory processes modelled as parameterised Gaussian random walks. For each entity, the resultant location following a migration process is calculated, using Pythagoras’ theorem, from two independent picks from a normal distribution with mean \( \mu_{\text{mig}} \) and standard deviation \( \sigma_{\text{mig}}=d_{\text{mig}} \). Positive and negative values picked from the distribution correspond to movement in opposite directions, namely positive and negative x and y directions, respectively. The mean of the distribution is therefore set to \( \mu_{\text{mig}}=0 \) to ensure that movement in opposite directions is equally likely. Parameter \( d_{\text{mig}} \) corresponds to the standard deviation, or width, of the normal distribution and specifies the range of values that the migration distance is most likely to take in each of the x and y directions. Parameter \( d_{\text{mig}} \) is treated as an unknown and a range of values are explored, constrained to a uniform prior range corresponding to movement of between 0 km and 4 km per year, with the value in each simulation randomly assigned from this prior range.

5.2.2.2.2 Fission / Extinction Processes

In addition to the migratory process undergone at each generation, each entity also undergoes a fission / extinction process. This process is analogous to genetic drift, and decreases the amount of overall diversity in the modelled domain.

At each generation, each entity undergoes a fission process with probability \( p_{\text{f/e}} \). Parameter \( p_{\text{f/e}} \) is treated as an unknown and a range of values are explored, constrained to a uniform prior range of between 0.01 and 0.2, with the value in each simulation randomly assigned from this prior range. Once an entity has undergone a fission process, two entities, the parent and offspring, are present in the next generation. In subsequent generations, the parent and offspring entities migrate and undergo fission processes independently.
Should the number of entities in the current generation exceed the maximum number of entities the modelled domain can sustain, set at 10,000, an extinction process occurs. If this is the case, a number of entities, equal to the difference between the current and maximum number of entities, are selected at random and deleted from the simulation.

### 5.2.2.2.3 Inter-Generation Craniometric Variation

Each entity in the modelled domain is assigned 50 craniometric traits, each of which varies slightly at each generation. This reflects the fact that craniometric traits are passed down from parent to offspring entities with some small changes in craniometric trait values occurring between generations, which results in inter-generation craniometric variation. This process is analogous to identity-by-descent with modification in population genetics and, as in population genetics, has the effect of increasing diversity.

The offspring entity therefore retains the craniometric traits of the parent entity, with some small changes in craniometric trait values. This variation is modelled according to a Gaussian distribution, such that the value of a particular trait in the next generation is picked from a normal distribution with mean equal to the value of that trait in the current generation and standard deviation equal to a parameterised value. The initial mean for each of the 50 craniometric traits is set to 1,000. The standard deviation, labelled $p_{\text{mut}}$, is treated as an unknown parameter and a range of values are explored, constrained to a uniform prior range of between 0.001 and 0.5, with the value in each simulation randomly assigned from this prior range.

### 5.2.2.3 Simulated Datasets

Each simulation, made up of the demographic and evolutionary processes described in section 5.2.2.2, results in a simulated dataset. For a particular simulation, this dataset is created by sampling the entities with a fixed probability, set to 0.0001. This sampling process is performed at each generation after the burn-in period of 1,500 generations (corresponding to 37,500 years). For each sampled entity, data recorded are the time of sampling in years, the entity’s location in the modelled domain, namely x and y values, and the measurement values of the entity’s 50 craniometric traits.

### 5.2.3 Summary of Relevant Results

Figure 5.2 and Figure 5.3 are example outputs of analyses of simulated data using the method proposed in section 5.2.1.2. The plots show correlations between simulated average migration distance (x-axis) and the value inferred by the estimator (y-axis) for different values of the rate
of mutation parameter \( p_{\text{mut}} \) (rows) and different rates of fission / extinction parameter \( p_{f/e} \) (columns) for simulations of 4,000 (Figure 5.2) and 8,000 (Figure 5.3) generations that produced extra correlation (i.e. \( EC > 0 \)). The colours on the scale bar to the right of each panel represent different levels of extra correlation.

Overall, the results of analyses on simulated data (i.e. data simulated by the spatio-temporally explicit simulation model described in section 5.2.2) indicate that the estimator proposed in section 5.2.1.2 strongly correlates with simulated migration rates, and would therefore be a useful tool for investigating past migratory activity using observed data. The results also indicate that the estimator recovers the underlying level of migratory activity more accurately when there is a high degree of extra correlation (i.e. a clear signal), thus indicating that both the estimator value and the extra correlation should be taken into account when analysing data using this method.

This study is on-going and a manuscript is currently being prepared for publication. While it is not for this thesis to discuss the details of the results, it is notable that initial analyses of observed craniometric data using this method indicate that the extent of Late Pleistocene migratory activities in Sub-Saharan Africa are lower than those outside; this makes sense considering that human populations have expanded outside of Africa to cover a large geographic area.
Figure 5.2: Correlation plots between simulated average migration distance (x-axis) and the value inferred by the estimator (y-axis) for different values of the rate of mutation parameter $p_{mut}$ (rows) and different rates of fission / extinction parameter $p_{f/e}$ (columns) for simulations of 4,000 generations that produced extra correlation (EC>0). The colours on the scale bar to the right of each panel represent different levels of extra correlation.
Figure 5.3: Correlation plots between simulated average migration distance (x-axis) and the value inferred by the estimator (y-axis) for different values of the rate of mutation parameter $p_{\text{mut}}$ (rows) and different rates of fission / extinction parameter $p_{f/e}$ (columns) for simulations of 8,000 generations that produced extra correlation ($EC > 0$). The colours on the scale bar to the right of each panel represent different levels of extra correlation.
Chapter 6: General Discussion

Summary

This chapter presents a general discussion of this thesis and the conclusions of the study. The chapter begins with a summary of the thesis and conclusions, followed by a critical evaluation of the approaches adopted, with the aim of highlighting how these approaches can contribute to understanding and interpretations of distributions of material culture.
Summary of Thesis and Conclusions

In this thesis, spatially explicit simulation modelling and Approximate Bayesian Computation (ABC) (Tavare et al. 1997; Beaumont et al. 2002) approaches have been used to explore how human demography and cultural transmission processes have shaped geographic patterns of Aurignacian material culture, as evidenced in the archaeological record.

In a recent study (Vanhaeren and d’Errico 2006), ethno-linguistic differentiation has been invoked as an explanation of the high degree of structuring that is seen in the spatial distribution of symbolic artefact types associated with the Aurignacian culture in Upper Palaeolithic Europe. The primary aim of this thesis was to test whether ethnic structuring is required to explain the spatial distribution of variation in this archaeological material culture data.

To do this, spatially explicit cultural transmission simulation models that generate expectations of a range of spatial statistics describing the distribution of shared ornament types have been developed (sections 2.1 and 4.1). By integrating this modelling framework with ABC to allow comparison between simulated and observed spatial statistics (calculated on the observed material culture data (Vanhaeren and d’Errico 2006)) hypotheses concerning group interaction dynamics for the period have been tested and compared.

The results of the analyses (section 4.2.2) do not support the hypothesis that Aurignacian populations in Early Upper Palaeolithic Europe were ethnically structured in a manner related to ornamental material culture. The spatially explicit simulation models and ABC analyses, conditioned on the data presented by Vanhaeren and d’Errico (2006), indicate that – in terms of fit between simulated and observed data – there is little difference between the simple scenario of cultural identity-by-descent with modification and isolation-by-distance, and the more complex one that, in addition, invokes cultural group interaction processes that would be expected if differences in material culture were marking differences between ethnic groups and homophilically driving interaction processes. These analyses therefore imply that the interpretive conclusions reached by Vanhaeren and d’Errico (2006) may not be correct, and that more explicit probing of the data, possibly using simulation modelling approaches such as those presented here, is a crucial step in order to be able to robustly draw such conclusions.

Additionally, these analyses indicate that data simulated under the bounded stepwise mutation (i.e. innovation in the terminology of cultural evolution theory) model provides a better fit to observed data than that simulated under the bounded non-stepwise (discrete) mutation model. The stepwise mutation model relaxes the assumption that the cultural trait data considered in this study is necessarily categorical, since categorical distinctions are often arbitrarily made and cultural change may in fact be scalable. Mutation under this model corresponds to a group being more likely to add an item to its cultural repertoire that is morphologically similar to one that is already present in its cultural repertoire than one that is very different, while the non-stepwise
(discrete) mutation model corresponds to a group being equally likely to add any item—regardless of similarity or difference to items already present in—to its cultural repertoire. This result suggests the existence of scaling of item similarity in material culture repertoires, where scaling of item similarity implies that the extent of similarity of a particular item to items already in the repertoire would have affected how that item was perceived.

Furthermore, and somewhat surprisingly, the model comparison analyses indicate a better fit between observed data and data simulated under a uniform distribution of carrying capacities than that simulated under distributions in which carrying capacities are conditioned on Palaeoclimate data. Although the model comparison analyses show preference for uniform distribution of carrying capacities, the differences in the relative marginal likelihood estimates for each model are not large (Figure 4.12 and Table 4.2). This suggests that there is insufficient power in the observed data to differentiate between these scenarios.

The posterior distributions of some model parameters, particularly probability of cultural mutation and migration distance, are considerably narrower than their corresponding prior ranges (Figure 4.14, Figure 4.15 and Table 4.3), which suggests that some indirect information on these parameters may be contained in the observed data and so the analysis can be deemed to be somewhat informative for these parameters. Compared to their prior ranges, the posterior estimates of these two parameters are relatively low, which suggests low rates of cultural mutation (posterior range: 0.0003-0.0402, mode=0.0020) and low migration distances (posterior range (km): 9-86, mode=16.2427) in Aurignacian populations. The estimate of the rate of cultural mutation is in good agreement with estimates from studies of different stylistic and functional aspects of Neolithic material culture (Linearbandkeramic culture: Kandler and Shennan 2013; arrowheads: Crema et al. 2014). The estimate of migration distance requires further investigation, however, it is interesting to note that the estimated value of this parameter is similar to the average distance per residential move estimated in ethnographic studies of hunter-gatherer societies (Binford 2001; Kelly 2013: 80-84). These ethnographic studies, however, observe multiple residential moves each year, which is clearly a much greater temporal resolution compared to the generational time step considered in the simulation models presented in this thesis, and may therefore not be comparable to the current estimate. Nevertheless, the ethnographic studies do indicate that limiting factors of the annual ranges of movement include climate, distribution of resources and foraging strategies, all of which are in part reflected by the underlying carrying capacities estimated from Palaeoclimate data in this study (section 4.1.1).

The parameter estimation analysis is uninformative for other parameters (Figure 4.14, Figure 4.15 and Table 4.3: probability of fission / extinction, interaction radius), as evidenced by posterior distributions that are similar to their corresponding prior ranges, indicating that there is little information contained in the observed data that relates to the values of these parameters. Nevertheless, allowing these parameters to vary within uniform prior ranges serves to integrate uncertainty in their values.
Robustness of / Confidence in Conclusions

It is important to consider the quality of the results reported in this thesis, and the subsequent robustness of conclusions drawn above, in a more general sense. There are two elements to consider: the quality of the model and the quality of the observed data.

Assuming that the modelling framework is sufficiently realistic and explicit to model the true processes that would have given rise to the observed data, and that the observed dataset is rich enough and contains the information on ethnic structuring that is required to differentiate between different processes that may have given rise to the observed distributions, the conclusions drawn above stand.

If, on the other hand, the data does not contain sufficient information about ethnic structuring to distinguish between the different processes, then this study provides an argument for improving the archaeology of the time period, as well as for further comparison. It is likely that the model comparison (section 4.2.2) and parameter estimation (section 4.2.3) analyses would be more informative if the temporal dimension of the observed data was resolved (i.e. chronological data for sites were available), since this would further inform on the performance of different models. It is important to note here that, if the observed data contained further information (e.g. if the data was temporally resolved), this would need to be reflected in the number and definition of summary statistics, in order to accurately capture this additional information and use it to inform on the performance of different models.

Similarly, the results presented here would gain further significance if they could be compared with those from the subsequent Gravettian and later cultural periods of the same region. Similarly, it may be possible to gain further insight into group interaction dynamics during the Aurignacian by comparing the results of the ornament analysis (Vanhaeren and d’Errico 2006) with patterns derived from similarities and differences between lithic assemblages at the same sites.

Such methods as adopted here may have more deductive power in richer material culture data sets, particularly those that also have a chronological dimension, such as the Neolithic personal ornamentation dataset recently analysed by Shennan et al. (2015).
Reflections on Simulation Modelling Approach

The simulation modelling and ABC approaches adopted here are relatively advanced in the context of modelling archaeological data; the first application of simulation modelling and ABC approaches solely to archaeological data was published last year by Crema et al. (2014).

Inferring processes that may have given rise to observed patterns in the spatial distribution of material culture data is challenging, partly due to the relatively poor quality of the archaeological data for the time period considered (which limits the inferences that can be made based on the data) and partly due to the effectively infinite number of processes that may have given rise to such distributions (equifinality; for a detailed explanation, see Premo (2010)).

Given the issues of equifinality and evolutionary variance (i.e. the underlying stochasticity of evolutionary processes), it would be relatively easy to define overly complex models that invoke parameters to explain aspects of the data that are in fact due to randomness (i.e. overfitting); it is therefore necessary to test simple models first and to penalise complexity (Kass and Raftery 1995). For the models developed in this thesis, every effort has been made to ensure that they are as simple as possible, in terms of invoking as few parameters as possible (i.e. making as few assumptions as possible), whilst retaining some semblance of realism.

However, simpler (i.e. less realistic) models than those presented in this thesis can also prove useful in exploring different proposed scenarios, in an attempt to more fully understand the processes modelled and the conditions under which such models are able to simulate the observed archaeological material culture patterns. It is equally useful and important to understand the limits of the assumptions of such models, which become apparent in scenarios where the simulated data fail to match the observed, and is something that could be explored further for the models presented in this thesis. Taking the exploratory approach with very simple models provides systematic tools for generating and testing alternative hypotheses, and for exploring the dynamics of each process modelled across the parameter space, to fully understand the differences between plausible scenarios that do and do not generate simulated data that reflects that observed in the archaeological record (Premo 2007).

In terms of model parameter values, given that there is little information in the archaeological record that can act as a direct proxy of the values that the model parameters may take, the value of each parameter in each simulation is randomly assigned from a uniform prior range and fixed for the duration of each simulation. While constraining each parameter to a uniform prior range serves to integrate uncertainty in their values, a shortcoming of this approach is that the value each parameter takes in any one simulation is constant over the length of that simulation. This may be an issue for some parameters, such as probability of fission / extinction, which may not be acting uniformly over the spatial and temporal ranges simulated, but may instead be dependent on other factors such as climatic variables. Modelling such dependences would be difficult without additional information and therefore, in the absence of this additional
information, fixing the parameter for the length of a simulation is a more parsimonious assumption.

As with any simulation model, many simplifying assumptions have to be made and the extent to which these simplifications may lead to erroneous conclusions remains unknown. In particular, while the use of the term ‘ethnicity’ in archaeology has a controversial history, and as a result of this has taken on many nuanced meanings (see section 1.2.3), in this context, a very simplified link between ethnicity and the nature of cultural interaction processes has been made: if ethnicity exists and is reflected by similarities and differences in groups’ material cultures, then it would be expected to have a homophilic drive on cultural interaction processes. This is in part justified since this thesis is not concerned with the cognitive processes involved in ethnic structuring – since intentional and non-intentional modification of cultural traits both collapse to modification and there is little information available in the archaeological record to distinguish between the two, these are not distinguished in the models presented in this thesis in order to prevent overfitting (i.e. invoking parameters to explain aspects of the data that are in fact due to randomness). Instead, this thesis considers the effects on cultural interaction processes if differences in material culture mark differences between ethnic groups and homophilically drive cultural interactions, and the feedback effects of this on material culture.

A related simplification is that of the cultural interaction processes that are at the heart of the hypothesis tests in this thesis (section 2.1.3.4), which attempt to represent a multitude of different possible interaction processes and their effects. Given the lack of data available from which the exact nature of the likely interaction processes can be deduced, this generalisation attempts to account for a range of different possible mechanisms that may have acted in shaping material culture data. However, this may be an overgeneralisation and a more explicit treatment of these processes may lead to different conclusions. That said, more realistic modelling of the range of possible processes would likely mean invoking many more parameters, for which little information is available in the archaeological record, and, as explained above, this thesis is not concerned with the cognitive processes involved in formulations of ethnicity.

It should be noted, however, that the approach to cultural interaction processes adopted here (i.e. ‘sharing’ and ‘conflict’ processes as described in sections 2.1.3.4.2 and 2.1.3.4.3, respectively) is only one way to model the outcome of inter-group interactions. Future work could consider alternative mechanisms of cultural boundary formation, such as anti-conformism, whereby culturally different groups become more different upon interaction, for example through moving their trait values in opposite directions. This interaction process would represent an expression of individual identity, which is potentially a form of assertive style as discussed in section 1.2.4.

In addition to the above simplifications, it is not necessarily accurate to assume that the aggregate data set used in this study represents cultural interaction processes acting uniformly over the approximately 13,000 years of the Aurignacian; it might instead represent a spurious
averaging of a variety of different processes and this is something that should be considered in future applications of this model.

Another consideration is that the culture transmission process used in this framework assumes neutrality, in that ornament types are not assumed to differentially affect group fitness (as mentioned in section 2.1.3.4.3, group fitness can be defined in a multitude of different ways, including, for example, as the average individual fitness in a group), or be more or less likely to be copied. In the past, the neutral model has been proposed as a null hypothesis for studying cultural evolution (Neiman 1995; Shennan and Wilkinson 2001; Bentley et al. 2004; Steele et al. 2010) and in some cases the authors were unable to reject neutrality using cultural transmission models (Neiman 1995; Steele et al. 2010). However, it is difficult to demonstrate that cultural variants are selectively neutral (Neiman 1995; Shennan and Wilkinson 2001; Steele et al. 2010). This is partly due to the lack of statistical methods available to test for deviations from neutrality, however Kandler and Shennan (2013) have recently proposed one such method that tests whether changes in the frequency distributions of cultural variants over time can be explained assuming the neutral model. Given these issues, it is important to be cautious when assuming neutrality, however there is no a priori reason why use of different ornament types should differentially affect group fitness.

A further suggestion for improving the modelling framework is to re-consider the way migratory processes are modelled. In this study, group migrations are modelled as Gaussian random walks of parameterised average distance (sections 2.1.2.1 and 4.1.3). A potential alternative is to model these processes as parameterised Lévy random walks. Lévy walks are a type of random walk in which movement distances follow power-law distributions, and studies (Brown et al. 2007; Raichlen et al. 2014) looking at foraging patterns in human hunter-gatherer populations have suggested that Lévy walks are the optimal movement pattern when foraging for heterogeneously located resources (with little or no prior knowledge of resource distribution patterns). With this in mind, migratory processes in this study could be modelled as parameterised Lévy random walks, with the distance that each group traverses in a migration process selected from parameterised power-law distributions.
Reflections on ABC Approach

A clear advantage of the ABC approach is the flexibility in accommodating different proposed models that allows the comparison of relative support for different competing hypotheses based on the same observed data, as well as for estimating parameters of interest (Bertorelle et al. 2010). The application of ABC does have limitations, and the analysis of highly dimensional datasets such as those presented here relies on both the choice of summary statistics used to describe the data and the acceptance threshold used to decide which simulations generate data most similar to the observed.

In this study, the simple ABC rejection algorithm, as described in section 1.3.4.1, was adopted over more complex approaches (for example locally weighted linear regression (Beaumont et al. 2002), ABC-GLM (Leuenberger and Wegmann 2010) and ABC-MCMC (Marjoram et al. 2003) as described in section 1.3.4.4) that aim to improve on the relatively poor efficiency of this algorithm. Whilst such approaches may improve computational efficiency, they result in the loss of an important feature of ABC, namely that simulations cannot be considered independent and the retained combinations of parameters are therefore not random – these approaches may therefore favour some parameter combinations over others. As a result, the outcomes of such approaches would be more difficult to interpret and so it was decided that the simple ABC rejection algorithm should be used in this study, especially during model development.

Given that the ABC approach does not use all of the information in the observed data and instead relies on summary statistics that capture its relevant features, the choice of appropriate summary statistics is pivotal. Ideally, the set of summary statistics used should be sufficient, meaning that all the information that is contained in the observed data is captured by these statistics. This can prove difficult however, since outside of overly simple models there is no way to guarantee that a chosen set of statistics is sufficient. While defining a sufficient set of summary statistics is in theory independent of the models considered, simulated data can prove useful in helping to refine the set of summary statistics used, for example by highlighting aspects of the observed data that are ambiguously described – this is particularly true of complex observed data sets with high dimensionality, such as that considered in this thesis (Vanhaeren and d'Errico 2006). It is therefore essential to re-evaluate summary statistic choice after the model has been developed. The first step towards this has been attempted by considering the correlations between summary statistics and model parameters (section 4.2.1.2), as well as between different summary statistics (section 4.2.1.1).

Following on from these correlation analyses, in order to increase confidence in the conclusions made here, the summary statistics used in this thesis (section 2.2.1) should be re-evaluated using data generated under the current modelling framework. The current statistics were developed at the onset of this project and tested using a much simpler predecessor of the current model. As a result, they may not be sufficiently informative, which may partly be the reason for the low correlations observed between summary statistic and model parameter values (section
4.2.1.2). However, the low correlations observed between different statistics suggest that each captures information about the observed data that has some degree of independence (section 4.2.1.1), and so using this set of summary statistics should in theory capture a large amount of information about the observed dataset.

Nevertheless, it may be necessary to define additional summary statistics to use with, or instead of, those used currently. For example, the mutual information statistics used (section 2.2.1.2) quantify similarity based on shared presences and shared absences. Since negative matches (i.e. shared absences) are common in sparse matrices with many absences (Shennan and Bentley 2008), as is the case of the observed data in this study, this measure may overestimate the dependence between variables in which common absences are predominant; it may therefore be necessary to use a different statistic to take this effect into account. One such statistic is the Jaccard distance, which ignores instances of negative matches and is therefore useful for analysing archaeological data where such matches are predominant (Shennan et al. 2015); to quantify the similarity between individual sites, the Jaccard distance is defined as the ratio between the sum of the number of traits present in one site but not in the other, and the sum of the number of traits that are present in one or both sites. There are some similarities between the Jaccard distance and the shared information statistic used in this thesis and they may therefore capture similar information about the observed data, in which case the inclusion of both may not be necessary. However, other statistics could potentially be introduced to attempt to capture aspects of the observed data that may not be represented by the current set of summary statistics.

It is important to note that including statistics that capture the same information about the observed data, or including many statistics that provide little information about the parameters of interest, would introduce stochastic noise and reduce the number of simulations retained (Beaumont et al. 2002). In such cases, it may be necessary to apply dimensional reductions to the statistics where possible in order to ensure that a particular aspect of the data is not over-represented in the analysis. To this end, some approaches have been proposed for determining the most informative of a large set of statistics (Joyce and Marjoram 2008; Wegmann et al. 2009), as described in section 1.3.4.2. Given the number of summary statistics (i.e. 12) and model parameters (i.e. 6) currently used in this study, such approaches were not applied, but may become a necessary additional step if many additional summary statistics are added in the future.

Another important step in this approach should be to address the degree to which the proposed summary statistics can distinguish between different models, if an effect were there to be found. Veeramah et al. (2012) have proposed a method for testing which of a proposed set of summary statistics hold the most power in distinguishing between proposed models, which is based on the assumptions that the multidimensional distribution of summary statistics is different for each proposed model, and that the observed statistic lies closer to the centre of one distribution than another. Formal methods for selecting the optimum set of summary statistics for parameter
estimation have also been suggested (see, for example Joyce and Marjoram (2008) and Wegmann et al. (2009)). Although such methods have not been implemented in this thesis, they could be included as an extra step in the process in order to allow the issue of power to be addressed more clearly.
Concluding Remarks

A simulation modelling approach is considerably more complex and laborious to implement compared to the interpretation of descriptive statistics or patterns in data alone. However, as has been stated before, this is a formal scientific approach that proposes a model with an explicit prediction of the distribution of data, and tests this formally by comparing the simulated data to the observed data for validation. Taking this approach necessitates reduced models and these, by definition, will never fully describe the complexity of the true processes that shaped the material culture data. However, the model building and testing process is not a closed one and various ways in which aspects of the current model could potentially be improved have been suggested. Nonetheless, the approach adopted here is explicit and transparent and therefore less likely to be influenced by the subjective biases that guide interpretation (Gerbault et al. 2014).

Despite the drawbacks as discussed above, explicit approaches for hypothesis testing are an essential tool when inferring past processes, since the observed data are only one of a number of possible outcomes of a set of stochastic processes. Simulation modelling approaches coupled with the ABC framework provide a means to obtain insights into the accumulated outcomes of iterated processes going on for hundreds or thousands of years. Interpretive approaches are necessary to postulate hypotheses about the processes that may have given rise to the observed data, however, as this thesis has shown, it is necessary to formally test the support by the observed data for these different hypotheses when inferring past processes.

To my knowledge, the approach reported here has not been attempted when considering archaeological evidence for ethnic structuring. Given that there is little or no representation in the literature of explicit simulation modelling approaches to questions of ethnic structuring, while interpretative approaches are well represented, this study attempts to begin to fill an important gap in the literature.
References


de Martonne E (1920) Traité de géographie physique: climat--hydrographie--relief du sol--biogéographie. A. Colin,

de Martonne E (1941) Nouvelle carte mondiale de l’indice d’aridité.


Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of
the Hadley Centre coupled model without flux adjustments. Clim Dynam 16 (2-3):147-168


